Metal Machining

Theory and Applications

Thomas Childs
University of Leeds, UK

Katsuhiro Maekawa
Ibaraki University, Japan

Toshiyuki Obikawa
Tokyo Institute of Technology, Japan

Yasuo Yamane
Hiroshima University, Japan

A member of the Hodder Headline Group
LONDON
Copublished in North, Central and South America
by John Wiley & Sons Inc.
New York-Toronto

First published in Great Britain in 2000 by
Arnold, a member of the Hodder Headline Group,
338 Euston Road, London NW1 3BH

http://www.arnoldpublishers.com

Copublished in North, Central and South America by
John Wiley & Sons Inc., 605 Third Avenue,
New York, NY 10158–0012

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN 0 340 69159 X
ISBN 0 470 39245 2 (Wiley)

1 2 3 4 5 6 7 8 9 10

Commissioning Editor: Matthew Flynn
Production Editor: James Rabson
Production Controller: Iain McWilliams
Cover Design: Mouse Mat Design

Typeset in 10/12 pt Times by Cambrian Typesetters, Frimley, Surrey
Printed and bound in Great Britain by Redwood Books Ltd.

What do you think about this book? Or any other Arnold title?
Please send your comments to feedback.arnold@hodder.co.uk

Contents

Preface

Improved manufacturing productivity, over the last 50 years, has occurred in the area of machining through developments in the machining process, in machine tool technology and in manufacturing management. The subject of this book is the machining process itself, but placed in the wider context of manufacturing productivity. It is mainly concerned with how mechanical and materials engineering science can be applied to understand the process better and to support future improvements.

There have been other books in the English language that share these aims, from a variety of viewpoints. *Metal Cutting Principles* by M. C. Shaw (1984, Oxford: Clarendon Press) is closest in spirit to the mechanical engineering focus of the present work, but there have been many developments since that was first published. *Metal Cutting* by E. M. Trent (3rd edn, 1991, Oxford: Butterworth-Heinemann) is another major work, but written more from the point of view of a materials engineer than the current book's perspective. *Fundamentals of Machining and Machine Tools* by G. Boothroyd and W. A. Knight (2nd edn, 1989, New York: Marcel Dekker) covers mechanical and production engineering perspectives at a similar level to this book. There is a book in Japanese, *Modern Machining Theory* by E. Usui (1990, Tokyo: Kyoritu-shuppan), that overlaps some parts of this volume. However, if this book, *Metal Machining*, can bear comparison with any of these, the present authors will be satisfied.

There are also more general introductory texts, such as *Manufacturing Technology and Engineering* by S. Kalpakjian (3rd edn, 1995, New York: Addison-Wesley) and *Introduction to Manufacturing Processes* by J. A. Schey (2nd edn, 1987, New York: McGraw-Hill) and narrower more specialist ones such as *Mechanics of Machining* by P. L. B. Oxley (1989, Chichester: Ellis Horwood) which this text might be regarded as complementing.

It is intended that this book will be of interest and helpful to all mechanical, manufacturing and materials engineers whose responsibilities include metal machining matters. It is, however, written specifically for masters course students. Masters courses are a major feature of both the American and Japanese University systems, preparing the more able twenty year olds in those countries for the transition from foundation undergraduate courses to useful professional careers. In the UK, masters courses have not in the past been popular, but changes from an elite to a mass higher education system are resulting in an increasingly important role for taught advanced level and continuing professional development courses.

It is supposed that masters course readers will have encountered basic mechanical and materials principles before, but will not have had much experience of their application. A feature of the book is that many of these principles are revised and placed in the machining context, to relate the material to earlier understanding. Appendices are heavily used to meet this objective without interrupting the flow of material too much.

It is a belief of the authors that texts should be informative in practical as well as theoretical detail. We hope that a reader who wants to know how much power will be needed to turn a common engineering alloy, or what cutting speed might be used, or what material properties might be appropriate for carrying out some reader-specific simulation, will have a reasonable chance either of finding the information in these pages or of finding a helpful reference for further searching.

The book is essentially organized in two parts. Chapters 1 to 5 cover basic material. Chapters 6 to 9 are more advanced. Chapter 1 is an introduction that places the process in its broader context of machine tool technology and manufacturing systems management. Chapter 2 covers the basic mechanical engineering of machining: mechanics, heat conduction and tribology (friction, lubrication and wear). Chapters 3 and 4 focus on materials' performance in machining, Chapter 5 describes experimental methods used in machining studies.

The core of the second part is numerical modelling of the machining process. Chapter 6 deals with mechanics developments up to the introduction of, and Chapters 7 and 8 with the development and application of, finite element methods in machining analysis. Chapter 9 is concerned with embedding process understanding into process control and optimization tools.

No book is written without external influences. We thank the following for their advice and help throughout our careers: in the UK, Professors D. Tabor, K. L. Johnson, P. B. Mellor and G . W. Rowe (the last two, sadly, deceased); in Japan, Professors E. Usui, T. Shirakashi and N. Narutaki; and Professor S. Ramalingam in the USA. More closely connected with this book, we also especially acknowledge many discussions with, and much experimental information given by, Professor T. Kitagawa of Kitami Institute of Technology, who might almost have been a co-author.

We also thank the companies Yasda Precision Tools KK, Okuma Corporation and Toyo Advanced Technologies for allowing the use of original photographs in Chapter 1, British Aerospace Airbus for providing the cover photograph, Mr G. Dean (Leeds University) for drafting many of the original line drawings and Mr K. Sekiya (Hiroshima University) for creating some of the figures in Chapter 4. One of us (it is obvious which one) thanks the British Council and Monbusho for enabling him to spend a 3 month period in Japan during the Summer of 1999: this, with the purchase of a laptop PC with money awarded by the Jacob Wallenberg Foundation (Royal Swedish Academy of Engineering Science), resulted in the final manuscript being less late than it otherwise would have been.

We must thank the publisher for allowing several deadlines to pass and our wives – Wendy, Yoko, Hiromi and Fukiko – and families for accepting the many working weekends that were needed to complete this book.

Thomas Childs, Katsuhiro Maekawa, Toshiyuki Obikawa, and Yasuo Yamane
England and Japan
September, 1999

1

Introduction

Machining (turning, milling, drilling) is the most widespread metal shaping process in mechanical manufacturing industry. Worldwide investment in metal-machining machine tools holds steady or continues to increase year by year, the only exception being in the worst of recessions. The wealth of nations can be judged by this investment. Figure 1.1 shows the annual expenditure on machine tools by each of the most successful countries – Germany, Japan and the USA. For each, it was between £1bn and £2bn (bn = 10^9) in the late 1970s. It fell abruptly in the world recession (the oil crisis) of 1981–82 and has now recovered to between £2bn and £3bn (all expressed in 1985 prices: £1 was then equivalent to 300¥ or $1.3). Figure 1.1 also shows similar trends (a growth over the last 20 years from

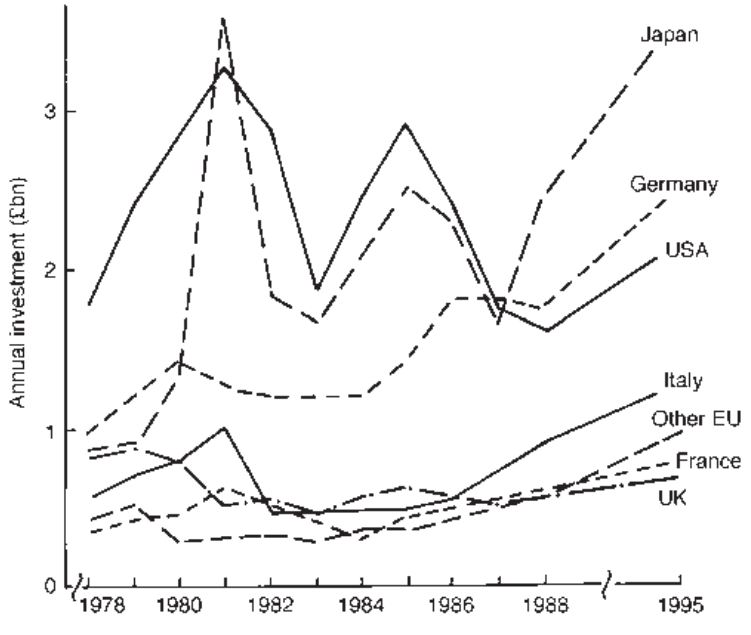


Fig. 1.1 International demand for machine tools, 1978–88, £bn at 1985 prices (from European community statistics 1988) and projected at that time to 1995

50% to 100% in annual expenditure) for the developed European Community countries. Only in the UK has there been a decline in investment. Over this period, investment in metal machining has remained at about three times the annual investment in metal forming machinery.

Investment has continued despite perceived threats to machining volume, such as the displacement of metal by plastics products in the consumer goods sector, and material wastefulness in the production of swarf (or chips) that has encouraged near-net (casting and forging) process substitution in the metal products sector. One reason is that metal machining is capable of high precision: part tolerances of 50 μm and surface finishes of 1 μm are readily achievable (Figure 1.2(a)). Another reason is that it is very versatile: complicated free-form shapes with many features, over a large size range, can be made more cheaply, quickly and simply (at least in small numbers) by controlling the path of a standard cutting tool rather than by investing considerable time and cost in making a dedicated moulding, forming or die casting tool (besides, machining is needed to make the dies for moulding, forging and die casting processes).

One measure of a part's complexity is the product of the number of its independent dimensions and the precision to which they must be made (Ashby, 1992). Figure 1.2(b) gives limits to the component size (weight units – a cube of steel of side 3 m weighs approximately 2×10^5 kg) and complexity of machining and its competitive processes. Complexity is defined by

$$C = n \log_2 (l/\Delta l) \tag{1.1}$$

where n is the number of the dimensions of the part and $\Delta l/l$ is the average fractional precision with which they are specified.

A third reason for the success of metal machining is that the need from competition to increase productivity, to hold market share and to find new markets, has led to large changes in machining practice. The changes have been of three types: advances in machine tools (machine technology), in the organization of machining (manufacturing systems) and in the cutting edges themselves (materials technology). Each new improvement in one area

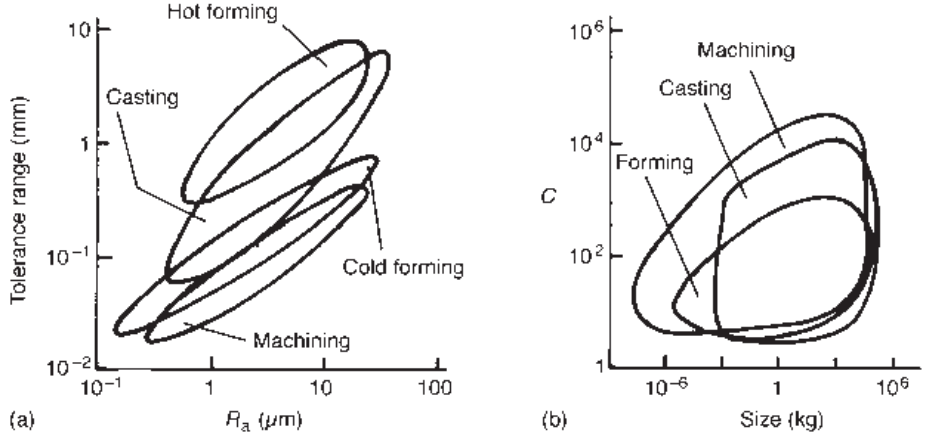


Fig. 1.2 (a) Typical accuracy and finish and (b) complexity and size achievable by machining, forming and casting processes, after Ashby (1992)

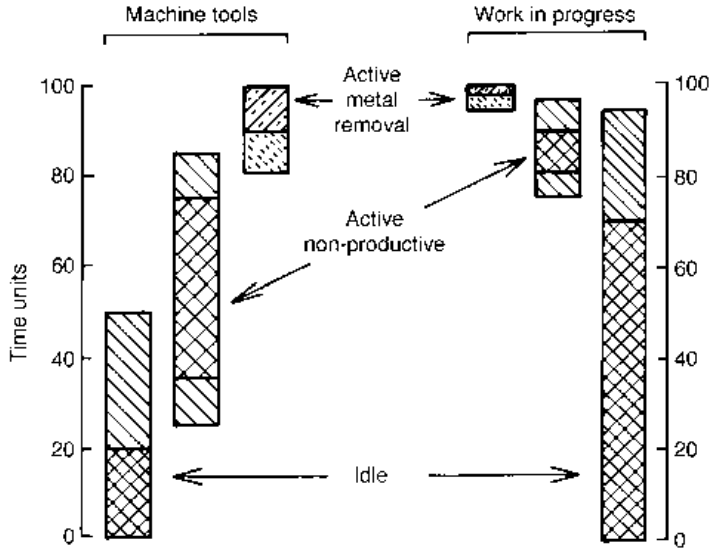


Fig. 1.3 Idle and active times in batch manufacturing, from surveys *circa* 1970

throws pressure on to another. It is worthwhile briefly to review the evolution of these changes, from the introduction of numerical controlled machine tools in the late 1950s to the present day, in order to place in its wider context the special content of this book (the consideration of the chip forming process itself), which is at the heart of machining.

1.1 Machine tool technology

In the early 1970s a number of surveys were carried out on the productivity of machine shops in the UK, Europe and the USA (Figure 1.3). As far as the machine tools were concerned it was found that they were actually productive, removing metal, for only 10 to 20% of the time: different surveys, however, gave different values. For 40 to 60% of the time the machine tools were in use but not productively: i.e. they were being set up for manufacture, or being loaded and unloaded, or during manufacture tools were being moved and positioned for cutting but they were not removing metal. For 20 to 50% of the time they were totally unused – idle.

As far as work in progress was concerned, batches of components typically spent from 70 to 95% of their time inactive on the shop floor. So overwhelming was the clutter of partly finished work that a component requiring several different operations for its completion, on different machine tools, might find these carried out at the rate of only one a week. From 10 to 20% of their time components were being positioned for machining and for only from 1 to 5% of the time was metal actually being removed.

From the late 1960s to the early 1970s both forms of waste – the active, non-productive and the idle times – began significantly to be attacked, the former mainly by developing machine tool technology and the latter by new forms of manufacturing organization.

1.1.1 Machine tool technology – mainly turning machines

From 1970 onwards, machine tools of new design started to be introduced in significant numbers into manufacturing industry, with the effect of greatly reducing the times for tool positioning and movement between cuts. These new, computer numerical control (CNC), designs stemmed directly from the development of numerically controlled (NC) machine tools in the 1950s. In traditional, mechanically controlled machine tools, for example the lathe in Figure 1.4, the coordination needed between the main rotary cutting motion of the workpiece and the feed motions of the tool is obtained by driving all motions from a single motor. The feed motions are obtained from the main motion via a gear box and a slender feed rod (or lead screw for thread cutting). With the exception of machines known as copying machines (which derive their feed motion by following a copy of a shape to be made) only simple feed motions are obtainable: on a lathe, for example, these are in the axial and radial directions – to machine a radius on a lathe requires the use of a form tool. In addition, the large amount of backlash in the mechanical chain requires time and a skilled operator to set the tool at the right starting point for a particular cut.

In a CNC machine tool, all the motions are mechanically separate, each driven by its own motor (Figure 1.4) and each coordinated electronically (by computer) with the others. Not only are much more complicated feed motions possible, for example a combined radial and axial feed to create a radius or to take the shortest path between two points at different axial and radial positions, but the requirement of coordination has led to the development of much more precise, backlash-free ball-screw feed drives. This precise numerical control of feed motions, with the ability also to drive the tools quickly between cuts, together with other reductions in set-up times (to be considered in Section 1.2), has approximately halved machine tool non-productive cycle time, relative to its pre-1970 levels.

This halving of time is indicated in Figure 1.5(a) (Figure 1.5(b) is considered in Section 1.1.2). A further halving of non-productive cycle time has been possible from about 1980 onwards, with the spread throughout all manufacturing industry of new types of machine tools that have become called *turning centres* (related to lathes) and *machining centres* (developed from milling machines). These new tools, first developed in the 1960s for mass production industry, individually can carry out operations that previously would have required several machine tools. For example, it is possible on a traditional lathe to present a variety of tools to the workpiece by mounting the tools on a turret. In a new turning centre, some of the tools may be power driven and the main power drive, usually used to rotate the workpiece in turning operations, may be used as a feed drive to enable milling and drilling as well as turning to be carried out on the one machine.

Figure 1.6 is an example of a keyway being milled in a flanged hollow shaft. Pitch circle holes previously drilled in the flange can also be seen. This part would have required three traditional machines for its manufacture: a lathe, a milling and a drilling machine, with three loadings and unloadings and three set-ups. It is the possibility of reducing loadings and set-ups that has led to the further halving of cycle times – although this figure is an average. Individual time savings increase with part complexity and the number of set-ups that can be eliminated. Centres are also much more expensive than more simple traditional machine tools and need to be heavily used to be cost effective. The implications of this for the development of metal cutting practice – a trend towards higher speed machining – will be developed in Section 1.4.

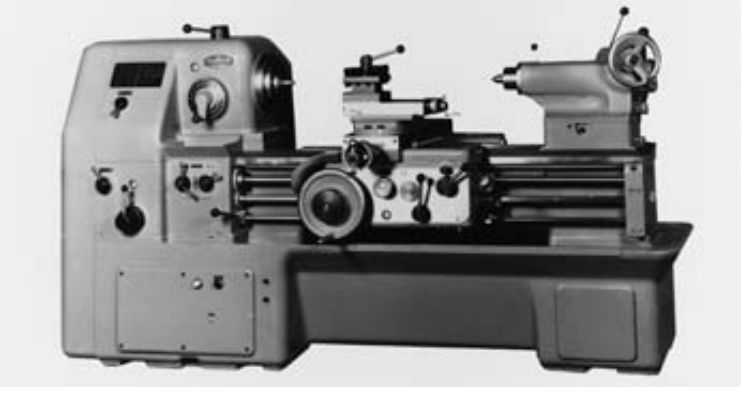


Fig. 1.4 A mechanically controlled lathe and (below) partly-built and complete views of a numerically controlled machine with individual feed drive motors

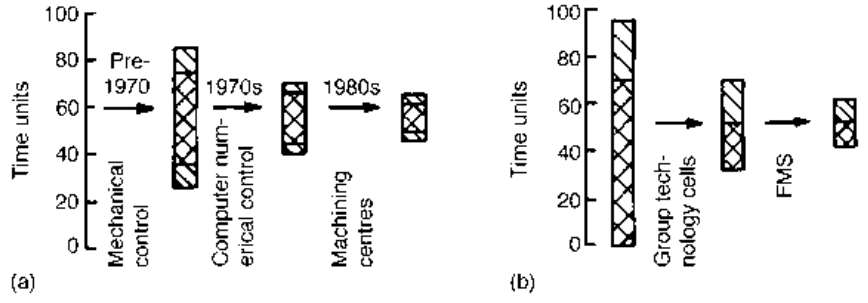


Fig. 1.5 Reductions from the levels shown in Figure 1.3 of (a) machine tool non-productive time and (b) work in progress idle time, due to better technology and organization

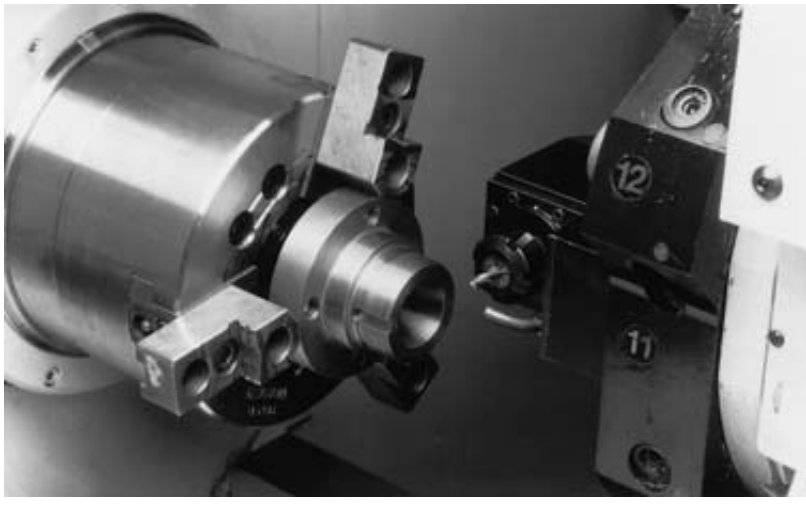


Fig. 1.6 A flanged shaft turned, drilled and milled in one set-up on a turning centre

The increased versatility of machine tools (based on turning operations as an example) has been briefly considered: the freedom given by CNC to create more complicated feed motions, both by path and speed control; and the evolution of multi-function machine tools (centres). The cost penalty has just been mentioned. As part of the continuing scene setting for the conditions in which metal cutting is carried out, which will be combined with systems and materials technology considerations in Section 1.4, some broad machine tool mechanical design and cost considerations will now be introduced – still in the context of turning.

Figure 1.7 sketches a turning operation, in which, in one revolution of the bar, the tool moves an axial distance f (the feed distance) to reduce the bar radius by an amount d (the depth of cut). The figure also shows the cutting force F_c acting on the tool, the diameter D at which the cutting is taking place and both the angular speed Ω at which the bar rotates and the consequent linear speed V (in later chapters this will be called U_{work}) at the diameter D. Material is removed, in the form of chips, at the rate fdV. (More detail of cutting terminology is given in Chapter 2).

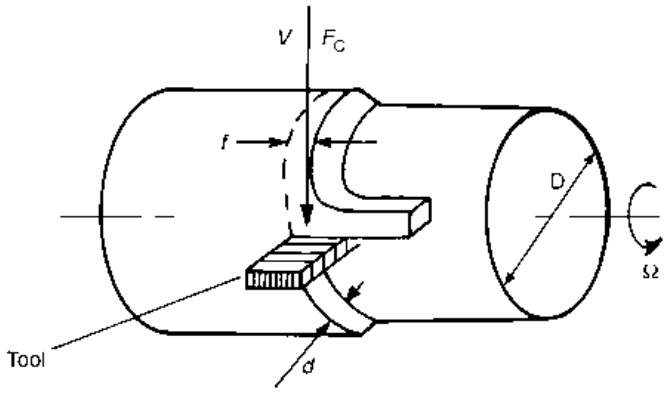


Fig. 1.7 The turning process – not to scale

The torque T and power P that the main drive motor must generate to support this turning operation is, by elementary mechanics

$$T = F_c\,(D/2) \equiv (F_c^* fd)(D/2) \tag{1.2a}$$

$$P = F_c V \equiv (F_c^* fd)V \text{ or } F_c^*\,(fdV) \tag{1.2b}$$

A new quantity F_c^* has been introduced. It is the cutting force per unit area of removed material. Called the specific cutting force, it depends to a first approximation mainly on the material being cut. Equation (1.2a) indicates that, for a constant area of cut fd, a turning machine should be fitted with a motor with a torque capacity proportional to the largest diameter being cut. It is shown later that for any combination of work and tool there is a preferred linear cutting speed V. Equation (1.2b) suggests that for a constant area of cut the required motor power should be independent of diameter cut. Observing what motors, with their torque and power capacities, are fitted to production machine tools can give insight into what duties the machine tools are expected to perform; and what forces the cutting tools are expected to withstand. This is considered next.

Machine tool manufacturers' catalogues show that turning machines are fitted with motors the torques and powers of which increase, respectively, with the square of and linearly with, the maximum work diameter. A typical catalogue specifies, among other things, the main motor power, the maximum rev/min at which the work rotates and the maximum diameter of work for which the machine is designed. Figure 1.8(a) plots the torque at maximum rev/min, obtained from $P = \Omega T$, against maximum design diameter, both on a log scale, for a range of mechanically controlled and CNC centre lathes and chucking turning centres (as illustrated in Figures 1.4 and 1.6 respectively). Apart from two sets of data marked 't', which are for lathes described as for training and which might be expected to be underdesigned relative to machines for production use, both the mechanical and CNC classes of machine show the same squared power law dependence of torque on maximum work diameter.

It seems that machines are designed to support larger areas of cut, fd, the larger the work diameter D. Not only are larger diameter workpieces stiffer and able to support larger forces (and hence areas of cut), but usually they require more material to be removed from them. A larger area of cut enables the time for machining to be kept within bounds. A

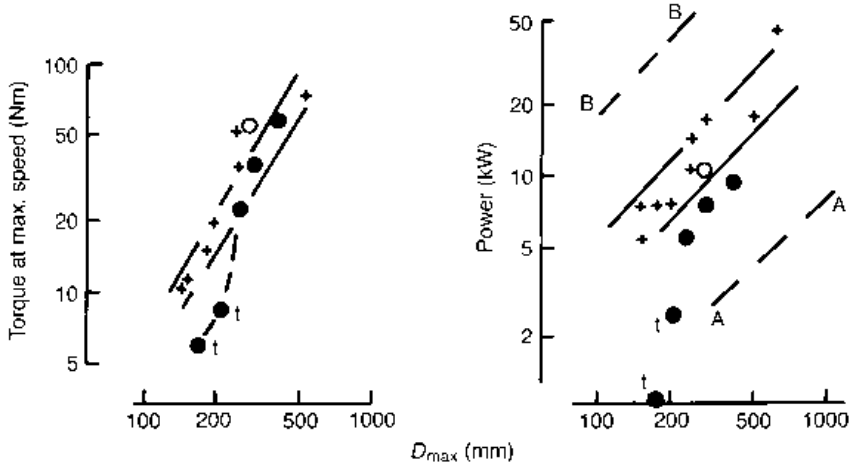


Fig. 1.8 Torque capacity at maximum speed and power of typical production mechanical (•) and basic CNC (o), lathes and turning centres (+) for steel machining, from manufacturers' catalogues

design specification that the maximum depth of cut d should increase in proportion to the maximum work diameter D would, from equation (1.2a), give the observed squared power law.

Design cutting forces may be deduced from the torque/diameter relationship shown in Figure 1.8(a). For example the lowest torque of 10 N m in Figure 1.8(a) would be caused by a cutting force of 140 N at the diameter of 145 mm, while the upper limit around 50 N m would be caused by 270 N at 365 mm. Of course, a workpiece will not be machined only at its maximum diameter. The highest rotational speeds are, in fact, used at the smallest machined diameters (to maintain a high linear speed). If features were machined at one tenth maximum diameter, the 10 N m and 50 N m torques would be generated by cutting forces of 1.4 kN and 2.7 kN. The turning machines represented in Figure 1.8 are, in fact, designed to generate cutting forces up to 2 or 3 kN. These are the forces to which the cutting tools are exposed.

Figure 1.8(b) shows designed power is proportional to maximum work diameter, consistent with equation (1.2b) if d is proportional to D. Further, the CNC machines have motors up to twice as powerful as mechanically controlled machines for a given work diameter. The top rotational speeds of CNC machines tend to be twice those of mechanically controlled ones, for example 4000 to 5000 rev/min as opposed to 2000 to 2500 rev/min for maximum work diameters around 250 mm. It is tempting to speculate that this is part of a trend to higher productivity through higher cutting speeds (Section 1.4). This may be partly true, but there is also another reason – it is due to the different characteristics of the motors used in mechanically and CNC controlled machines. The main drive of a mechanically controlled lathe runs at constant speed, and different work rotational speeds are obtained through a gear box. Apart from gear box losses, the motor can deliver a constant power to the work, independent of work speed. A CNC main drive motor is a variable speed motor with, as illustrated in Figure 1.9, a power capacity that drops off at low rotation speeds, i.e. when turning at maximum bar diameter. To compensate for this, a motor with a higher power at high rotational speeds must be employed.

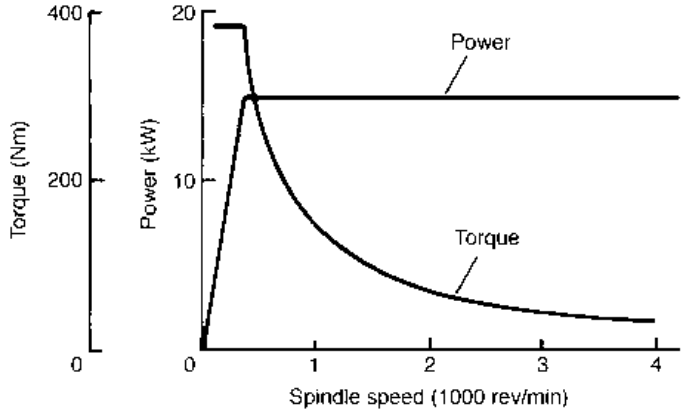


Fig. 1.9 The torque and power characteristics of a typical 15 kW AC variable speed motor used in CNC turning machines

The cutting speeds V at which the machine tools are expected to operate can be deduced from the available power and the expected cutting forces at high rotation speeds, i.e. at small cutting diameters. Continuing the example above, of a cutting force range of 1.4 kN to 2.7 kN; associating these with powers from 5 kW to 20 kW (Figure 1.8(b)), gives cutting speeds from 215 m/min to 450 m/min. It will be seen later (Section 1.3 and Chapter 3) that speeds in the range 100 to 1000 m/min are indeed practical for turning steels with cutting tools made from cemented carbides (tungsten and titanium carbides bonded by cobalt), which are the workhorse tools of today.

The dissipation of up to 5 to 20 kW through cutting tools results in them becoming very hot: 1000˚C is not unusual (this is justified later). For the tools to carry kN forces (or rather the associated stresses, approaching 1 GPa) at such temperatures requires high temperature strength. It is this that ultimately limits the productivity of cutting tools. Obsolete machine tools – from the 1960s and earlier – were provided with lower power motors (line A–A in Figure 1.8(b)) because they were designed for use with less productive tools made from high speed steels, with a lower high-temperature strength than cemented carbides. Some modern machine tools, designed for use with ceramic tooling and higher cutting speeds, are being fitted with higher power motors (line B–B in Figure 1.8(b)).

These 'facts of life' of the turning process – forces up to 2 or 3 kN and cutting speeds up to 1000 m/min – are set by the material properties of the work and tool materials as well as the mechanics of the process. Later chapters will be devoted to the details of why these 'facts of life' are so. They, and the functional versatility considered earlier, determine the price of turning machine tools. Machines must have a sufficient bulk and mass to be stiff and stable when cutting the high speed rotating mass of the workpiece. Figure 1.10(a) shows, for the same machine tools as in Figure 1.8, how their masses increase in proportion to motor power (the maximum workpiece lengths are in the range 500 mm to 1 m; machine mass increases with workpiece length as well as diameter capacity). Mass turns out to be one practical measure of value in a machine tool, the other being versatility. Figure 1.10(b) shows the list price of machine tools (without tax) as a function of mass (the data were gathered in 1990).

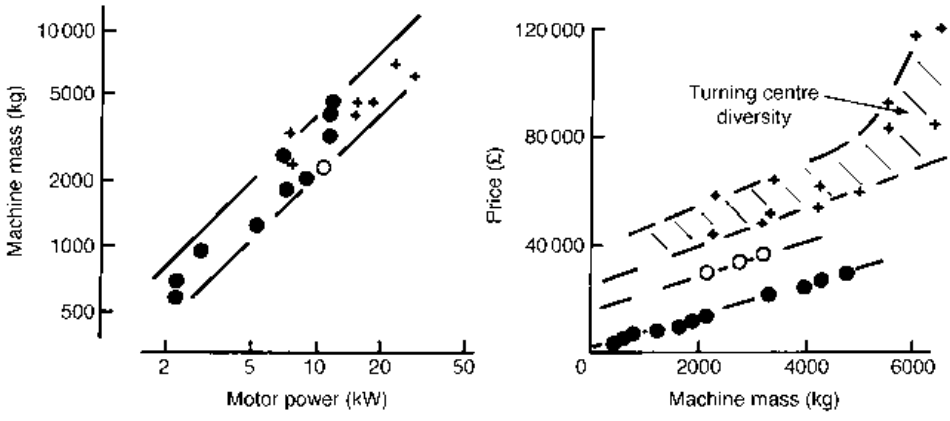


Fig. 1.10 Mass/power and price/mass relationships for turning machines

Here and later in the Chapter, prices and costs have been collected in the UK, during the early 1990s. A decision has been made to leave the information in units of UK£, unadjusted for inflation. An approximate conversion to values in the USA may be made at UK£1 = US$1; and to values in Japan at UK£1 = ¥200. These are not general exchange rates but equivalent purchasing rates.

Mechanically controlled centre lathes vary in price from around £3000 to £30 000 as their mass increases from 500 kg to 5000 kg. Changing to CNC controlled main and feed drives (the 1970s development of Figure 1.5(a)) displaces the price/mass relation upwards by about £15 000, while the further development of increased functionality of turning centres displaces the relation upwards by at least a further £15 000 to £20 000. There is a wide range of turning centre prices per unit mass, reflecting the wide range of complexity that can be built in to such a machine in a manner tailored to suit the needs of the parts being machined on it. The more specialized the turning centre, the more productive it can be: the degree of investment that is worthwhile will depend on whether a manufacturer can keep it occupied. The most specialized tend to be used with robotic loading and unloading systems (see Section 1.2). The prices in Figure 1.10(b) do not include such external materials handling devices.

1.1.2 Milling and drilling machines

Up to this point, the description of machine tool development has been in terms of the turning process. Before moving to consider the role of manufacturing organization in influencing the machining process, it is interesting to consider the parallel development of milling machine tools and machining centres. As with turning machines, there have been two stages of development: a post-1970 stage, which saw the substitution of mechanically controlled machines by their CNC equivalents; and a post-1980 stage, which has, in addition, seen the development of more versatile machining centres. Figure 1.11 compares the annual UK investment in mechanical and CNC turning and milling machines around the 1980 watershed. Pre-1980, the purchase of mechanically controlled machines was holding steady, with roughly twice the investment in turning as in milling machines. At the same

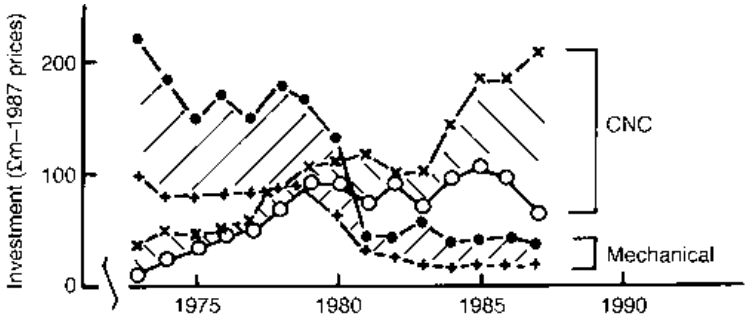


Fig. 1.11 Annual UK investment in mechanically and CNC controlled turning (•,o) and milling (+,x) machines, from UK government statistics

time, investment in CNC machines was growing, equally spread between turning and milling. Post-1980, investment in mechanically controlled machines collapsed and that in CNC turning machines held steady, while CNC milling machine investment increased to the stage where it was twice that of turning machines. This increase was mainly due to the influence of machining centres.

At first sight it is surprising that pre-1980 investment in substituting mechanically controlled for CNC-controlled milling machines equalled that for lathes, because there is less to be gained from reducing non-productive cycle times. The obvious difference between turning and milling processes is that, in turning, the main power is used to rotate an essentially cylindrical workpiece, with feed motions applied to the tool; whereas in milling the main power rotates a cutting tool, with the prismatic workpiece undergoing feed motions. Milling cutting tools have many cutting edges, and are more complicated than turning tools (Figure 1.12) and each edge cuts only intermittently. The cost of the tools makes it prudent to remove metal more slowly, and vibrations set up by the intermittent tool contacts reinforce this. The longer cutting times make the non-productive time less significant.

However, investment in milling machines in the pre-1980 period was not only in order to take advantage of the reduced non-productive time due to numerical control. A revolution was taking place, not only in machine control but also in machine structure. When mechanical feed drives were replaced by individual ball-screw feed drives, it was found that the accuracy of the cut was no longer limited by the accuracy of the drive but by elastic deflection of the milling machine frame. The introduction of CNC control led directly to a mechanical redesign of milling machines in order to produce machines of higher stiffness and hence accuracy. Figure 1.13 compares the new type of design with the earlier one. In addition, the freedom to vary x–y feed motions simultaneously to create curved feed paths opened up the possibilities for free-form shape generation by milling that existed before only with difficulty.

After 1980, machining centres attacked the long set-up and tool change times associated with milling. The number of set-ups was reduced by developing machines with more degrees of freedom in their motions than before. In addition to x,y table motions and z spindle motions, machines were built in which the spindle could be tilted. Automatic tool change magazines were developed. Automatically interchangeable work tables were also

Fig. 1.12 Examples of turning and milling solid, brazed and insert tools

devised so that setting up of one part could be carried out while another part was being machined. In an extreme form, it was possible to pre-prepare parts on a carousel worktable, such that, with magazine tool changing, a milling machining centre could be loaded with enough work and tools to keep it running overnight without attention from an operator. These changes, much greater than the changes in the development of turning centres from lathes, explain the greater investment in milling than turning in the post-1980 period as shown in Figure 1.11. Figure 1.14 shows an example of a new design of machine with a tiltable spindle and interchangeable worktables. Figure 1.15 shows a detail of a tool change magazine.

As far as process mechanics is concerned, equations (1.2) for torque and power can be applied to milling if D is interpreted as the diameter of the cutting tool and fdV remains the volume removal rate. However, torque and power are not limited by workpiece stiffness. It is the stiffness or strength of the cutter spindle that is important. The polar second moment of area J of a shaft is proportional to D to the fourth power, and surface stress in a shaft varies as TD/J. The torque T to create a given surface stress thus increases as D^3. The torque to create a given angular twist of the spindle also increases as D^3, if spindle length increases in proportion to D. A torque increases as D^3 if cutting force increases as D^2. For a given cutting speed, from equation (1.2b), the machine power to provide that force would also increase as D^2. Manufacturers' catalogues show that milling machine tools do have different power-to-capacity relations than turning machine tools, which can be explained on the basis that spindle failure or deflection limits their use, as just outlined. They also have different mass to power characteristics. However, the price of milling machines per unit mass is similar to turning machines. All this is developed in Figure 1.16.

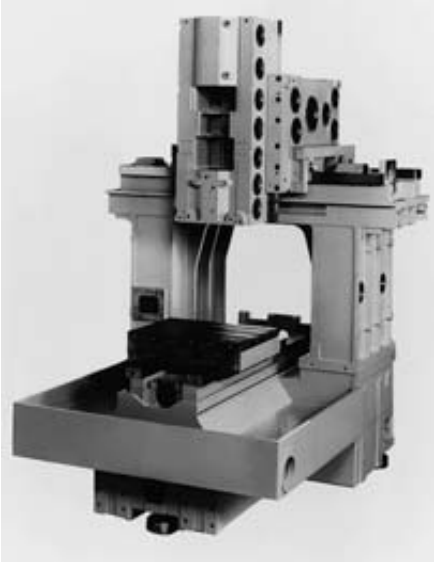


Fig. 1.13 A traditional – column and knee – design and (right and below) partly-built and complete views of a modern (bed) design of milling machine

In Figures 1.16(a) and (b) the capacity of a milling machine is measured by its cross-traverse capacity. This defines maximum workpiece size in a similar manner to defining the capacity of a turning centre by maximum work diameter (Figure 1.8). Figures 1.16(a) and (b) show that torque and power increase as cross-traverse cubed and squared respectively. An assumption that machines are designed to accommodate larger diameter cutters in proportion to workpiece size yields the D^3 and D^2 relations derived in the previous paragraph.

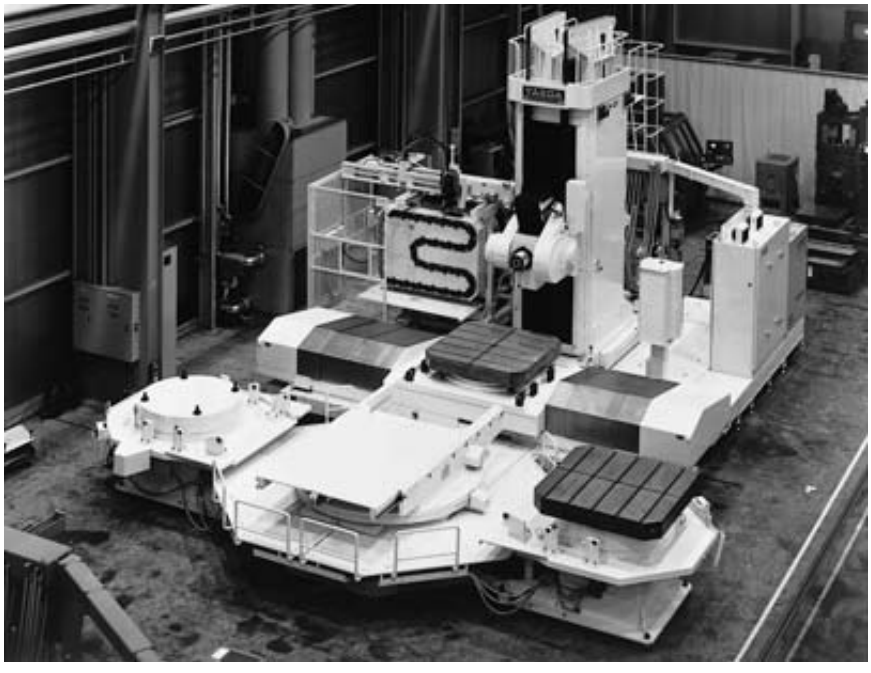


Fig. 1.14 A 5-axis milling machine with interchangeable work tables

If Figure 1.16(b) is compared with Figure 1.8(b) it is seen that for given workpiece size (cross-traverse or work diameter) a milling machine is likely to have from one fifth to one half the power capacity of a turning machine, depending on size. This means that milling machines are designed for lower material removal rates than are turning machines, for a given size of work. Figure 1.16(c), when compared with Figure 1.10(a), shows that milling machines are up to twice as massive per unit power as turning machines, reflecting the greater need for rigidity of the (more prone to vibration) milling process. Figure 1.16(d), admittedly based on a rather small amount of data, shows little difference in price between milling and turning machines when compared on a mass basis. Combining all these relationships, the price of a milling machine is about 2/3 that of a turning machine for a 200 mm size workpiece but rises to 1.5 times the price for 1000 mm size workpieces. The consequences for economic machining of these different capital costs, as well as the different removal rate capacities that stem from the different machine powers, are returned to in Section 1.4.

The D^3 and D^2 torque and power relationships found for milling machines are also observed, approximately, for drilling machines. In this case, size capacity can be directly related to the maximum drill diameter for which the machine is designed. Motor torques and powers, from catalogues, typically vary from 1 N m to 35 N m and from 0.2 kW to 4 kW as the maximum drill diameter that a drilling machine can accept rises from 15 mm to 50 mm. The ranges of torques and powers just quoted are respectively 20% and 10% of the ranges typically provided for milling machines (Figure 1.16). In drilling deep holes, there is a real danger of breaking the tools by applying too much torque, so machine capacity is purposely reduced. Drilling machines also have much less mass per unit power than

Fig. 1.15 A milling machine tooling magazine

milling machines: there is less tendency for vibration and the axial thrust causes less distortion than the side thrusts that occur on a milling cutter. The prices of drilling machines are negligible compared with milling or turning. On the other hand, the low power availability implies a much lower material removal rate capacity. It is perhaps a saving grace of the drilling process that not much material is removed by it. This too is taken up in Section 1.4.

1.2 Manufacturing systems

The attack on non-productive cycle times described in the previous section has resulted in machine tools capable of higher productivity, but they are also more expensive. If they had been available in the late 1960s, they would have been totally uneconomic as the manufacturing organization was not in place to keep them occupied. The flow of work in progress was not effectively controlled, so that batches of components could remain in a factory totally idle for up to 95% of the time, and even the poorly productive machines that were then common were idle for up to 50% of the time (Figure 1.3). Manufacturing technology has, in fact, evolved hand in hand with manufacturing system organization, sometimes one pushing and the other pulling, sometimes vice versa.

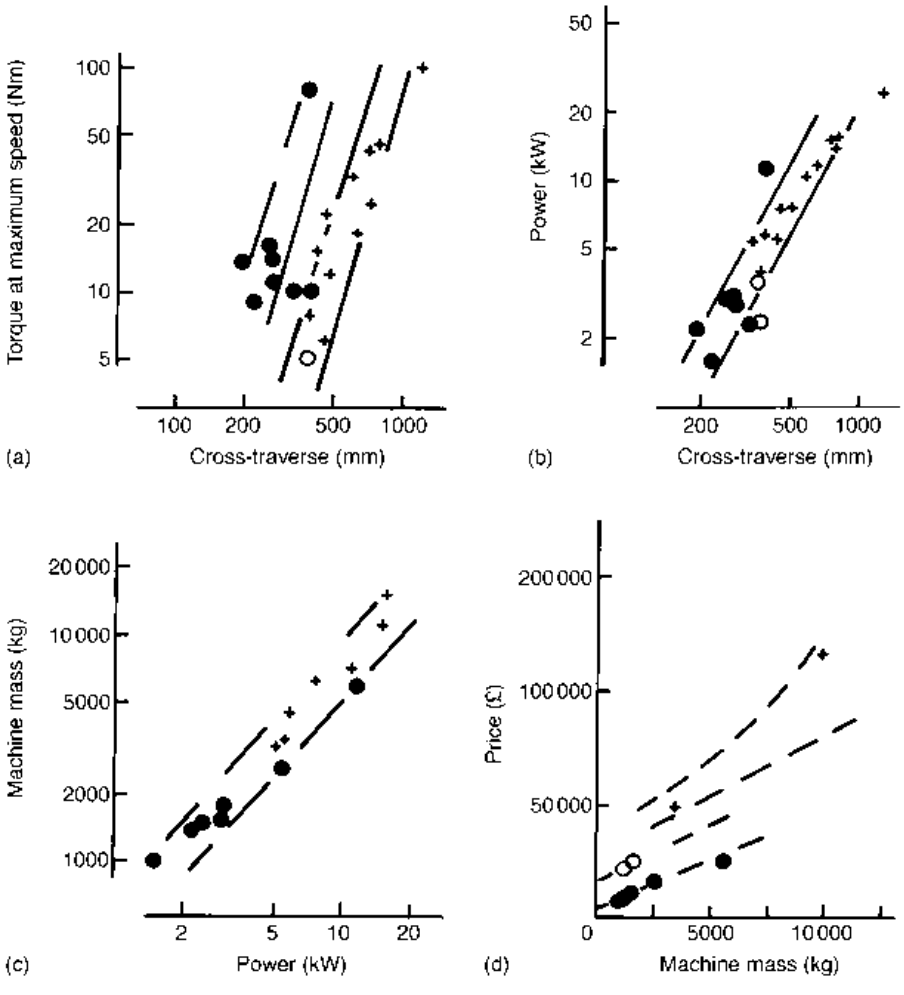


Fig. 1.16 (a) Torque and (b) power as a function of cross-traverse capacity and (c) mass/power and (d) price/mass relations, from manufacturers' catalogues, for mechanical (•) and basic CNC (o) milling machines and centres (+)

In the late 1960s there were two standard forms of organizing the machine tools in a machine shop. At one extreme, suitable for the dedicated production of one item in long runs – for example as might occur in converting sheet metal, steel bar, casting metal, paint and plastics parts into a car (Figure 1.17) – machine tools were laid out in flow lines or transfer lines. One machine tool followed another in the order in which operations were performed on the product. Such dedication allowed productivity to be gained at the price of flexibility. It was very costly to create the line and to change it to accommodate any change in manufacturing requirements.

At the other extreme, and by far the more common, no attempt was made to anticipate the order in which operations might be performed. Machine tools were laid out by type of process: all lathes in one area, all milling machines in another, all drills in another, and so on. In this so-called jobbing shop, or process oriented layout, different components were

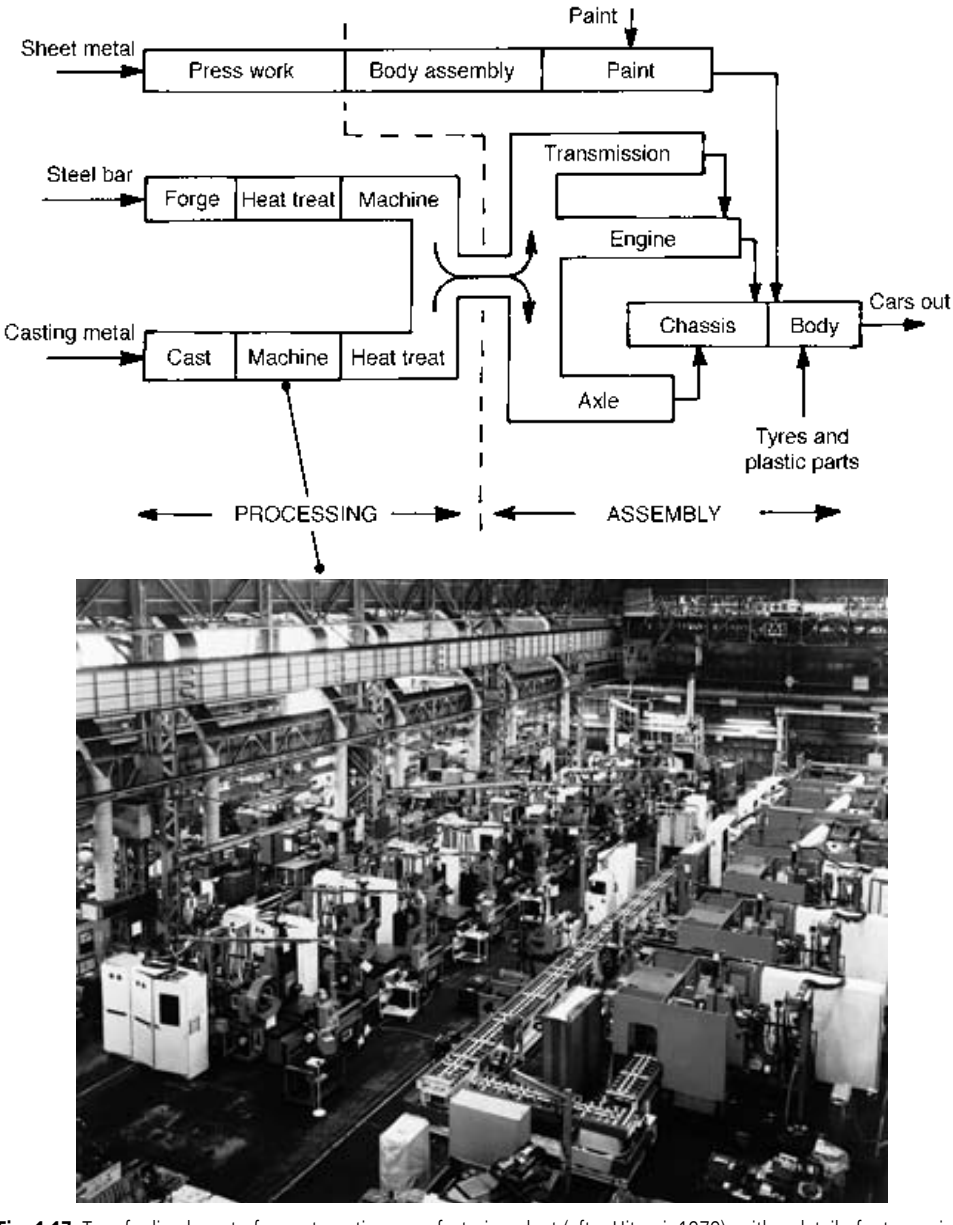


Fig. 1.17 Transfer line layout of an automotive manufacturing plant (after Hitomi, 1979), with a detail of a transmission case machining line

manufactured by carrying them from area to area as dictated by the ordering of their operations. It resulted in tortuous paths and huge amounts of materials handling – a part could travel several kilometres during its manufacture (Figure 1.18). It is to these circumstances that the survey results in Figure 1.3 apply.

It is now understood that there are intermediate layouts for manufacturing systems,

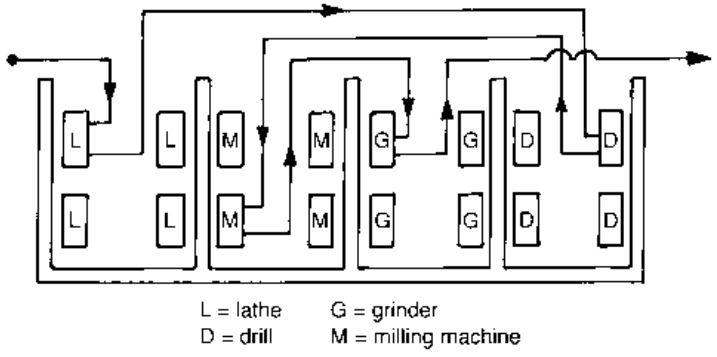


Fig. 1.18 Materials transfers in a jobbing shop environment (after Boothroyd and Knight, 1989)

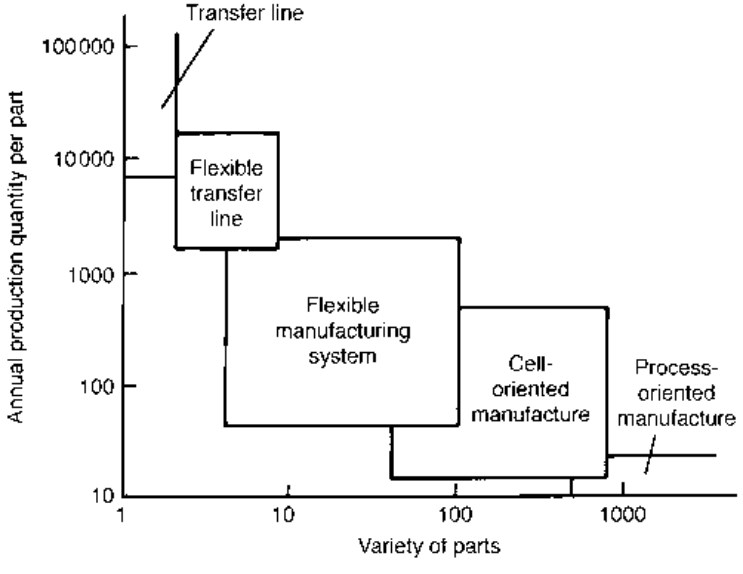


Fig. 1.19 The spectrum of manufacturing systems (after Groover and Zimmers, 1984)

appropriate for different mixes of part variety and quantity (Figure 1.19). If a manufacturer's spectrum of parts is of the order of thousands made in small batches, less than 10 to 20 or even one at a time, then planning improved materials handling strategies is probably not worthwhile. The large amounts of materials handling associated with job shop or process oriented manufacture cannot be avoided. Investment in highly productive machine tools is hard to justify. Such a manufacturer, for example a general engineering workshop tendering for sub-contract prototype work from larger companies, may still have some mechanically controlled machines, although the higher quality and accuracy attainable from CNC control will have forced investment in basic CNC machines. (As a matter of fact, the large jobbing shop is becoming obsolete. Its low productivity cannot support a large overhead, and smaller, perhaps family based, companies are emerging, offering specialist skills over a narrow manufacturing front.)

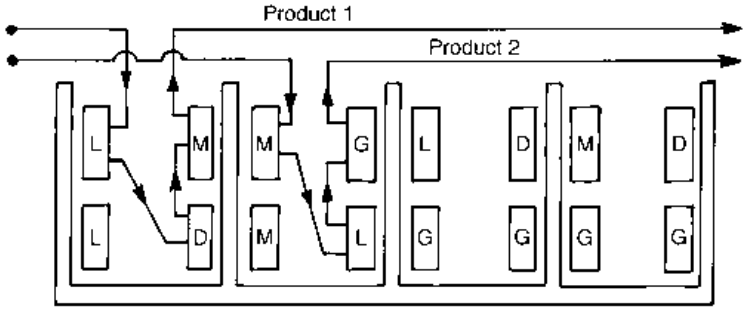


Fig. 1.20 Reduced materials flow through cell-oriented organizations and group technology (after Boothroyd and Knight, 1989)

If part variety reduces, perhaps to the order of hundreds, and batch size increases, again to the order of hundreds, it begins to pay to organize groups or cells of machine tools to reduce materials handling (Figure 1.20). The classification of parts to reduce, in effect, their variety from the manufacturing point of view is one aspect of the discipline of Group Technology. Almost certainly the machine tools in a cell will be CNC, and perhaps the programming of the machines will be from a central cell processor (direct numerical control or DNC). A low level of investment in turning or machining centre type tools may be justified, but it is unlikely that automatic materials handling outside the machine tools (robotics or automated guided vehicles – AGVs) will be justifiable. Cell-oriented manufacture is typically found in companies that own products that are components of larger assemblies, for example gear box, brakes or coupling manufacturers.

As part variety reduces further and batch size increases, say to tens and thousands respectively, the organization known as a flexible manufacturing system becomes justifiable. Heavy use can be justified of turning and/or machining centres and automatic handling between machine tools. Flexible manufacturing systems are typically found in companies manufacturing high value-added products, who are further up the supply chain than the component manufacturers for whom cell-oriented manufacture is the answer. Examples are manufacturers of ranges of robots, or the manufacturers of ranges of machine tools themselves (Figure 1.21). (Figure 1.19 also identifies a flexible transfer line layout – this could describe, for example, an automotive transfer line modified to cope with several variants of cars.)

The work in progress idle time (Figure 1.3) that has been the driver for the development of manufacturing systems practice has been reduced typically by half in circumstances suitable for cell-oriented manufacture and by a further half again in flexible manufacturing systems (Figure 1.5(b)), which is in balance with the increased capacity to remove metal of the machine tools themselves (Figure 1.5(a)).

1.3 Materials technology

The third element to be considered in parallel with machine technology and manufacturing organization, for its contribution to the evolution of machining practice, is the properties of the cutting edges themselves. There are three issues to be introduced: the material

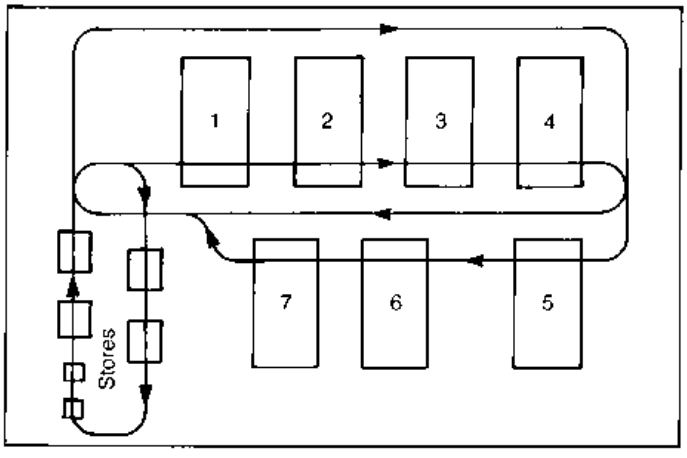


Fig. 1.21 Flexible manufacturing system layout

properties of these cutting edges that limit the material removal rates that can be achieved by them; how they are held in the machine tool, which determines how quickly they may be changed when they are worn out; and their price.

1.3.1 Cutting tool material properties

The main treatment of materials for cutting tools is presented in Chapter 3. As a summary, typical high temperature hardnesses of the main classes of cutting tool materials (high speed steels, cemented carbides and cermets, and alumina and silicon nitride ceramics; diamond and cubic boron nitride materials are introduced in Chapter 3) are shown in Figure 1.22. The temperatures that have been measured on tool rake faces during turning various work materials at a feed of 0.25 mm are shown in Figure 1.23. If the work material removal rate that can be achieved by a cutting tool is limited by the requirement that its hardness must be maintained above some critical level (to prevent it collapsing under the stresses caused by contact with the work), it is clear that carbide tools will be more productive than high speed steel tools; and ceramic tools may, in some circumstances, be more productive than carbides (for ceramics, toughness, not hardness, can limit their use). Also, copper alloys will be able to be machined more rapidly than ferrous alloys and than titanium alloys.

Tools do not last forever at cutting speeds less than those speeds that cause them to collapse. This is because they wear out, either by steady growth of wear flats or by the accumulation of cracks leading to fracture. Failure caused by fracture disrupts the machining process so suddenly that conditions are chosen to avoid this. Steady growth of wear eventually results in cutting edges having to be replaced in what could be described as preventative maintenance. It is an experimental observation that the relation between the lifetime T of a tool (the time that it can be used actively to machine metal) and the cutting speed V can be expressed as a power law: $VT^n = C$. It is common to plot experimental life/speed observations on a log-log basis, to create the so-called Taylor life curve. Figure 1.24 is a representative example of turning an engineering low alloy steel at a feed of

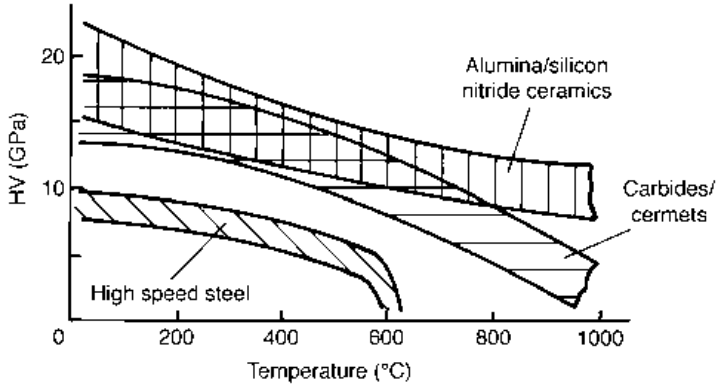


Fig. 1.22 The hardness of cutting tool materials as a function of temperature

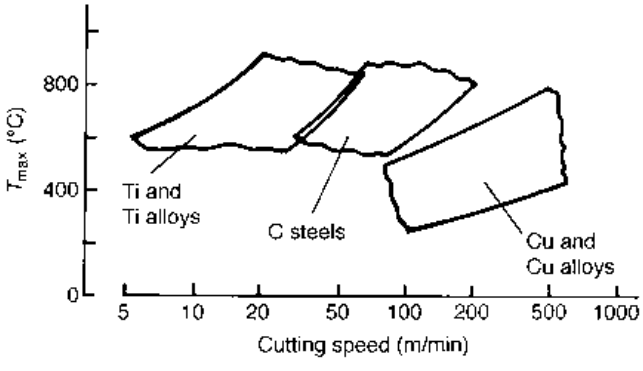


Fig. 1.23 Maximum tool face temperatures generated during turning some titanium, ferrous and copper alloys at a feed of 0.25 mm (after Trent, 1991)

0.25 mm with high speed steel, a cemented carbide and an alumina ceramic tool (the data for the ceramic tool show a fracture (chipping) range). Over the straight line regions (on a log-log basis), and with T in minutes and V in m/min

for high speed steel	$VT^{0.15} = 30$	(1.3a)
for cemented carbide	$VT^{0.25} = 150$	(1.3b)
for alumina ceramic	$VT^{0.45} = 500$	(1.3c)

These representative values will be used in the economic considerations of machining in Section 1.4. A more detailed consideration of life laws is presented in Chapter 4. The constants n and C in the life laws typically vary with feed as well as cutting speed; they also depend on the end of life criterion, reducing as the amount of wear that is regarded as allowable reduces. At the level of this introductory chapter treatment, it is not straightforward to discuss how the constants in equations (1.3) may differ between turning, milling and drilling practice. It will be assumed that they are not influenced by the machining process. Any important consequences of this assumption will be pointed out where relevant.

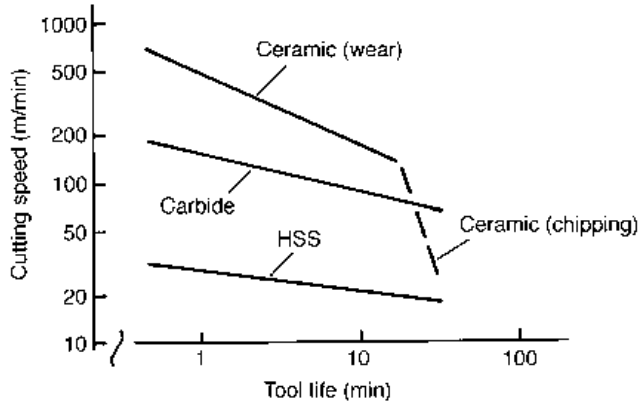


Fig. 1.24 Representative Taylor tool life curves for turning a low alloy steel

1.3.2 Cutting tool costs

Apart from tool lifetime, the replacement cost of a worn tool (consumable cost) and the time to replace a worn-out tool are important in machining economics. Machining economics will be considered in Section 1.4. Some different forms of cutting tool have already been illustrated in Figure 1.12. High speed steel (HSS) tools were traditionally ground from solid blocks. Some cemented carbide tools are also ground from solid, but the cost of cemented carbide often makes inserts brazed to tool steel a cheaper alternative. Most recently, disposable, indexable, insert tooling has been introduced, replacing the cost and time of brazing by the cheaper and quicker mechanical fixing of a cutting edge in a holder. Disposable inserts are the only form in which ceramic tools are used, are the dominant form for cemented carbides and are also becoming more common for high speed steel tools. Typical costs associated with different sizes of these tools, in forms used for turning, milling and drilling, are listed in Table 1.1.

There are three sorts of information in Table 1.1. The second column gives purchase prices. It is the third column, of more importance to the economics of machining, that gives the tool consumable costs. A tool may be reconditioned several times before it is thrown away. The consumable cost C_t is the initial price of the tool, plus all the reconditioning costs, divided by the number of times it is reconditioned. It is less than the purchase price (if it were more, reconditioning would be pointless). For example, if a solid or brazed tool can be reground ten times during its life, the consumable cost is one tenth the purchase price plus the cost of regrinding. If an indexable turning insert has four cutting edges (for example, if it is a square insert), the consumable cost is one quarter the purchase price plus the cost of resetting the insert in its holder (assumed to be done with the holder removed from the machine tool). If a milling tool is of the insert type, say with ten inserts in a holder, its consumable cost will be ten times that of a single insert.

In Table 1.1, a range of assumptions have been made in estimating the consumable costs: that the turning inserts have four usable edges and take 2 min at £12.00/hour to place in a holder; that the HSS milling cutters can be reground five times and cost £5 to £10 per regrind; that the solid carbide milling cutters can also be reground five times but the brazed carbides only three times, and that grinding cost varies from £10 to £20 with

Table 1.1 Typical purchase price, consumable cost and change time for a range of cutting tools (prices from UK catalogues, *circa* 1990, excluding discounts and taxes)

Tool type and size, dimensions in mm.		Typical purchase price, £.	Tool consumable cost C_t, £.	Tool change time t_{ct} , min.
Turning				
solid HSS, 6 x 8 x 100		≈ 6	0.50	Time depends on machine
Brazed carbide		–	2.00	tool: for example 5 min.
carbide insert, plain				for solid tooling on
12 x 12 x 4		2.50–5.00	1.00–1.60	mechanical or simple CNC
25 x 25 x 7		7.50–10.50	2.30–3.00	lathe; 2 min for insert tooling
carbide insert, coated				on simple CNC lathe; 1 min
12 x 12 x 4		3.00–6.00	1.10–1.90	for insert tooling on turning
25 x 25 x 7		9.00–11.20	2.65–3.20	centre
ceramic insert, plain				
12 x 12 x 4		4.50–9.00	1.50–2.70	
25 x 25 x 7		13.50–17.00	3.80–4.65	
cubic boron nitride		50–60	–	
polycrystalline diamond		60–70	–	
Milling				
solid HSS	∅6	9–14	7–8	Machine dependent, for
	∅25	30–60	13–20	example 10 min for
	∅100	100–250	30–60	mechanical machine; 5 min
solid carbide	∅6	18–33	14–17	for simple CNC mill; 2 min
	∅12	40–80	23–31	for machining centre
	∅25	200–400	60–100	
brazed carbide	∅12	≈ 50	≈ 27	
	∅25	≈ 75	≈ 40	
	∅50	≈ 150	≈ 70	
carbide inserts, ∅ > 50 plain, per insert		as turning price	as turning, per insert	
Drilling				–
solid HSS	∅3	≈ 1 – 3	≈ 1.00	
	∅6	≈ 1.5 – 5	≈ 1.25	
	∅12	≈ 3 – 8	≈ 1.50	
solid carbide	∅3	≈ 7	≈ 3.00	
	∅6	≈ 15	≈ 3.75	
	∅12	≈ 60	≈ 4.50	

cutter diameter; and that drilling is similar to milling with respect to regrind conditions. There is clearly great scope for these costs to vary. The interested reader could, by the methods of Section 1.4, test how strongly these assumptions influence the costs of machining.

To extend the range of Table 1.1, some data are also given for the price and consumable costs of coated carbide, cubic boron nitride (CBN) and polycrystalline diamond (PCD) inserts. Coated carbides (carbides with thin coatings, usually of titanium nitride, titanium carbide or alumina) are widely used to increase tool wear resistance particularly in finishing operations; CBN and PCD tools have special roles for machining hardened steels (CBN) and high speed machining of aluminium alloys (PCD), but will not be considered further in this chapter.

Finally, Table 1.1 also lists typical times to replace and set tool holders in the machine tool. This tool change time is associated with non-productive time (Figure 1.3) for most machine tools but, for machining centres fitted with tool magazines, tool replacement in the magazine can be carried out while the machine is removing metal. For such centres,

non-productive tool change time, associated with exchanging the tool between the magazine and the main drive spindle, can be as low as 3 s to 10 s. Care must be taken to interpret appropriately the replacement times in Table 1.1.

1.4 Economic optimization of machining

The influences of machine tool technology, manufacturing systems management and materials technology on the cost of machining can now be considered. The purpose is not to develop detailed recommendations for best practice but to show how these three factors have interacted to create a flow of improvement from the 1970s to the present day, and to look forward to the future. In order to discuss absolute costs and times as well as trends, the machining from tube stock of the flanged shaft shown in Figure 1.6 will be taken as an example. Dimensions are given in Figure 1.25. The part is created by turning the external diameter, milling the keyway, and drilling four holes. The turning operation will be considered first.

1.4.1 Turning process manufacturing times

The total time, t_{total}, to machine a part by turning has three contributions: the time t_{load} taken to load and unload the part to and from a machine tool; the time t_{active} in the machine tool; and a contribution to the time taken to change the turning tool when its edge is worn out. t_{active} is longer than the actual machining time t_{mach} because the tool spends some time moving and being positioned between cuts. t_{active} may be written t_{mach}/f_{mach}, where f_{mach} is the fraction of the time spent in removing metal. If machining N parts results in the tool edge being worn out, the tool change time t_{ct} allocated to machining one part is t_{ct}/N. Thus

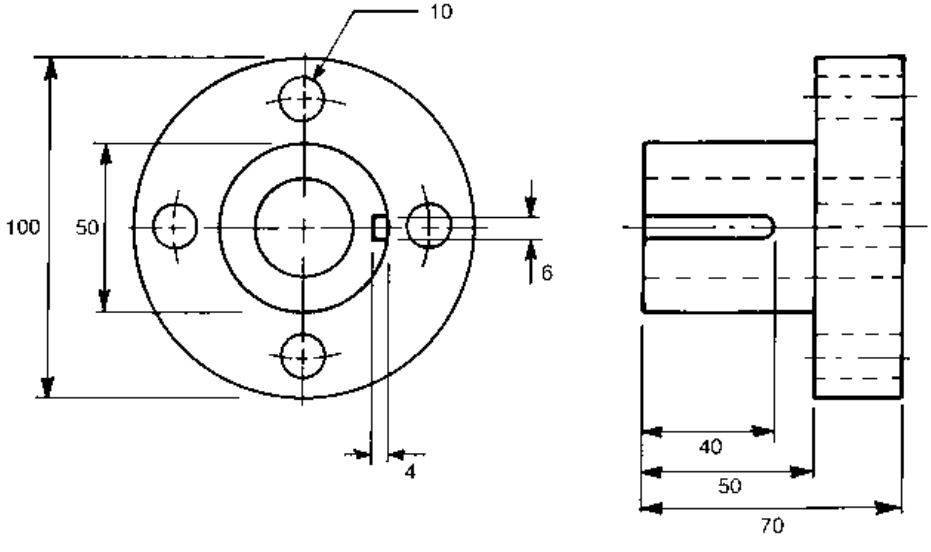


Fig. 1.25 An example machined component (dimensions in mm)

$$t_{total} = t_{load} + \frac{t_{mach}}{f_{mach}} + \frac{t_{ct}}{N} \tag{1.4}$$

It is easy to show that as the cutting speed of a process is increased, t_{total} passes through a minimum value. This is because, although the machining time decreases as speed increases, tools wear out faster and N also decreases. Suppose the volume of material to be removed by turning is written V_{vol}, then

$$t_{mach} = \frac{V_{vol}}{fdV} \tag{1.5}$$

The machining time for N parts is N times this. If the time for N parts is equated to the tool life time T in equation (1.3) (generalized to $VT^n = C$), N may be written in terms of n and C, f, d, V_{vol} and V, as

$$N = \frac{fdC^{1/n}}{V_{vol}V^{(1-n)/n}} \tag{1.6}$$

Substituting equations (1.5) and (1.6) into equation (1.4):

$$t_{total} = t_{load} + \frac{1}{f_{mach}} \frac{V_{vol}}{fdV} + \frac{V_{vol}V^{(1-n)/n}}{fdC^{1/n}} t_{ct} \tag{1.7}$$

Equation (1.7) has been applied to the part in Figure 1.25, as an example, to show how the time to reduce the diameter of the tube stock from 100 mm to 50 mm, over the length of 50 mm, depends on both what tool material (the influence of n and C) and how advanced a machine technology is being used (the influence of f_{mach} and t_{ct}). In this example, V_{vol} is 2.95×10^5 mm^3. It is supposed that turning is carried out at a feed and depth of cut of 0.25 mm and 4 mm respectively, and that t_{load} is 1 min (an appropriate value for a component of this size, according to Boothroyd and Knight, 1989). Times have been estimated for high speed steel, cemented carbide and an alumina ceramic tool material, in solid, brazed or insert form, used in mechanical or simple CNC lathes or in machining centres. n and C values have been taken from equation (1.3). The f_{mach} and t_{ct} values are listed in Table 1.2. The variation of f_{mach} with machine tool development has been based on active non-productive time changes shown in Figure 1.5(a). t_{ct} values for solid or brazed and insert cutting tools have been taken from Table 1.1. Results are shown in Figure 1.26.

Figure 1.26 shows the major influence of tool material on minimum manufacturing

Table 1.2 Values of f_{mach} and t_{ct}, min, depending on manufacturing technology

Tool form	Machine tool development		
	Mechanical	Simple CNC	Turning centre
Solid or brazed	$f_{mach} = 0.45$; $t_{ct} = 5$	$f_{mach} = 0.65$; $t_{ct} = 5$	
Insert		$f_{mach} = 0.65$; $t_{ct} = 2$	$f_{mach} = 0.85$; $t_{ct} = 1$

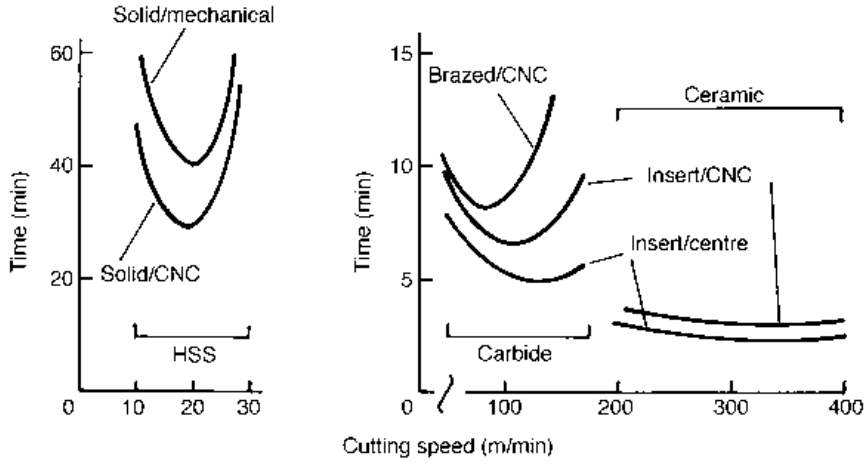


Fig. 1.26 The influence on manufacturing time of cutting speed, tool material (high speed steel/carbide/ceramic) and manufacturing technology (solid/brazed/insert tooling in a mechanical/simple CNC/turning centre machine tool) for turning the part in Figure 1.25

time: from around 30 min to 40 min for high speed steel, to 5 min to 8 min for cemented carbide, to around 3 min for alumina ceramic. The time saving comes from the higher cutting speeds allowed by each improvement of tool material, from 20 m/min for high speed steel, to around 100 m/min for carbide, to around 300 m/min for the ceramic tooling.

For each tool material, the more advanced the manufacturing technology, the shorter the time. Changing from mechanical to CNC control reduces minimum time for the high speed steel tool case from 40 min to 30 min. Changing from brazed to insert carbide with a simple CNC machine tool reduces minimum time from 8 min to 6.5 min, while using insert tooling in a machining centre reduces the time to 5 min. Only for the ceramic tooling are the times relatively insensitive to technology: this is because, in this example, machining times are so small that the assumed work load/unload time is starting to dominate.

It is always necessary to check whether the machine tool on which it is planned to make a part is powerful enough to achieve the desired cutting speed at the planned feed and depth of cut. Table 1.3 gives typical specific cutting forces for machining a range of materials. For the present engineering steel example, an appropriate value might be 2.5 GPa. Then, from equation 1.2(b), for $fd = 1$ mm^2, a power of 1 kW is needed at a cutting speed of 25 m/min (for HSS), 5 kW is needed at 120 m/min (for cemented carbide) and 15 kW

Table 1.3 Typical specific cutting force for a range of engineering materials

Material	F_c^*, GPa	Material	F_c^*, GPa
Aluminium alloys	0.5–1.0	Carbon steels	2.0–3.0
Copper alloys	1.0–2.0	Alloy steels	2.0–5.0
Cast irons	1.5–3.0		

is needed around 400 m/min (for ceramic tooling). These values are in line with supplied machine tool powers for the 100 mm diameter workpiece (Figure 1.8).

1.4.2 Turning process costs

Even if machining time is reduced by advanced manufacturing technology, the cost may not be reduced: advanced technology is expensive. The cost of manufacture C_p is made up of two parts: the time cost of using the machine tool and the cost C_t of consuming cutting edges. The time cost itself comprises two parts: the charge rate M_t to recover the purchase cost of the machine tool and the labour charge rate M_w for operating it. To continue the turning example of the previous section:

$$C_p = (M_t + M_w)t_{total} + \frac{V_{vol}V^{(1-n)/n}}{fdC^{1/n}}\ C_t \tag{1.8}$$

The machine charge rate

M_t is the rate that must be charged to recover the total capital cost C_m of investing in the machine tool, over some number of years Y. There are many ways of estimating it (Dieter, 1991). One simple way, leading to equation (1.9), recognizes that, in addition to the initial purchase price C_i, there is an annual cost of lost opportunity from not lending C_i to someone else, or of paying the interest on C_i if it has been borrowed. This may be expressed as a fraction f_i of the purchase price. f_i typically rises as the inflation rate of an economy increases. There is also an annual maintenance cost and the cost of power, lighting, heating, etc associated with using the machine, that may also be expressed as a fraction, f_m, of the purchase price. Thus

$$C_m = C_i\,(1 + [f_i + f_m]Y) \tag{1.9}$$

Earnings to set against the cost come from manufacturing parts. If the machine is active for a fraction f_o of n_s 8-hour shifts a day (n_s = 1, 2 or 3), 250 days a year, the cost rate M_t for earnings to equal costs is, in cost per min

$$M_t = \frac{C_i}{120\,000 f_o n_s}\left[\frac{1}{Y} + (f_i + f_m)\right] \tag{1.10}$$

Values of f_o and n_s are likely to vary with the manufacturing organization (Figure 1.19). It is supposed that process and cell oriented manufacture will usually operate two shifts a day, whereas a flexible manufacturing system (FMS) may operate three shifts a day, and that f_o varies in a way to be expected from Figure 1.5(b). Table 1.4 estimates, from equation (1.10), a range of machine cost rates, assuming $Y = 5$, $f_i = 0.15$ and $f_m = 0.2$. Initial costs C_i come from Figure 1.9, for the machine powers indicated and which have been shown to be appropriate in the previous section. In the case of the machining centres, a capacity to mill and drill has been assumed, anticipating the need for that later. Some elements of the table have no entry. It would be stupid to consider a mechanically controlled lathe as part of an FMS, or a turning centre in a process oriented environment. Some elements have been filled out to enable the cost of unfavourable circumstances to be estimated: for example, a turning centre operated at a cell-oriented efficiency level.

Table 1.4 Cost rates, M_t, £/min, for turning machines for a range of circumstances

Machine type		C_i, £	Manufacturing system			
			Process-oriented	Cell-oriented	FMS	
			$f_o = 0.5$ $n_s = 2$	$f_o = 0.75$; $n_s = 2$	$f_o = 0.85$; $n_s = 2$	$n_s = 3$
Mechanical	1 kW	6000	0.028			
Simple CNC	1 kW	20000	0.092	0.060		
	5 kW	28000	0.13	0.086		
	15 kW	50000	0.23	0.15		
Turning centre	5 kW	60000		0.18	0.16	0.11
	15 kW	120000		0.37	0.33	0.22

The labour charge rate

M_w is more than the machine operator's wage rate or salary. It includes social costs such as insurance and pension costs as a fraction f_s of wages. Furthermore, a company must pay all its staff, not only its machine operators. M_w should be inflated by the ratio, r_w, of the total wages bill to that of the wages of all the machine operator (productive) staff. If a worker's annual wage is C_a, and an 8-hour day is worked, 220 days a year, the labour cost per minute is

$$M_w = \frac{C_a}{110\,000}(1 + f_s)r_w \tag{1.11}$$

Table 1.5 gives some values for C_a = £15 000/year, typical of a developed economy country, and $f_s = 0.25$. r_w varies with the level of automation in a company. Historically, for a labour intensive manufacturing company, it may be as low as 1.2, but for highly automated manufacturers, such as those who operate transfer and FMS manufacturing systems, it has risen to 2.0.

Example machining costs

Equation (1.8) is now applied to estimating the machining costs associated with the times of Figure 1.26, under a range of manufacturing organization assumptions that lead to different cost rates, as just discussed. These are summarized in Table 1.6. Machine tools have been selected of sufficient power for the type of tool material they use. M_t values have been extracted from Table 1.4, depending on the machine cost and the types of manufacturing organization of the examples. M_w values come from Table 1.5. Tool consumable costs are taken from Table 1.1. Two-shift operation has been assumed unless otherwise indicated. Results are shown in Figure 1.27.

Table 1.5 Range of labour rates, £/min, in high wage manufacturing industry

Manufacturing organization		
Labour intensive	Intermediate	Highly automated
0.20	0.27	0.34

Table 1.6 Assumptions used to create Figures 1.26 and 1.27. * indicates three shifts

	Time influencing variables					
	Cutting tool	Machine tool/ power, kW	Manufacturing organization	M_t, [£/min]	M_w, [£/min]	C_t, [£]
a	solid HSS	mechanical/1	process oriented	0.028	0.20	0.50
b	solid HSS	basic CNC/1	cell-oriented	0.060	0.27	0.50
c	brazed carbide	basic CNC/5	cell-oriented	0.086	0.27	2.00
d	insert carbide	basic CNC/5	cell-oriented	0.086	0.27	1.50
e	insert carbide	centre CNC/5	FMS	0.165	0.34	1.50
e*	insert carbide	centre CNC/5	FMS*	0.110	0.34	1.50
f	insert ceramic	basic CNC/15	cell-oriented	0.15	0.27	2.50
g*	insert ceramic	centre CNC/15	FMS*	0.22	0.34	2.50

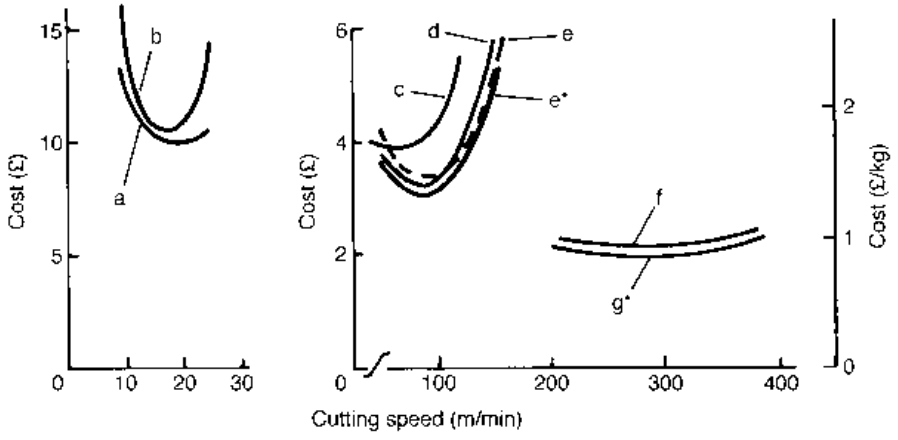


Fig. 1.27 Costs associated with the examples of Figure 1.26 , a–g as in Table 1.6

Figure 1.27 shows that, as with time, minimum costs reduce as tool type changes from high speed steel to carbide to ceramic. However, the cost is only halved in changing from high speed steel to ceramic tooling, although the time is reduced about 10-fold. This is because of the increasing costs of the machine tools required to work at the increasing speeds appropriate to the changed tool materials.

The costs associated with the cemented carbide insert tooling, curves d, e and e* are particularly illuminating. In this case, it is marginally more expensive to produce parts on a turning centre working at FMS efficiency than on a simple (basic) CNC machine working at a cell-oriented level of efficiency, at least if the FMS organization is used only two shifts per day (comparing curves d and e). To justify the FMS investment requires three shift per day (curve e*).

To the right-hand side of Figure 1.27 has been added a scale of machining cost per kg of metal removed, for the carbide and ceramic tools. The low alloy steel of this example probably costs around £0.8/kg to purchase. Machining costs are large compared with materials costs. When it is planned to remove a large proportion of material by machining, paying more for the material in exchange for better machinability (less tool wear) can often be justified.

Up to this point, only a single machining operation – turning – has been considered. In most cases, including the example of Figure 1.25 on which the present discussion is based, multiple operations are carried out. It is only then, as will now be considered, that the organizational gains of cell-oriented and FMS organization bring real benefit.

1.4.3 Milling and drilling times and costs

Equations (1.7) and (1.8) for machining time and cost of a turning operation can be applied to milling if two modifications are made. A milling cutter differs from a turning tool in that it has more than one cutting edge, and each removes metal only intermittently. More than one cutting edge results in each doing less work relative to a turning tool in removing a given volume of metal. The intermittent contact results in a longer time to remove a given volume for the same tool loading as in turning. Suppose that a milling cutter has n_c cutting edges but each is in contact with the work for only a fraction α of the time (for example $\alpha = 0.5$ for the 180° contact involved in end milling the keyway in the example of Figure 1.25). The tool change time term of equation (1.7) will change inversely as n_c, other things being equal. The metal removal time will change inversely as (αn_c):

$$t_{total} = t_{load} + \frac{1}{f_{mach}} \left(\frac{V_{vol}}{\alpha n_c f d V} + \frac{V_{vol} V^{(1-n)/n}}{n_c f d C^{1/n}} \right) t_{ct} \tag{1.12}$$

Cost will be influenced indirectly through the changed total time and also by the same modification to the tool consumable cost term as to the tool change time term:

$$C_p = (M_t + M_w) t_{total} + \left(\frac{V_{vol} V^{(1-n)/n}}{n_c f d C^{1/n}} \right) C_t \tag{1.13}$$

For a given specific cutting force, the size of the average cutting force is proportional to the group $[\alpha n_c f d]$. Suppose the machining times and costs in milling are compared with those in turning on the basis of the same average cutting force for each – that is to say, for the same material removal rate – first of all, for machining the keyway in the example of Figure 1.25; and then suppose the major turning operations considered in Figures 1.26 and 1.27 were to be replaced by milling.

In each case, suppose the milling operation is carried out by a four-fluted solid carbide end mill ($n_c = 4$) of 6 mm diameter, at a level of organization typical of cell-oriented manufacture: the appropriate turning time and cost comparison is then shown by results 'brazed/CNC' in Figure 1.26 and 'c' in Figure 1.27.

For the keyway example, $\alpha = 0.5$ and thus for $[\alpha n_c f d]$ to be unchanged, f must be reduced from 0.25 mm to 0.125 mm (assuming d remains equal to 4 mm). Then the tool life coefficient C (the cutting speed for 1 min tool life) is likely to be increased from its value of 150 m/min for $f = 0.25$ mm. Suppose it increases to 180 m/min. Suppose that for the turning replacement operation, the end mill contacts the work over one quarter of its circumference, so $\alpha = 0.25$. Then f remains equal to 0.25 mm for the average cutting force to remain as in the turning case, and C is unchanged. Table 1.7 lists the values of the various coefficients that determine times and costs for the two cases. Their values come from the previous figures and tables – Figure 1.16 (milling machine costs), Table 1.1 (cutting tool data) and equations (1.10) and (1.11) for cost rates.

Table 1.7 Assumptions for milling time and cost calculation examples

Quantity	Keyway operation, $[\alpha n_c fd] = 1$ mm^2	Replacement turning operation (i), $[\alpha n_c fd] = 1$ mm^2	Replacement turning operation (ii), $[\alpha n_c fd] = 0.5$ mm^2
V_{vol} [mm^3]	960	295000	295000
f_{mach}	0.65	0.65	0.65
$n_c fd$ [mm^2]	2	4	2
C [m/min]	180	150	180
n	0.25	0.25	0.25
t_{ct} [min]	5	5	5
t_{load} [min]	2	2	2
M_m [£/min]	0.092	0.092	0.092
M_w [£/min]	0.27	0.27	0.27
C_t [£]	15	15	15

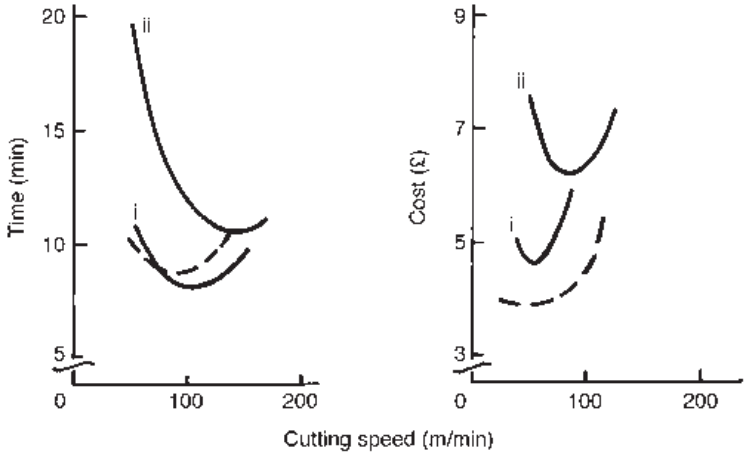


Fig. 1.28 Times and costs to remove metal by milling, for the conditions i and ii of Table 1.7 compared with removing the same metal by turning (- - -)

If milling were carried out at the same average force level as turning, peak forces would exceed turning forces. For this reason, it is usual to reduce the average force level in milling. Table 1.7 also lists (in its last column) coefficients assumed in the calculation of times and costs for the turning replacement operation with average force reduced to half the value in turning.

Application of equations (1.12) and (1.13) simply shows that for such a small volume of material removal as is represented by the keyway, time and cost is dominated by the work loading and unloading time. Of the total time of 2.03 min, calculated near minimum time conditions, only 0.03 min is machining time. At a cost of £0.36/min, this translates to only £0.011. Although it is a small absolute amount, it is the equivalent of £1.53/kg of material removed. This is similar to the cost per weight rate for carbide tools in turning (Figure 1.27).

In the case of the replacement turning operation, Figure 1.28 compares the two sets of data that result from the two average force assumptions with the results for turning with

a brazed carbide tool. When milling at the same average force level as in turning (curves 'i'), the minimum production time is less than in turning, but the mimimum cost is greater. This is because fewer tool changes are needed (minimum time), but these fewer changes cost more: the milling tool consumable cost is much greater than that of a turning tool. However, if the average milling force is reduced to keep the peak force in bounds, both the minimum time and minimum cost are significantly increased (curves 'ii'). The intermittent nature of milling commonly makes it inherently less productive and more costly than turning.

The drilling process is intermediate between turning and milling, in so far as it involves more than one cutting edge, but each edge is continuously removing metal. Equations (1.12) and (1.13) can be used with $\alpha = 1$. For the example concerned, the time and cost of removing material by drilling is negligible. It is the loading and unloading time and cost that dominates. It is for manufacturing parts such as the flanged shaft of Figure 1.25 that turning centres come into their own. There is no additional set-up time for the drilling operation (nor for the keyway milling operation).

1.5 A forward look

The previous four sections have attempted briefly to capture some of the main strands of technology, management, materials and economic factors that are driving forward metal machining practice and setting challenges for further developments. Any reader who has prior knowledge of the subject will recognize that many liberties have been taken. In the area of machining practice, no distinction has been made between rough and finish cutting. Only passing acknowledgement has been made to the fact that tool life varies with more than cutting speed. All discussion has been in terms of engineering steel workpieces, while other classes of materials such as nickel-based, titanium-based and abrasive silicon-aluminium alloys, have different machining characteristics. These and more will be considered in later chapters of this book.

Nevertheless, some clear conclusions can be drawn that guide development of machining practice. The selection of optimum cutting conditions, whether they be for minimum production time, or minimum cost, or indeed for combinations of these two, is always a balance between savings from reducing the active cutting time and losses from wearing out tools more quickly as the active time reduces. However, the active cutting time is not the only time involved in machining. The amounts of the savings and losses, and hence the conditions in which they are balanced, do not depend only on the cutting tools but on the machine tool technology and manufacturing system organization as well.

As far as the turning of engineering structural steels is concerned, there seems at the moment to be a good balance between materials and manufacturing technology, manufacturing organization and market needs, although steel companies are particularly concerned to develop the metallurgy of their materials to make them easier to machine without compromising their required end-use properties. The main activities in turning development are consequently directed to increasing productivity (cutting speed) for difficult to machine materials: nickel alloys, austenitic stainless steels and titanium alloys used in aerospace applications, which cause high tool temperatures at relatively low cutting speeds (Figure 1.23); and to hardened steels where machining is trying to

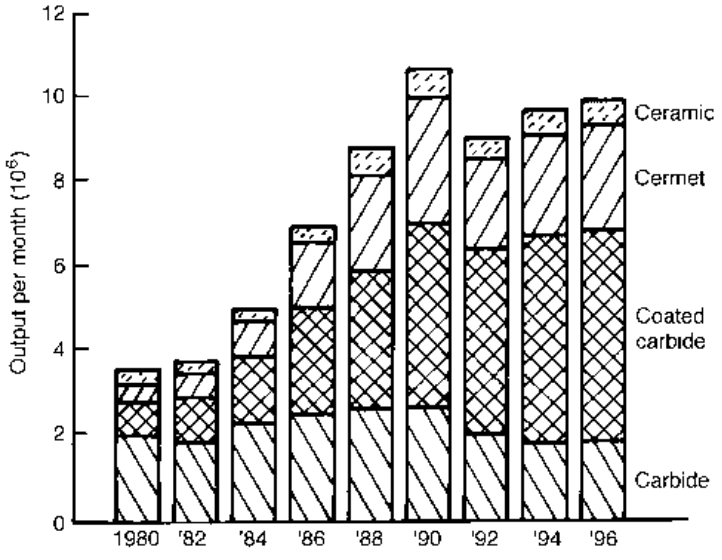


Fig. 1.29 Sales of insert cutting tips in Japan, 1980 to 1996

compete with grinding processes. Attention is also being paid to environmental issues: how to machine without coolants, which are expensive to dispose of to water treatment plant.

Developments in milling have a different emphasis from turning. As has been seen, the intermittent nature of the milling process makes it inherently more expensive than turning. A strategy to reduce the force variations in milling, without increasing the average force, is to increase the number of cutting edges in contact while reducing the feed per edge. Thus, the milling process is often carried out at much smaller feeds per edge – say 0.05 to 0.2 mm – than is the turning process. This results in a greater overall cutting distance in removing a unit volume of metal and hence a greater amount of wear, other things being equal. At the same time, the intermittent nature of cutting edge contact in milling increases the rate of mechanical and thermal fatigue damage relative to turning. The two needs of cutting tools for milling, higher fatigue resistance and higher wear resistance than for similar removal rates in turning, are to some extent incompatible. At the same time, a productivity push exists to achieve as high removal rates in milling as in turning. All this leads to greater activity in milling development at the present time than in turning development.

Perhaps the biggest single and continuing development of the last 20 years has been the application of Surface Engineering to cutting tools. In the early 1980s it was confidently expected that the market share for newly developed ceramic indexable insert cutting tools (for example the alumina tools considered in Section 1.4) would grow steadily, held back only by the rate of investment in the new, more powerful and stiffer machine tools needed for their potential to be realized. Instead, it is a growth in ceramic (titanium nitride, titanium carbide and alumina) coated cutting tools that has occurred. Figure 1.29 shows this. It is always risky being too specific about what will happen in the future.

References

Ashby, M. F. (1992) *Materials Selection in Mechanical Design*. Oxford: Pergamon Press.
Boothroyd, G. and Knight, W. A. (1989) *Fundamentals of Machining and Machine Tools*, 2nd edn. New York: Dekker.
Dieter, G. E. (1991) *Engineering Design*, 2nd edn. New York: McGraw-Hill.
Groover, M. P. and Zimmers, E. W. (1984) CAD/CAM. New York: Prentice Hall.
Hitomi, K. (1979) *Manufacturing Systems Engineering*. London: Taylor & Francis.
Trent, E. M. (1991) *Metal Cutting*, 3rd edn. Oxford: Butterworth-Heinemann.

2

Chip formation fundamentals

Chapter 1 focused on the manufacturing organization and machine tools that surround the machining process. This chapter introduces the mechanical, thermal and tribological (friction, lubrication and wear) analyses on which understanding the process is based.

2.1 Historical introduction

Over 100 years ago, Tresca (1878) published a visio-plasticity picture of a metal cutting process (Figure 2.1(a)). He gave an opinion that for the construction of the best form of tools and for determining the most suitable depth of cut (we would now say undeformed chip thickness), the minute examination of the cuttings is of the greatest importance. He was aware that fine cuts caused more plastic deformation than heavier cuts and said this was a driving force for the development of more powerful, stiffer machine tools, able to make heavier cuts. At the same meeting, it was recorded that there now appeared to be a mechanical analysis that might soon be used – like chemical analysis – systematically to assess the quality of formed metals (in the context of machining, this was premature!).

Three years later, Lord Rayleigh presented to the Royal Society of London a paper by Mallock (Mallock, 1881–82). It recorded the appearance of etched sections of ferrous and non-ferrous chips observed through a microscope at about five times magnification (Figure

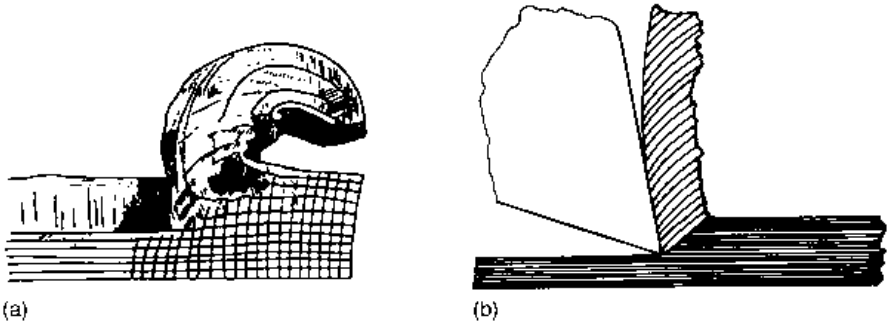


Fig. 2.1 Early chip observations by (a) Tresca (1878) and (b) Mallock (1881–82)

2.1(b)). Mallock was clear that chip formation occurred by shearing the metal. He argued that friction between the chip and tool was of great importance in determining the deformation in the chip. He commented that lubricants acted by reducing the friction between the chip and the tool and wrote that the difficulty is to see how the lubricant gets there. He also wrote down equations for the amount of work done in internal shear and by friction between the chip and tool. Surprisingly, he seemed unaware of Tresca's work on plasticity and thought that a metal's shear resistance was directly proportional to the normal stress acting on the shear plane. As a result, his equations gave wrong answers. This led him to discount an idea of his that chips might form at a thickness that minimized the work of friction. With hindsight, he was very close to Merchant's law of chip formation, which in fact had to wait another 60 years for its formulation (Section 2.2.4).

Tresca's and Mallock's papers introduce two of the main elements of metal cutting theory, namely plasticity and the importance of the friction interaction between chip and tool. Tresca was also very clear about the third element, the theory of plastic heating, but his interest in this respect was taken by reheating in hot forging, rather than by machining. In his 1878 paper, he describes tests that show up to 94% conversion of work to heat in a forging, and explicitly links his discussion to the work of Joule.

In machining, the importance of heating for tool life was being tackled practically by metallurgists. A series of developments from the late 1860s to the early 1900s saw the introduction of new steel alloy tools, with improved high temperature hardness, that allowed higher and higher cutting speeds with correspondingly greater productivities. A classic paper (Taylor, 1907) describes the early work, from 1881 onwards, on productivity optimization through improved tool materials (high speed steels) and their best use.

Thus, the foundations of machining theory and practice were laid between around 1870 and 1905. At this stage, with the minor exception of Mallock's work, the emphasis was on observing rather than predicting behaviour. This remained the case for the next 30 years, with huge collections of machinability (force and tool life) data (for example, Boston, 1926; Herbert, 1928), and of course the introduction of even more heat resistant cemented carbide tools. By the late 1920s, there was so much data that the need for unifying theories was beginning to be felt. Herbert quotes Boston (1926) as writing: 'If possible, a theory of metal cutting which underlies all types of cutting should be developed. . . . All this is a tremendous problem and should be undertaken in a big way.'

The first predictive stage of metal cutting studies started about the late 1930s–mid-1940s. The overriding needs of the Second World War may have influenced the timing, and probably the publication, of developments but also created opportunities by focusing the attention of able people onto practical metal plasticity issues. This first phase, up to around 1960/65, was, in one sense, a backwards step. The complexity of even the most straightforward chip formation – for example the fact that most chips are curled (Figure 2.1) – was ignored in an attempt to understand why chips take up their observed thicknesses. This is the key issue: once the chip flow is known, forces, stresses and temperatures may all be reasonably easily calculated. The most simple plastic flow leading to the formation of straight chips was assumed, namely shear on a flat shear plane (as described in more detail later in this chapter). The consequent predictions of chip thickness, the calculations of chip heating and contemporary developments in tribology relevant to understanding the chip/tool interaction are the main subjects of this chapter.

This first stage was not successful in predicting chip thickness, only in describing its consequences. It became clear that the flow assumptions were too simple; so were the

chip/tool friction law assumptions; and furthermore, that heating in metal cutting (and the high strain rates involved) caused in-process changes to a metal's plastic shear resistance that could not be ignored. From the mid-1960s to around 1980 the main focus of mechanics research was exploring the possibilities and consequences of more realistic assumptions. This second phase of predictive development is the subject of Chapter 6. By the 1980s it was clear that numerical methods were needed to analyse chip formation properly. The development of finite element methods for metal cutting are the subject of Chapter 7 and detailed researches are introduced in Chapter 8.

The rest of this chapter is organized into three main sections: on the foundations of mechanics, heating and tribology relevant to metal machining. Appendices 1 to 3 contain more general background material in these areas, relevant to this and subsequent chapters. Anyone with previous knowledge may find it is not necessary to refer to these Appendies, at least as far as this chapter is concerned.

2.2 Chip formation mechanics

The purpose of this section is to bring together observations on the form of chips and the forces and stresses needed to create them. The role of mechanics in this context is more to aid the description than to be predictive. First, Section 2.2.1 describes how chip formation in all machining processes (turning, milling, drilling and so on) can be described in a common way, so that subsequent sections may be understood to relate to any process. Section 2.2.2 then reports on the types of chips that have been observed with simple shapes of tools; and how the thicknesses of chips have been seen to vary with tool rake angle, the friction between the chip and the tool and with the work hardening behaviour of the machined material. Section 2.2.3 describes how the forces on a tool during cutting may be related to the observed chip shape, the friction between the chip and the tool and the plastic flow stress of the work material. It also introduces observations on the length of contact between a chip and tool and on chip radius of curvature; and discusses how contact length observations may be used to infer how the normal contact stresses between chip and tool vary over the contact area. Sections 2.2.2 and 2.2.3 only describe what has been observed about chip shapes. Section 2.2.4 introduces early attempts, associated with the names of Merchant (1945) and Lee and Shaffer (1951), to predict how thick a chip will be, while Section 2.2.5 brings together the earlier sections to summarize commonly observed values of chip characteristics such as the specific work of formation and contact stresses with tools. Most of the information in this section was available before 1970, even if its presentation has gained from nearly 30 years of reflection.

2.2.1 The geometry and terminology of chip formation

Figure 2.2 shows four examples of a chip being machined from the flat top surface of a parallel-sided metal plate (the work) by a cutting tool, to reduce the height of the plate. It has been imagined that the tool is stationary and the plate moves towards it, so that the cutting speed (which is the relative speed between the work and the tool) is described by U_{work}. In each example, U_{work} is the same but the tool is oriented differently relative to the plate, and a different geometrical aspect of chip formation is introduced. This figure illustrates these aspects in the most simple way that can be imagined. Its relationship to the

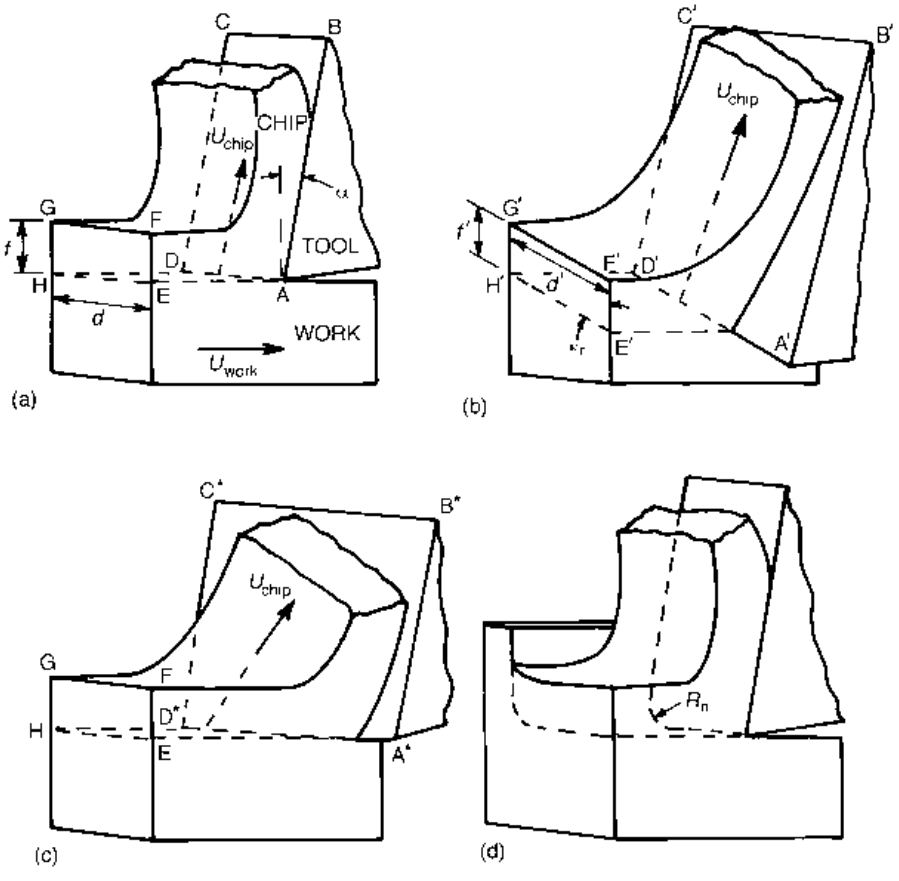


Fig. 2.2 (a and b) Orthogonal, (c) non-orthogonal and (d) semi-orthogonal chip formation.

turning milling and drilling processes is developed after first describing what those aspects are.

Orthogonal and non-orthogonal chip formation

In Figure 2.2(a) the cutting edge AD of the plane tool rake face ABCD is perpendicular to the direction of U_{work}. It is also perpendicular to the side face of the plate. As the tool and work move past one another, a volume of rectangular section EFGH is removed from the plate. The chip that is formed flows with some velocity U_{chip}, which is perpendicular to the cutting edge. All relative motions are in the plane normal to the cutting edge. In this condition, cutting is said to be orthogonal. It is the most simple circumstance. Apart from at the side faces of the chip, where some bulging may occur, the process geometry is fully described by two-dimensional sections, as in Figure 2.1(b).

It may be imagined that after reducing the height of the plate by the amount HG, the tool may be taken back to its starting position, may be fed downwards by an amount equal to HG, and the process may be repeated. For this reason the size of HG is called the feed, *f*, of the process. The dimension HE of the removed material is known as the depth of cut,

d. Figure 2.2(a) also defines the tool rake angle α as the angle between the rake face and the normal to both the cutting edge and U_{work}. (α is, by convention, positive as shown.)

When, as in Figure 2.2(a), the cutting edge is perpendicular to the side of the plate, its length of engagement with the plate is least. If it is wished to spread the cutting action over a longer edge length (this reduces the severity of the operation, from the point of view of the tool), the edge may be rotated about the direction of the cutting velocity. This is shown in Figure 2.2(b). AD from Figure 2.2(a) is rotated to A′D′. As long as the edge stays perpendicular to U_{work}, the chip will continue to flow perpendicular to the cutting edge and the cutting process remains orthogonal. However, the cross-sectional shape of the removed work material is changed from the rectangle EFGH to the parallelogram E′F′G′H′. If the amount of rotation is described by the angle κ_r between E′F′ and E′H′, the length of cutting edge engagement increases to $d' = d/\sin\kappa_r$ and the thickness of the removed layer, f', known as the uncut chip thickness, reduces to $f\sin\kappa_r$. κ_r is called the major cutting edge angle, although it and other terms to be introduced have different names in different machining processes – as will be considered later. The uncut chip thickness is more directly important to chip formation than is the feed because, with the cutting speed, it strongly influences the temperature rise in machining (as will be seen in Section 2.3).

In Figure 2.2(b), rotation of the cutting edge causes the chip flow direction to be inclined to the side of the plate. Another way of achieving this is to rotate the cutting edge in the plane ADHE (Figure 2.2(a)) so that it is no longer perpendicular to U_{work}. In Figure 2.2(c) it is shown rotated to A*D*. The section of removed material EFGH stays rectangular but U_{chip} becomes inclined to the cutting edge.

Neither U_{work} nor U_{chip} are perpendicular to the cutting edge. Chip formation is then said to be non-orthogonal. The angle of rotation from AD to A*D* is called the cutting edge inclination angle, λ_s. The mechanics of non-orthogonal chip formation are more complicated than those of orthogonal chip formation, because the direction of chip flow is not fixed by λ_s.

Finally, Figure 2.2(d) shows a situation in which the cutting edge AD is lined up as in Figure 2.2(a), but it does not extend the full width of the plate. In practice, as shown, the cutting edge of the tool near point D is rounded to a radius R_n – the tool nose radius. Because the cutting edge is no longer straight, it is not possible for the chip (moving as a rigid body) to have its velocity U_{chip} perpendicular to every part of the cutting edge. Even if every part of the cutting edge remains perpendicular to U_{work}, the geometry is not orthogonal. This situation is called semi-orthogonal. If $R_n << d$, the semi-orthogonal case is approximately orthogonal.

Turning

The turning process has already been introduced in Chapter 1 (Figure 1.7). In that case, orthogonal chip formation with a 90° major cutting edge angle was sketched. Figure 2.3 shows a non-orthogonal turning operation, with a major cutting edge angle not equal to 90°. The feed and depth of cut dimensions are also marked. In this case, the cutting speed U_{work} equals $\pi D\Omega$ m/min (if the units of D and Ω are m and rev/min).

In turning, the major cutting edge angle is also known by some as the approach angle, and the inclination angle as the back rake. The rake angle of Figure 2.2(a) can be called the side rake. Table 2.1 summarizes these and other alternatives. (See, however, Chapter 6.4 for more comprehensive and accurate definitions of tool angles.)

The uncut chip thickness in turning, f', is $f\sin\kappa_r$. It is possible to reach this obvious

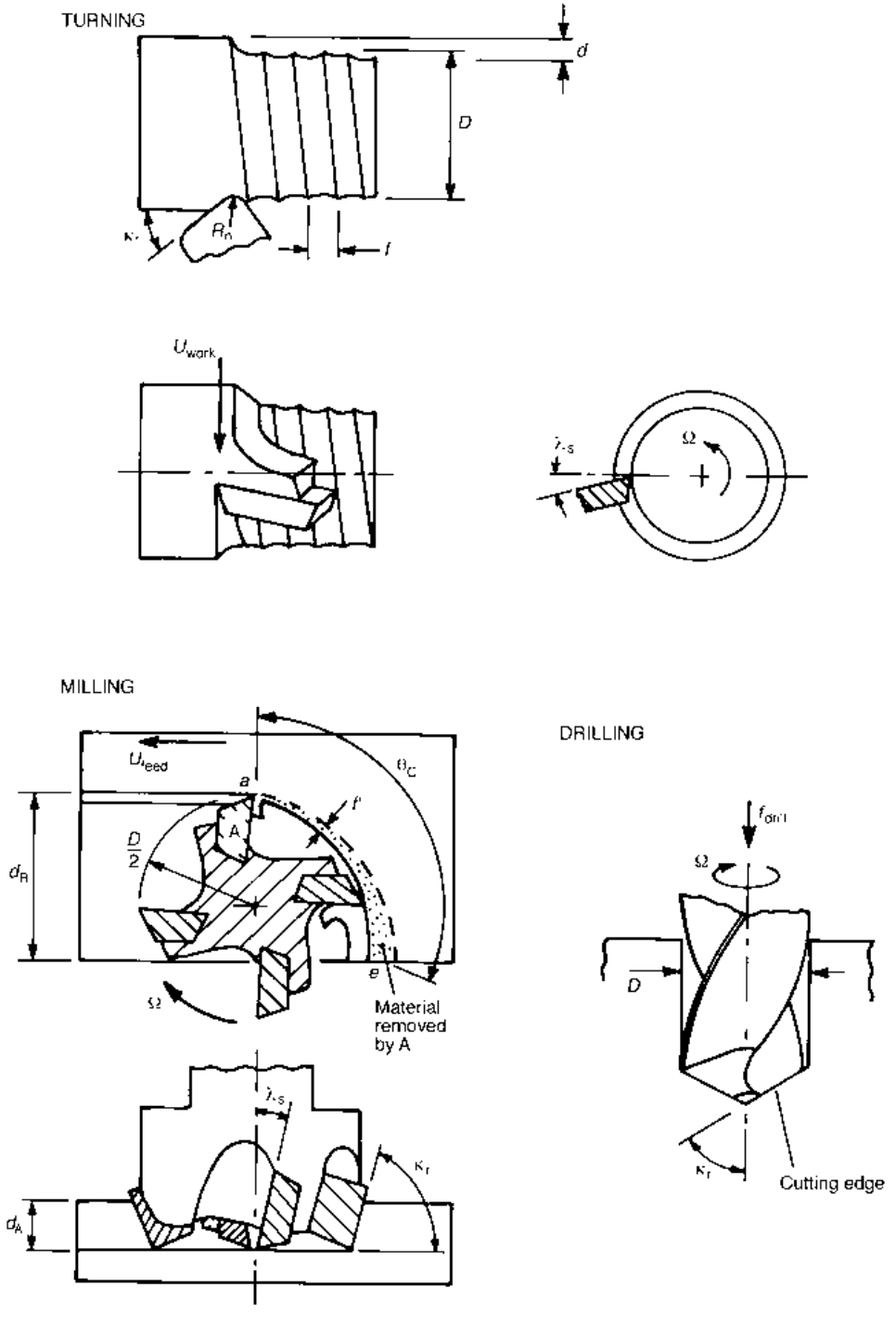


Fig. 2.3 Turning, milling and drilling processes

Table 2.1 Some commonly encountered near-alternative chip formation terms (see Chapter 6.4 for a more detailed consideration of three-dimensional tool geometry)

	Equivalent name in		
General name and symbol	Turning	Milling	Drilling
Rake angle, α	Side rake angle	Radial rake angle	Rake angle
Inclination angle, λ_s	Back rake angle	Axial rake angel	Helix angle
Major cutting edge angle, κ_r	Approach angle	Entering angle	Point angle
Feed	Feed per rev.	Feed per edge	Feed per rev.
Depth of cut	Depth of cut	Axial depth of cut	Hole radius

conclusion in a rather more general way which, although it has no merit for turning, becomes useful for working out the uncut chip thickness in a milling process. Equation (2.1a) is a statement of that more general way. It is a statement that the volume removed from the work is the volume swept out by the cutting edge. In turning, the volume removed per unit time is fdU_{work}. The distance that the cutting edge sweeps through the work in unit time is simply U_{work}. The truth of equation (2.1a) is obvious.

$$f' = \frac{\text{Volume removed per unit time}}{\text{Distance swept out by cutting edge per unit time}} \frac{\sin \kappa_r}{d} \tag{2.1a}$$

Milling

There are many variants of the milling process, described in detail by Shaw (1984) and Boothroyd and Knight (1989). Figure 2.3 shows face milling (and could also represent the end milling process). A slab is reduced in thickness by an amount d_A over a width d_R by movement at a linear rate U_{feed} normal to the axis of a rotating cutter. d_A is called the axial depth of cut and d_R is the radial width of cut. The cutter has N_f cutting edges (in this example, $N_f = 4$) on a diameter D and rotates at a rate Ω. Each cutting edge is shown with a major cutting edge angle κ_r and inclination angle λ_s, although in milling these angles are also known as the entering angle and the axial rake angle (Table 2.1). For some cutters, with long, helical, cutting edges, the axial rake angle is further called the helix angle. The cutting speed, as in turning, is $\pi D\Omega$.

In Figure 2.3, the cutter is shown rotating clockwise and travelling through the work so that a cutting edge A enters the work at a and leaves at e. A chip is then formed from the work with an uncut chip thickness increasing from the start to the end of the edge's travel. If the cutter were to rotate anticlockwise (and its cutting edges remounted to face the other way), a cutting edge would enter the work at e and leave at a, and the uncut chip thickness would decrease with the edge's travel.

In either case, the average uncut chip thickness can be found from (2.1a). The work volume removal rate is $d_A d_R U_{feed}$. The distance swept out by one cutting edge in one revolution of the cutter is the arc length ae, or $(D/2)\theta_C$, where θ_C can be determined from D and d_R. The distance swept out by N_f edges per unit time is then $N_f\Omega(D/2)\theta_C$. d in equation (2.1a) is d_A. Substituting all these into equation (2.1a) gives

$$f'_{av.,milling} = \frac{2d_R U_{feed}}{N_f \Omega D \theta_C} \sin \kappa_r \tag{2.1b}$$

The relation between the uncut chip thickness's average and maximum values depends on the detailed path of the cutting edge through the work. In the case shown in Figure 2.3 in which the uncut chip thickness near a is zero, the maximum value at e is twice that of equation (2.1b), but there are other circumstances (in which neither at entry nor exit is the cutting edge path nearly tangential to the cut surface) in which the maximum and average values can be almost equal.

Table 2.1 contains the term 'feed per edge'. This is the distance moved by the work for every cutting edge engagement. It is $U_{\text{feed}}/(N_{\text{f}}\Omega)$. The ratio of the uncut chip thickness to this differs from the value $\sin\kappa_{\text{r}}$ that is the ratio in turning.

Drilling

Finally, Figure 2.3 also shows a drilling process in which a hole (diameter D) is cut from an initially blank plate. The simpler case (from the point of view of chip formation) of enlarging the diameter of a pre-existing hole is not considered. The figure shows a two-flute (two cutting edges) drill with a major cutting edge angle κ_{r} (in drilling called the point angle). The inclination angle in drilling is usually zero. The depth of cut is the radius of the hole being drilled. The axial feed of a drill is usually described, as in turning, as feed per revolution.

Drilling has an intermediate position between milling and turning in the sense that, although a drill has more than one cutting edge (usually two), each edge is engaged continuously in the work. The special feature of drilling is that the cutting speed varies along the cutting edge, from almost zero near the centre of the drill to the circumferential speed of the drill at its outer radius. The uncut chip thickness can be obtained from equation (2.1a). The volume removed per revolution of the drill is $(\pi D^2/4)f$. The distance per revolution swept out by N_{f} cutting edges, at the average radius $(D/4)$ of the drill, is $(\pi D/2)N_{\text{f}}$. Substituting these, and $d \equiv D/2$, into equation (2.1a) gives

$$f'_{\text{drilling}} = \frac{f}{N_{\text{f}}} \sin\kappa_{\text{r}} \tag{2.1c}$$

This, as in the case of turning, could have been obtained directly.

On feed, uncut chip thickness and other matters

The discussion around Figure 2.2 introduced some basic terminology, but it is clear from the descriptions of particular processes that there are many words to describe the same function, and sometimes the same word has a different detailed meaning depending on the process to which reference is being made. *Feed* is a good example of the latter. In turning and drilling, it means the distance moved by a cutting edge in one revolution of the work; in milling it means the distance moved by the work in the time taken for each cutting edge to move to the position previously occupied by its neighbour. However, in every case, it describes a relative displacement between the cutting tool and work, set by the machine tool controller.

Feed and *depth of cut* always refer to displacements from the point of view of machine tool movements. *Uncut chip thickness* and *cutting edge engagement length* are terms closely related to feed and depth of cut, but are used from the point of view of the chip formation process. It is a pity that the terms uncut chip thickness and cutting edge engagement length are so long compared with feed and depth of cut.

In the case of turning with a 90° major cutting edge or approach angle, there is no difference between feed and uncut chip thickness nor between depth of cut and cutting edge engagement length. Further, the cutting speed is the same as the work speed U_{work}. In the remainder of this book, chips will be described as being formed at a cutting speed U_{work}, at a feed f and depth of cut d – meaning at an uncut chip thickness f and a cutting edge engagement length d. This is correct only for turning, as just described. The reader, however, should be able to convert that convenient terminology to the description of other processes, by the relations that have been developed here.

2.2.2 Chip geometries and influencing factors

Figure 2.1 shows views of chips observed more than 100 years ago. Figure 2.4 shows more modern images, photographs taken from polished and etched quick-stop sections (in the manner described in Chapter 5). It shows the wide range of chip flows that are free to be formed, depending on the material and cutting conditions. All these chips have been created in turning tests with sharp, plane rake face cutting tools. Steady or continuous chip formation is seen in Figure 2.4(a) (as has been assumed in Figure 2.2). This example is for 70/30 brass, well known as an easy to machine material. Some materials, however, can form a more segmented, or saw tooth, chip (e.g. stainless steel – Figure 2.4(b)). Others do not have sufficient ductility to form continuous chips; discontinuous chips are formed instead. Figures 2.4(c) (for a brass made brittle by adding lead) and 2.4(d) (for a mild steel cut at very low cutting speed) are, respectively, examples of discontinuous chips showing a little and a lot of pre-failure plastic distortion. In other cases still (mild steel at an intermediate cutting speed – Figure 2.4(e)) work material cyclically builds up around, and breaks away from, the cutting edge: the chip flows over the modified tool defined by the shape of the built-up edge. The built-up edge has to withstand the loads and temperatures generated by the chip formation. As cutting speed, and hence the temperature, increases, the built-up edge cannot survive (or does not form in the first place): Figure 2.4(f) (mild steel at higher speed) shows the thin layer of build-up that can exist to create a chip geometry that does not look so different from that of Figure 2.4(a).

This chapter will be concerned with only the most simple type of chip formation – continuous chip formation (Figures 2.4(a) and (f)) by a sharp, plane rake face tool. Further, only the orthogonal situation (Section 2.2.1) will be considered. The role of mechanics is to show how the force and velocity boundary conditions at the chip – tool interface and the work material mechanical properties determine the flow of the chip and the forces required for cutting. For continuous chip formation, determining the flow means at least determining the thickness of the chip, its contact length with the tool and its curvature: none of these are fixed by the tool shape alone. In fact, determining the chip shape is the grand challenge for mechanics. Once the shape is known, determining the cutting forces is relatively simple; and determining the stresses and temperatures in the work and tool, which influence tool life and the quality of the machined surface, is only a little more difficult.

The main factors that affect the chip flow are the rake angle of the tool, the friction between the chip and the tool and the work hardening of the work material as it forms the chip. Some experimental observations that establish typical magnitudes of the quantities involved will now be presented, but first some essential notation and common simplifications to the flow (to be removed in Chapter 6) will be introduced. Figure 2.5(a) is a sketch of Figure 2.4(a). It shows the chip of thickness t being formed from an undeformed layer

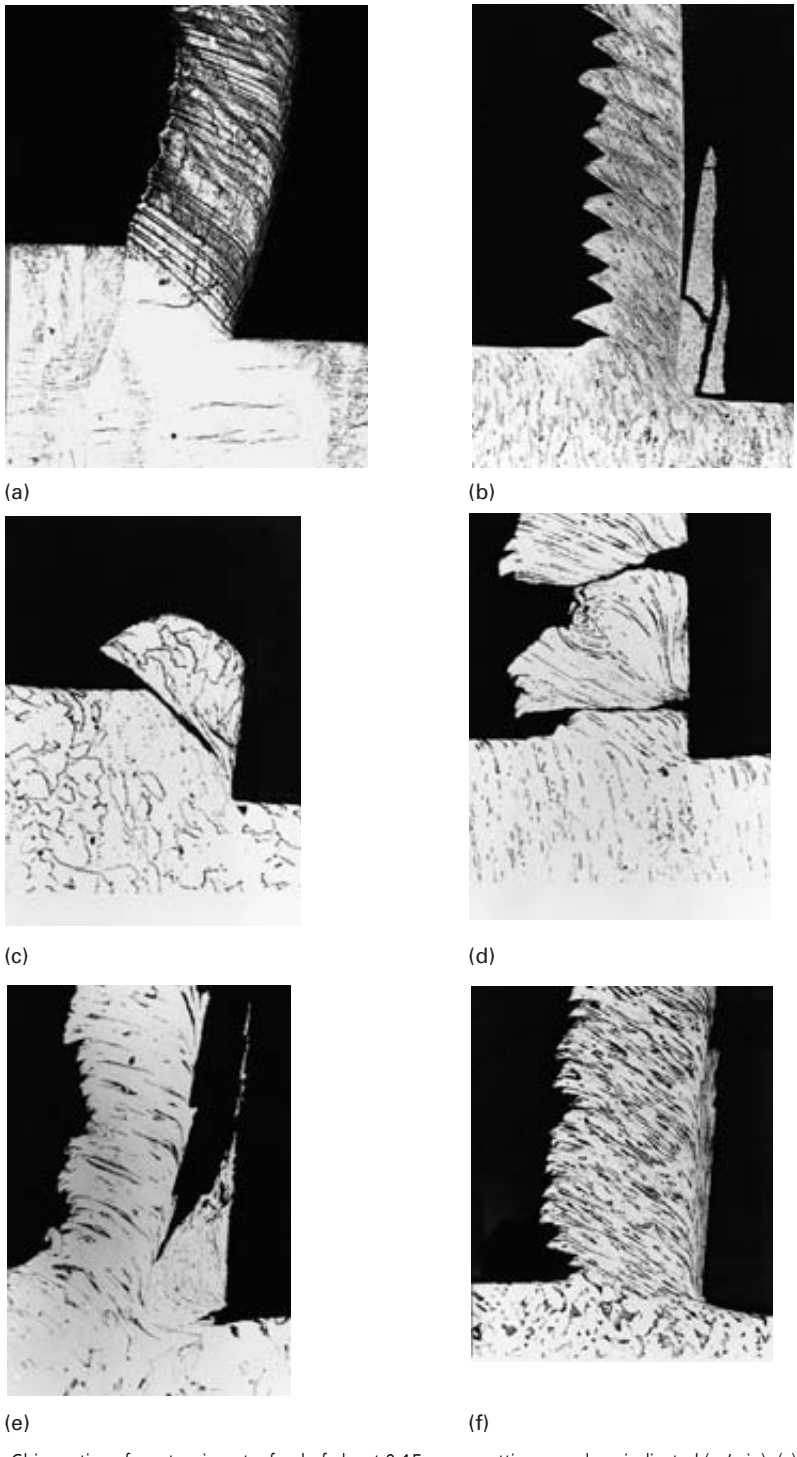


Fig. 2.4 Chip sections from turning at a feed of about 0.15 mm – cutting speeds as indicated (m/min): (a) 70/30 brass (50), (b) austenitic stainless steel (30), (c) leaded brass (120): (d) mild steel (5), (e) mild steel (25), (f) mild steel (55)

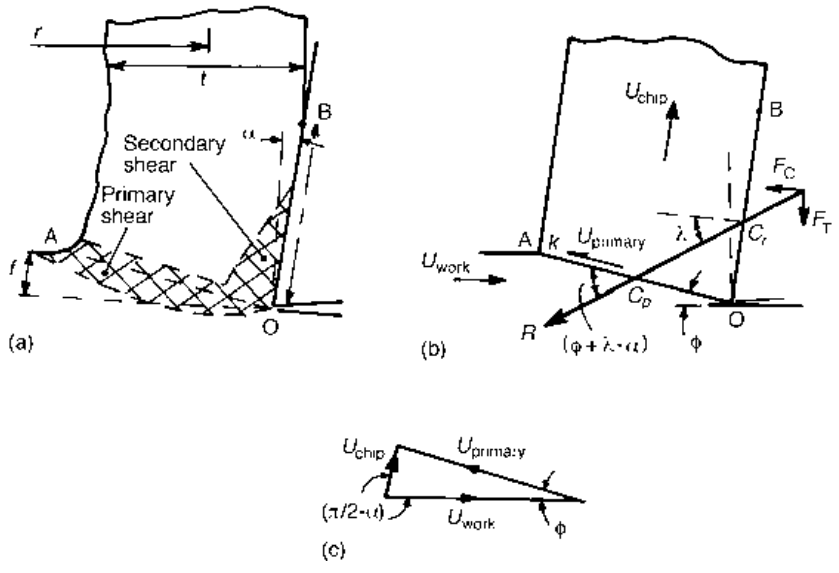


Fig. 2.5 Chip flow (a) sketched from Figure 2.4(a); (b) simplified and (c) its velocity diagram

of thickness f (the feed) by a tool of rake angle α. The contact length with the tool, OB, is l and the chip radius is r. Regions of plastic flow are identified by the hatched markings. The main deformation zone, known as the primary shear zone, exists around the line OA. Further strain increments are frequently detectable next to the rake face, in the secondary shear zone. A simplified flow (Figure 2.5(b)) replaces the primary zone by a straight surface, the shear plane OA and neglects the additional deformations in the secondary zone (although the region might still be at the plastic limit). Figure 2.5(b) shows OA inclined at an angle ϕ to the cutting speed direction. ϕ is called the shear plane angle. As the length of the shear plane OA can be obtained either from (f/sin ϕ) or from (t/cos($\phi - \alpha$)),

$$\frac{t}{f} = \frac{\cos(\phi - \alpha)}{\sin \phi} \tag{2.2}$$

Figure 2.5 also identifies the velocity change, U_{primary}, that occurs on the primary shear plane, which converts U_{work} to U_{chip}. It further shows the resultant force R responsible for the flow, inclined at the friction angle λ to the rake face normal (tan λ = the friction coefficient μ) and thus at ($\phi + \lambda - \alpha$) to OA. It also introduces other quantities referred to later.

The magnitude of U_{primary}, and of the resulting U_{chip}, relative to U_{work}, can be found from the velocity diagram for the simplified flow (Figure 2.5(c)):

$$\frac{U_{\text{primary}}}{\cos \alpha} = \frac{U_{\text{chip}}}{\sin \phi} = \frac{U_{\text{work}}}{\cos(\phi - \alpha)} \tag{2.3}$$

The shear strain that occurs as the chip is formed is the ratio of the primary shear velocity to the component of the work velocity normal to the shear plane. The equivalent strain is

$1/\sqrt{3}$ times this (Appendix 1). Combining this with equations (2.3) and (2.2), the equivalent strain is:

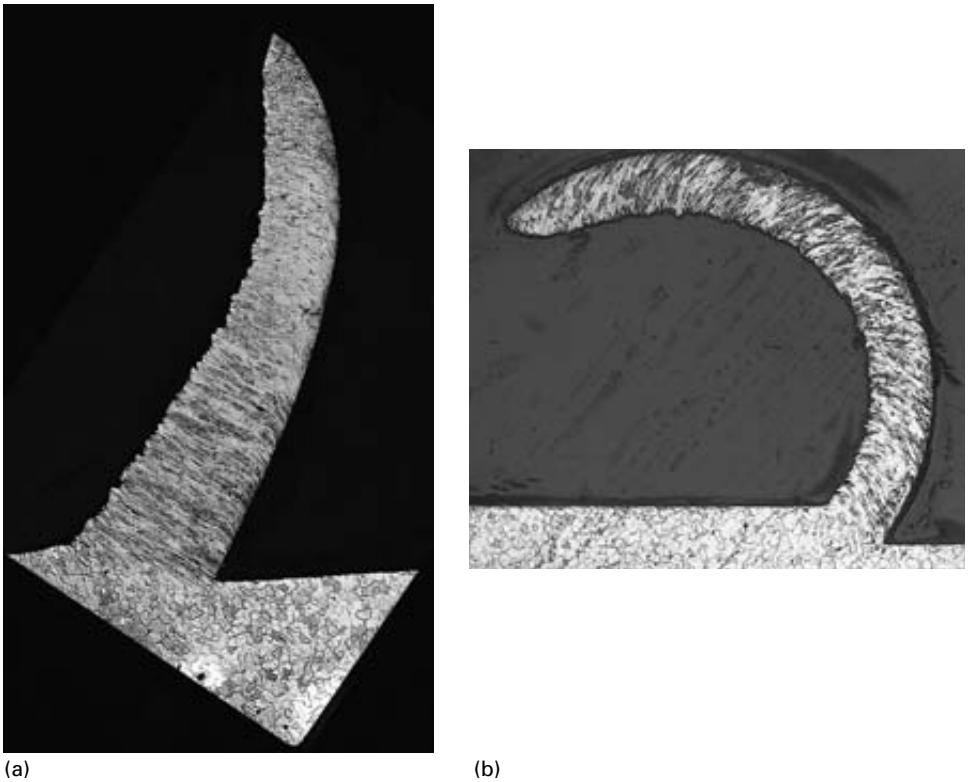
$$\bar{\varepsilon} \equiv \frac{\gamma}{\sqrt{3}} = \frac{U_{\text{primary}}}{\sqrt{3}U_{\text{work}}\sin\phi} = \frac{\cos\alpha}{\sqrt{3}\sin\phi\cos(\phi-\alpha)} = \frac{\cos\alpha}{\sqrt{3}\cos^2(\phi-\alpha)}\frac{t}{f} \tag{2.4a}$$

Thus, the severity of deformation is determined by α, $(\phi - \alpha)$ and the chip thickness ratio (t/f). The ratio $\cos\alpha/\cos^2(\phi - \alpha)$, as will be seen, is almost always in the range 0.9 to 1.3. So

$$\bar{\varepsilon} \approx (0.5 \text{ to } 0.75)(t/f) \tag{2.4b}$$

Mallock's (1881–82) observation that chip thickness is strongly influenced by lubrication has already been mentioned. Figure 2.6 dramatically illustrates this. It is a quick-stop view of iron cut by a 30° rake angle tool at a very low cutting speed (much less than 1 m/min). In an air atmosphere the chip formed is thick and straight. Adding a lubricating fluid causes the chip to become thinner and curled. In this case, adding the lubricant caused the friction coefficient between the chip and tool to change from 0.57 to 0.25 (Childs, 1972).

The lubricating fluid used in this study was carbon tetrachloride, CCl_4, found by early



(a) (b)

Fig. 2.6 Machining iron at low speed: (a) dry (in air) and (b) with carbon tetrachloride applied to the rake face

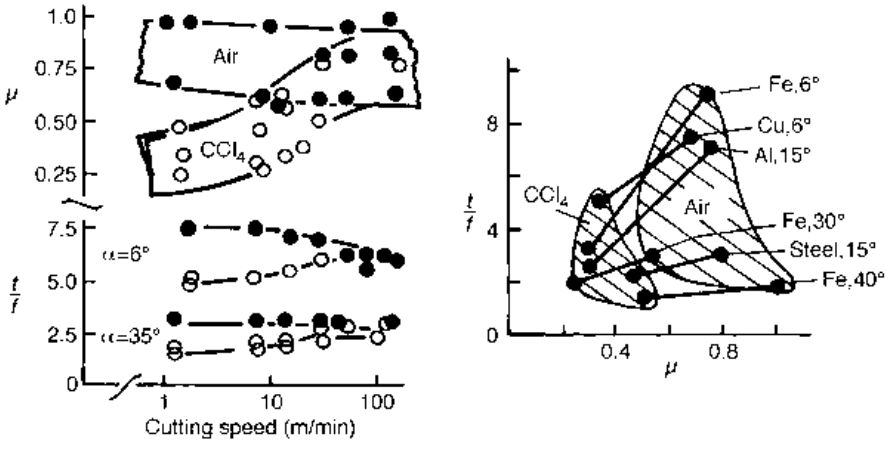


Fig. 2.7 (a) Collected data on the machining of copper, dry (•) and lubricated (o); and (b) lubricant effects for a range of conditions at cutting speeds around 1 m/min

researchers to be one of the most effective friction reducing fluids. However, it is toxic and not to be recommended for use today. In addition, CCl_4 only acts to reduce friction at low cutting speeds. Figure 2.7 brings together results from several sources on the cutting of copper. It shows, in Figure 2.7(a), friction coefficients measured in air and CCl_4 atmospheres at cutting speeds from 1 to 100 m/min, at feeds between 0.1 and 0.25 mm and with cutting tools of rake angle 6° to 40°. At the higher speeds the friction-reducing effect of the CCl_4 has been lost. Mallock was right to be puzzled by how the lubricant reaches the interface between the chip and tool. How lubricants act in metal cutting is considered further in Section 2.4.2.

The range of friction coefficients in Figure 2.7(a) for any one speed and lubricant partly comes from the range of rake angles to which the data apply. Higher friction coefficients are associated with lower rake angles. Figure 2.7(a) also shows how both lubricating fluid and rake angle affect the chip thickness ratio. Both low friction and high rake angles lead to low chip thickness ratios. General experience, for a range of materials and rake angles, is summarized in Figure 2.7(b). In the context of metal cutting, low friction coefficients and chip equivalent strains (from equation 2.4(b)) are 0.25 to 0.5 and 1 to 3 respectively; whereas high friction coefficients and strains are from 0.5 to 1 (and in a few cases higher still) and up to 5.

High work hardening rates are also found experimentally to lead to higher chip thickness ratios – although it is difficult to support this statement in a few lines in an introductory section such as this. The reason is that it is difficult to vary work hardening behaviour without varying the friction coefficient. One model material, with a friction coefficient more constant than most, is α-brass (70%Cu/30%Zn). Figure 2.8(a) shows the work hardening characteristics of this metal. The chips from work material pre-strained, for example to point C, may expect to be work hardened to their maximum hardness by machining. The friction coefficients and chip thickness ratios obtained when forming chips from variously pre-strained samples, with a 15° rake angle high speed steel tool, at feeds around 0.2 mm and cutting speeds from 1 to 50 m/min are shown in Figure 2.8(b) (Childs *et al.*, 1972). Anticipating a later section, the measure of work hardening used as the independent

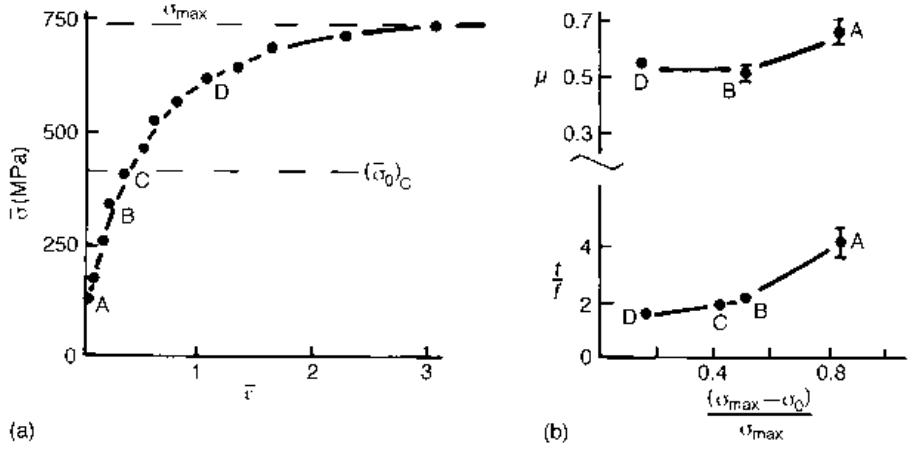


Fig. 2.8 (a) The work hardening of 70/30 brass and (b) friction coefficients and chip thickness ratios measured for samples pre-strained by amounts A to D as marked

variable in Figure 2.8(b) is the ratio of the increase in equivalent stress to the maximum equivalent stress caused by machining. For materials D, C and B, thicker chips occur the greater is the work hardening, despite a constant friction coefficient. Material A shows a thicker chip still, but its friction coefficient is marginally increased too. Comparing Figures 2.8(b) and 2.7(b), changes in work hardening and friction coefficient have similar influences on chip thickness ratio.

Thus, rake angle, friction and work hardening are established as all influencing the chip formation. To make further progress in describing the mechanical conditions of machining, the constraints of force and moment equilibrium must be introduced.

2.2.3 Force and moment equilibrium

Cutting and thrust forces

The resultant force R has already been introduced in Figure 2.5(b). Its inclination to the primary shear plane is, from geometry, $(\phi + \lambda - \alpha)$. From the previous section, the shear stress k on the shear plane is expected to be that of the fully work hardened material. Resolving R onto the shear plane, dividing it by the area of the plane and equating the result to k leads to

$$R = \frac{kfd}{\sin\phi\cos(\phi + \lambda - \alpha)} \tag{2.5a}$$

where d is the width of the shear plane (depth of cut) out of the plane of Figure 2.5. The cutting and thrust force components, F_c and F_t, also defined in Figure 2.5, are

$$F_c = \frac{kfd\cos(\lambda - \alpha)}{\sin\phi\cos(\phi + \lambda - \alpha)}; \qquad F_t = \frac{kfd\sin(\lambda - \alpha)}{\sin\phi\cos(\phi + \lambda - \alpha)} \tag{2.5b}$$

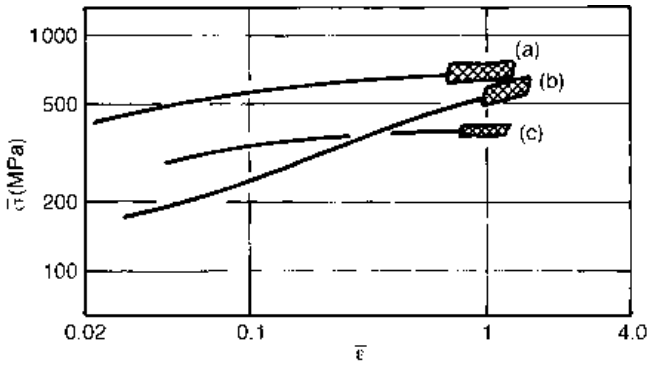


Fig. 2.9 Equivalent stress-strain data for (a) a mild steel, (b) an aluminium alloy; and (c) an α-brass obtained from compression testing (—) and values from metal cutting tests (hatched), after Kobayashi and Thomsen (1959)

Alternatively, k may be directly related to F_c and F_t:

$$kfd = (F_c \cos \phi - F_t \sin \phi)\sin \phi \tag{2.5c}$$

Many experimental studies of continuous chip formation have confirmed these relations. Indeed, departures are a clear indication of a breakdown in the assumptions, for example of the presence of a built-up edge changing the tool geometry. One particularly thorough study was carried out by Kobayashi and Thomsen (1959), measuring forces and chip thicknesses in the machining of ferrous and non-ferrous metals, and using equation (2.5c) to estimate k. Figure 2.9 shows their results converted to equivalent stress ($\bar{\sigma} = k\sqrt{3}$), compared with data obtained from compression testing.

Chip/tool contact lengths

The contact length between the chip and tool, as well as the chip thickness, is of interest in metal cutting. Chip moment equilibrium may be applied to relate the contact length to the chip thickness. Figure 2.5(b) shows the resultant force R passing through the centres of pressure C_p and C_r on the primary shear plane and rake face respectively. Zorev (1966) introduced the length ratios $m = OC_p/OA$ and $n = OC_r/OB$: from the moment equilibrium about O, contact length l and chip thickness t are related by

$$l = \frac{m}{n} t[\mu + \tan(\phi - \alpha)] \tag{2.6}$$

Zorev gives experimental results obtained from turning a large range of carbon steels (0.12 to 0.83%C) and low alloy engineering steels, at feeds from 0.15 to 0.5 mm and cutting speeds from 15 to 300 m/min, that agree well with equation (2.6) if (m/n) is taken to be in the range 3.5 to 4.5. However, the contact length is a difficult quantity to measure, and even to define. Zorev himself commented that the 45% of the contact length furthest from the cutting edge may carry only 15% of the rake face load. Other researchers have obtained lower values for (m/n). Figure 2.10(a) shows Zorev's mean value data as the solid line, with observations by the present authors obtained from restricted contact and split tool tests. (m/n) values as low as 1.25 have been observed, and values of 2 are common. To put

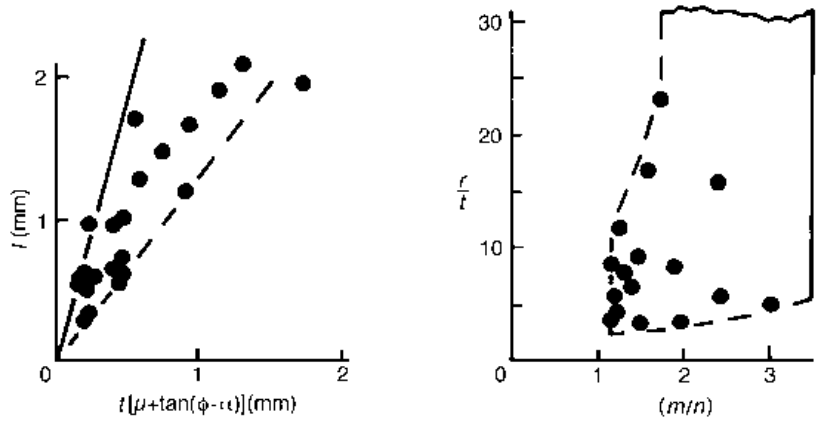


Fig. 2.10 Chip/tool contact length and chip radius observations. (a) Measured dependence of chip/tool contact length on chip thickness; and (b) wide variations of dimensionless chip radius (r/t) with (m/n)

these values in perspective, a uniform pressure along the shear plane and a triangular pressure distribution along the rake face (with a peak at the cutting edge) would give (m/n) = 1.5. The range of (m/n) found in practice suggests that different materials machine with different pressure distributions along the shear plane or rake face, or both.

Chip/tool contact pressures

The question of what contact pressure distributions exist between the chip and tool and on the primary shear plane will be covered in later chapters in some detail. However, the elementary mechanics considerations here may be developed to give some insight into possible contact pressure distributions. The procedure is first to consider the primary shear plane pressure distribution and the associated likely range of the parameter m. m and m/n together then enable the size range of n to be deduced. Different values of n are associated with different tool contact pressure distributions.

First of all, suppose that the contact pressure is not uniform along the primary shear plane OA (Figure 2.5(a)), but falls from a maximum value at A to a lower value at O. Oxley (1989) pointed out that this will be the case for a work hardening material. Figure 2.11(a) is developed from his work. It shows the shear plane OA imagined as a parallel-sided zone of width w and length s ($s = f/\sin\phi$). Work hardening results in the shear stress k_{max} at the exit to the zone being more than that k_o at the entry. A force balance on the hatched region establishes that p must reduce towards O, from some maximum value p_s at the free surface. When the shear zone is parallel sided, the reduction is uniform with distance from A. At O the reduction has become $(s/w)\Delta k$, where Δk is ($k_{max} - k_o$). The average pressure is half the sum of the pressures at A and O. The ratio of the average pressure to the shear stress on the shear plane is equal to the tangent of the angle between the resultant force R and the shear plane. This is $\tan(\phi + \lambda - \alpha)$. It follows that

$$\tan(\phi + \lambda - \alpha) = \frac{p_s}{k_{max}} - \frac{1}{2}\,\frac{s}{w}\,\frac{\Delta k}{k_{max}} \tag{2.7}$$

Further, by taking moments about the cutting edge O, m can be expressed in terms of p_s, (s/w) and $(\Delta k/k_{max})$, as shown in equation (2.8a). Then (s/w) can be eliminated with the help of equation (2.7), as shown in (2.8b)

$$m = \frac{1}{2}\left[1 + \frac{1}{6}\,\frac{(\Delta k/k_{max})(s/w)}{p_s/k_{max} - 1/2(\Delta k/k_{max})(s/w)}\right] \tag{2.8a}$$

$$m = \frac{1}{3} + \frac{1}{6}\,\frac{p_s/k_{max}}{\tan(\phi + \lambda - \alpha)} \tag{2.8b}$$

Data exist to test equation (2.7) and hence to deduce values of m. It is commonly observed that $(\phi + \lambda - \alpha)$ varies from material to material (a range of data will be given in Section 2.2.4). It reduces the more the material work hardens. Figure 2.11(b) shows, as solid circles, data obtained from the same set of tests that led to Figure 2.8, while the open circles are for steels, aluminium alloys and brass (from the work of Kobayashi and Thomsen, 1959). The data support equation (2.7), with $p_s/k_{max} \approx 1.4 \pm 0.2$. From equation (2.8b), with $p_s/k_{max} \approx 1.4$ and with $\tan(\phi + \lambda - \alpha)$ varying from 0.6 to 1.4, values of m from 0.5 to 0.72 are obtained.

The gradient of –1 in Figure 2.11(b) implies $s/w = 2$. This is less than expected, given quick-stop views of how narrow is the shear zone. For example, the hatched primary shear region of Figure 2.5(a) has $s/w \approx 4$. However, other studies (considered in Chapter 6) have suggested values for s/w as small as 2.6. It all depends how carefully one defines where are the edges of the zone. For now it is enough to point out that the shear plane model approximation clearly loses some essential detail of force analysis in machining, even though it has a use in obtaining a range of values of m.

The range of m from 0.5 to 0.72 is not wide compared with the variation of (m/n) from 1.25 to 3.8 (equation (2.6) and Figure 2.10(a)). It seems that n is a more variable quantity

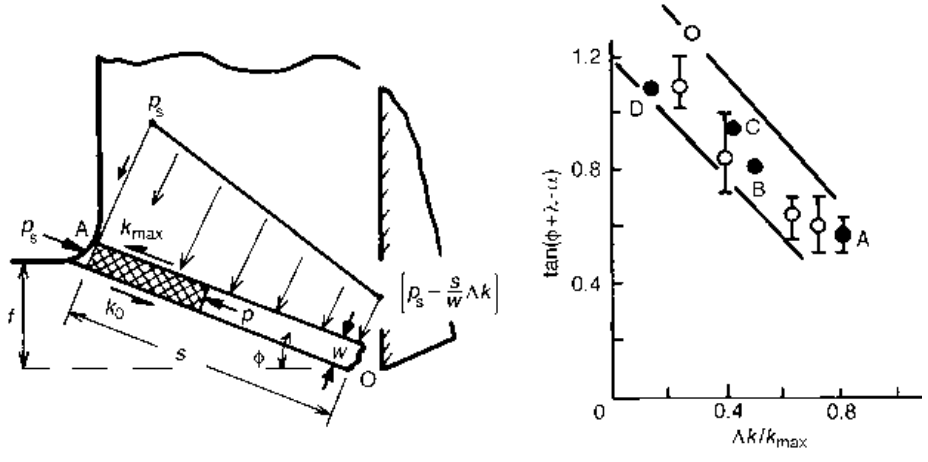


Fig. 2.11 (a) The primary shear region modelled as a parallel-sided zone of thickness w, showing pressure variations due to work hardening; and (b) observed reductions of $\tan(\phi + \lambda - \alpha)$ with increasing work hardening

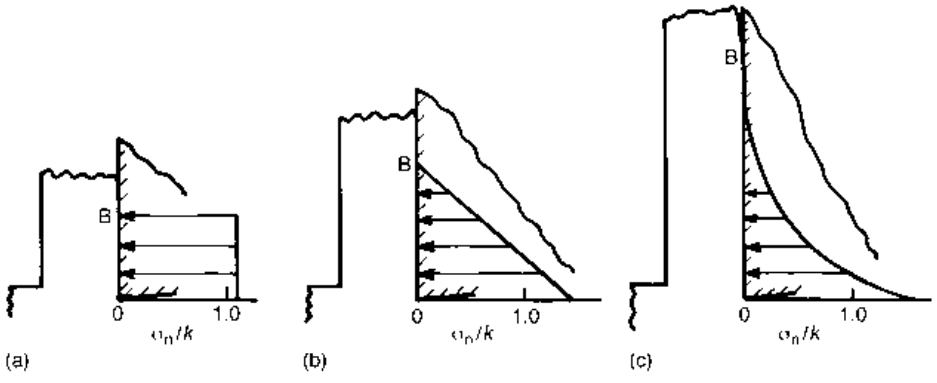


Fig. 2.12 A range of possible rake face contact pressure distributions σ_n/k, for $\alpha = 0°$, $\phi = 10°$ and $\lambda = 35°$ and $n =$ (a) 1/2, (b) 1/3 and (c) 1/6

than m. The common (m/n) value of 2 (from Figure 2.10(a)) is consistent with $n \approx 0.3$. This would be expected of a triangular distribution of contact pressure between the chip and tool. However, extreme n values are derived from 0.55 to 0.15. The former describes an almost uniform contact pressure on the rake face, while the latter corresponds almost to a fourth power law variation. These pressure distributions lead to different peak pressures at the cutting edge. Figure 2.12 shows, for the arbitrary example of $\alpha = 0^\circ$, $\phi = 10^\circ$ and $\lambda = 35^\circ$, three chips identical but for their contact length and pressure distribution with the tool. The pressures have been calculated, relative to k, from the tool forces and chip/tool contact length, by combining equations (2.2), (2.5) and (2.6). Figure 12(c) is associated with a peak contact stress with the cutting edge 50% larger than that for Figure 12(a). What contact stress distribution actually occurs is clearly relevant to tool failure and is considered further in subsequent sections and chapters.

Chip radii

Chip curvature has been ignored in simplifying the description of chip flow in Figure 2.5. However, comments may still be made on what is observed. First of all, lubricated chip flows are almost invariably highly curled. A good example is seen in Figure 2.6. Free-machining steels (containing MnS or MnS and Pb) also give tightly curled chips in their free-machining speed range (even in the absence of built-up edge formation). Beyond this, there seems to be no generalization that can be made, or relationship derived between chip curvature and other machining parameters.

One reason is that chip curvature is very sensitive to external interference, for example from interaction of the chip with the tool holder or from collision with the workpiece. Even if care is taken to avoid such real (and common) considerations, there are no simple laws governing chip curvature. For example, it could be imagined that chips with long contacts with the tool relative to their thickness might have larger radii than chips with shorter contact lengths. Figure 2.10(b) collects data on dimensionless chip radius (the radius relative to the chip thickness) and (m/n). It includes results from machining brass and iron (already referenced in Figures 2.6 and 2.8) and low carbon non-free and free cutting steels which have already featured in Figure 2.10(a). There is certainly no single valued relation between (r/t) and (m/n) although widely spaced boundaries can be drawn around the data.

No better relationship has ever been found, for machining with plane-faced tools. The reason for this is easy to understand. Qualitatively, a curled chip may be regarded as shorter (more compressed) at its inner radius than at its outer radius. Only rarely are chips so tightly curled that $(r/t) < 5$; even then the variation in compression from the chip centre-line to its inner and outer radii is only ± 0.1, i.e. $t/(2r)$. Average chip equivalent strains (equation 2.4(b)) are typically greater than 1. Thus, the modifications to flow associated with curvature are secondary relative to the magnitude of the flow itself. The sort of factors that could affect chip radius are variations of friction along the chip/tool contact length and the roundness of the cutting edge, and also the work hardening behaviour and variations of work hardening behaviour through the thickness of the chip (most chips are formed from surfaces which themselves have previously been strained by machining).

2.2.4 Shear plane angle prediction

The previous section gives data that show that chip thickness, and hence shear plane angle, depends on tool rake angle, friction and work hardening; and it records how forces and tool stresses can be estimated if shear plane angle, rake angle and friction angle are known. In this section, early attempts, by Merchant (1945) and Lee and Shaffer (1951), to predict the shear plane angle are introduced. Both attempted to relate shear plane angle to rake angle and friction angle, and ignored any effects of work hardening.

Merchant suggested that chip thickness may take up a value to minimize the energy of cutting. For a given cutting velocity, this is the same as minimizing the cutting force (equation (2.5(b)) with respect to ϕ. The well-known equation results:

$$\phi = \pi/4 - (\lambda - \alpha)/2 \tag{2.9}$$

Lee and Shaffer proposed a simple slip line field to describe the flow (see Appendix 1 and Chapter 6 for slip line field theory). For force equilibrium of the free chip, it requires that the pressure on the primary shear plane is constant along the length of the shear plane and equal to k. If $(p/k) = 1$ and $\Delta k = 0$ are substituted in equation (2.7), Lee and Shaffer's result is obtained:

$$\phi = \pi/4 - (\lambda - \alpha) \qquad \text{or} \qquad (\phi - \alpha) = \pi/4 - \lambda \tag{2.10}$$

Neither equation (2.9) nor (2.10) is supported by experiment. Although they correctly show a reducing ϕ with increasing λ and reducing α, each predicts a universal relation between ϕ, λ and α and this is not found in practice. However, they stimulated much experimental work from which later improvements grew.

It is common practice to test the results of experiments against the predictions of equations (2.9) and (2.10) by plotting the results as a graph of ϕ against $(\lambda - \alpha)$. It is an obvious choice for testing equation (2.9); and equation (2.9) was the first of these to be derived. As far as equation (2.10) is concerned, an equally valid choice would be to plot $(\phi - \alpha)$ against λ. Different views of chip formation are formed, depending on which choice is taken. The first choice may be regarded as the machine-centred view: $(\lambda - \alpha)$ is the angle between the resultant force on the tool and the direction of relative motion between the work and tool. The second choice gives a process-centred view: $(\phi - \alpha)$ is the complement of the angle between the shear plane and the tool rake face. Figures 2.13 and 2.14 present selected experimental results according to both views.

The data in Figure 2.13 (from Shaw, 1984) were obtained by machining a free-cutting

steel at a low cutting speed (0.025 m/min), with high speed steel tools with rake angles from 0° to 45°. A range of cutting fluids were applied to create friction coefficients from 0.13 to 1.33. When the results are plotted as commonly practised (Figure 2.13(a)), data for each rake angle lie on a straight line, with a gradient close to 0.75, half way between the expectations of equations (2.9) and (2.10). When the process-centred view is taken (Figure 2.13(b)), an almost single relation is observed between the friction coefficient and $(\phi - \alpha)$.

Figure 2.14 collects data at higher, more practical, cutting speeds for turning a range of ferrous, aluminium and copper alloys (Eggleston *et al.*, 1959; Kobayashi and Thomsen, 1959). Both parts of the figure show each material to have its own characteristic behaviour. Both show that annealed steel machines with a lower shear plane angle than the same steel cold-rolled. Figure 2.14(b) marginally groups the data in a smaller area than does Figure 2.14(a). Certainly part b emphasizes the range of friction angles, common to all the materials, from 25° to 40° (friction coefficient from 0.47 to 0.84). As this book is machining-process centred, the view of part b is preferred.

Figure 2.15 gathers more data on this basis. Figure 2.15(a) shows that free-cutting steels

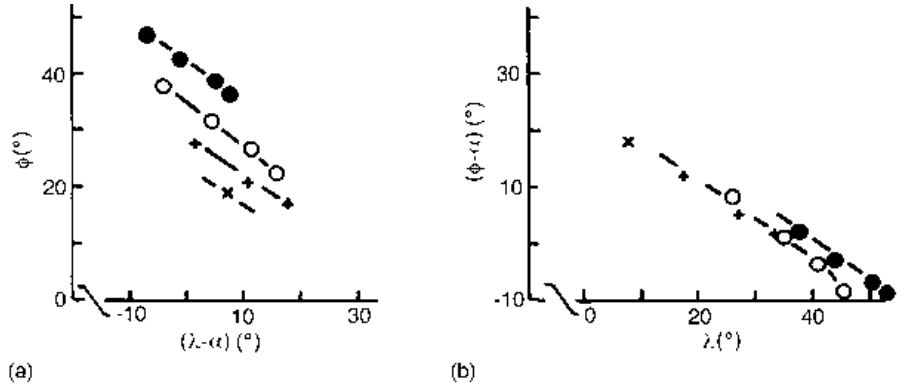


Fig. 2.13 ϕ–λ–α relationships for low speed turning of a free cutting steel with tools of different rake angle (0°x, 16°+, 30°o, 45°•), varying friction by selection of cutting fluid: (a) ϕ versus $(\lambda–\alpha)$ and (b) $(\phi–\alpha)$ versus λ (after Shaw, 1984)

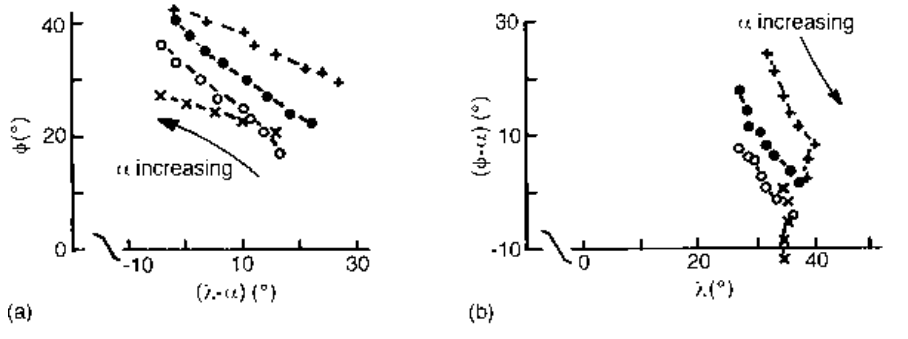


Fig. 2.14 ϕ–λ–α relationships for normal production speed turning by high speed steel tools, with rake angles from 5° to 40°, of cold rolled (•) and annealed (o) free cutting steel, an aluminium alloy (+) and an α-brass (×): (a) ϕ versus $(\lambda–\alpha)$ and (b) $(\phi–\alpha)$ versus λ (data from Eggleston *et al.*, 1959)

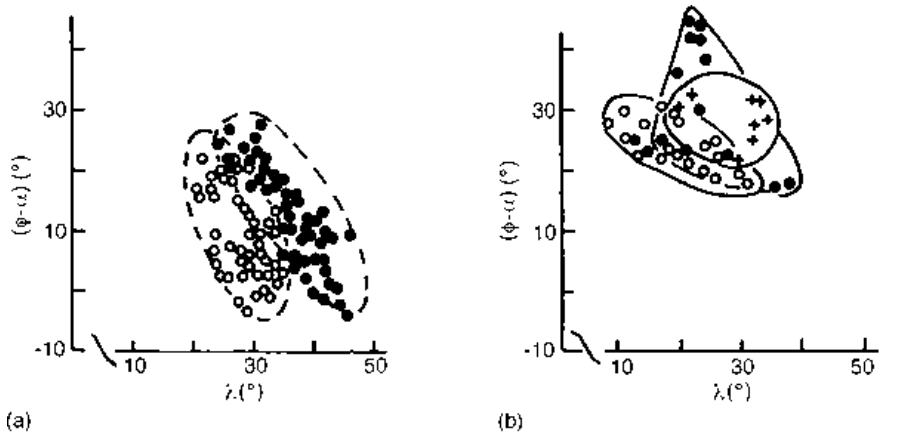


Fig. 2.15 ϕ–λ–α relationships compared for (a) free-machining (o) and non-free machining (•) carbon and low alloy steels; and (b) austenitic stainless and high manganese steels (o), nickel-chromium heat resistant (•) and titanium alloys (+) turned by cemented carbide and ceramic tooling

generally have lower friction coefficients (from 0.36 to 0.70) than non-free-cutting steels (from 0.47 to 1.00) when turned with high speed steel or cemented carbide tools (Childs, 1980a). Figure 2.15(b) extends the data to the machining of difficult materials such as austenitic stainless and high manganese steels, nickel-chromium and titanium alloys, by carbide and ceramic tools. Friction angles remain in the same range as for other materials but larger differences between shear plane and rake angle are found. Care must be taken in interpreting this last observation. Not only are lower rake angles used for the difficult to machine materials (from +10° to –5° for the data in the figure), biasing the data to larger ($\phi - \alpha$), but these materials also give serrated chips. The data in Figure 2.15(b) are averaged over the cycle of non-steady chip formation.

2.2.5 Specific energies and material stress levels in machining

In the preceding sections, basic force and moment equilibrium considerations have been used, with experimental observations, to establish the mechanical conditions of continuous chip formation. With the exception of the Merchant and Lee and Shaffer laws, prediction of chip shape has not been attempted. Predictive mechanics is left to Chapters 6 and after. In this section, by way of a summary, some final generalizations are made, concerning the energy used to form chips, and the level of contact stresses on the tool face.

The work done per unit machined volume, the specific work, in metal cutting is $F_c/(fd)$. The dimensionless specific work, may be defined as $F_c/(kfd)$. Equation (2.11) takes equation (2.5b) and manipulates it to

$$\frac{F_c}{kfd} = \frac{\cos(\lambda - \alpha)}{\sin\phi\cos(\phi + \lambda - \alpha)} \equiv \frac{1}{\tan\phi} + \tan(\phi + \lambda - \alpha) \tag{2.11}$$

From Figures 2.13 to 2.15, the range of observed ($\phi + \lambda - \alpha$) is 25° to 55° (except for the nickel-chromium and titanium alloys); and the range of λ is 20° to 45°. With these

numbers, the non-dimensional specific work may be calculated for a range of rake angles. Figure 2.16(a) gives, for rake angles from 0° to 30°, bounds to the specific work for $\tan(\phi + \lambda - \alpha)$ from 0.5 to 1.5 and for $\lambda = 20°$ to 45°. It summarizes the conflicts in designing a machining process for production. For a high rake angle tool ($\alpha = 30°$), specific work is relatively low and insensitive to changes in ϕ and λ. In such conditions an easily controlled and high quality process could be expected; but only high speed steel tools are tough enough to survive such a slender edge geometry (at least in sharp-edged, plane rake face form). At the other extreme ($\alpha = 0°$), cutting edges can withstand machining stresses, but the specific work is high and extremely sensitive to small variations in friction or shear plane angle. In practice, chamfered and grooved rake faces are developed to overcome these conflicts, but that is for a later chapter.

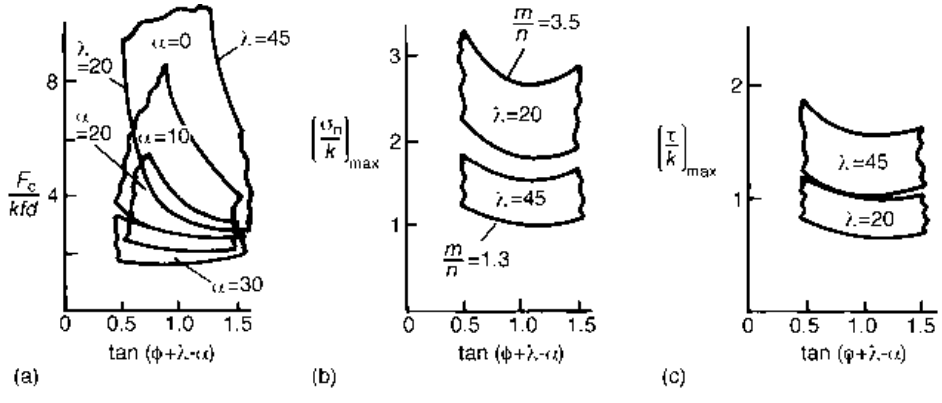


Fig. 2.16 Ranges of (a) dimensionless specific cutting force, (b) maximum normal contact stress and (c) maximum friction stress, for observed ranges of ϕ, λ, α (°) and m/n

Of the total specific work, some is expended on primary shear deformation and some on rake face friction work. The specific primary shear work, U_p, is the product of shear force $kfd/\sin\phi$ and velocity discontinuity on the plane (equation (2.3)). After 'non-dimensionalizing' with respect to kfd,

$$\frac{U_p}{kfd} = \frac{\cos\alpha}{\sin\phi\cos(\phi - \alpha)} \tag{2.12}$$

which is the same as the shear strain γ of equation (2.4a). The percentage of the primary work to the total can be found from the ratio of equation (2.12) to (2.11). For the same ranges of numbers as used in Figure 2.16(a), the percentage ranges from more than 80% when $\tan(\phi + \lambda - \alpha) = 0.5$, through more than 60% when $\tan(\phi + \lambda - \alpha) = 1.0$, to as little as 50% when $\tan(\phi + \lambda - \alpha) = 1.5$. The distribution of work between the primary shear region and the rake face is important to considerations of temperature increases in machining. Temperature increases are the subject of Section 2.3.

Finally, equations (2.5) can be used to determine the normal and friction forces on the tool face, and can be combined with equations (2.6) and (2.2) for the contact length between the chip and tool, in terms of the feed, to create expressions for the average normal and friction contact stresses on the tool:

$$\left(\frac{\sigma_n}{k}\right)_{av.} = \frac{n}{m}\,\frac{2\cos^2\lambda}{\sin 2(\phi+\lambda-\alpha)}; \qquad \left(\frac{\tau_n}{k}\right)_{av.} = \frac{n}{m}\,\frac{2\cos\lambda\,\sin\lambda}{\sin 2(\phi+\lambda-\alpha)} \tag{2.13}$$

In Section 2.2.3, the influence of m/n on contact stress distribution was considered, leading to Figure 2.12. The same considerations can be applied to deriving the peak contact stresses associated with the average stresses of equations (2.13). Figures 2.16(b) and (c) show ranges of peak normal and friction stress for the same data as given in Figure 2.16(a), for the practically observed range of m/n from 1.3 to 3.5. Peak normal stress ranges from one to three times k. Peak friction stress is calculated to be often greater than k. This, of course, is not physically realistic. The loads in machining are so high, and the lubrication so poor, that the classical law of friction – that friction stress is proportional to normal stress – breaks down near the cutting edge. Section 2.4 gives alternative friction modelling, first widely disseminated by Shaw (1984).

It has already been mentioned that the focus of this introductory mechanics section is descriptive and not predictive. However, the earliest predictive models for shear plane angle have been introduced – equations (2.9) and (2.10). In most cases, they give upper and lower bounds to the experimental observations. It may be asked what is the need for better prediction? The answer has two parts. First, as shown in Figure 2.16(a), the specific forces in machining (and hence related characteristics such as temperature rise and machined surface quality) are very sensitive to small variations in shear plane angle, for commonly used values of rake angle. Secondly, the cutting edge is a sacrificial part in the machining process, with an economic life often between 5 and 20 minutes (see Chapter 1). Small variations in mechanical characteristics can lead to large variations in economic life. It is the economic pressure to use cutting edges at their limit that drives the study of machining to ever greater accuracy and detail.

2.3 Thermal modelling

If all the primary shear work of equation (2.12) were converted to heat and all were convected into the chip, it would cause a mean temperature rise ΔT_1 in the chip

$$\Delta T_1 = \frac{k}{\rho C}\,\frac{\cos\alpha}{\sin\phi\,\cos(\phi-\alpha)} \equiv \frac{k\gamma}{\rho C} \tag{2.14}$$

where ρC is the heat capacity of the chip material. Table 2.2 gives some typical values of $k/(\rho C)$. Given the magnitudes of shear strains, greater than 2, that can occur in machining (Section 2.2), it is clear that significant temperature rises may occur in the chip. This is without considering the additional heating due to friction between the chip and tool. It is important to understand how much of the heat generated is convected into the chip and what are the additional temperature rises caused by friction with the tool.

The purpose of this section is to identify by simple analysis and observations the main parameters that influence temperature rise and their approximate effects. The outcome will be an understanding of what must be included in more complicated numerical models (the subject of a later chapter) if they are also to be more accurate. Thus, the simple view of chip formation, that the primary and secondary shear zones are planar, OA and OB of

Table 2.2 Mechanical and physical property data for machining heating calculations

Work material	Carbon/low alloy steels	Copper alloys	Aluminium alloys	Ni-Cr alloys	Titanium alloys
k [MPa]	400–800	300–500	120–400	500–800	500–700
ρC [MJ/m^3]	≈ 3.5	≈ 3.5	≈ 2.5	≈ 4.0	≈ 2.2
$k/\rho C$	110–220	85–140	50–160	120–200	220–320
ΔT_1 [°C][a]	230–470	180–300	110–340	250–430	470–680
K_{work} [W/m K]	25–45	100–400	100–300	15–20	6–15
K_{tool}[b] [W/m K]	20–50	80–120	100–500	80–120	50–120
K*	0.5–2	0.2–1	0.3–5	4–8	3–20

[a] ΔT_1 for $\gamma \approx 2.5$ and $\beta = 0.85$; [b] tool grades appropriate for work materials.

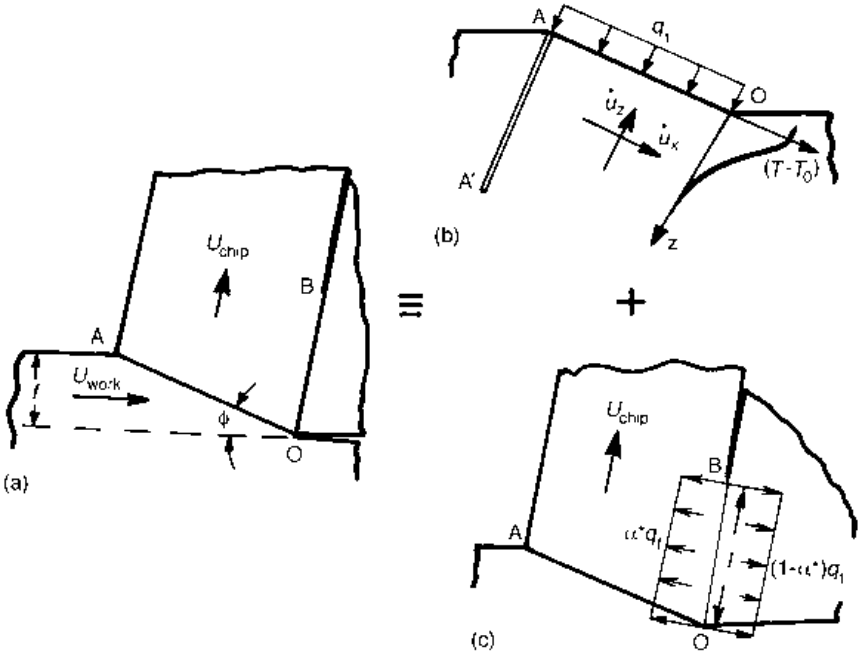


Fig. 2.17 (a) Work, chip and tool divided into (b) work and (c) chip and tool regions, for the purposes of temperature calculations

Figure 2.17(a), will be retained. Convective heat transfer that controls the escape of heat from OA to the workpiece (Figure 2.17(b)) is the focus of Section 2.3.1. How friction heat is divided between the chip and tool over OB (Figure 2.17(c)) and what temperature rise is caused by friction is the subject of Section 2.3.2. The heat transfer theory necessary for all this is given in Appendix 2.

2.3.1 Heating due to primary shear

The fraction of heat generated in primary shear, β, that flows into the work material is the main quantity calculated in this section. When it is known, the fraction $(1 - \beta)$ that is carried into the chip can also be estimated. The temperature rise in the chip depends on it.

Figure 2.17(b) shows a control volume AA′ fixed in the workpiece. The movement of the workpiece carries it both towards and past the shear plane with velocities $\dot{u}_z$ and $\dot{u}_x$, as shown. $\dot{u}_z = U_{\text{work}} \sin\phi$ and $\dot{u}_x = U_{\text{work}} \cos\phi$. When the control volume first reaches the shear plane (as shown in the figure), it starts to be heated. By the time the control volume reaches the cutting edge (at O), some temperature profile along z is established, also as shown in the figure. The rate of escape of heat to the work (per unit depth of cut), by convection, is then the integral over z of the product of the temperature rise, heat capacity of the work and the velocity $\dot{u}_x$:

$$Q_{\text{convected to work}} = \int_0^\infty \dot{u}_x (T - T_o) \rho C \, \mathrm{d}z \tag{2.15a}$$

The temperature profile $(T - T_0)$ is given in Appendix 2.3.1: once a steady state temperature is reached along Oz

$$Q_{\text{convected to work}} = \int_0^\infty \frac{\dot{u}_x}{\dot{u}_z} q_1 \mathrm{e}^{-\dot{u}_z z/\kappa} \, \mathrm{d}z \equiv \frac{\dot{u}_x q_1 \kappa}{\dot{u}_z^2} \tag{2.15b}$$

where q_1 is the shear plane work rate per unit area and κ is the thermal diffusivity of the work material. The total shear plane heating rate per unit depth of cut is the product of q_1 and the shear plane length, $q_1(\mathrm{f}/\sin\phi)$. The fraction β of heat that convects into the work is the ratio of equation (2.15b) to this. After considering that equation (2.15b) is a maximum estimate of heat into the work (the steady temperature distribution might not have been reached), and also after substituting for values of $\dot{u}_x$ and $\dot{u}_z$ in terms of U_{work}

$$\beta \leq \frac{\kappa}{U_{\text{work}} f \tan\phi} \tag{2.16}$$

According to equation (2.16), the escape of heat to the work is controlled by the thermal number $[U_{\text{work}} f \tan\phi/\kappa]$. This has the form of the Peclet number, familiar in heat transfer theory (Appendix A2.3.2). The larger it is, the less heat escapes and the more is convected into the chip. A more detailed, but still approximate, analysis has been made by Weiner (1955). Equation (2.16) agrees well with his work, provided the primary shear Peclet number is greater than 5. For lower values, equation (2.16), considered as an equality, rapidly becomes poor.

Figure 2.18(a) compares Weiner's and equation (2.16)'s predictions with experimental and numerical modelling results collected by Tay and reported by Oxley (1989). Weiner's result is in fair agreement with observation. β varies only weakly with $[U_{\text{work}} f \tan\phi/\kappa]$: a change of two orders of magnitude, from 0.1 to 10, is required of the latter to change β from 0.9 to 0.1. There is evidence that as $[U_{\text{work}} f \tan\phi/\kappa]$ increases above 10, β becomes limited between 0.1 and 0.2. This results from the finite width of the real shear plane. The implication from Figure 2.18(a) is that numerical models of primary shear heating need only include the finite thickness of the shear zone if $[U_{\text{work}} f \tan\phi/\kappa] > 10$, and then only if $(1 - \beta)$, the fraction of heat convected into the work, needs to be known to better than 10%.

Figure 2.18(b) takes the mean observed results in Figure 2.18(a) and, for $\phi = 25^\circ$, converts them to relations between U_{work} and f that result in $\beta = 0.15$ and 0.3, for $\kappa = 3$, 12 and 50 mm^2/s. These values of κ are representative of heat resistant alloys (stainless steels,

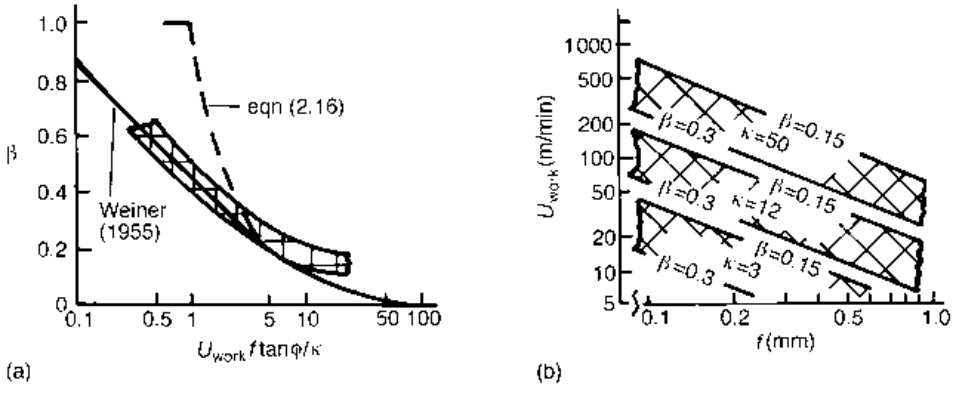


Fig. 2.18 (a) Theoretical (—, ---) and observed (hatched region) dependence of β on $[U_{work} f \tan\phi/\kappa]$; (b) iso-$\beta$ lines (β = 0.15 and 0.3) mapped onto a (U_{work}, f) plane for κ = 3, 12 and 50 mm^2/s and ϕ = 25°

nickel and titanium alloys), carbon and low alloy steels, and copper and aluminium alloys respectively. The speed and feed combinations that result coincide with the speed/feed ranges that are used in turning and milling for economic production (Chapter 1). In turning and milling practice, $\beta \approx 0.15$ is a reasonable approximation (actual variations with cutting conditions are considered in more detail in Chapter 3). A fraction of primary shear heat $(1 - \beta)$, or 0.85, then typically flows into the chip. The ΔT_1 of Table 2.2 give primary zone temperature rises when $\phi \approx 25^\circ$ and $\beta = 0.85$. For carbon and low alloy steels, copper and Ni-Cr alloys, these rises are less than half the melting temperature (in K): plastic flow stays within the bounds of cold working. However, for aluminium and titanium alloys, temperatures can rise to more than half the melting temperature: microstructural changes can be caused by the heating. Given that the primary shear acts on the workpiece, these simple considerations point to the possibility of workpiece thermal damage when machining aluminium and titanium alloys, even with sharp tools.

The suggested primary shear temperature rise in Table 2.2 of up to 680°C for titanium alloys is severe even from the point of view of the edge of the cutting tool. The further heating of the chip and tool due to friction is considered next.

2.3.2 Heating due to friction

The size of the friction stress τ between the chip and the tool has been discussed in Section 2.2.5. It gives rise to a friction heating rate per unit area of the chip/tool contact of $q_f = \tau U_{chip}$. Of this, some fraction α^* will flow into the chip and the remaining fraction $(1 - \alpha^*)$ will flow into the tool. The first question in considering the heating of the chip is what is the value of α^*?

The answer comes from recognizing that the contact area is common to the chip and the tool. Its temperature should be the same whether calculated from the point of view of the flow of heat in the tool or from the flow of heat in the chip. Exact calculations lead to the conclusion that α^* varies from point to point in the contact. Indeed so does q_f. To cope with such detail is beyond the purpose of this section. Here, an approximate analysis is developed to identify the physically important properties that control the average value of α^*

and to calculate the average temperature rise in the contact. It is supposed that q_f and α^* are constant over the contact, and that α^* takes a value such that the average contact temperature is the same whether calculated from heat flow in the tool or the chip. Figure 2.17(c) shows the situation of q_f and α^* constant over the contact length l between the chip and tool. The contact has a depth d (depth of cut) normal to the plane of the figure.

As far as the tool is concerned, there is heat flow into it over the rectangle fixed on its surface, of length l and width d. Appendix A2.2.5 considers the mean temperature rise over a rectangular heat source fixed on the surface of a semi-infinite solid. To the extent that the nose of the cutting tool in the machining case can be regarded as a quadrant of a semi-infinite solid, equation (A2.14) of Appendix 2 can be applied to give

$$(T - T_0)_{\text{av.tool contact}} = s_f \frac{(1 - \alpha^*)\tau_{av} U_{chip} l}{K_{tool}} \tag{2.17}$$

where T_0 is the ambient temperature, K is thermal conductivity and s_f is a shape factor depending on the contact area aspect ratio (d/l): for example, its value increases from 0.94 to 1.82 as d/l increases from 1 to 5.

As far as the chip is concerned, it moves past the heat source at the speed U_{chip}. Its temperature rise is governed by the theory of a moving heat source. This is summarized in Appendix A2.3. When the Peclet number $U_{chip}l/(4\kappa)$ is greater than 1, heat conducts a small distance into the chip compared with the chip thickness, in the time that an element of the chip passes the heat source. In this condition, equation (A2.17b) of Appendix 2 gives the average temperature rise due to friction heating. Remembering that the chip has already been heated above ambient by the primary shear,

$$(T - T_0)_{\text{av.chip contact}} = (1 - \beta) \frac{k\gamma}{(\rho C)_{work}} + 0.75 \frac{\alpha^* \tau_{av} U_{chip} l}{K_{work}} \left(\frac{\kappa_{work}}{U_{chip} l} \right)^{1/2} \tag{2.18}$$

Equating (2.17) to (2.18) leads, after minor rearrangement, to an expression for α^*:

$$\alpha^* \frac{\tau_{av}}{(k\gamma)} \frac{U_{chip} l}{\kappa_{work}} \left[0.75 \left(\frac{\kappa_{work}}{U_{chip} l} \right)^{1/2} + s_f \frac{K_{work}}{K_{tool}} \right] = s_f \frac{\tau_{av}}{(k\gamma)} \frac{(\rho C)_{work}}{K_{tool}} U_{chip} l - (1 - \beta) \tag{2.19}$$

τ_{av} is related to k, l to f and U_{chip} to U_{work} by functions of ϕ, λ, α and (m/n), as described previously, by combining equations (2.2), (2.3), (2.6) and (2.13). γ is also a function of ϕ and α. After elimination of τ_{av}, l and U_{chip} in favour of k, f and U_{work}, equation (2.19) leads to

$$\alpha^* \left[1 + \frac{0.75}{s_f} \frac{K_{tool}}{K_{work}} \left(\frac{n}{m} \frac{\cos\lambda \cos(\phi - \alpha)\tan\phi}{\sin(\phi + \lambda - \alpha)} \right)^{1/2} \left(\frac{\kappa_{work}}{U_{work} f \tan\phi} \right)^{1/2} \right] \tag{2.20a}$$

$$= 1 - \frac{(1 - \beta)}{s_f [U_{work} f \tan\phi / \kappa_{work}]} \left(\frac{K_{tool}}{K_{work}} \right) \frac{\cos\alpha \cos(\phi + \lambda - \alpha)}{\sin\lambda \cos\phi}$$

The manipulation has introduced the thermal number $[U_{work} f \tan\phi/\kappa_{work}]$. β depends on this too (Figure 2.18(a)). If typical ranges of ϕ, λ, α and (m/n), from Figures 2.10, 2.14 and 2.15 are substituted into equation (2.20a), the approximate relationship is found

$$\alpha^* \left[1 + \frac{(0.45 \pm 0.15)}{s_f} \left(\frac{K_{tool}}{K_{work}} \right) \left(\frac{\kappa_{work}}{U_{work} f \tan \phi} \right)^{1/2} \right] \tag{2.20b}$$

$$\approx 1 - \frac{(1.35 \pm 0.5)}{s_f} \frac{(1-\beta)}{[U_{work} f \tan \phi/\kappa_{work}]} \left(\frac{K_{tool}}{K_{work}} \right)$$

Figure 2.19(a) shows predicted values of α^* when observed β values from Figure 2.18(a) and the mean value coefficients 0.45 and 1.35 are used in equation (2.20b). A strong dependence on $[U_{work} f \tan\phi/\kappa_{work}]$ and the conductivity ratio $K^* = K_{tool}/K_{work}$ is seen, and a smaller but significant influence of the shape factor s_f. Predictions are only shown for $[U_{work} f \tan\phi/\kappa_{work}] > 0.5$: at lower values the assumption behind equation (2.18), that $U_{chip} 1/(4\kappa)$ is greater than 1, is invalid; and anyway friction heating becomes small and is not of interest. As a matter of fact, the assumption starts to fail for $[U_{work} f \tan\phi/\kappa_{work}] < 5$. Figure 2.19(a) contains a small correction to allow for this, according to low speed moving heat source theory (see Appendix A2.3.2).

Figure 2.19(a) reinforces the critical importance of the relative conductivities of the tool and work. When the tool is a poorer conductor than the work ($K^* < 1$), the main proportion of the friction heat flows into the chip. As K^* increases above 1, this is not always so. Indeed, a strong possibility develops that $\alpha^* < 0$. When this occurs, not only does all the friction heat flow into the tool, but so too does some of the heat generated in primary shear. The physical result is that the chip cools down as it flows over the rake face and the hottest part of the tool is the cutting edge. When $\alpha^* > 0$, the chip heats up as it passes over the tool: the hottest part of the tool is away from the cutting edge.

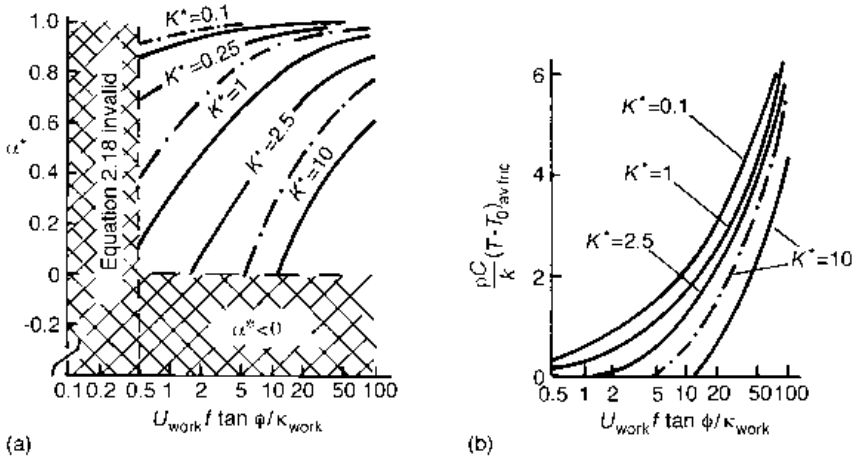


Fig. 2.19 Dependence of (a) α^* and (b) friction heating mean contact temperature rise on $[U_{work} f \tan\phi/\kappa_{work}]$, $K^* = K_{tool}/K_{work}$ from 0.1 to 10; $s_f = 1$ (—) and 2 (- · -)

From physical property data in Chapter 3, tool conductivities range from 20 to 50 W/m K, for P grade cemented carbides, high speed steels, cermets, alumina and silicon nitride based tools; to 80 to 120 W/m K for K grade carbides; up to ≈ 100 to 500 W/m K for polycrystalline diamond tools. Table 2.2 gives typical ranges of K^* for different groups of work materials, assumed to be cut with recommended tool grades (for example P grade carbides for carbon and low alloy steels, K grade carbides for non-ferrous materials, the possibility of polycrystalline diamond for aluminium alloys). The heat resistant Ni-Cr and Ti alloys (and austenitic stainless steels would be included in this group) are distinguished from the carbon/low alloy steels, copper and aluminium alloys by their larger K^* values. Particularly for the Ti alloys, there is a high possibility that α^* may be less than zero.

The analytical modelling that leads to Figure 2.19(a) is only approximate (because it deals only in average rake face quantities). Its value though is more than its quantitative results. It gives guidance on what is important to be included in more detailed numerical models. For example, in conditions in which $\alpha^* \approx 0$, small changes of operating conditions may have a large effect on the observed tool failure mode, from edge collapse when $\alpha^* < 0$ to cratering type failures as $\alpha^* > 0$ and the hottest part of the tool moves from the cutting edge. Figure 2.19(a) shows that the speed and feed at which $\alpha^* = 0$ for a particular work and tool combination will vary with the shape factor s_f. To study such conditions numerically would certainly require three-dimensional modelling.

Once α^* is determined, the temperature rise associated with it can be found. The second term on the right-hand side of equation (2.18) is the friction heating contribution to the average temperature of the chip/tool contact. After applying the same transformation and substitution of typical values of ϕ, λ, α and (m/n) that led to equation (2.20b)

$$(T - T_0)_{\text{av.friction}} = \frac{k}{(\rho C)_{\text{work}}} \alpha^* \, (0.7 \pm 0.2) \, [U_{\text{work}} f \tan \phi / \kappa_{\text{work}}]^{1/2} \qquad (2.21)$$

Figure 2.19(b) shows the predicted dependence of non-dimensional temperature rise on $[U_{\text{work}} f \tan\phi/\kappa_{\text{work}}]$, after substituting values of α^* from Figure 2.19(a) in equation (2.21).

In this section, an approximate approach has been taken to estimating the temperature rise in the primary shear zone and the average temperature rise on the rake face of the tool. One final step may be taken, to aid a comparison with observations and to summarize the limitations and value of the approach. The moving heat source theory in Appendix 2 concludes that for a uniform strength fast moving heat source and α^* constant over the contact, the maximum temperature rise due to friction is 1.5 times the average rise. The absolute maximum contact temperature between the chip and tool can thus be found from the sum of the primary shear heating (with β from Figure 2.18) and 1.5 times the temperature rise from equation (2.21) or from Figure 2.19(b). Equation (2.22) summarizes this.

$$\frac{(\rho C)_{\text{work}}}{k} (T - T_o)_{\text{max chip contact}} = (1 - \beta)\gamma + (1 \pm 0.3)\alpha^* [U_{\text{work}} f \tan \phi / \kappa_{\text{work}}]^{1/2} \qquad (2.22)$$

Examples of how temperature rises vary with cutting speed have been calculated from this, for a range of work material types. They are shown as the solid lines in Figure 2.20(a). Mean values of $k/(\rho C)$, K and κ for the different groups of work materials have been used, and have been taken from Table 2.2. Typical values of $\gamma = 2.5$ and $\phi = 25^\circ$ have

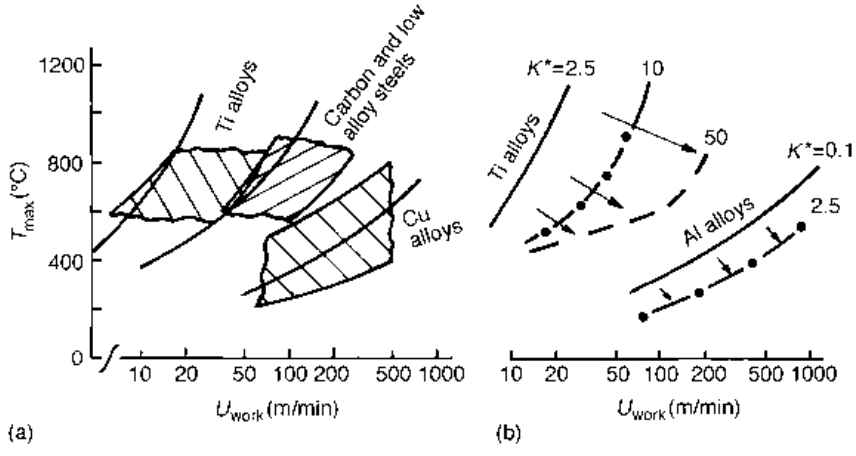


Fig. 2.20 (a) Predicted (—) and observed (hatched) dependence of maximum rake face temperature on cutting speed; (b) predicted influence of tool conductivity change

been arbitrarily chosen. A feed of 0.25 mm and K_{tool} = 30 W/m K (typical of a high speed steel tool and needed to assign a value to K^*) have been chosen so that a comparison can be made with the experimental results summarized by Trent (1991), which are shown as the hatched regions in the figure. These are the same results that were introduced in Chapter 1 (Figure 1.23). They are maximum temperatures deduced from observations of microstructural changes in tool steels, used to turn different titanium, ferrous and copper alloys at a feed of 0.25 mm.

The calculated results for the copper alloys fall in the middle of the experimentally observed range, but those for the titanium and ferrous alloys are close to the maximum observed temperatures. The overestimate for the titanium alloys arises mainly from the use of the mean value coefficient of 1.0 in the second term on the right of equation (2.22), rather than its lower limit of 0.7. For the ferrous alloys, the experimental measurements were probably for materials with k_{work} less than the mean value of 600 MPa assumed in the calculations. The overlap between theory, with all its simplifying, two-dimensional, steady state and other approximations, and experiments is enough to support the following conclusions. Temperature rise in metal machining depends most sensitively on the ratio of primary shear flow stress to heat capacity $k/(\rho C)$, on the shear strain γ and on $[U_{work} f \tan\phi/\kappa_{work}]$. The latter not only occurs explicitly in equation (2.22) but also controls the values of α^* and β. Of next importance are the ratio of tool to work conductivity, K^*, and the shape factor s_f. These also affect α^*, but are more important in some conditions than others. The tool rake angle and chip/tool friction coefficient mainly have an indirect influence on temperature, through their effect on γ and ϕ, although they are also the cause of the range of ± 0.3 around the mean value coefficient of 1.0 in the friction heating term of equation (2.22); and only practical values of rake angle have been considered in estimating that coefficient.

Equation (2.22), with Figures 2.18 and 2.19, is valuable for the understanding it gives of heat transfer in metal machining. It suggests ways that temperatures may be reduced, in conditions in which direct testing is difficult. For example, Figure 2.20(b) shows the predicted decrease in maximum rake face temperature for machining a titanium alloy on

changing from a cutting tool with K = 30 W/mK (K^* = 2.5), to one with K = 120 W/m K (K^* = 10) – K-type carbides are preferred to P-type for machining titanium alloys; and finally to one with K = 500 W/m K (K^* = 50) – polycrystalline diamond (PCD) tools are successfully used to machine titanium alloys; and for machining an aluminium alloy on increasing K_{tool} from 30 to 750 W/m K – another typical value for PCD tools (depending on grade). The reduced temperature with high thermal conductivity tools is one reason for choosing them – but the conductivity must be high relative to that of the work. Of course, an increase in tool conductivity, although it will reduce the rake face temperature, may, as a result of the changed balance of the ratio of heat flow into the tool to its conductivity, lead to higher flank face temperatures. If this were a problem, it might be overcome by the development of composite tools with a graded composition and thermal conductivity, from rake to flank region. Thus, equation (2.22) is qualitatively good enough to drive choices and development of tooling. It is not, however, quantitatively sufficient for the prediction of tool life. At the high temperatures shown in Figure 2.20, tool mechanical wear and failure properties, and also work plastic flow resistance, can be so sensitive to temperature that the uncertainties in the predictions of equation (2.22) are too large. These uncertainties come from the initial assumption of a uniform heat source over the chip/tool contact and the ± 30% uncertainty in the coefficient of its friction heating term. Furthermore, K can vary significantly with temperature over the temperature ranges that occur in machining: consequently, what values should be used in equation (2.22)? As was concluded in Section 2.2, the use of tools in a sacrificial mode drives the need for better, numerical modelling.

2.4 Friction, lubrication and wear

Up to this point, it has been assumed that the friction stress on the rake face is proportional to the normal stress. In other words the friction stress is related to the normal stress by a friction coefficient μ or friction angle λ ($\tan \lambda = \mu$). That has led to deductions from measurements (Figure 2.16) of peak normal stresses on the rake face of between one and three times k, and of peak friction stresses of up to almost twice k. The last is not believable, because a metal is not able to transmit a shear stress greater than its own shear flow stress. In this section, a closer look will be taken at the friction conditions and laws at the rake face. A closer look will also be taken at how the rake face may be lubricated. One of the first questions raised (Section 2.1) was how might a lubricant penetrate between the chip and the tool; and experimental results (Figure 2.7) suggest the answer is: only with difficulty. Finally, the subject of tool wear will be raised in the context of what is known about wear from general tribological (friction, lubrication and wear) studies.

2.4.1 Friction in metal cutting

One way to study the contact and friction stresses on the rake face is by direct measurement. However, this is difficult because the stresses are large and the contact area is small. Apart from some early experiments in which lead was cut with photoelastic polymeric tools (for example Chandrasekeran and Kapoor, 1965), the main experimental method has used a split cutting tool (Figure 2.21). Two segments of a tool are separately mounted, at least one part on a force measuring platform, with a small gap between them of width g.

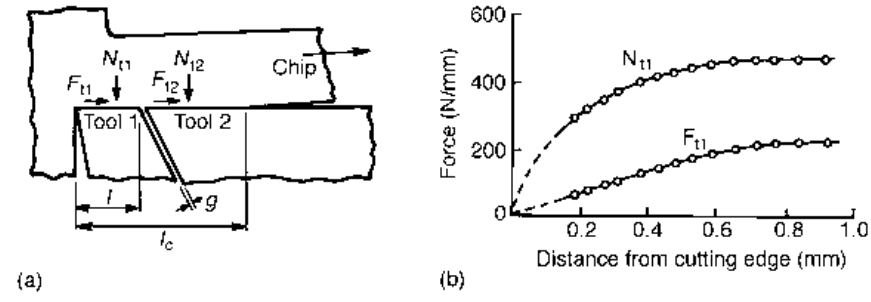


Fig. 2.21 (a) Schematic split tool, and (b) force measurements from it

This gap must be small enough that the chip flowing over the rake face does not extrude into the gap and large enough that the two parts of the tool do not touch as a result of any deflections caused by the forces. In Figure 2.21(a), the gap is shown a distance l from the cutting edge. When the length l is changed, for example as a result of grinding away the clearance face of the tool, the forces measured on the parts 1 and 2 of the tool also change. Figure 2.21(b) shows the increase with l of forces per unit width of cutting edge (depth of cut) on the front portion (Part 1) of the tool. The contact stresses on the rake face can be obtained from the rate of change of force with l:

$$\sigma_n = \frac{dN_{tl}}{dl}; \qquad \tau = \frac{dF_{tl}}{dl} \tag{2.23}$$

Use of the technique is limited to cutting work materials that do not break the split tool: to date, the upper limit of materials' Vickers hardness for success is about 3 GPa. There is a minimum value of l, below which the front tool becomes too fragile. In Figure 2.21(b) that value is about 0.2 mm, but measurements down to 0.1 mm have been claimed.

Split tool data are shown in Figure 2.22, for conditions listed in Table 2.3. In the figure, the contact stresses have been normalized by the shear stress k acting on the primary shear plane, calculated from equation (2.6c). The distance l from the cutting edge has been normalized by the chip thickness t. In most cases, the normal stress rises to a peak at the cutting edge, as suggested in Figures 2.11(b) and (c). However, in two cases (for aluminium and copper), the rise in normal stress towards the cutting edge is capped by a plateau. Peak normal stresses range from $0.7k$ to $2.5k$.

Friction stress also rises towards the cutting edge, but is always capped at a value $\leq k$. When friction stress is replotted against normal stress, or rather τ/k versus σ_n/k, as in the bottom panels of the figure, the two are seen to be proportional at low normal stress levels (in the region of contact farthest from the cutting edge) but at high normal stresses (near the cutting edge) the friction stress becomes independent of normal stress. (In the bottom right panel of Figure 2.22, the comments Elastic/Transition/Plastic, with the labels p_E/k = 0 or 1, are discussed later.)

The low stress region constant of proportionality μ ($\tau = \mu\sigma_n$) and the plateau stress ratio value m ($\tau = mk$) are listed in Table 2.3. These are also defined in the inset to Figure 2.23. These data are just examples. They demonstrate that on the rake face the friction stress is not everywhere proportional to the normal stress. At high normal stresses, the friction

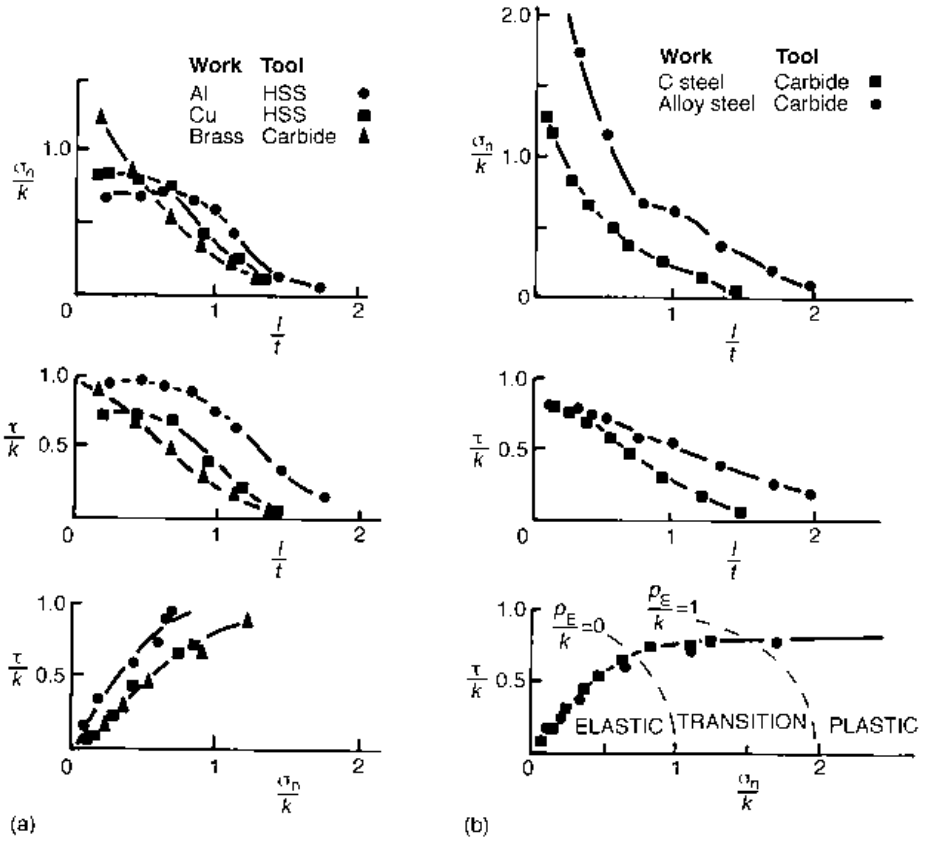


Fig. 2.22 Derived rake face stresses, (a) non-ferrous and (b) ferrous work materials

Table 2.3 Materials, conditions and sources of the data in Figure 2.22

Work/tool materials	α^o	U_{work} [m/min]	f [mm]	k [MPa]	μ	m	Data derived from
Al/HSS	20	50	0.2	130	1.4	0.95	Kato *et al.* (1972)
Cu/HSS	20	50	0.2	335	0.9	0.75	Kato *et al.* (1972)
Brass/carbide	30	48	0.3	450	0.9	0.95	Shirakashi and Usui (1973)
C steel/carbide	10	46	0.3	600	1.3	0.8	Shirakashi and Usui (1973)
Low alloy steel/ Carbide	0	100	0.2	600	1.3	0.8	Childs and Maekawa (1990)

stress approaches the shear flow stress of the work material; at low normal stress, the friction coefficient, from 0.9 to 1.4, is of a size that indicates very poor, if any, lubrication.

Recently, the split tool technique has been added to by measuring the temperature distribution over the rake face (see Chapter 5). Figure 2.23 contains data obtained by the authors on the dependence of m and μ on contact temperature. The data are for a 0.45%C plain carbon steel (•), 0.45%C and 0.09%C resulphurized free machining carbon steels (o) and a 0.08%C resulphurized and leaded free machining carbon steel (x), machined at cutting speeds from 50 to 250 m/min and feeds of 0.1 and 0.2 mm, by zero rake angle

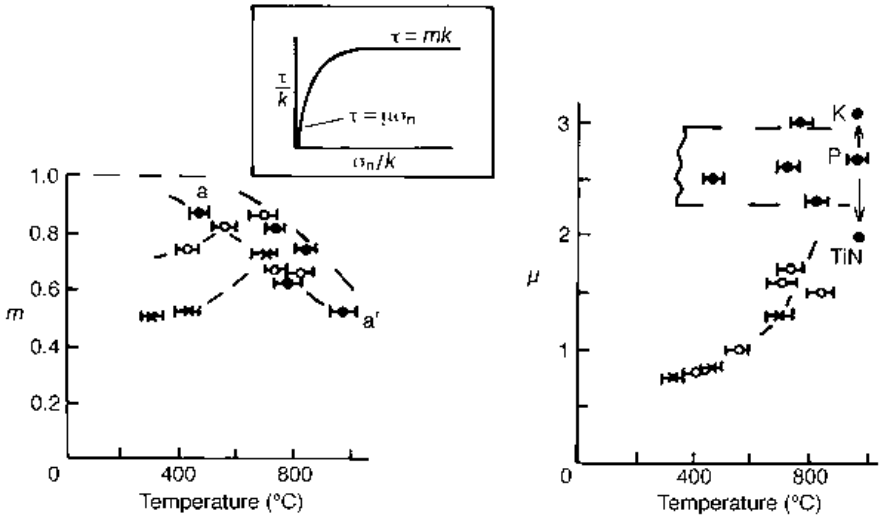


Fig. 2.23 m and μ variations with temperature for a plain carbon (•); a resulphurized (o); and a resulphurized and leaded (x) steel (authors' data)

tools, P grade carbides (see Chapter 3) unless otherwise stated. Adding sulphur to steel in small amounts results in the formation of manganese sulphide inclusions. These and lead can act as solid lubricants between the chip and tool.

Figure 2.23(a) shows two trends for the variations of m with temperature. First, there is a general trend aa′ for m to reduce with increasing temperature, from around 0.9 at 400˚C to as low as 0.5 at 1000˚C. However, m is also reduced at low temperatures by the presence of the free machining additives. Figure 2.23(b), for the variation of μ with temperature, also shows two trends. The plain carbon steel shows a friction coefficient independent of temperature. In this case it is a very high value (compared with the data in Table 2.3) between 2 and 3. In one case, marked P, the tool was changed first to a K-grade carbide (K) and then to a TiN cermet (TiN): this changed μ as shown. The free machining steels show a friction coefficient increasing rapidly with temperature, from around 0.7 at 300˚C towards 2 at 800˚C.

In order to simulate the machining process, it would be desirable to be able to model both the form of the variation of τ with σ_n, in terms of the coefficients m and μ; and also to understand what determines the values of m and μ.

The most simple friction model is to neglect altogether the low stress variation of τ with σ_n, to write

$$\tau = mk \tag{2.24a}$$

This is the approach taken in slip line field modelling (Chapter 6). A next best approximation is

$$\begin{aligned} \tau &= \mu\sigma_n \quad \text{if} \quad \mu\sigma_n < mk \\ \tau &= mk \quad \text{if} \quad \mu\sigma_n \geq mk \end{aligned} \tag{2.24b}$$

Shaw (1984, Ch.10) – who was one of the earliest researchers to study machining friction conditions – and also Shirakashi and Usui (1973), empirically blended the low stress into the high stress behaviour by an exponential function. In the present notation

$$\tau = mk\left(1 - \exp\left[-\frac{\mu\sigma_n}{mk}\right]\right) \tag{2.24c}$$

By noting that $e^{-x} \approx (1 - x)$ when x is small and positive, and tends to zero as x becomes large, it may be verified that equation (2.24c) approaches (2.24b) at extreme values of σ_n/k. The rate of change of τ with σ_n at intermediate levels of σ_n/k may be varied by the further empirical modification of equation (2.24d), where n^* is an exponent that in practice is found to vary between 1 and about 3:

$$\tau = mk\left(1 - \exp\left[-\left(\frac{\mu\sigma_n}{mk}\right)^{n^*}\right]\right)^{1/n^*} \tag{2.24d}$$

All of the forms (equations (2.24b) to (2.24(d)) have been used in finite element modelling of machining.

The form of the friction law

Why does the friction law have the form that it does? Figure 2.24(a) shows a chip sliding over a segment of the tool face of area A_n. The interface is imagined to be rough, so that contact with the chip may not occur over the whole area A_n but only over the high spots, or asperities. The contact then has a smaller real area, A_r. It is this real area of contact that transmits the contact forces. Suppose that it has a shear strength s. Then the friction force across it is

$$F = A_r s \tag{2.25a}$$

The nominal friction stress τ is F/A_n:

$$\tau = \frac{A_r}{A_n} s \tag{2.25b}$$

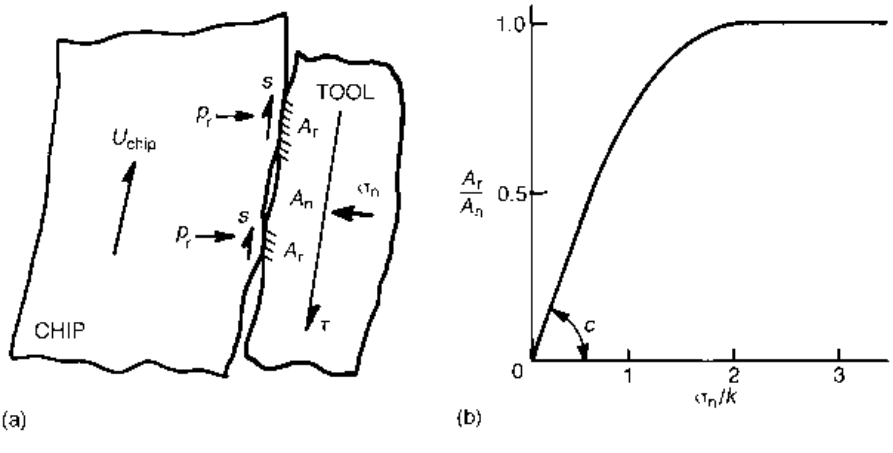


Fig. 2.24 (a) The sliding contact of a chip on a tool; (b) schematic dependence of A_r/A_n on σ_n/k

and its size relative to k, τ/k, is

$$\frac{\tau}{k} = \frac{A_r}{A_n}\,\frac{s}{k} \tag{2.25c}$$

It is easy to imagine that the degree of contact A_r/A_n increases with the nominal contact stress severity σ_n/k. Figure 2.24(b) shows a schematic variation. A_r/A_n is proportional to σ_n/k ($= c\sigma_n/k$) at low values of A_r/A_n (say $A_r/A_n < 0.5$). It becomes constant, equal to 1.0, at high values of σ_n/k. When these variations are substituted in equation (2.25c),

$$\frac{\tau}{k} = c\,\frac{\sigma_n}{k}\,\frac{s}{k} \qquad \text{if} \qquad \frac{A_r}{A_n} < 0.5 \tag{2.26a}$$

$$\frac{\tau}{k} = \frac{s}{k} \qquad \text{if} \qquad \frac{A_r}{A_n} = 1.0 \tag{2.26b}$$

This is the form of equation (2.24b). μ is identified as $c(s/k)$ and m as (s/k).

Degree of contact laws for metal machining

Theoretically deduced ranges of actual variation of A_r/A_n with σ_n/k are shown in Figure 2.25. They depend on how the chip asperity displacements, caused by the real contact stresses, are accommodated by the chip. At the lightest loadings, when both an asperity and the chip below it remain elastic (range EE), displacements are taken up by elastic compression. If the asperity becomes plastic but the chip below it remains elastic (range PE), plastically displaced material flows to the free surface round the contact. If the chip below the contact also becomes plastic (range PP), the asperity may sink into the chip. In equation

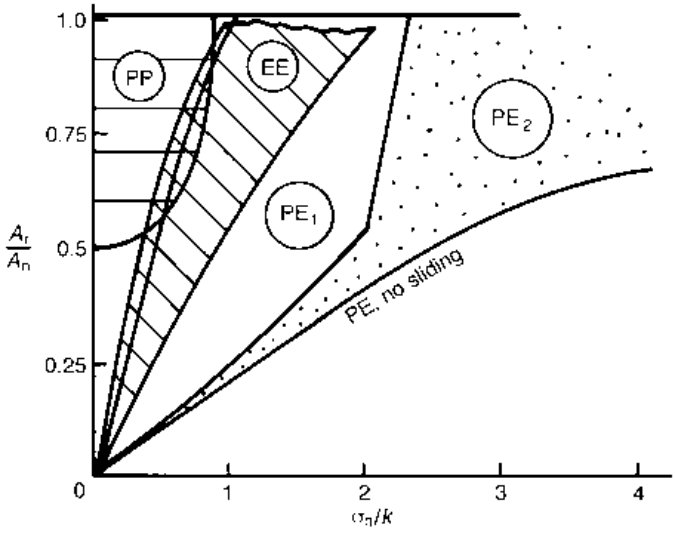


Fig. 2.25 Ranges of possible degrees of contact: PP = plastic asperity on plastic chip; PE = plastic asperity on elastic chip; EE = elastic asperity on elastic chip

(2.26), the resulting values of c range from 0.2 to greater than 1.0; values of σ_n/k at which A_r/A_n reaches 1.0 range from almost zero up to almost 10.

Thus, which regime occurs at the rake face and what are its laws strongly affects the friction laws in metal machining. Appendix 3 contains a review of contact mechanics relevant to these regimes, and the following sections summarize it. However, it must be acknowledged that understanding of this does not yet exist in sufficient detail to be able quantitatively to predict friction laws from first principles. The following sections may be omitted at a first reading.

Plastic asperities on a plastic chip – and the size of m

When asperities sink into the chip, how they do so depends not on the local conditions at the contact, but on the bulk plastic flow field. The lower is the hydrostatic stress in the bulk flow field, the more easily the asperity sinks. In Figure 2.25, the region PP has been drawn for likely hydrostatic stress values in the secondary shear zone of metal machining. A_r/A_n will certainly be unity when $\sigma_n/k > 1$, if the contact is in a plastically stressed region of the chip.

Whether the contact is in a plastically stressed part of the chip can be judged from the local values of τ/k, σ_n/k and the relative size of the plastic field hydrostatic stress level p_E/k. In Figure 2.22 (bottom right panel) the elastic–plastic borderline (from Appendix 3) is superimposed on the contact stress distribution for two values of p_E/k. Almost the whole of the plateau friction stress region is in the plastic (secondary shear) region, with $\sigma_n/k > 1$. The plateau region is consistent with $A_r/A_n = 1$.

The values of m, Table 2.3, may then be interpreted as the ratio of the shear flow strength at the chip tool interface to the primary shear flow stress of the chip material. The example results of Figure 2.23(a) suggest two causes for m being less than 1. First, if there is no solid lubricant phase in the work material, m can approach 1 at low temperatures, but reduces as temperature increases. The proposition that m is controlled by the shear flow stress of the chip material in the high temperature conditions in which it finds itself on the rake face is now well accepted. In equations (2.24) it is common to put $m = 1$ and to redefine k as the local, rather than the primary shear plane, shear stress. The problem is to determine how the local shear flow stress varies with temperature at the high strains and strain rates experienced in the secondary shear zone. This advanced topic is returned to in Chapters 6 and 7.

Secondly, it is clear from Figure 2.23(a) that it is possible, at least in the case of steel with manganese sulphide and lead, to lubricate the interface. Then m is a measure of shear stress of the solid lubricant relative to the chip, or of the fraction of the interface covered by the lubricant. Although there is good qualitative understanding, there is not at the moment a model that can predict how – changes with changes in the distribution of manganese sulphide and lead: experimentally determined values must be used in machining simulations.

Asperities on an elastic foundation – and the size of μ

The bottom right panel of Figure 2.22 also indicates that, at low contact stresses, certainly in this case when $\sigma_n/k < 0.7$, the chip beneath the asperities is elastic. Whether the asperites are elastically or plastically stressed then depends on the roughness of the tool face and on the level of s/k that exists. Appendix 3 introduces the concept of the plasticity index $(E^*/k)\Delta_q$, where k is the local shear stress of the asperity, E^* is an average Young's modulus for the

asperity and tool material: $1/E^* = 1/E_{\text{asperity}} + 1/E_{\text{tool}}$ and Δ_q is the root mean square slope of the surface roughness of the tool face. When s/k is less than 0.5, an asperity is totally elastic if the plasticity index is less than 5 and totally plastic if it is greater than 50. As s/k increases to 1, these critical values of the plasticity index reduce. In the large s/k conditions of metal machining, an asperity is expected to be fully plastic if

$$\Delta_q \geq 10 \frac{k_{\text{local}}}{E^*} \tag{2.27}$$

Typical roughnesses of insert cutting tool rake faces are shown in Figure 2.26 (the much larger vertical than horizontal magnification of these profiles should be pointed out). Trace (a) is typical of CVD (chemical vapour deposition – see Chapter 3) coated inserts and (b) of ground inserts. Sometimes inserts are better finished, trace (c). Table 2.4 lists the ranges of measured tool face roughnesses for inserts of these three types found in the authors' workshops. It also records values of $10k_{\text{local}}/E^*$ calculated from data in Table 2.3, taking $k_{\text{local}} = mk$, E for high speed steel tools to be 210 GPa and for cemented carbides to be 550 GPa. In all cases, for the coated and the ground tools, the relative sizes of Δ_q and $10k_{\text{local}}/E^*$ cause the asperities to be plastically stressed (although the brass/carbide case is marginal).

Consequently, in the majority of machining applications, the asperities in the lightly loaded region, where the chip is leaving the rake face contact, are plastically stressed. Then, the relations between A_r/A_n and σ_n/k are as expected from the regions marked PE_1

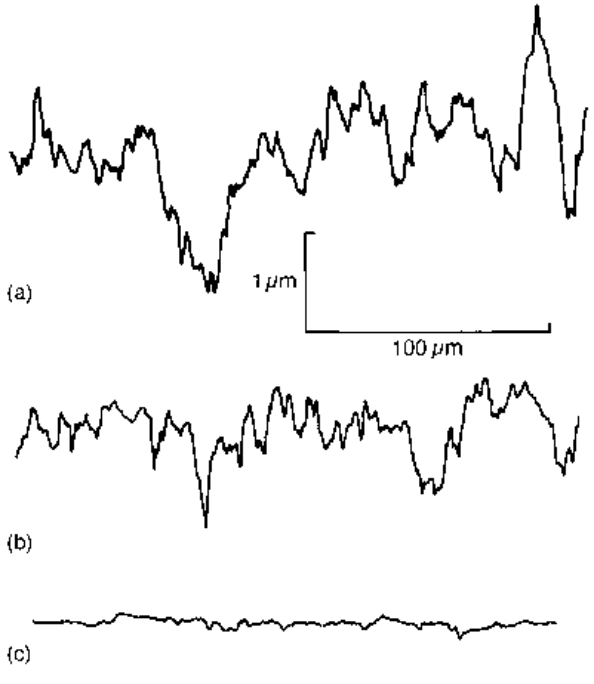


Fig. 2.26 Roughness profiles from (a) CVD coated, (b) ground and (c) superfinished insert tools

Table 2.4 Tool surface roughness and contact stress severity data

Tool finish	Roughness data R_a [μm]	Δ_q, °	$10k_{local}/E^*$, ° Al/ HSS	Cu/ HSS	Brass/ Carbide	Steel/ Carbide
CVD coated	0.2–0.5	3–7	1.2	1.9	2.8	1.8
Ground	0.1–0.25	2–4	1.2	1.9	2.8	1.8
Super-finished	0.03	0.4	1.2	1.9	2.8	1.8

and PE_2 in Figure 2.25. PE_1 represents theoretical analyses (Appendix 3) when the roughness is imagined to be on the tool surface and PE_2 when it is imagined to be on the chip. However, for large values of s/k_{local}, both regions have almost the same upper boundary, with c (equation (2.26)) approximately equal to 1. One would then expect

$$\mu \approx \frac{s}{k_{local}} \tag{2.28}$$

In those circumstances, when μ is measured to be < 1, this seems to be a reasonable relation. For example, in Figure 2.23, for the free machining steels when the rake face temperature is below 600˚C, μ is roughly the same as the ratio of m for the steel to that for the plain carbon steel at the same temperature. However, equation (2.28) cannot explain observations of $\mu > 1$, of the sort recorded in Figure 2.23(b) for the non-free machining steel or for the free machining steels above 600˚C.

Friction coefficients greater than 1.0

The plastic contact mechanics modelling reviewed in Appendix 3, which leads to $c \leq 1$, for the most part assumes that the asperity does not work harden and that the load on the asperity is constant through its make and break life cycle. In the final section of Appendix 3 there is a brief speculation about departures from these assumptions that could lead to larger values of c and to $\mu > 1$. All proposals require the shear strength of the junction to be maintained while the normal stress is unloaded. It is certain that, for this to occur, the strongest levels of adhesion must exist between the asperities and the tool. The freshly formed, unoxidized, nature of the chip surface, created by the parting of the chip from the work typically less than 10^{-3}s before it reaches the end of the contact length, and the high temperatures reached at high cutting speeds, are just the conditions that could promote strong adhesion (or friction welding). However, there is, at the moment, no quantitative theory to relate friction coefficients greater than 1 to the underlying asperity plastic properties and state of the interface.

The proper modelling of friction is crucial to the successful simulation of the machining process. This section, with Appendix 3, is important in setting current knowledge in a contact mechanics framework, but there is still work to be done before friction in metal machining is fully understood.

2.4.2 Lubrication in metal cutting

The previous section has emphasized the high friction conditions that exist between a chip and tool, in the absence of solid lubricants. The conditions that lead to high friction are

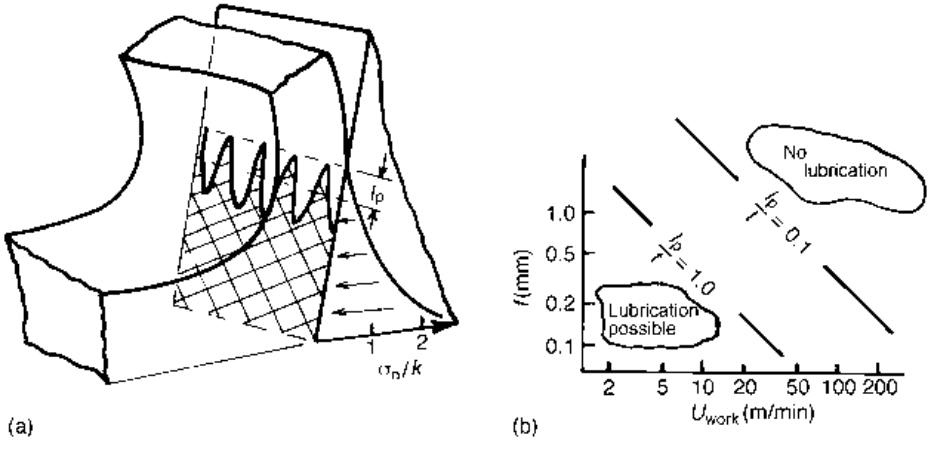


Fig. 2.27 (a) Defining the penetration distance l_p of the lubricant into the rear of the contact region and (b) derived feed/speed regions of complete and negligible penetration, for oxygen

high cutting speeds – for steels, speeds greater than around 100 m/min when the feed rate is 0.1 to 0.2 mm. However, earlier in this chapter (Figure 2.7) liquid lubrication was demonstrated at low cutting speeds; and one of the earliest questions asked of metal cutting (Section 2.1) was how can lubricant penetrate the rake face contact?

The question can now be asked in the context of the contact mechanics of the previous section. Figure 2.27 shows, somewhat schematically, the contact between the chip and tool. The hatched region represents the real area of contact, covering 100% of the contact near the cutting edge, where the normal stress is high, and reducing to zero towards the end of the contact. It is now generally agreed that neither gaseous nor liquid lubricants can penetrate the 100% real contact region, but they can infiltrate along the non-contact channels at the rear of the contact. These channels may typically be from half to one chip thickness long, depending on the normal contact stress distribution (Figure 2.22). Their height depends on the surface roughness of the cutting tool, but is typically 0.5 to 1 μm (Figure 2.26). If the lubricant reacts with the chip to reduce friction in the region of the channels, the resistance to chip flow is reduced, the primary shear plane angle increases, the chip becomes thinner and unpeels from the tool. Thus, a lubricant does not have to penetrate the whole contact: by attacking at the edge, it can reduce the whole. So the question becomes: what is the distance l_p (Figure 2.27) that a gas or liquid can penetrate along the channels? The following answer, for the penetration of gaseous oxygen and liquid carbon tetrachloride along channels of height h, is based on work by Williams (1977).

It is imagined that the maximum penetration results from a balance of two opposing transport mechanisms: the motion of the chip carrying the gas or liquid out of the contact and the pressures driving them in. For a gas, absorption on to the back of the freshly formed chip is the mechanism of removal from the contact. The absorption creates a gas pressure gradient along the channel which drives the gas in. Williams identified two mechanisms of inward flow, based on the kinetic theory of gases: viscous (Poiseuille) flow at high gas vapour pressure and Knudsen flow at low pressures, when the mean free path of the gas is greater than the channel height h. He showed that l_p (mm) is inversely

proportional to the chip velocity U_{chip}(m/min) with the constant of proportionality depending on the gas molecular diameter, obtained from its molecular weight M and its density in the liquid state ρ_{liquid} (kg/m^3), on its vapour pressure p_v (Pa), its viscosity η (Pa s) absolute temperature θ_T and on the height h (μm). For a channel much wider than its height

$$l_p U_{chip} = \text{the larger of } 3.3 \times 10^{-10} \frac{h^3}{\eta} \frac{p_v^2}{\theta_T} \left(\frac{M}{\rho_{liquid}} \right)^{2/3} \text{(Poiseuille)} \quad (2.29a)$$

or

$$0.71 h^2 p_v \left(\frac{M}{\theta_T^3 \rho_{liquid}^4} \right)^{1/6} \text{(Knudsen)} \quad (2.29b)$$

For oxygen, at its normal partial pressure in air of $\approx 2 \times 10^4$ Pa, and $M = 32$, $\rho_{liquid} = 1145$ kg/m^3, $\eta = 20 \times 10^{-6}$ Pa s, $\theta_T = 293$ and for $h = 0.5$ μm,

$$l_p U_{chip} = 3.4 \quad (2.30a)$$

This is about half the value given by Williams, because of different assumptions about the cross-sectional shape of the channels; and it does depend strongly on the assumed value of h.

Because of volume conservation, the product of U_{chip} and chip thickness t is the same as of U_{work} and feed f. Equation (2.30a) can therefore be modified to

$$\left(\frac{l_p}{t} \right) (f U_{work}) = 3.4 \quad (2.30b)$$

At feeds and speeds for which l_p/t is calculated to be > 1, total penetration of oxygen into the channels is expected. When $l_p/t < 0.1$, penetration may be considered negligible. Figure 2.27 marks these regions as possibly lubricated, and not lubricated, respectively. It is important because it shows a size effect for the effectiveness of lubrication. Williams (1977) also considered the penetration of liquids into the contact, driven by capillary forces and retarded by shear flow between the chip and the tool. For carbon tetrachloride liquid (which also has a significant vapour phase contribution to its penetration) he concluded the limiting feeds and speeds for lubrication were about the same as for oxygen.

Although it is certain that there can be no lubrication in the 'no lubrication' region of Figure 2.27, it is not certain that there will be lubrication in the 'possible lubrication' region. Whatever penetrates the channels must also have time to react and form a low friction layer. The time to react has also been studied by Williams (Wakabayashi *et al.*, 1995). It seems that this, rather than the ability to penetrate the channels, can be the controlling step for effective lubrication.

It is not the purpose of this section to expand on the effectiveness of different lubricating fluids for low speed applications. This has been covered elsewhere, for example Shaw (1984). Rather, it is to gain an understanding of the inability of liquids or gases to influence the contact at high cutting speeds. The reason why cutting fluids are used at high speeds is to cool the work material and to flush away swarf.

2.4.3 Wear in metal cutting

Finally, the sliding of the chip over the rake face, and of the work past the flank, causes the tool to wear away. Tool wear will be considered in detail in Chapter 4. Here, the purpose is briefly to review knowledge of wear from other studies, to create a standard to which tool wear can be related.

One of the most simple types of wear test is a pin on disc test (Figure 2.28). A cylindrical pin of cross-section A is pressed with a load W against a rotating disc which has some sliding speed U against the pin. The rate of loss of height, h, of the pin is measured against time. Usually there is an initial, running-in, time of high wear rate, before a constant, lower, rate is established. A common observation is that, in the steady state, the wear volume rate, $A\mathrm{d}h/\mathrm{d}t$ in this example, is proportional to W and the sliding speed. Archard's wear law (Archard and Hirst, 1956) may be written

$$\frac{\mathrm{d}h}{\mathrm{d}t} = k_{\mathrm{swr}} \frac{W}{A} U \equiv k_{\mathrm{swr}} \sigma_{\mathrm{n}} U \tag{2.31a}$$

where the constant of proportionality k_{swr} is called the specific wear rate and has units of inverse pressure. (In the wear literature k_{swr} is written k, but k has already been used in this book for a metal's shear flow stress.)

The proportionality of wear rate to load and speed is perhaps obvious. However, Archard considered the mechanics of contact to establish likely values for k_{swr}. He considered two types of contact, abrasive and adhesive (Figure 2.29) – the terminology is expanded on in Appendix 3. In the abrasive case, the disc surface consists of hard, sharp conical asperities (as might be found on abrasive papers or a grinding wheel). They dig

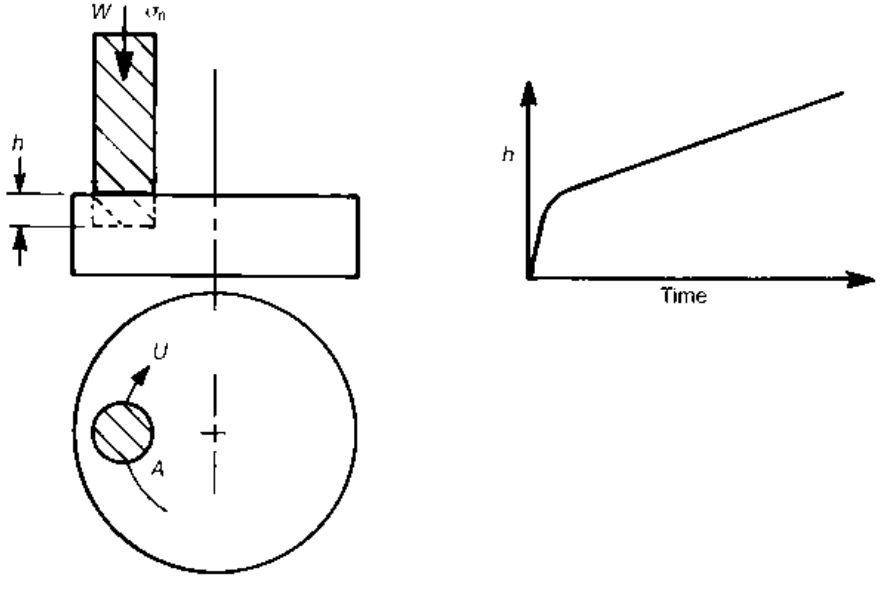


Fig. 2.28 A pin on disc wear test and a typical variation of pin height with time

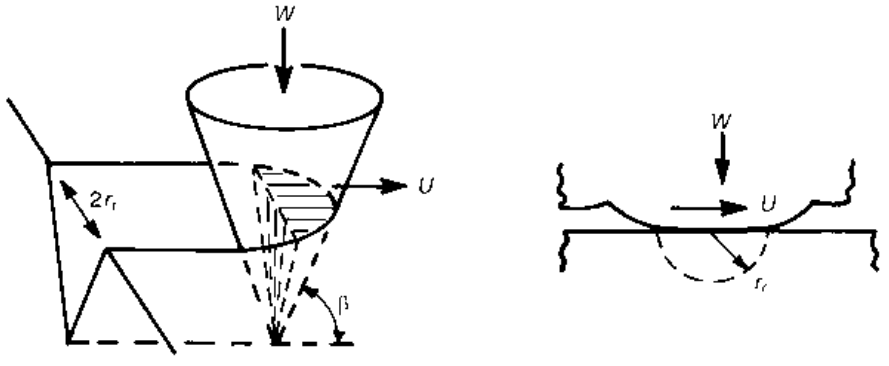


Fig. 2.29 Schematic views of abrasive and adhesive wear mechanisms

into the softer pin to create a number of individual real contacts, each of width $2r_r$. As a result of sliding, a scratch is formed of depth $r_r \tan\beta$, where β is the slope of the cones. If it is supposed that all the scratch volume becomes wear debris, the volume wear per unit time is $Ur_r^2 \tan\beta$. At the time Archard was writing, the analogy was made between the indentation of the cone into the flat and a hardness test, to relate the contact width to the load W on the cone. Noting that, during sliding, the load W is supported on the semicircle of area $\pi r_r^2/2$, r_r^2 was equated to $(2/\pi)(W/H)$, where H is approximately the Vickers or Brinell hardness of the softer surface. By substituting this into the expression for the scratch volume and summing over the large number of scratches that contribute to the wear process, it is easy to convert equation (2.31a) to the form of (2.31b), where a dimensionless wear coefficient K has been introduced instead of the specific wear rate k_{swr}, with a magnitude as written for this abrasive example.

A similarly simple model for adhesive wear (also Figure 2.29) assumes that a hemispherical wear particle of radius r_r is torn from the surface every time an asperity slides a distance $2r_r$, and that the real contact pressure is also H. It leads to the adhesive wear estimate of K also being included in equation (2.31b)

$$\frac{dh}{dt} = \frac{K}{H}\sigma_n U; \quad K = \frac{2\tan\beta}{\pi} \text{ for abrasive wear} \qquad = \frac{1}{3} \text{ for adhesive wear} \tag{2.31b}$$

If these equations were being derived today, there would be discussion as to whether the real contact pressure was H (equivalent to $5k$) or only to k (Section 2.4.1 and Appendix 3). However, such discussion is pointless. It is found that the K values so deduced are orders of magnitude different from those measured in experiments. Actual wear mechanisms are not nearly as severe as imagined in these examples. Different asperity failure mechanisms are observed, depending on the surface roughness, through the plasticity index already introduced in Section 2.4.1 and on the level of adhesion expressed as s/k or m. Figure 2.30 is a wear mechanism map showing what failure mode occurs in what conditions. It also shows what ranges of K are typical of those modes (developed from Childs, 1980b, 1988).

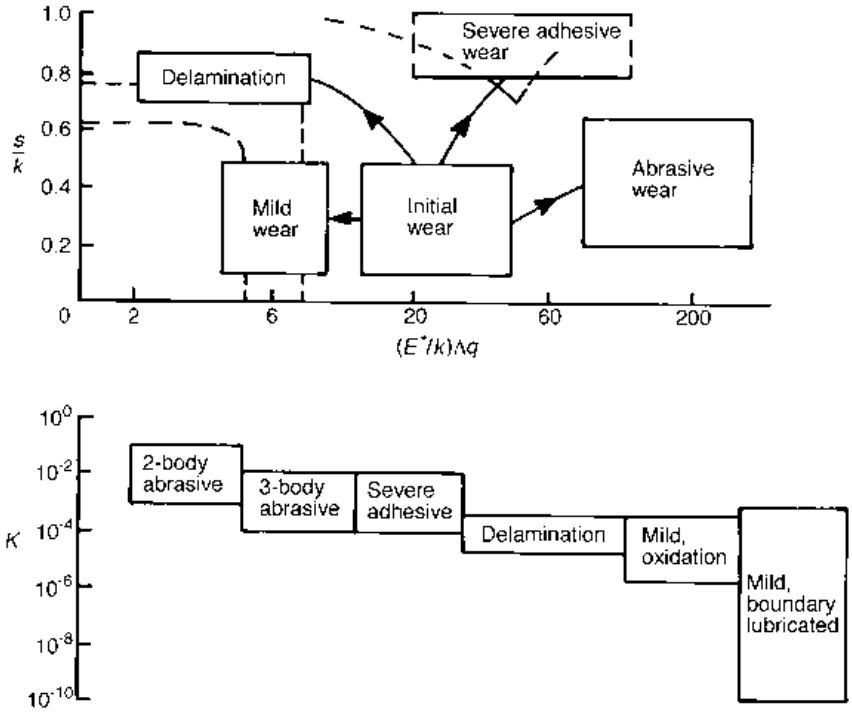


Fig. 2.30 A wear mechanism map

The initial wear region is the running-in regime of Figure 2.28. Surface smoothing occurs until the contacting asperities deform mainly elastically. If the surface adhesion is small (mild wear region), material is first oxidized before it is removed – values of K from 10^{-4} to 10^{-10} are measured (all the data are for experiments in air, nominally at room temperature). At higher adhesions subsurface fatigue (delamination) is found, with K around 10^{-4}. Sometimes, running-in does not occur and surfaces do tear themselves apart (severe adhesive wear), but even then K is found to be only 10^{-2} to 10^{-3}, compared with the value of 1/3 predicted above. Finally, if abrasive conditions do exist, K is found between 10^{-1} and 10^{-4}, depending on whether the abrasive is fixed on one surface (2-body) or is loose (3-body).

What is the relevance of this to metal machining? In Chapter 1, it was described how the economics of machining lead to the use of, for example, cemented carbide tools at cutting speeds and feeds such that the tools last only 5 to 10 minutes before wearing out. Definitions of wear-out differ from application to application, but common ones are that the flank wear length is less than 300 μm, or that the depth of any crater on the rake face is less than 60 μm. Figure 2.31(a) shows a worn tool, with crater depth h_c and flank depth wear h_f. h_f is related to the length of the wear land by tan γ, where γ is the flank clearance angle. Figures 2.31(b) and (c) are examples of wear measured for a low alloy steel at a feed of 0.12 mm and a cutting speed of 225 m/min, which is near the economic speed. For the flank, $dh_f/dt \approx 2$ μm/min; for the crater example $dh_c/dt \approx 7$ μm/min. Supposing the contact stress level is characterized by $\sigma_n/k \approx 1$, and noting that $H \equiv 5k$, values of K, from equation 2.31(b), are 4×10^{-8} on the flank, up to 3×10^{-7} on the rake (the speed of the chip was

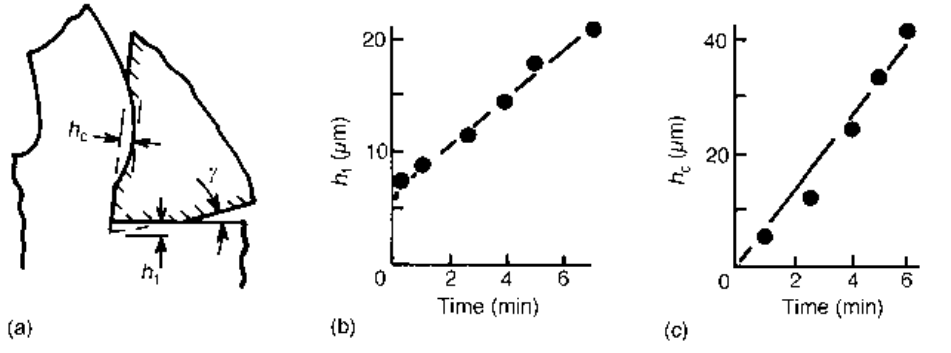


Fig. 2.31 (a) Flank and crater tool wear regions, with typical (b) flank and (c) crater wear observations

half that of the work). Considering that s/k is large in machining, these values are smaller than expected from the general wear testing experience summarized in Figure 2.29. (There is another point: the proportionality between dh/dt and σ_n/k in equation (2.31) is only established for conditions in which $A_r/A_n < 0.5$. Values larger than this occur over much of the tool contacts in machining. However, the uncertainty that this places in the deduced values of K is not likely to alter the orders of magnitude deduced for its values.)

There is one point to be made: the K values in Figure 2.30 are appropriate for the wear of the chip and work by the tool, rather than of the tool by the chip or work! In Figure 2.30, the plasticity index is, in effect, the ratio of the work material's real contact stress to its shear flow stress. To use the map to determine wear mechanisms in the tool, it seems appropriate to redefine the index as the ratio of the contact stress in the work to the tool material's shear flow stress. For typical tool materials (HV = 10 GPa to 15 GPa) and work materials (say HV = 2.5 GPa), this would effectively reduce the plasticity index value for the tool about fivefold relative to the work. For typical work plasticity index values of about 20 (Table 2.4), this would place the tool value at about 4, in the elastic range of Figure 2.30. The mechanisms available for tool wear are likely to be fatigue and chemical reaction (oxidation) with the atmosphere.

This conclusion is based on a continuum view of contact mechanics. In practice, work materials contain hard abrasive phases and tool materials contain relatively soft binding phases, so abrasion occurs on a microstructural scale. The transfer of work material to the tool, by severe adhesive wear, can also increase the tool stresses. At the temperature of cutting, chemical reactions can occur between the tool and work material as well as with the atmosphere. The story of abrasive, mechanical fatigue, adhesive and reaction wear of cutting tools is developed in Chapter 4.

2.5 Summary

The sections of this chapter have established the severe mechanical and thermal conditions typical of machining. A certain amount of factual information has been gathered and deductions made from it, but for the most part this has been at the level of observation. Predictive mechanics is taken up in the second half of this book, from Chapters 6 onwards.

First however, materials aspects of, and experimental techniques for, machining studies are introduced in Chapters 3 to 5.

References

Archard, J. F. and Hirst, W. (1956) The wear of metals under unlubricated conditions. *Proc. Roy. Soc. Lond.* **A236**, 397–410.

Boothroyd, G. and Knight, W. A. (1989) *Fundamentals of Machining and Machine Tools*. New York: Marcel Dekker.

Boston, O. W. (1926) A research in the elements of metal cutting. *Trans. ASME* **48**, 749–848.

Chandrasekeran, H. and Kapoor, D. V. (1965) Photoelastic analysis of tool-chip interface stresses. *Trans ASME J. Eng. Ind.* **87B**, 495–502.

Childs, T. H. C. (1972) The rake face action of cutting lubricants. *Proc. I. Mech. E. Lond.* **186**, 717–727.

Childs, T. H. C. (1980a) Elastic effects in metal cutting chip formation. *Int. J. Mech. Sci.* **22**, 457–466.

Childs, T. H. C. (1980b) The sliding wear mechanisms of metals, mainly steels. *Tribology International* **13**, 285–293 .

Childs, T. H. C. (1988) The mapping of metallic sliding wear. *Proc. I. Mech. E. Lond.* **202 Pt. C**, 379–395.

Childs, T. H. C. and Maekawa, K. (1990) Computer aided simulation of chip flow and tool wear. *Wear* **139**, 235–250.

Childs, T. H. C., Richings, D and Wilcox, A. B. (1972) Metal cutting: mechanics, surface physics and metallurgy. *Int. J. Mech. Sci.* **14**, 359–375.

Eggleston, D. M., Herzog, R. and Thomsen, E. G. Some additional studies of the angle relationships in metal cutting. *Trans ASME J. Eng. Ind.* **81B**, 263–279.

Herbert, E. G. (1928) Report on machinability. *Proc. I. Mech. E. London* **ii**, 775–825.

Kato, S., Yamaguchi, Y. and Yamada, M. (1972) Stress distribution at the interface between chip and tool in machining. *Trans ASME J. Eng. Ind.* **94B**, 683–689.

Kobayashi, S. and Thomsen, E. G. (1959) Some observations on the shearing process in metal cutting. *Trans ASME J. Eng. Ind.* **81B**, 251–262.

Lee, E. H. and Shaffer, B. W. (1951) The theory of plasticity applied to a problem of machining. *Trans. ASME J. Appl. Mech.* **18**, 405–413.

Mallock, A. (1881–82) The action of cutting tools. *Proc. Roy. Soc. Lond.* **33**, 127–139.

Merchant, M. E. (1945) Mechanics of the metal cutting process. *J. Appl. Phys.* **16**, 318–324.

Oxley, P. L. B. (1989) *Mechanics of Machining*. Chichester: Ellis Horwood.

Shaw, M. C. (1984) *Metal Cutting Principles*, Ch. 13. Oxford: Clarendon Press.

Shirakashi T. and Usui, E. (1973) Friction characteristics on tool face in metal machining. *J. JSPE* **39**, 966–972.

Taylor, F. W. (1907) On the art of cutting metals. *Trans. ASME* **28**, 31–350.

Trent, E. M. (1991) *Metal Cutting*, 3rd edn., Ch.9. Oxford: Butterworth Heinemann.

Tresca, H. (1878) On further applications of the flow of solids. *Proc. I. Mech. E. Lond.* pp. 301–345 and plates 35–47.

Wakabayashi, T., Williams, J. A. and Hutchings I. M. (1995) The kinetics of gas phase lubrication in the orthogonal machining of an aluminium alloy. *Proc. I Mech. E. Lond.* **209Pt.J**, 131–136.

Weiner, J. H. (1955) Shear plane temperature distribution in orthogonal cutting. *Trans ASME* **77**, 1331–1341.

Williams, J. A. (1977) The action of lubricants in metal cutting. *J. Mech. Eng. Sci*. **19**, 202–212.

Zorev, N. N. (1966) *Metal Cutting Mechanics*. Oxford: Pergamon Press.

3

Work and tool materials

In Chapter 2, the emphasis is on the mechanical, thermal and friction conditions of chip formation. The different work and tool materials of interest are introduced only as examples. In this chapter, the materials become the main interest. Table 3.1 summarizes some of the main applications of machining, by industrial sector and work material group, while Table 3.2 gives an overview of the classes of tool materials that are used. In Section 3.1 data will be presented of typical specific forces, tool stresses and temperatures generated when machining the various work groups listed in Table 3.1. In Section 3.2 the properties of the tools that resist those stresses and temperatures will be described.

A metal's machinability is its ease of achieving a required production of machined components relative to the cost. It has many aspects, such as energy (or power) consumption, chip form, surface integrity and finish, and tool life. Low energy consumption, short (broken) chips, smooth finish and long tool life are usually aspects of good machinability. Some of these aspects are directly related to the continuum mechanical and thermal conditions of the

Table 3.1 Some machining activities by work material alloy and industrial sector

Alloy system	General engineering	Auto-motive	Aerospace	Process engineering	Information technology
Carbon and alloy steels	Structures fasteners, power train, hydraulics	Power train, steering, suspension, hydraulics fasteners	Power train, control and landing gear	Structures	Printer spindles and mechanisms
Stainless steels	For corrosion resistance	–	Turbine blades	For corrosion resistance	–
Aluminium	Structures	Engine block and pistons	Airframe spars, skins	For corrosion resistance	Scanning mirrors, disc substrates
Copper	–	–	–	For corrosion resistance	–
Nickel	–	–	Turbine blades and discs	Heat exchangers, and corrosion resistance	–
Titanium	–	–	Compressor/ airframe	Corrosion resistance	–

Table 3.2 Recommended tool and work material combinations

	Soft non-ferrous (Al, Cu)	Carbon/ low alloy steels	Hardened tool and die steels	Cast iron	Nickel -based alloys	Titanium alloys
High speed steel	O/⊗	O/⊗	x	⊗/x	⊗/x	⊗/x
Carbide (inc. coated)	O	√/O	⊗	√/O	√	O
Cermet	⊗/x	√	x	⊗	x	x
Ceramic	x	√/O	O	√	√/O	x
cBN	⊗/x	x	√	√/O	O	O
PCD	√	x	x	x	x	√

√ good; O all right in some conditions; ⊗ possible but not advisable; x to be avoided.

Table 3.3 Mechanical, thermal and materials factors affecting machinability

Main tools for study	Process variables	Machinability attribute
	Cutting speed and feed	Chip form
	Tool shape	Tool forces
Applied mechanics	Work mechanical and thermal properties	Power consumption
and thermal analysis	Tool thermal properties	Tool stresses and temperatures
	Tool failure properties	Tool failure
	Chip/tool friction laws	Surface integrity and finish
Materials engineering	Work/tool wear interactions	Tool wear and life

machining process. In principle, they may be predicted by mechanical and thermal analysis (but at the current time some are beyond prediction). Other aspects, principally tool life, depend not only on the continuum surface stresses and temperatures that are generated but also on microstructural, mechanical and chemical interactions between the chip and the tool. Table 3.3 summarizes these relations and the principal disciplines by which they may be studied (perhaps chip/tool friction laws should come under both the applied mechanics and materials engineering headings?). This chapter is mainly concerned with the work material's mechanical and thermal properties, and tool thermal and failure properties, which affect machinability. Tool wear and life are so important that a separate chapter, Chapter 4, is devoted to these subjects.

3.1 Work material characteristics in machining

According to the analysis in Chapter 2, cutting and thrust forces per unit feed and depth of cut, and tool stresses, are expected to increase in proportion to the shear stress on the primary shear plane, other things being equal. This was sometimes written k and sometimes k_{max}.

Forces also increase the smaller is the shear plane angle and hence the larger is the strain in the chip. The shear plane angle, however, reduces the larger is the strain hardening in the primary shear region, measured by $\Delta k/k_{max}$ (equation (2.7)). Thus, k_{max} and $\Delta k/k_{max}$ are likely to be indicators of a material's machinability, at least as far as tool forces and stresses and power consumption are concerned. Figure 3.1 gathers information on the typical values of these quantities for six different groups of work materials that are important in machining practice. The data for steels exclude quench hardened materials as, until

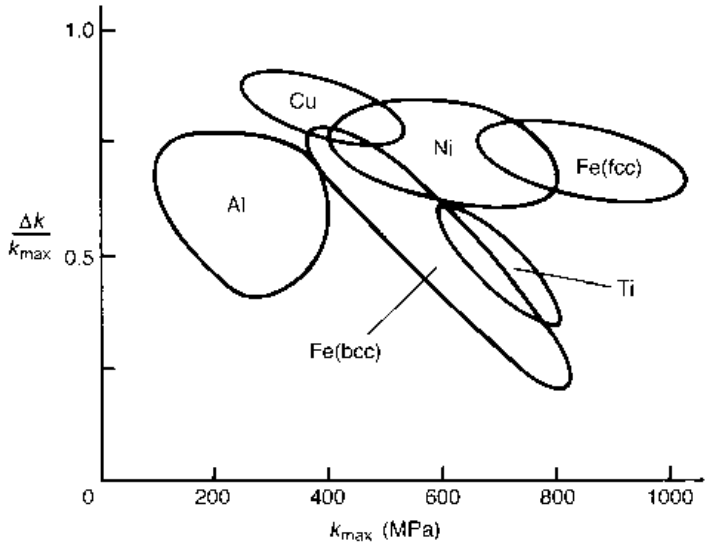


Fig. 3.1 Shear stress levels and work hardening severities of initially unstrained, commonly machined, aluminium, copper, iron (b.c.c. and f.c.c.), nickel and titanium alloys

recently, these were not machinable. The data come from compression testing at room temperature and at low strain rates of initially unworked metal. The detail is presented in Appendix 4.1. Although machining generates high strain rates and temperatures, these data are useful as a first attempt to relate the severity of machining to work material plastic flow behaviour. A more detailed approach, taking into account variations of material flow stress with strain rate and temperature, is introduced in Chapter 6.

Work heating is also considered in Chapter 2. Temperature rises in the primary shear zone and along the tool rake face both depend on $fU_{work}\tan\phi/\kappa_{work}$. Figure 3.2(a) summarizes the conclusions from equation (2.14) and Figures 2.17(a) and 2.18(b). In the primary shear zone the dimensionless temperature rise $\Delta T(\rho C)/k$ depends on $fU_{work}\tan\phi/\kappa_{work}$ and the shear strain γ. Next to the rake face, the additional temperature rise depends on $fU_{work}\tan\phi/\kappa_{work}$ and the ratio of tool to work thermal conductivity, K^*. Figure 3.2(b) summarizes the typical thermal properties of the same groups of work materials whose mechanical properties are given in Figure 3.1. The values recorded are from room temperature to 800°C. Appendix 4.2 gives more details.

Figures 3.1 and 3.2 suggest that the six groups of alloys may be reduced to three as far as the mechanical and thermal severity of machining them is concerned. Copper and aluminium alloys, although showing high work hardening rates, have relatively low shear stresses and high thermal diffusivities. They are likely to create low tool stresses and low temperature rises in machining. At the other extreme, austenitic steels, nickel and titanium alloys have medium to high shear stresses and work hardening rates and low thermal diffusivities. They are likely to generate large tool stresses and temperatures. The body centred cubic carbon and alloy steels form an intermediate group.

The behaviours of these three different groups of alloys are considered in Sections 3.1.3 to 3.1.5 of this chapter, after sections in which the machining of unalloyed metals is

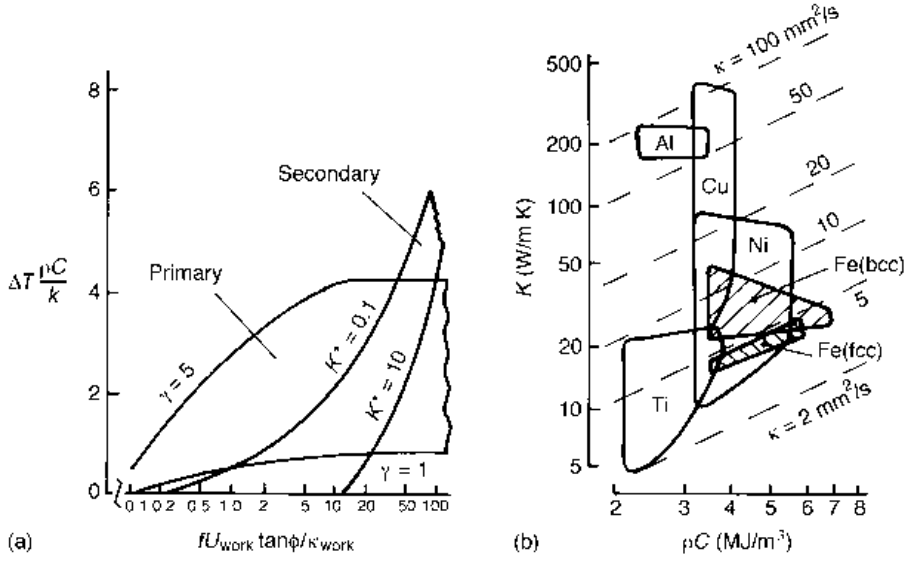


Fig. 3.2 Thermal aspects of machining: (a) a summary of heating theory and (b) thermal property ranges of Al, Cu, Fe, Ni and Ti alloys

described. It will be seen that these groups do indeed give rise to three different levels of tool stress and temperature severity. This is demonstrated by presenting representative experimentally measured specific cutting forces (forces per unit feed and depth of cut) and shear plane angles for these groups as a function of cutting speed. Then, primary shear zone shear stress k, average normal contact stress on the rake face $(\sigma_n)_{av}$ and average rake face contact temperature $(T_{rake})_{av}$ are estimated from the cutting data. A picture is built up of the stress and temperature conditions that a tool must survive in machining these materials.

The primary shear plane shear stress is estimated from

$$k = \frac{(F_c \cos\phi - F_T \sin\phi)\sin\phi}{fd} \tag{3.1}$$

The average normal contact stress on the tool rake face is estimated from the measured normal component of force on the rake face, the depth of cut and the chip/tool contact length l_c:

$$(\sigma_n)_{av} = \frac{F_c \cos\alpha - F_T \sin\alpha}{l_c d} \tag{3.2}$$

l_c is taken, from the mean value data of Figure 2.9(a), to be

$$l_c = 1.75f \frac{\cos(\phi - \alpha)}{\sin\phi} [\mu + \tan(\phi - \alpha)] \tag{3.3}$$

Finally, temperatures are estimated after the manner summarized in Figure 3.2.

The machining data come mainly from results in the authors' possession. The exception are data on the machining of the aluminium alloy Al2024 (Section 3.1.2), which are from results by Kobayashi and Thomsen (1959). The data on machining elemental metals come from the same experiments on those metals considered by Trent in his book (Trent, 1991).

3.1.1 Machining elemental metals

Although the elemental metals copper, aluminium, iron, nickel and titanium have little commercial importance as far as machining is concerned (with the exception of aluminium used for mirrors and disk substrates in information technology applications), it is interesting to describe how they form chips: what specific forces and shear plane angles are observed as a function of cutting speed. The behaviour of alloys of these materials can then be contrasted with these results. Figure 3.3 shows results from machining at a feed of 0.15 mm with high speed steel (for copper and aluminium) and cemented carbide (for iron, nickel and titanium) tools of 6° rake angle.

At the lowest cutting speeds (around 30 m/min), except for titanium, the metals machine with very large specific forces, up to 8 GPa for iron and nickel and around 4 GPa for copper and aluminium. These forces are some ten times larger than the expected shear flow stresses of these metals (Figure 3.1) and arise from the very low shear plane angles, between 5° and 8°, that occur. These shear plane angles give shear strains in the primary shear zone of from 7 to 12. As cutting speed increases to 200 m/min, the shear plane angles increase and the specific forces are roughly halved. Further increases in speed cause much less variation in chip flow and forces. The titanium material is an exception. Over the whole speed range, although decreases of specific force and increases of shear plane angle with cutting speed do occur, its shear plane angle is larger and its specific forces are

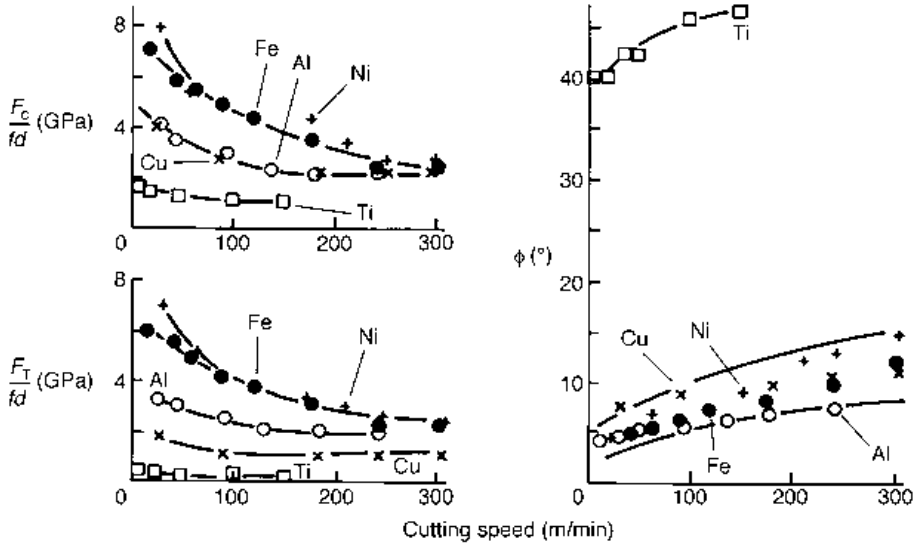


Fig. 3.3 Cutting speed dependence of specific forces and shear plane angles for some commercially pure metals (f = 0.15 mm, α = 6°)

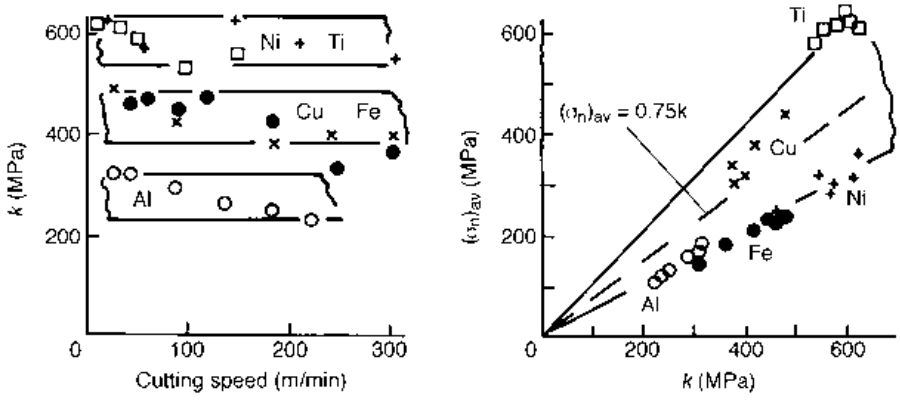


Fig. 3.4 Process stresses, derived from the observations of Figure 3.3

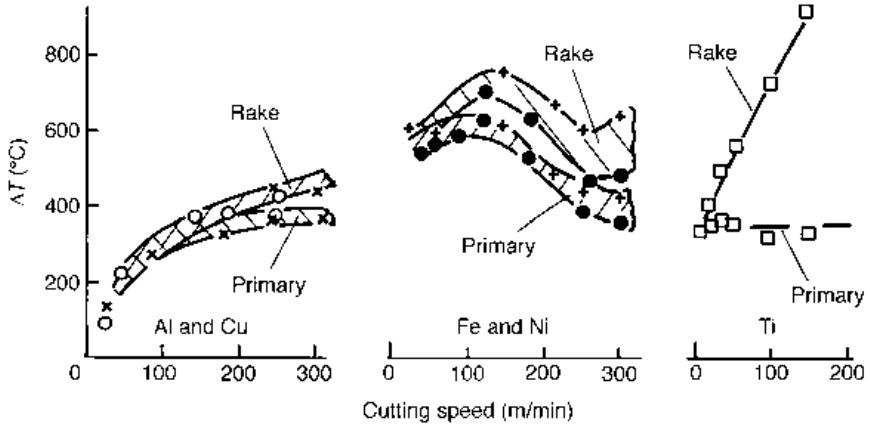


Fig. 3.5 Temperatures estimated from the observations of Figure 3.3

smaller than for the other, more ductile, metals. A reduction in forces and an increase in shear angle with increasing speed, up to a limit beyond which further changes do not occur, is a common observation that will also be seen in many of the following sections.

Although the forces fall with increasing speed, the process stresses remain almost constant. Figure 3.4 shows aluminium to have the smallest primary shear stress, k, followed by copper, iron, nickel and titanium.

The estimated average normal stresses $(\sigma_n)_{av}$ lie between $0.5k$ and $1.0k$. This would place the maximum normal contact stresses (which are between two and three times the average stress) in the range k to $3k$. This is in line with the estimates in Chapter 2, Figure 2.15.

The different thermal diffusivities of the five metals result in different temperature variations with cutting speed (Figure 3.5). For copper and aluminium, with κ taken to be 110 and 90 mm^2/s respectively (Appendix 4.2), $fU_{work}\tan\phi/\kappa_{work}$ hardly rises to 1, even at the cutting speed of 300 m/min. Figure 3.2 suggests that then the primary shear temperature rise dominates the secondary (rake) heating. The actual increase in temperature shown in

Figure 3.5 results from the combined effect of increasing fraction of heat flowing into the chip and reducing shear strain as cutting speed rises.

Iron and nickel, with κ taken to be 15 and 20 mm^2/s respectively, machine with $fU_{work}\tan\phi/\kappa_{work}$ in the range 1 to 10 in the conditions considered. In Figure 3.5, the primary shear and average rake face temperatures are distinctly separated. Over much of the speed range, the temperature actually falls with increasing cutting speed. This unusual behaviour results from the reduction of strain in the chip as speed increases.

Finally, titanium, with κ taken to be 7.5 mm^2/s, machines with $fU_{work}\tan\phi/\kappa_{work}$ from 7 to 70. The rake face heating is dominant and a temperature in excess of 800°C is estimated at the cutting speed of 150 m/min.

3.1.2 Effects of pre-strain and rake angle in machining copper

In the previous section, the machining of annealed metals by a 6° rake angle tool was considered. Both pre-strain and an increased rake angle result in reduced specific cutting forces and reduced cutting temperatures, but have little effect on the stressses on the tool. These generalizations may be illustrated by the cutting of copper, a metal sufficiently soft (as also is aluminium) to allow machining by tools of rake angle up to around 40°. Figure 3.6 shows examples of specific forces and shear plane angles measured in turning annealed and heavily cold-worked copper at feeds in the range 0.15 to 0.2 mm, with high speed steel tools of rake angle from 6° to 35°. Specific forces vary over a sixfold range at the lowest cutting speed, with shear plane angles from 8° to 32°.

The left panel of Figure 3.7 shows that the estimated tool contact stresses change little with rake angle, although they are clearly larger for the annealed than the pre-strained material. The right-hand panel shows that the temperature rises are halved on changing from a 6° to 35° rake angle tool. These observations, that tool stresses are determined by

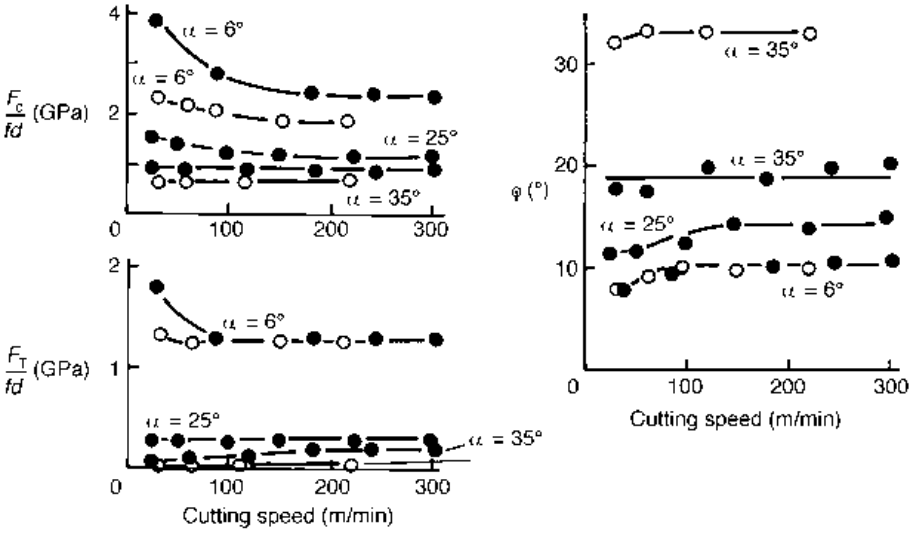


Fig. 3.6 Specific force and shear plane angle variations for annealed (•) and pre-strained (o) commercially pure copper (f = 0.15 to 0.2 mm, α = 6° to 35°)

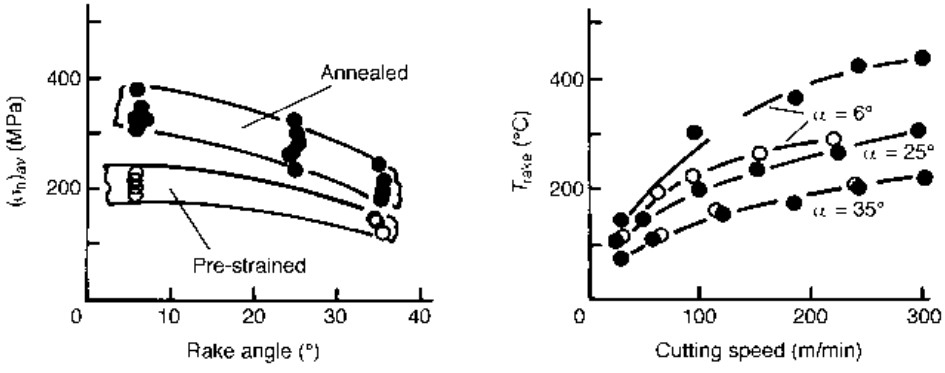


Fig. 3.7 Average rake face contact stresses and temperatures, from the results of Figure 3.6

the material being cut and do not vary much with the cutting conditions, while temperatures depend strongly on both the material being cut and the cutting conditions, is a continuing theme that will be developed for metal alloys in the following sections.

3.1.3 Machining copper and aluminium alloys

It is often found that alloys of metals machine with larger shear plane angles and hence lower specific forces than the elemental metals themselves. Sometimes a strong reason is a lower value of the strain hardening parameter $\Delta k/k_{max}$, at other times the chip/tool friction (as indicated by the friction coefficient) is less; and at others again it is not at all obvious why this should be so. But even when the specific forces are lower, the tool contact stress can be higher. In this section, examples of machining two copper and one aluminium alloy are taken to illustrate this.

Figure 3.8 records the behaviours of a CuNi and a CuZn alloy. The CuNi alloy, with 80%Ni, might better be considered as a Ni alloy. However, it machines at a higher shear plane angle at a given cutting speed than either copper or nickel, despite its strain-hardening characteristic being similar to or more severe than either of these (Appendix 4.1). The CuZn alloy (an α-brass) is a well-known very easy material to machine. Its shear plane angle is twice as large as that of Cu, despite having a similar strain-hardening characteristic (Appendix 4.1 again) and an apparently higher friction interaction with the tool (as judged by the relative sizes of its specific thrust and cutting forces). (Figure 3.8 describes the machining of an annealed brass. After cold-working, even higher shear plane angles, and lower specific forces are obtained.) These two examples are ones where the reason for the easier machining of the alloys compared with the elemental metals is not obvious from their room temperature, low strain rate mechanical behaviours.

Figure 3.9 shows machining data for an aluminium alloy. In this case the variation of behaviour with rake angle is shown. At a rake angle and speed comparable to that shown in Figure 3.3, the shear plane angle is five times as large and the specific cutting force is half as large for the alloy as for pure Al. In this case both the strain-hardening and friction factors are less for the alloy than for pure Al.

For both the copper and aluminium alloy examples, the primary shear plane shear stress and the average rake contact stresses are similar to, or slightly larger than, those for the

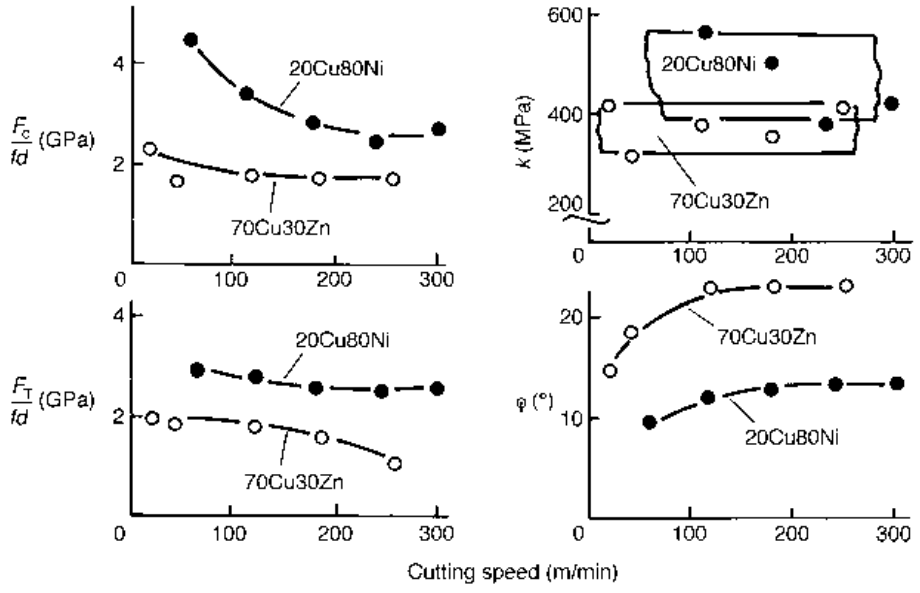


Fig. 3.8 Observed and calculated machining parameters for two copper alloys (f = 0.15 mm, α = 6°)

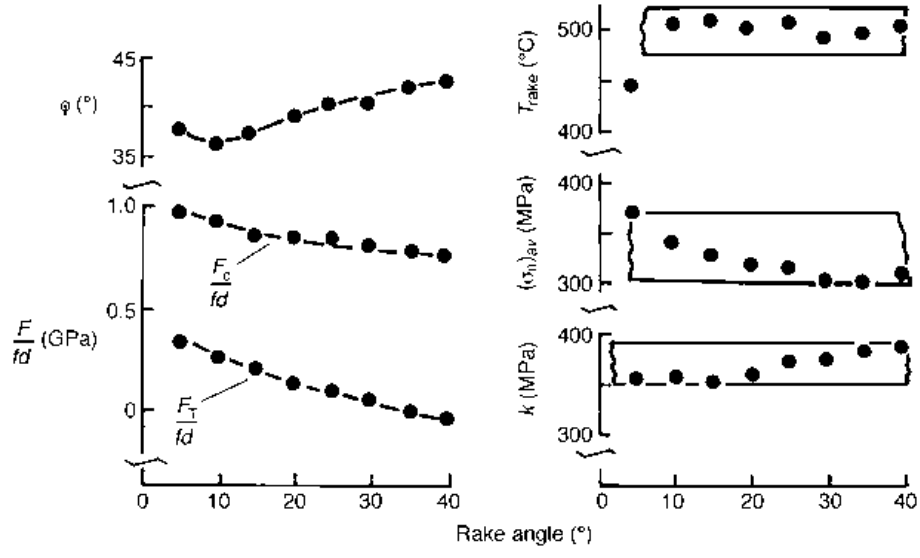


Fig. 3.9 Machining parameter variation with rake angle for Al2024-T4 alloy, at a cutting speed of 175 m/min and f = 0.25 mm

elemental metals. Figure 3.8 shows only the values of k, but $(\sigma_n)_{av}$ may be calculated to be $\approx 0.6k$. Figure 3.9 shows both k and $(\sigma_n)_{av}$. It also shows that, in this case, the estimated rake face temperature does not change as the rake angle is reduced. This is different from the observations recorded in Figure 3.7: perhaps the maximum temperature is limited by melting of the aluminium alloy?

Table 3.4 Approximate ranges of k and $(\sigma_n)_{av}$ estimated from machining tests

	Alloy system					
Stress (MPa)	Al	Cu	Fe(bcc)	Fe(fcc)	Ni	Ti
k	200–400	300–550	350–750	500–800	550–850	550–700
$(\sigma_n)_{av}$	120–370	150–400	200–550	400–700	300–800	600–700

The choice in Figure 3.9 of showing how machining parameters vary with rake angle has been made to introduce the observation that, in this case, at a rake angle of around 35° the thrust force passes through zero. Consequently, such a high rake angle is appropriate for machining thin walled structures, for which thrust forces might cause distortions in the finished part.

However, the main point of this section, to be carried forward to Section 3.2 on tool materials, is that the range of values estimated for k follows the range expected from Figure 3.1 and the estimated values of $(\sigma_n)_{av}$ range from 0.5 to 1.0k. This is summarized in Table 3.4 which also contains data for the other alloy systems to be considered next.

3.1.4 Machining austenitic steels and temperature resistant nickel and titanium alloys

The austenitic steels, NiCr, and Ti alloys are at the opposite extreme of severity to the aluminium and copper alloys. Although their specific forces are in the same range and their shear plane angles are higher, the tool stresses and temperatures (for a given speed and

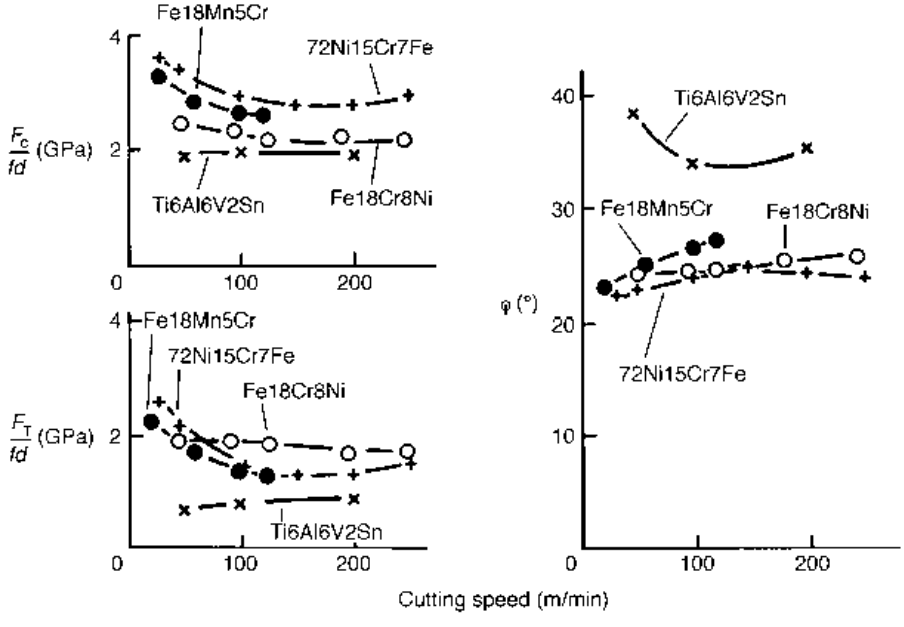


Fig. 3.10 Specific force and shear plane angle variations for some austenitic steels, nickel-chromium and titanium alloys (f = 0.1 to 0.2 mm, α = 0° to 6°)

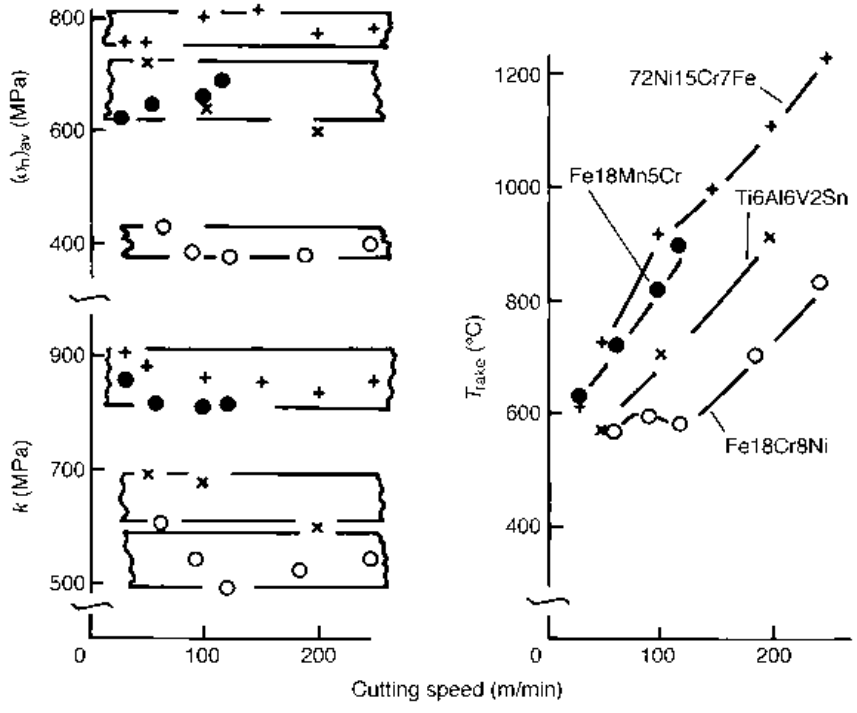


Fig. 3.11 Process stresses and temperatures derived from (and symbols as) Figure 3.10

feed) that they generate are significantly higher. Figure 3.10 presents observations for two austenitic steels, a NiCr and a Ti alloy. One of the austenitic steels (the 18Cr8Ni material) is a common stainless steel. The 18Mn5Cr material, which also contains 0.47C, is an extremly difficult to machine creep and abrasion resistant material. The NiCr alloy is a commercial Inconel alloy, X750. In all cases the feed was 0.2 mm except for the Ti alloy, for which it was 0.1 mm. The rake angle was 6° except for the NiCr alloy, for which it was 0°. Specific cutting forces are in the range 2 to 4 GPa. Thrust forces are mainly between 1 and 2 GPa. Shear plane angles are mainly greater than 25°. In most cases, the chip formation is not steady but serrated. The values shown in Figure 3.10 are average values. Figure 3.11 shows stresses and temperatures estimated from these. The larger stresses and temperatures are clear.

3.1.5 Machining carbon and low alloy steels

Carbon and alloy steels span the range of machinability between aluminium and copper alloys on the one hand and austentic steels and temperature resistant alloys on the other. There are two aspects to this. The wide range of materials' yield stresses that can be achieved by alloying iron with carbon and small amounts of other metals, results in their spanning the range as far as tool stressing is concerned. Their intermediate thermal conductivities and diffusivities result in their spanning the range with respect to temperature rise per unit feed and also cutting speed.

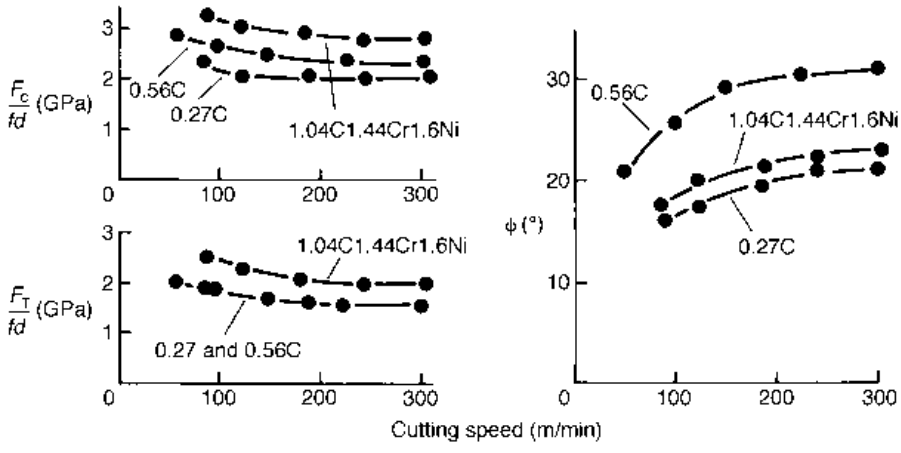


Fig. 3.12 Representative specific force and shear plane angle variations for hot rolled carbon and alloy steels (f = 0.15 mm, α = 6°)

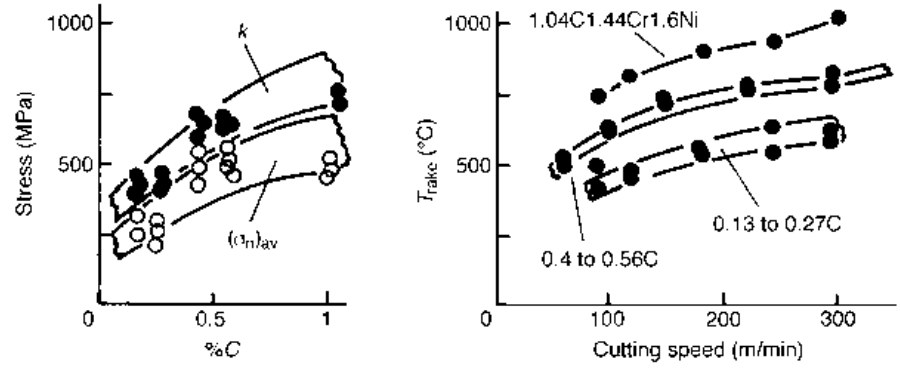


Fig. 3.13 Process stresses and temperatures derived from Figure 3.12

Figure 3.12 shows typical specific force and shear plane angle variations with cutting speed measured in turning steel bars that have received no particular heat treatment other than the hot rolling process used to manufacture them. At cutting speeds around 100 m/min the specific forces of 2 to 3 GPa are smaller than those for pure iron (Figure 3.3), but as speed increases, the differences between the steels and pure iron reduce. In the same way as for many other alloy systems, the shear plane angles of the ferrous alloys are larger than for the machining of pure iron.

In the hot rolled condition, steels (other than the austenitic steels considered in the previous section) have a structure of ferrite and pearlite (or, at high carbon levels, pearlite and cementite). For equal coarsenesses of pearlite, the steels' hardness increases with carbon content. The left panel of Figure 3.13 shows how the estimated k and $(\sigma_n)_{av}$ values from the data of Figure 3.12 increase with carbon content. Additional results have been included, for the machining of a 0.13C and a 0.4C steel. An increase of both k and $(\sigma_n)_{av}$ with %C is clear. The right panel of the figure likewise shows that the increasing carbon

content gives rise to increasing temperatures for a given cutting speed. This comes from the increasing shear stress levels.

This completes this brief survey of the stresses and temperatures generated by different alloy groups in machining. Tool stresses are mainly controlled by the metal being machined and vary little with cutting conditions (although the tool rake face area over which they act changes with speed and, obviously, also with feed). Temperatures, on the other hand, depend not only on the material being machined (both through stress levels and thermal properties) but also on the speeds and feeds used.

3.1.6 Machining with built-up edge formation

In the previous section, data were presented mainly for cutting speeds greater than 100 m/min. This is because, at slightly lower cutting speeds, at the feeds considered, those steels machine with a built-up edge (BUE). In Chapter 2, photographs were shown of BUE formation. Figure 3.14 shows, for a 0.15C steel, what changes in specific force and shear plane angle are typically associated with this. In this example, the largest BUE occurred at a cutting speed close to 25 m/min. There, the specific forces passed through a minimum and the shear plane angle through a maximum. Qualitatively, this may be explained by the BUE increasing the effective rake angle of the cutting tool.

Built-up edge formation occurs at some low speed or other for almost all metal alloys. It offers a way of relieving the large strains (small shear plane angles) that can occur at low speeds, but at the expense of worsening the cut surface finish. For those alloys that do show BUE formation, the cutting speed at which the BUE is largest reduces as the feed increases. Figure 3.15 gathers data for three ferrous alloys and one Ni-Cr creep resistant alloy (Nimonic 80). One definition of high speed machining is machining at speeds above those of built-up-edge formation. These are the conditions mostly focused on in this book.

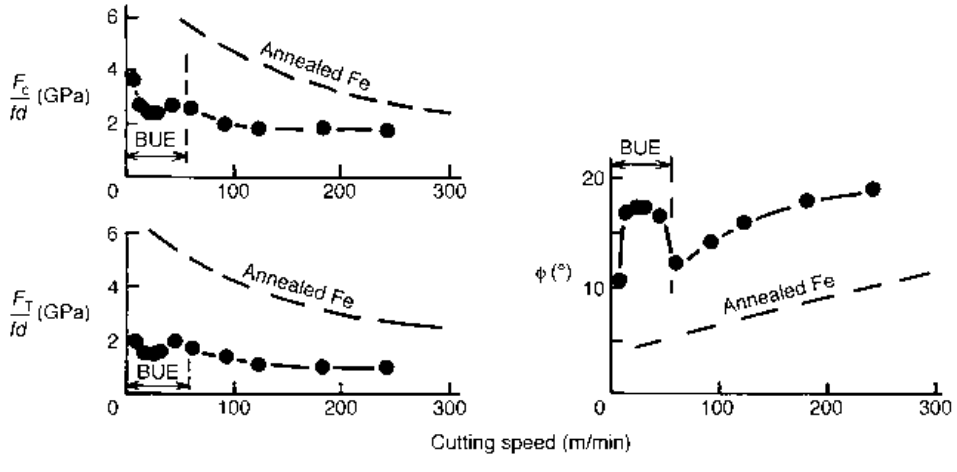


Fig. 3.14 Characteristics of built-up edge (BUE) formation (0.15C steel, $f = 0.15$ mm, $\alpha = 6°$)

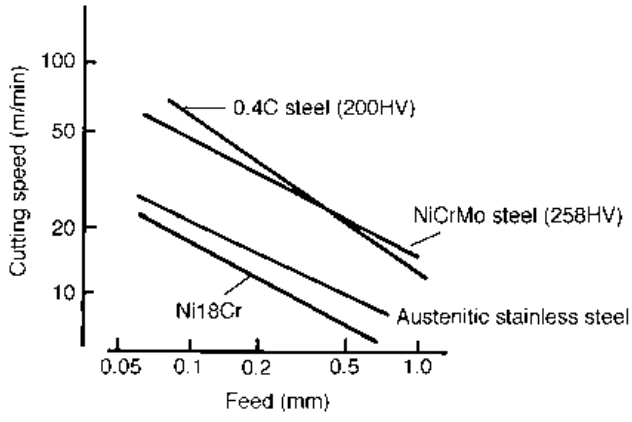


Fig. 3.15 Speed and feed dependence of built-up edge formation, after Trent (1991)

3.1.7 Free-cutting alloys

It is possible to make minor changes to the composition of alloys that result in major improvements in their machinability. The data considered up to this point have not been for such alloys. The effects of such composition changes will now be introduced, by considering first of all the machining of free-cutting low carbon steels.

Most carbon steels contain manganese, controlled at a level of around 1%, and sulphur as an impurity, up to a level of around 0.05%. One of the non-metallic inclusions that exists is manganese sulphide, MnS. If the sulphur is increased to 0.2% to 0.3% and the manganese is also increased (typical values are 1–1.5%), the amount of MnS is increased and becomes important. It can, in some conditions, form a layer over the chip/tool contact that can reduce chip/tool friction and hence ease chip formation. Lead (Pb) can also be added, commonly at a level of around 0.25%. It can further lubricate the contact. The magnitude of the friction change has already been introduced, in Section 2.4 (Figure 2.22). The action (of MnS forming a layer in the contact area) is specific to high speed steels and cutting tools containing Ti, that is to say cemented carbides (or cermets) containing TiC or mixed TiC/TaC; and to tools coated by TiN or TiC. The lubrication is only effective over a certain contact temperature range and hence depends on the cutting speed and feed. Figure 3.16 shows a typical effect of this lubricating action. The specific forces and shear plane angles observed in turning a MnS and a Pb-MnS free-cutting low carbon (0.08 to 0.09C) steel are compared with those for a similar non-free-cutting steel. At cutting speeds between 20 m/min and 75 m/min (at the feeds considered) the shear plane angles of the free-cutting materials are double and the specific forces around half of those for the non-free cutting steel (the built-up-edge is much smaller and more stable too). As cutting speed increases up to 200 m/min for the MnS steel and to 300 m/min for the Pb-MnS steel, these differences between the free- and non-free-cutting steels become insignificant. Although there is clear benefit in reduced forces from the free-cutting steels, there is no reduction in the tool normal contact stresses. For all the steels in Figure 3.16, k values are estimated between 400 MPa and 450 MPa (in line with Figure 3.13). $(\sigma_n)_{av}$ values around 300 MPa are estimated for the non-free-cutting steel (also in line with Figure 3.13), but values from 350 MPa to 400 MPa are estimated for the free-cutting steels.

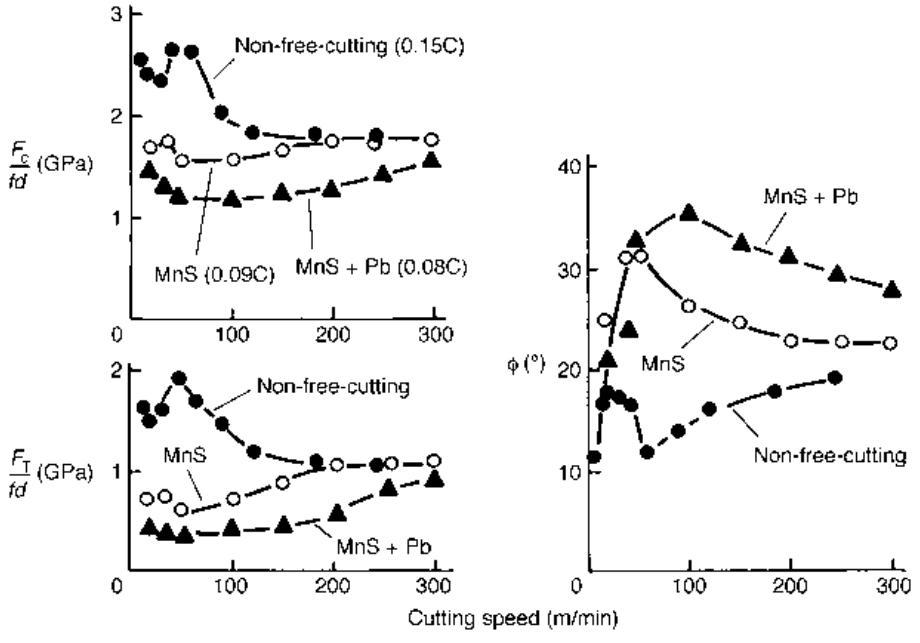


Fig. 3.16 Representative specific force and shear plane angle variations for low carbon free-machining steels turned by a steel cutting grade of carbide tool (f = 0.1 to 0.15 mm, α = 6°)

These free-cutting steels have a great commercial importance. They enable small diameter, intricate, parts such as spacers, screwed profiles and small electric motor spindles to be machined with a good surface finish and with less energy consumption than the equivalent non-free-cutting steel, in the speed range where the non-free-cutting steel would suffer from the poor finish associated with built-up edge formation. The free-cutting steels are, however, less tough than their non-free-cutting equivalents and are not used in applications in which the transmission of tensile stresses is critical. Semi-free-cutting grades of steel have been developed to compromise between machinability and strength requirements. These have been developed by control of the wide variety of non-metallic inclusions that can be created during the deoxidation of steel melts, as considered next.

Free oxygen in steel is removed from the melt most simply by adding small amounts of aluminium, silicon or calcium, to form alumina, silica or calcium oxides. Alumina is hard and abrasive and is certainly detrimental to tool life in machining. The addition of silicon and calcium can result in softer inclusions. It has been found that if, in addition, small amounts of sulphur (relative to the 0.2% to 0.3% used in free-cutting steels) are added, complex layers containing calcium, manganese and sulphur can build up on the rake face of tools. Again, the tools have to contain titanium. These layers have relatively small effects in altering specific forces and shear plane angles, but can significantly influence tool life. Typical quantities of calcium are 0.002% and of sulphur 0.03 to 0.1% (with silicon from 0.2 to 0.3%). The topics of tool wear and life are developed more fully in Chapter 4. Here, Figure 3.17 shows differences in the machining of a low alloy steel (nominally 0.4C1Cr0.2Mo), produced without and with small additions of Ca and S as just described.

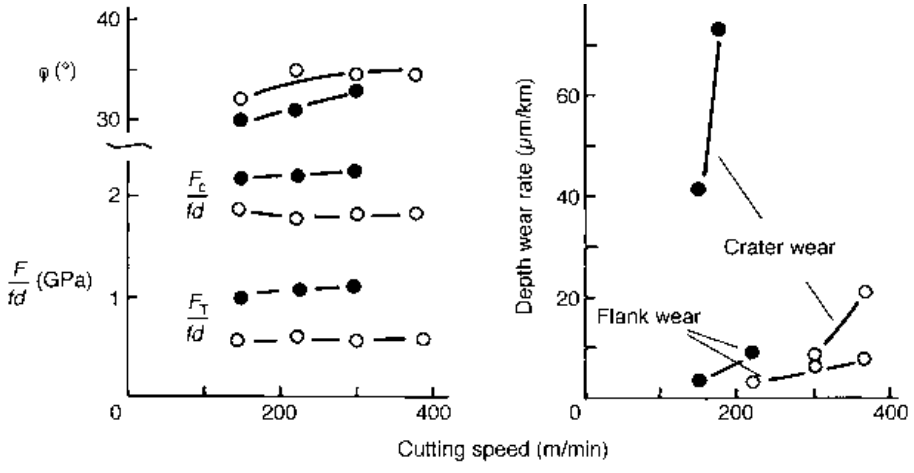


Fig. 3.17 Machining characterisitcs of a low alloy (•) and a semi-free-cutting low alloy (o) steel (f = 0.25 mm, α = 6°)

The tool was an uncoated steel cutting grade (P-type) carbide. Although differences can be seen between the specific forces and shear plane angles for these materials, the estimated rake contact normal stresses and temperatures are estimated to be hardly different for the two. Yet the tool wear rates, particularly the crater wear rates, are hugely different.

In Figure 3.17, there is at least some visible change in specific forces and shear plane angle brought about by controlling the deoxidation process. In other cases, for example by adding a small amount of calcium but no extra sulphur, changes in tool life can be produced with no change at all in chip form and forces. A study with this conclusion, for machining a 0.45% carbon steel, has been published by Sata *et al.* (1968). The reader is reminded of the comment at the start of this chapter, that stresses and temperatures define the continuum conditions to which the cutting tool is subjected, but life (other than immediate failure) depends, in addition, on the work material's microstructure and chemical interactions with the tool.

This section has considered only free-cutting and semi-free-cutting steels. Free-cutting versions of other alloys are also manufactured. The best known are leaded copper and aluminium alloys, but the purpose of the lead is different from that considered so far. Up to 1% or 2% lead causes embrittlement of chips and hence aids chip control and disposability as well as reducing specific forces.

3.1.8 Summary

Section 3.1 mentioned the variety of specific forces and shear plane angles that are commonly observed in machining aluminium, copper, ferrous, nickel and titanium alloys. It has sought to establish that the average contact stresses that a tool must withstand depend mainly on the material being machined, through the level of that material's shear flow stress and hardly at all on the cutting speed and feed nor on the tool rake angle. Table 3.4 lists the range of these stresses. Peak contact stresses may be two to three times as large as the average values recorded in the table. In contrast, the temperatures that a tool must withstand do depend on cutting speed and feed and rake angle, and on the work material's

thermal properties: diffusivity, conductivity and heat capacity. By both thermal and stress severity criteria, the easiest metals to machine are alumimium alloys and copper alloys. The most difficult to machine are austenitic steels, nickel heat resistant alloys and titanium alloys. Ferritic and pearlitic steels lie between these extremes, with stresses and temperatures increasing with carbon content and hardness.

Beyond that, this section has been mainly descriptive, particularly with respect to reporting what shear plane angles have been measured for the different alloys. This remains the main task of predictive mechanics.

The next section, on tool material properties, complements this one, in describing the properties of tool materials that influence and enable the tools to withstand the machining-generated stresses and temperatures.

3.2 Tool materials

The main classes of tool materials have already been listed in Table 3.2 as carbides and cermets, high speed steels, ceramics based on alumina and silicon nitride, and the superhard materials polycrystalline diamond and cubic boron nitride (single crystal diamonds are also used for the finishing of IT mirror and disc substrate products). Details of the various materials within these groups are given in Appendix 6. It is recommended that the descriptive parts of Appendix 6 be read briefly, before continuing. The largest amount of space is given to dividing the carbides and cermets into sub-groups depending on whether the carbides are mainly tungsten carbide (WC) or a mixture of mainly WC with titanium and tantalum carbides (TiC/TaC) and on whether they are cemented together mainly with cobalt (Co) or a mixture of Co and nickel (Ni). In the following sections, the main purpose is to compare the properties of these different groups, and to understand why which groups are used in what circumstances.

3.2.1 Tool mechanical property minimum requirements

The sizes of the shear stresses k or k_{max} have been considered in Section 3.1. From now on, k or k_{max} will be written k_{work}, to distinguish work from tool properties. Section 3.1 has established that the majority of work materials are machined with a shear stress k_{work} measured on the primary shear plane between 200 MPa and 800 MPa and that the average normal contact stress on the tool face ranges between 0.5 and 1 k_{work}. In fact, only hardened steels, not considered in the previous sections, but which are increasingly machined by the superhard polycrystalline cubic boron nitride (PcBN), are likely to yield values of k_{work} greater than 800 MPa. In Chapter 2 it was suggested that peak normal contact stresses (at the cutting edge) may be two to three times as large as the average stress; that is to say, in the range 1 to 3 k_{work}. This is supported by split-tool contact stress measurements (Figure 2.21). Split-tool measurements have also given tool rake face friction stresses τ from 0.5 to 1 k_{work}, depending on rake face temperature (Figure 2.22). These loadings are summarized in Figure 3.18(a).

Figure 3.18(b) also shows some other possible loadings. When a tool enters a cut, a finite displacement is required before the chip is fully developed. Initially the contact can look more like an indentation. Then, the peak normal stress may be as large as $5k_{work}$ (this is approximately the Vickers Hardness, or *HV*, value). Because the sliding of the chip over

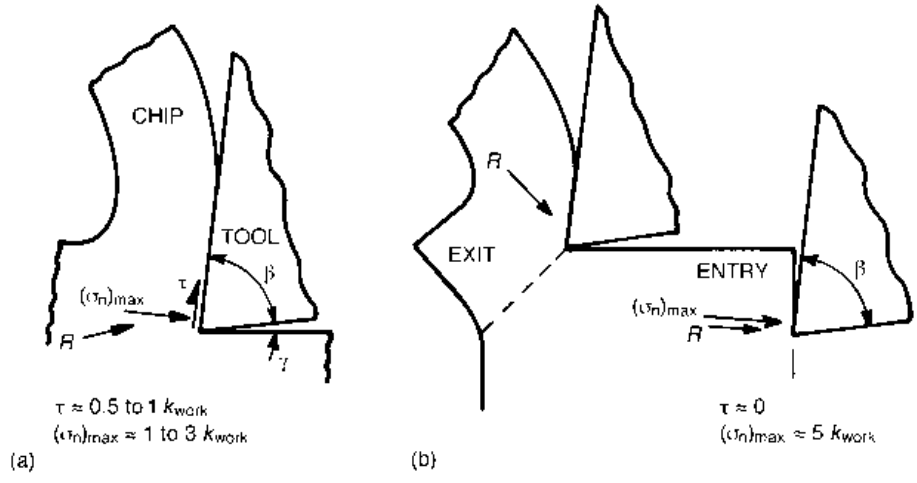


Fig. 3.18 Tool loads in (a) steady and (b) work entry and exit conditions

the rake is not established, τ may be close to zero and the direction of the resultant force R on the tool will be closer to the rake face normal than later on. At the end of a cut (at exit), the way in which the chip is pushed off the work to form a burr may result in the direction of R differing even more from its steady state direction. The questions are: what tool hardness is required to stop it yielding under the action of the contact stresses; what fracture resistance is required to stop it breaking?

The answers to both questions depend on how large is the tool included (or wedge) angle β (defined in Figure 3.18). It is qualitatively obvious that the smaller is β, the larger will be the maximum shear stress in the tool generated by the contact stresses, so the harder it must be to avoid yielding. Similarly, the smaller is β, the larger will be the maximum tensile stress on the rake face caused by bending of the tool edge region, so the tougher must be the tool to avoid fracture. An approximate analysis outlined in Appendix 5 shows that the entry condition (Figure 3.18) is more severe on the tool than the steady state. (The exit condition may be more severe still but has not been considered because it is more difficult to define the stress conditions.) Figure 3.19 summarizes its conclusions, in terms of required tool Vickers Hardness and Tensile Rupture Strength (TRS). TRS is a measure of fracture resistance usually determined experimentally by the maximum tensile stress that a bar of material can support without breaking in bending. Whether or not it is the best measure (fracture toughness K_{IC} may be fundamentally more sound) is open to discussion. It is, however, a practical measure: as will be seen in Section 3.2.2, there is more information available for TRS than there is for K_{IC} values of tool materials.

The left-hand panel of Figure 3.19 shows the relationship between minimum HV, β and k_{work}. For example (as shown for the double line), a material defined by k_{work} = 600 MPa, machined by a tool for which $\beta = 90^\circ$, requires a tool of $HV \geq 7.5$ GPa for tool yielding to be avoided. Similarly, the right-hand panel shows, for the same example, that the tool's TRS must be greater than between 1 and 2 GPa to avoid fracture.

Resistance to yielding and fracture depends on β but a tool's geometry is more usually defined by its rake angle α. The rake angle values along the top of the figure assume that the clearance angle $\gamma = 5^\circ$ (Figure 3.18(a)). It then can be seen that for k_{work} in the range

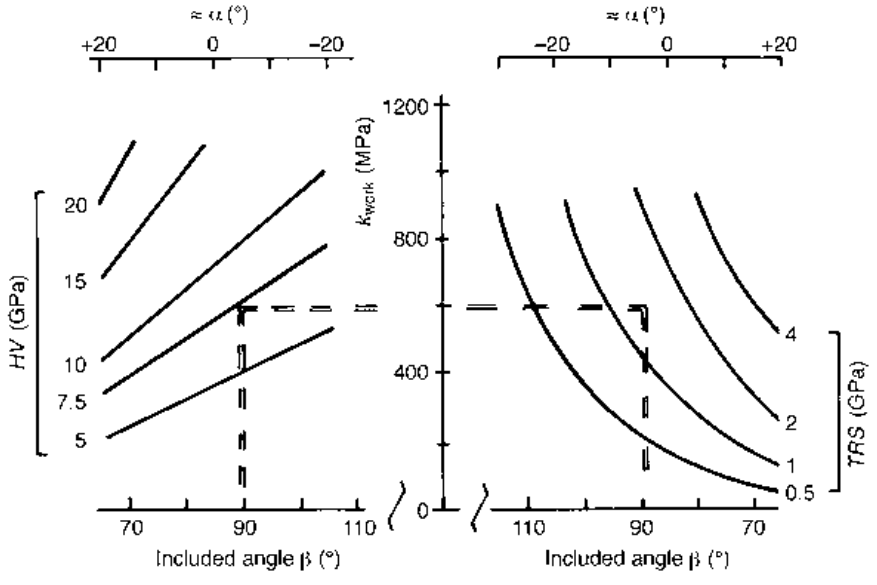


Fig. 3.19 Minimum tool HV and TRS values needed to machine metals, defined by their k_{work} values, with tools of wedge angle $\beta°$

200 to 1000 MPa and α between ± 20°, minimum tool hardnesses from 5 to 20 GPa and TRS values from 0.5 to 5 GPa are required. These are the ranges that practical tool materials do have.

3.2.2 Room temperature tool hardness and fracture resistance

Figure 3.20 gathers room temperature tool hardness and TRS data from a variety of sources, some published (Trent, 1991; Brookes, 1992) but also from manufacturers' information. It presents a snap-shot in time. For the well established high speed steels (HSS) and cemented carbides and cermets, there is high confidence that major property improvements will not occur in the future. That may not be the case for the other materials, particularly the PcBN group. The figure includes (towards its top left corner) the line $HV = 3TRS$. The tensile yield stress of a material is expected to be $\approx HV/3$, so above that line, a tool would be expected to show some ductile flow before fracture. Below that line is the region of predominantly elastic fracture. The figure also records (in a column to the right) the ranges of K_{IC} values that have been recorded, as an alternative to the TRS values. It can be seen that there is not an exact one-to-one relation between K_{IC} and TRS.

Only the HSS materials are so ductile that they are predominantly above the 'yield before fracture' border. The sub-micrometre (ultra fine grained) carbide materials almost reach that state at room temperature (and certainly do so at higher temperatures). Among the ceramic materials, those based on silicon nitride reach higher toughnesses than those based on alumina, with the exception of aluminas reinforced with silicon carbide (SiC) whiskers. Among the aluminas, aluminas combined with TiC (called black ceramics or black aluminas because of their colour) or reinforced with SiC whiskers, are harder than

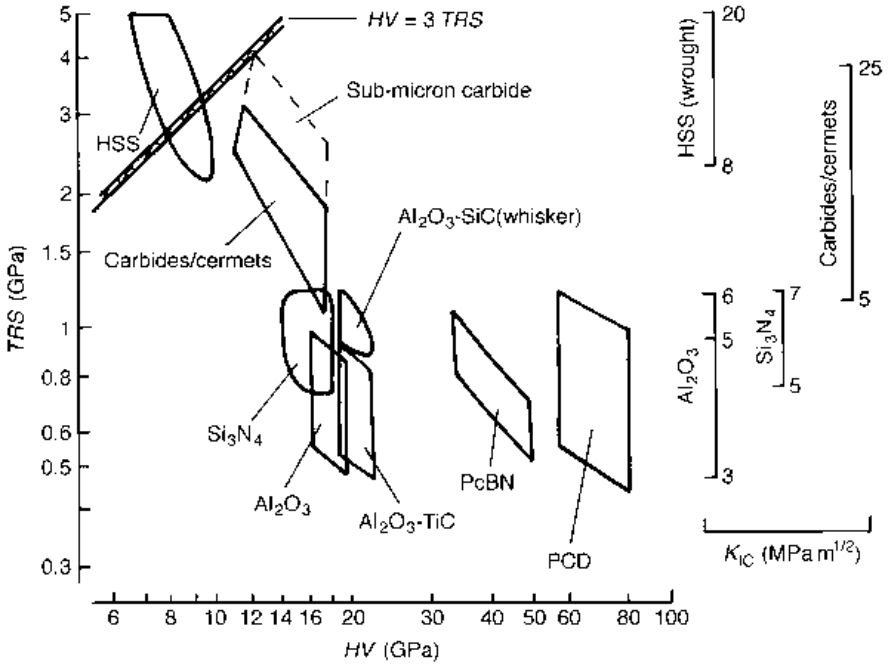


Fig. 3.20 Room temperature TRS and HV ranges of commercial uncoated cutting tool materials

the white aluminas (aluminas without TiC or SiC). At the present time, polycrystalline diamond (PCD) and PcBN have been developed to similar toughnesses as the aluminas and silicon nitride based materials, but are substantially harder.

3.2.3 Room temperature tool thermal and elastic properties

In Chapter 2, tool thermal conductivity was emphazied as influencing the steady state temperature rise in machining. In transient conditions, heat capacity is also important because, with conductivity, it determines thermal diffusivity κ and the rate of penetration of heat into the tool. Other thermal properties are important too, principally the thermal expansion coefficient α_e. With the tool's elastic Young's modulus E, α_e affects thermal stresses in the tool. The thermal expansion relative to that of coatings on the tool is also important. That is one of the factors that influence how well the coatings adhere to the tool (considered in Section 3.2.7).

Thermal shock resistance also affects a tool's performance. This composite property has several definitions. One is the ratio of TRS to $E\alpha_e$. It has units of ˚C, and it is the temperature change on cooling that would generate a tensile thermal stress equal to the TRS, if the thermal strain were not allowed to relax. Another definition is the product of (TRS/$E\alpha_e$) and the thermal conductivity K. A large thermal conductivity reduces the temperature gradients in a tool during cooling. It is also argued that K_{IC} should replace TRS and κ should replace K in these definitions. However, that does not change the rankings of tool groups with respect to thermal shock resisitance.

Table 3.5 summarizes the ranges of thermal and elastic properties of tool materials that

Table 3.5 Thermal and elastic properties of tool materials at room temperature

Tool type	K [W/m K]	ρC [MJ/m^3]	α_e [10^{-6}K^{-1}]	E [GPa]	TRS/($E\alpha_e$) [°C]
Diamond	600–2000	2.0	3.1	960–990	–
PCD	100–550	2.0**	3.8–4.2	620–840	140–540
PcBN	≈100*	1.9–2.1	4.7–4.9	680–710	150–340
K-carbide	75–120	3.0–3.4	4.5–6.0	550–650	390–925
P-carbide	25–55	4.0–4.1	5.8–6.8	490–560	390–840
Cermet	11–35	2.4–2.7	6.7–7.8	390–420	480–740
Al_2O_3	10–35	3.2–3.6	7.9–8.0	380–390	145–330
Al_2O_3/TiC	10–22	3.8–4.0	7.6–8.0	370–395	180–330
Al_2O_3/SiC(wh.)	10–35	≈3.4*	7.0–7.5	345–425	300–500
Si_3N_4/Sialon	15–30	2.1–2.3	3.2–3.6	280–320	650–1500
HSS	19–24	3.6–3.8	12–13	220–240	940–1740

*: information from limited data; **: assumed as for diamond.

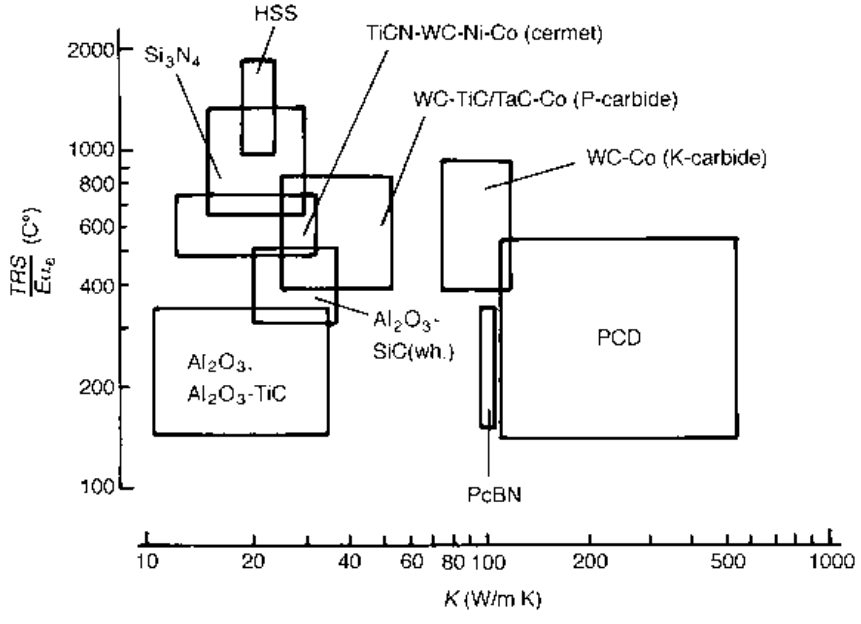


Fig. 3.21 Tool materials' characterization by thermal conductivity and shock resistance

have been reported at room temperature (with the exception of α_e values that tend to be measured as mean values, for example from room temperature to some typical high temperature). Variations with temperature are considered in Section 3.2.4.

The thermal shock parameter in Table 3.5 is TRS/($E\alpha_e$). TRS × K/($E\alpha_e$) can be deduced from Figure 3.21 which shows how the different tool groups are distinguished by thermal conductivity and shock resistance. The thermal shock resistance ranking is broadly the same as the TRS ranking in Figure 3.20, except that the Si_3N_4-based ceramics show a clear advantage over the other ceramic materials, and indeed over the carbides and cermets. This is due to the relatively low thermal expansion and Young's modulus of the Si_3N_4-based

ceramics. However, this advantage is not so clear if thermal shock resistance is considered to be (TRS)$K/(E\alpha_e)$. The low thermal conductivity of the silicon nitride based ceramics increases the temperature gradients that they are subjected to in practice. As alumina-SiC(whisker) ceramics have developed, the silicon nitride ceramics have found themselves competitively squeezed between these with respect to mechanical shock (TRS) resistance and the carbides with respect to thermal shock.

3.2.4 Tool property changes with temperature

Changes of tool behaviour with temperature are of three main types. First, all materials have some maximum temperature above which, for some reason, their composition or microstructure becomes unstable. If that temperature is exceeded by too much, the tool behaviour may be described as failing; but if it is exceeded only a little, rapid wear may be what is observed. Secondly, below the temperatures at which this degradation occurs, a tool's mechanical properties, such as hardness and resistance to fracture, may vary with temperature. Generally, a tool's reduction of hardness with temperature is of major importance to its use. Finally, and of less importance, thermal and elastic properties change, usually only slightly, with temperature.

Thermal stability

There are three main ways in which high temperatures cause a tool to degrade. One is by reaction with the atmosphere, usually oxidation. Secondly, a tool's microstructure will start to change above some critical temperature. Thirdly, tools may interact strongly with particular work materials. Table 3.6 summarizes some of the critical temperatures for the first two circumstances. Oxidation is not often critical for failure. In turning, the hottest tool regions are generally shielded from oxygen by the chip contact (although there is some exposure around the edges). There is more opportunity for oxidation in interrupted cutting conditions such as milling. These considerations are of more importance to wear (Chapter 4) than to failure. Structural change is more critical to failure. High speed steels soften rapidly as their structures over-temper, at temperatures from 550˚C upwards, depending on their composition. The microstructure of the binder phase of WC-Co changes with time at temperatures over 900˚C: a brittle phase, a mixed W–Co carbide known as the η-phase, forms as a result of WC dissolving in the cobalt binder (Santhanam *et al.*, 1990). Its formation is very slow at 900˚C: it does not become severe until 950˚C.

Table 3.6 Tool material oxidation and structural change temperature ranges

	Temperature range (°C) for:	
Tool material	Oxidation	Structural change (and nature of change)
High speed steel	–	> 600 (over-tempering)
WC-Co carbide	> 500	> 900–950 (solution of WC in Co)
Mixed carbides/cermets	> 700	–
Ceramics	–	> 1350–1500* (intergranular liquids)
PcBN	–	> 1100–1350 (change to hexagonal form)
PCD	> 900	> 700 (change to graphite)

*: very composition dependent – these temperatures indicate what is achievable.

Table 3.7 Tool/work chemical or adhesive interaction severities

Tool materials	Interactions with		
	Ni–Cr heat resistant alloys	Carbon steels	Ti alloys
WC-Co carbide	weak	strong	moderate
WC-TiC-TaC-Co carbide	weak	moderate	very strong
Ti(C,N)-Ni-Co cermet	moderate	weak	very strong
Al_2O_3 ceramic	weak	none*	very strong
Al_2O_3/TiC ceramic	weak	none	very strong
Al_2O_3/SiC(wh.) ceramic	weak/moderate	moderate	very strong
Si_3N_4 based ceramics	weak/moderate	strong	very strong
PcBN	none	weak	moderate
PCD	moderate	moderate	moderate

*: but Al_2O_3 can react with non-metallic silicate inclusions in steel.

This phase is easier to avoid with WC-TiC-TaC-Co carbides. Ceramic cutting tools undergo a sudden loss of strength if the temperature rises to a level at which grain boundary phases, often associated with sintering agents, become liquid; for example, around 1350˚C for Si_3N_4 and above 1500˚C for Al_2O_3. Finally, hard, cubic, boron nitride reverts to its soft hexagonal form, and diamond reverts to graphite at temperatures above 1100˚ and 700˚C respectively. All these temperatures should be regarded as approximate only, to indicate a ranking of thermal resistance.

High temperature interactions between tool and work materials are considered in Table 3.7. It should be possible to give critical temperatures for the onset of severe chemical reactions, based on knowledge of the phase diagrams for the materials involved. But tool performance depends on adhesion to the chip as well. Table 3.7 is based on common experience of the conditions of high speed machining, at a qualitative level. On the whole, a particular tool would not be used to machine a metal with which it had a strong or very strong interaction. Thus, there is a clear link between Table 3.7 and Table 3.2. Neither alumina nor silicon nitride ceramics are recommended for titanium alloys because of the very strong adherence of Al_2O_3 to and solubility of Si in Ti; but they are recommended for Ni-Cr heat resistant alloys because they are relatively inert in contact with these. Table 3.7 distinguishes between the suitability of WC-Co and other carbides and cermets for the machining of steels: WC-Co is not used at high cutting speeds because of rapid crater wear. In Table 3.2 all carbides (and coated carbides) are considered together, with the result that they are described as both good and all right for cutting steels.

However, there are differences between the tables. In part, these differences stem from the fact that a tool is chosen not only for its inertness with a work material but also because of its resistance to mechanical failure. However, just considering inertness, there is one (at first sight) surprising difference between the two. PCD is described as interacting moderately with all the alloys in Table 3.7 but is recommended for machining only Ti alloys in Table 3.2. In fact, it reacts strongly with Ti, but only over a certain temperature. Below that temperature there is a low adherence between the two. As already indicated in Chapter 2 (Figure 2.20) the high thermal conductivity of PCD tools helps the machining temperature to be kept low.

These considerations of the limiting conditions of tool use now give way to a description of tool properties in less thermally severe situations.

Mechanical property changes

Below the limiting temperatures of the previous paragraph, all tool materials become softer as their temperature increases. The left-hand panel of Figure 3.22 gives representative data, mainly from manufacturers' sources, of the reduction of Vickers Hardness with temperature. The right-hand panel replots this and further results as hardness at temperature relative to hardness at room temperature. As a first approximation, the relative hardnesses of all tool materials vary with temperature in the same way, up to 500˚C. At higher temperatures, the reduction in relative hardness with temperature falls into ranges depending on the tool material type. The hardness of high speed steels falls most rapidly. The carbides and cermets form the next group. The alumina ceramics, PcBN and PCD all soften relatively at the same rate. The silicon nitride base ceramics are the most temperature resistant group: without this quality, they would hardly find use as cutting tools at all.

On the other hand, tensile rupture stress varies only slightly with temperature, up to the tool's limiting usefulness temperature. Figure 3.23 gathers representative data for a range of commercial tool materials. TRS at elevated temperatures is generally within ± 25% of its room temperature value.

Thermo-elastic property changes

For completeness of information, this paragraph considers how the thermal conductivity, heat capacity, Young's modulus and thermal expansion coefficient of tool materials change with temperature. Such changes, in practice, have only a minor influence on a tool's performance.

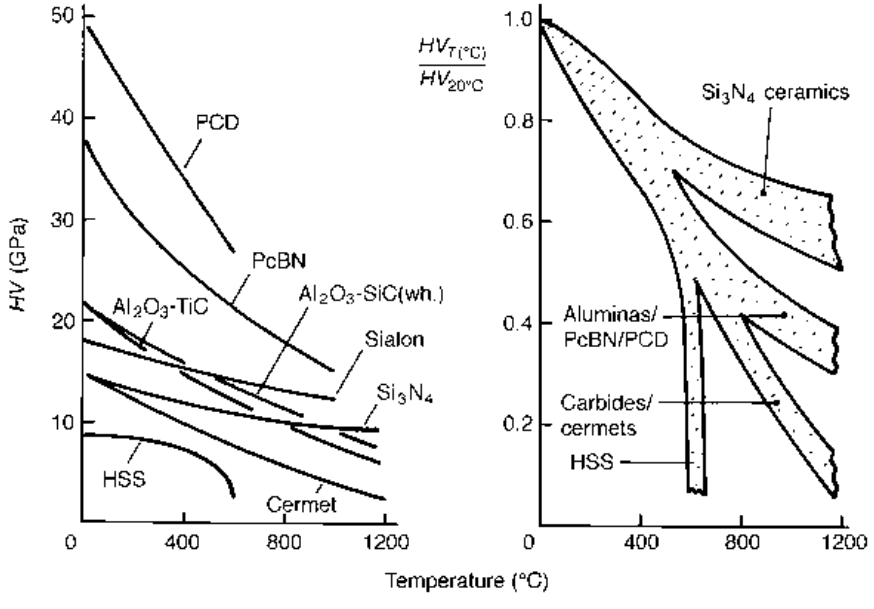


Fig. 3.22 Tool hardness changes with temperature: (a) representative values and (b) expressed relative to hardness at room temperature

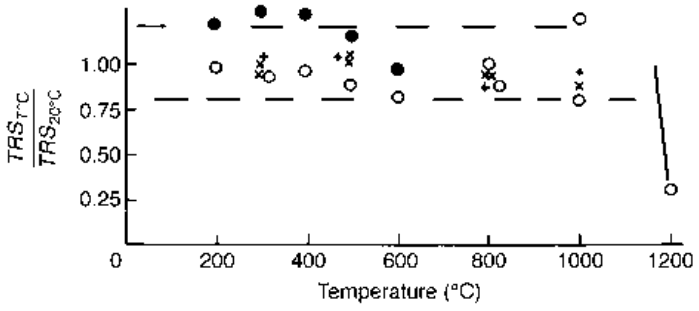


Fig. 3.23 Representative values of TRS relative to TRS at 20°C for HSS(•), carbide/cermet (x), Al_2O_3(o) and Si_3N_4(+) based tool materials

Figure 3.24(a) presents representative values of thermal conductivity. The changes that occur with temperature are less than the differences between one tool group and another. In principle, changes with temperature influence the partition of heat between the chip and tool (Chapter 2.3), but in fact these changes are only rarely large enough to have a significant effect. Conductivity also can influence the thermal stresses in a tool. If, however, in Figure 3.21, high temperature instead of room temperature conductivity values are used to rank a tool's thermo-elastic behaviour, the relative positions of the different tool groups in the figure are changed only slightly.

Figure 3.24(b) presents data on how heat capacity, Young's modulus and expansion coefficient vary with temperature, relative to their room temperature values. The heat capacity of all materials rises, and so the diffusivity falls, with temperature. The only effect of this would be – in interrupted cutting conditions such as milling – marginally to increase the time to establish a steady temperature field in the tool. As far as changes of Young's modulus and the expansion coefficient are concerned, the former falls and the latter rises with temperature. The product of the two remains almost unchanged. As TRS does not change much with temperature, neither does TRS/($E\alpha_e$).

3.2.5 Tool property changes with cyclic loading

In a milling operation in which, for example, the cutter is rotating at 1000 rev/min, each cutting edge receives 2×10^4 impacts in a 20 min cutting period. Cutting edges may experience fluctuating forces even in turning if the chip formation process is unsteady. For example, when turning cast iron, a discontinuous chip is formed almost every time the tool moves the feed distance. The number of cyclic loadings in a time t is then $U_{work}t/f$. For $U_{work} = 200$ m/min, $t = 20$ min and $f = 0.1$ mm, the number of cycles is 4×10^7. Force fluctuations in built-up-edge conditions occur at similar or slightly lower frequencies. Consequently, there is an interest in knowing how a tool material may survive in fatigue conditions, up to around 10^4 to 10^8 loading cycles.

There is not much published information on the tensile fatigue of cutting tool materials. Figure 3.25 presents some sample cyclic loading – life data for a range of tool materials, mainly from manufacturers' information, obtained from four-point bending conditions. Tensile stresses of around 0.5TRS will produce failure in the order of 10^6 to 10^8 loading cycles.

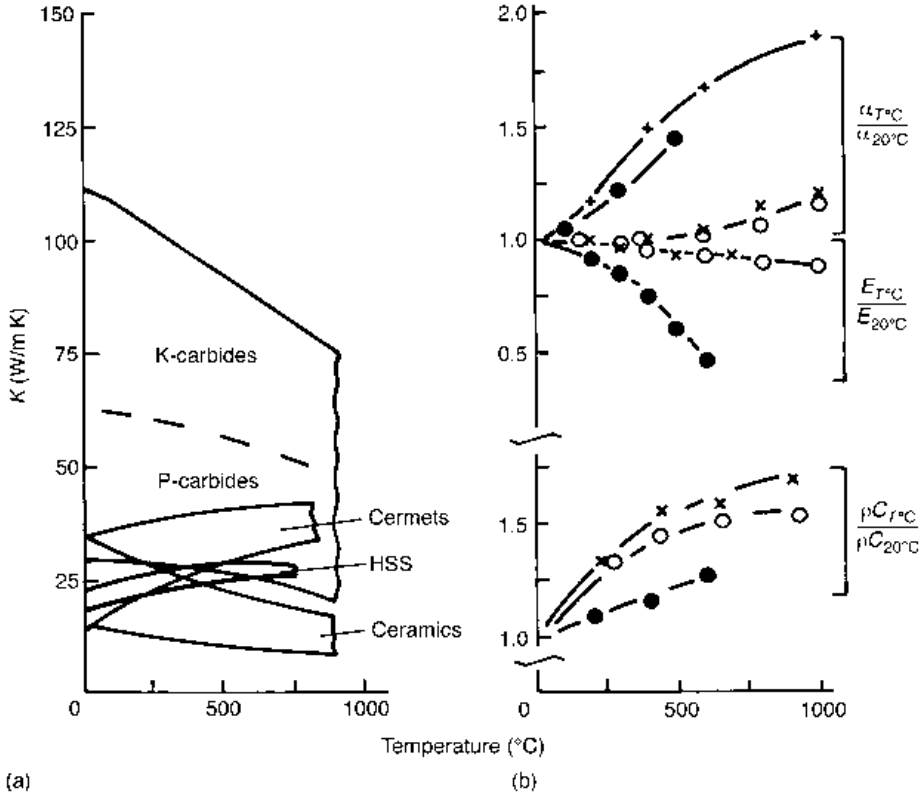


Fig. 3.24 The dependence on temperature of (a) thermal conductivity and (b) relative heat capacity, Young's modulus and thermal expansion coefficient, for HSS(•), carbide/cermet (x), Al_2O_3(o) and Si_3N_4(+) based tool materials

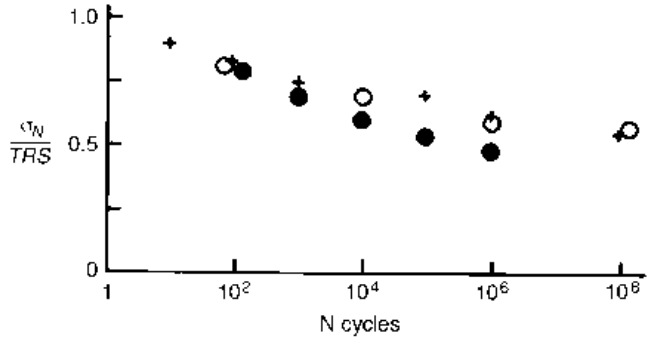


Fig. 3.25 Representative bending fatigue behaviour at room temperature of three tool materials: HSS(•), Al_2O_3(o) and Si_3N_4(+) based

3.2.6 Interim summary

The previous section suggests that, to avoid failure by fatigue, in a typical tool life time, a tool and the tool geometry should be selected to maintain the maximum tensile stress, caused by the cutting forces, at less than half the tool's TRS. In turning and milling operations, productivity demands that the chosen feeds and speeds are as large as temperature rises in the tool allow. Figure 3.22 (right-hand panel) suggests that the Vickers Hardness of a tool material at its operating temperature may be in the region of 0.4 to 0.6 of its room temperature value. A broad generalization is that a tool material and its geometry should be selected so that its loading brings it only half way to its room temperature plastic yielding. In this section, the tool material property data of Section 3.2 will be integrated with the work material property data of Section 3.1 to lead to predictions of how large the wedge angle of a tool should be to prevent failure by plastic yielding or fracture, depending on the tool material and the work material. When immediate tool failure by plastic yielding or fracture is avoided, tool life is determined by the gradual development of damage, which is the subject of Chapter 4.

Figure 3.19 provides a basis for analysing whether a work material, characterized by a particular shear flow stress k_{work} on the primary shear plane, will cause a tool of wedge angle β to yield or fracture. If the *HV* and TRS ranges of particular groups of tool materials are superimposed on to this figure, it is converted to one that can be used to assess how particular tools, characterized by their material properties and wedge angle, will perform.

From the initial considerations in this section, the working ranges of *HV* and TRS for a particular tool material have been considered to be half their room temperature values. These values have been taken from Figure 3.20 and superimposed on to Figure 3.19, to create Figure 3.26.

As an example of the use of Figure 3.26, consider the machining of a work material for which k_{work} = 600 MPa. Following the double-dashed line in the figure, if the material were machined by a cemented carbide of mid-range *HV* and TRS, β would need to be

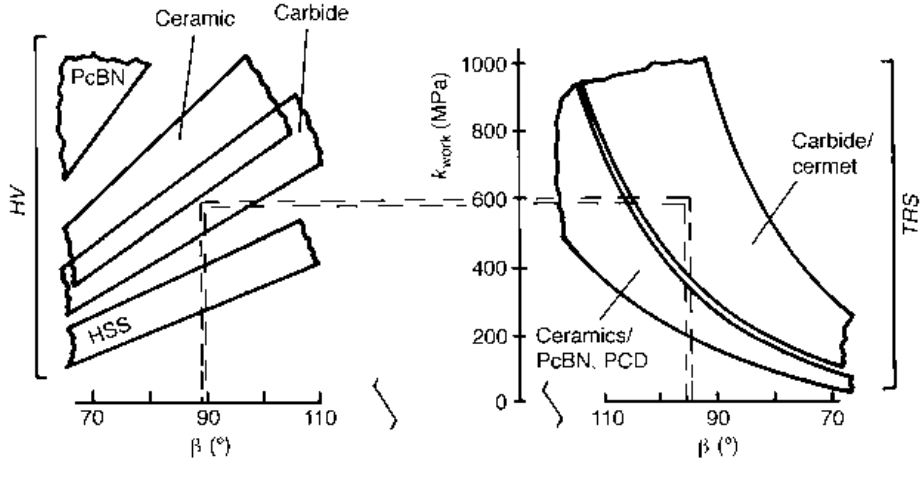


Fig. 3.26 A way to estimate minimum tool wedge angles β to avoid failure of a given tool material (specified by HV and TRS), acted on by stresses characterized by k_{work}, developed from Figure 3.19

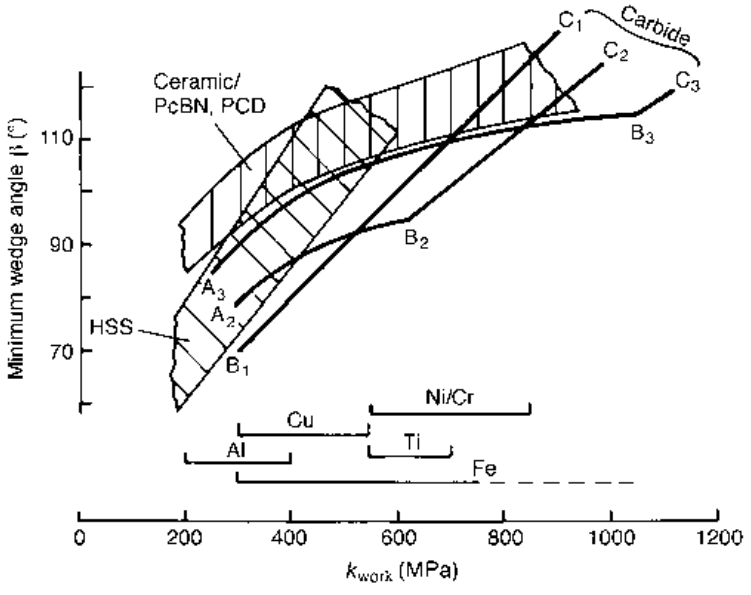


Fig. 3.27 Minimum values of β to prevent tool plastic or brittle failure, derived from Figure 3.26

chosen to be at least 90° to avoid plastic yielding and to be at least 95° to avoid fracture. In this case, the performance of the tool is limited by fracture. (The TRS range of HSS tools has been omitted from Figure 3.26, as it is found that fracture never limits the performance of these tools.)

If the minimum values of β to avoid failure are estimated from Figure 3.26 for all groups of tool materials, for all realistic values of k_{work} (from 200 MPa to 1200 MPa – the latter limit being appropriate for hardened steels), a picture can be created of how tool wedge angles should be chosen for different materials' combinations. The results of such an exercise are shown in Figure 3.27. It shows how minimum β values increase with k_{work}, for high speed steel, cemented carbide and ceramic, PcBN and PCD tools. The ranges of k_{work} appropriate to Al, Cu, Fe, Ti and Ni/Cr alloys have been taken from Section 3.1. The next few paragraphs discuss its results in more detail.

The β limits for HSS tools are found always to be determined by plastic failure. The figure suggests that for the lowest k_{work} aluminium alloy and the hardest high speed steel, β can be as small as 60°. However, for the hardest copper alloys, β should increase to 110°. At the opposite extreme, the β limits for ceramic, PcBN and PCD tools are always determined by fracture. For the lowest k_{work} aluminium alloy, the range of β is from 85° to 95°. For k_{work} = 800 MPa, the range is from around 110° to 130°. The response of cemented carbide tools is between these extremes. The behaviour of the softest, toughest, carbides (line B_1C_1 in the figure), is limited by plastic failure. The hardest, most brittle, carbides, on the other hand, are almost entirely limited by fracture. Line $A_3B_3C_3$ represents such a carbide. The portion A_3B_3 represents a fracture limit. Only for k_{work} > 1050 MPa (the portion B_3C_3), is the tool limited by its hardness. The line $A_2B_2C_2$ represents the behaviour of a mid-range carbide, for k_{work} < 620 MPa limited by brittleness, for k_{work} > 620 MPa limited by hardness.

The conclusions from Figure 3.27 of how large the wedge angle of a tool should be are broadly born out in practice. High speed steel tools with rake angles as large as 30° may be used to machine aluminium alloys. Ceramic tools frequently have negative rakes or negative chamfers (see Section 3.2.8) of –15° to –30°. Cemented carbides become limited by their hardness once work material shear flow stress increases above 800 MPa. For larger flow stresses, ceramic tools become more attractive because of their greater hardness. However, general experience suggests that, quantitatively, the β limits of Figure 3.27 are too large, by perhaps 5° to 10°. Nevertheless, the figure usefully guides the choice of tool materials and their shape to avoid mechanical failure.

Finally, it must be written that all the considerations of this section have been in terms of plane rake faced tools. In practice, cutting edges are strengthened against failure by edge preparations that include radiusing and chamfering. The minimum wedge angles of this section should more properly be interpreted as local to the cutting edge. The topic of tool cutting edge geometry is more fully considered in Section 3.2.8.

3.2.7 Tool coatings

The microstructure that gives a tool its required bulk hardness and toughness may not be the best to give the rake and clearance surfaces the best wear resistance. Cemented carbide tools illustrate this very well. The toughest and hardest can be made from WC-Co (K-grade) materials – and WC-Co has the highest thermal conductivity and is the cheapest material too; but WC-Co suffers from severe crater wear when cutting steels at high speed. Originally, this led to steels being machined by WC-TiC-TaC-Co (P-grade) materials, but these are inherently less tough, so higher cobalt contents are needed – leading to less hard grades. Fortunately, tool geometry is also available to be modified – so a satisfactory solution can be found to the machining of steels with cemented carbides. Nowadays, the solution is to use coated tools – their bulk optimized to resist failure and their surfaces coated to resist wear. The field of endeavour that seeks to optimize bulk and surface properties by coating is known as *Surface Engineering*. Various estimates indicate that, currently, from around 70% of cemented carbides sold for turning and 25% for milling, to up to 80% of all cemented carbides, are coated. Whatever the real figure, it is a clear majority of cemented carbides. In this section, the nature and choice of coatings are briefly considered, as well as the variety of manufacturing processes that lead to different qualities and applications. The main focus will be coated carbides, but high speed steel tools are also frequently coated (Hoyle, 1988), and there are possibilities of coating ceramic tool materials (Komanduri and Samanta, 1989; Santhanam and Quinto, 1994). These will also be mentioned.

Coating materials and properties

Coatings should be harder than the cemented carbides themselves, in order to give benefit in resisting abrasive wear, must be more inert to resist chemical wear, and must adhere well to the substrate. The three most common materials that satisfy these criteria (others will be mentioned later) are TiN, TiC and Al_2O_3. It is commonly said that TiC is the hardest and therefore best in resisting abrasive wear, that Al_2O_3 is the most inert, and TiN is a good all purpose material. In fact, the choice of coating depends on its use, as indicated in Table 3.8. The flank wear information comes from (Santhanam *et al*., 1990) and that for crater wear from an industry source. Information on WC-Co is included for comparison. The

Table 3.8 The ranking of coating materials for flank and crater wear resistance

Rank (1 = best)	Flank wear resistance* at a cutting speed (m/min) of		Crater wear resistance in turning		
	150	275	Carbon steels	Stainless steels	Ti alloys
1	TiC	Al_2O_3	Al_2O_3	TiN	(WC-Co)
2	TiN	TiC	TiN	TiC	TiC
3	Al_2O_3	TiN	TiC	Al_2O_3	TiN
4	(WC-Co)	(WC-Co)	(WC-Co)	(WC-Co)	Al_2O_3

* turning 0.45%C steel at a feed of 0.4 mm/rev.

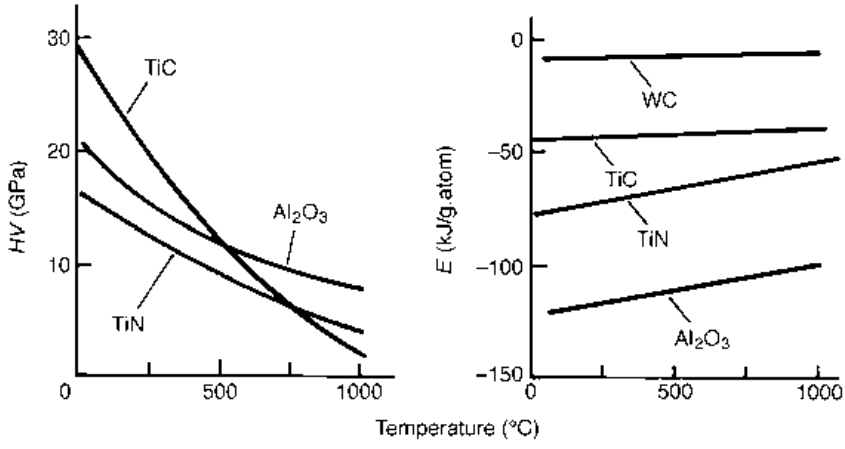


Fig. 3.28 The temperature dependence of hardness and standard free energy of formation of some coating materials, from Santhanam and Quinto (1994)

hardest material is in fact the best for flank wear resisitance at the lower cutting speed and the most inert (against steel) is the best for crater wear resistance. However, coating material properties change with temperature; and factors other than abrasion resistance and inertness are important too.

Figure 3.28 shows how the hardness and free energies of formation of coating materials vary with temperature (and Table 3.9 gives other property data). TiC is only hardest at room temperature. Above 600˚C, Al_2O_3 is the hardest. As cutting speed increases, so does the flank temperature. Additionally, the cutting process is likely to become steadier, with smaller force fluctuations. This also favours the use of Al_2O_3, which is more brittle than the other coatings (it is not easy to put a number to this – so there is no data on fracture behaviour in Table 3.9). These two factors account for the changes in flank wear resistance rankings with cutting speed of the coating materials.

The free energy of formation of a compound is the internal energy change associated with its creation from its elements, for example for the creation of Al_2O_3 from aluminium and atomic oxygen. The more negative it is, the more stable is the compound. Figure 3.28 confirms that Al_2O_3 is more stable than TiN by this measure, and TiN is, in turn, more stable than TiC or WC. However, the ranking for crater wear resistance only follows this order (Table 3.8) for turning carbon steels. For stainless steels, Al_2O_3 is reduced to the

Table 3.9 Thermal and elastic properties of cemented carbides and their coatings

Material	α_e [$10^{-6}K^{-1}$]	Young's modulus [GPa]	Thermal conductivity [W/m K]	
			100°C	1000°C
TiC	7.4–7.7	≈ 450	24–33	38–41
TiN	9.4	≈ 250	19–21	25–26
Al_2O_3	8.4–9.0	≈ 400	20–28	6–7.5
WC-Co*	4.5–6.0	550–650	75–120	50–75
WC-TiC-TaC-Co*	5.8–6.8	490–560	25–55	20–50
Co	≈ 12	≈ 180	70	–

* from Table 3.5.

third rank. This again illustrates the importance of mechanical effects as well as chemical effects in wear. The relatively brittle Al_2O_3 cannot stand up to the strong force fluctuations caused by serrated chip formation when machining stainless steels. The complete reversal of rank order relative to steels when turning Ti alloys just indicates the care that must be taken when applying thermodynamic principles to wear processes. Although Figure 3.28 gives free energies of the formation of coating materials from their elements, it says nothing about the free energies of other compounds that may be formed by reactions with titanium.

The practical conclusion is that all three coatings are useful. By the time other factors are considered, which stem from different manufacturing processes and the possibility of creating coatings in which all three materials exist in consecutive layers, it is easy to understand that surface engineering creates a large opportunity to optimize a tool for a particular operation. It becomes a marketing judgement whether to offer a tool specialized for a narrow use or one generalized for a broad application range.

CVD coatings

The earliest tool coatings were made by chemical vapour deposition (CVD). In the CVD process, inserts to be coated are placed in a hydrogen reducing atmosphere furnace – with the hydrogen typically at about 10% of atmospheric pressure. Gases containing the coating elements are added to the atmosphere and circulated through the furnace and over the inserts. The coatings are formed on the surfaces of the inserts, by chemical reactions between the gases, depending on the temperature of the surfaces. Typical temperatures for the formation of TiC, TiN and Al_2O_3 on the surfaces of cemented carbides are around 1000°C. Some of the furnace atmospheres and the reactions that lead to the coatings are:

$$TiCl_4(gas) + CH_4(gas) + H_2(gas) \Rightarrow TiC(solid) + 4HCl(gas) + H_2(gas)$$

$$2TiCl_4(gas) + N_2(gas) + 4H_2(gas) \Rightarrow 2TiN(solid) + 8HCl(gas)$$

$$2AlCl_3(gas) + 3CO_2(gas) + 3H_2(gas) \Rightarrow Al_2O_3(solid) + 6HCl(gas) + 3CO(gas)$$

Coating rates are around 1 μm/hr (for good performance coatings) and changing the reactive gases throughout the process can lead to the build-up of coatings with different compositions throughout their depth.

The first coating to be commercialized (in the early 1970s) was TiC on WC-Co. TiC is

a natural component of cemented carbides and it was found that its adhesion to the substrate was stronger than that of the other coating materials. Its thermal expansion coefficient (Table 3.9), although greater than that of the cemented carbide substrate, is closest to it. As the substrate cools down, tensile thermal strains are set up in the coating, but they are less than those that would be set up in the other coating materials. Even today, with multi-layer coatings common, TiC is frequently chosen for the layer closest to the substrate, because of its good adhesion and thermal expansion coefficient match. When other coating materials are built-up on it, the order tends to be that of increasing thermal expansion coefficient, to minimize thermal strains. Thus, common multi-layer coatings are:

$$TiC + Al_2O_3$$

$$TiC + Al_2O_3 + TiN$$

$$TiC + Ti(C,N) + TiN$$

where Ti(C,N) is a further coating type formed by a gradual change from CH_4 to N_2 in the reacting gases.

The thicker the coating, the longer its life may be expected to be. If the coating is too thick, however, it will lose the toughening reinforcement that it gains from its substrate. The tensile thermal strain resulting from its thermal mismatch with the substrate will further contribute to its failure. Even if cracks in a coating, caused by thermal strains, do not cause the coating to break away from the substrate, the cracks are sources of stress concentration that lead to a lowering of the substrate's resistance to fracture. For these reasons, practical CVD films are typically more than 4 μm but less than 12 μm thick.

The nature of their chemical formation results in their surfaces being quite rough. The roughness profile of a coated tool in Figure 2.24 is for a CVD coating. Its R_a is 0.5 μm. Further, the coating builds-up rather irregularly on a sharp cutting edge – it is common practice to hone the cutting edges of CVD coated tools, to the region of 40 μm to 70 μm. Such radii, particularly when increased by the thickness of the CVD coating are large compared with feeds of around 0.1 mm typical of finishing operations. Thus, the original CVD coatings were better suited to general and roughing turning operations than to finishing operations. Additionally, their inherent tensile residual stresses made them less able to stand intermittent cutting conditions as occurred in milling. CVD coatings were used in turning more than in milling.

Tool substrate compositions

These limitations have, to some extent, been reduced by the development of substrate compositions especially to support the coatings – it might be called *Subsurface Engineering*. Returning to a consideration of the adhesion of TiC to substrate material, although it adheres very well to WC-Co, it is possible, if close control of the process is not maintained, to damage the substrate microstructure during manufacture. An alternative to the reaction of $TiCl_4$ with CH_4 to form TiC is

$$TiCl_4(gas) + C(\text{from substrate}) + 2H_2(gas) \Rightarrow TiC(solid) + 4HCl(gas)$$

If this occurs, the loss of carbon from the substrate can result in the brittle η-phase (Section 3.2.4) forming. Some manufacturers have avoided this problem by developing coated tools on P-type substrates (the carbon content in these is less critical, and it is easier to avoid the

η-phase). In any case, using a P-type substrate would give added life to a tool used to cut steels once the coating wore off or if it failed. To compensate for loss of toughness on changing from K- to P-type, such substrates were typically manufactured with a greater %Co for a given grade of duty than if they were uncoated. Thus began the development of special substrate compositions for coated tools. By the early 1980s, substrates were being manufactured with surface layers containing from 1.5 to 3 times the amount of cobalt to that in the bulk, and from 10 μm to 30 μm thick, on near WC-Co bulk compositions. Toughness is maintained near the surface without reducing the hardness of the bulk. Considering the high thermal expansion coefficient of cobalt (Table 3.9), the surface layer of the substrate is better thermally matched to the coating materials and thermal strains are reduced. CVD-coated tools began to find uses in interrupted turning and light milling operations.

Considering the thicknesses of both the coatings and modified substrate surface layers, the composition (and hence thermal and mechanical properties) of CVD-coated tools can vary over depths of up to around 40 μm. This is not insignificant relative to the size of the stressed and heated regions during cutting. Detailed understanding of the interactions between the graded surface compositions and the mechanical and thermal fields generated in machining, leading to still further improvements in tool design, continues to develop.

PVD coatings

An alternative process for manufacturing coatings is Physical Vapour Deposition (PVD). It is similar to CVD in its productivity (in its basic form, deposition rates are also around 1 μm/hr) but requires substrates to be heated only to a few 100˚C, say 500˚C, so coatings can be deposited without the need to guard against unfavourable changes to the substrate. In contrast to CVD, in which the metallic elements of the coating are obtained from gases at around 10% atmospheric pressure, in PVD the metallic elements are obtained from solids in a high vacuum chamber environment. There are many variants of the process but all involve establishing a large electric potential difference (of the order of kV) between the substrate and a solid source of elements to be deposited on the substrate; and creating a glow discharge plasma between the two, typically with argon gas at low pressure. Material is evaporated from the source (by some form of heating or bombardment), is ionized in the plasma and is accelerated towards and adheres to the substrate. The source may have the composition of the material of the coating, or more commonly it may be a metal – for example titanium. In the latter case, for example in forming a TiN coating, nitrogen gas is also admitted to the plasma. The Ti ions combine with the nitrogen, to condense as TiN on the substrate.

The microstructure and properties of the coating are controlled by the substrate temperature and the deposition rate. It has been found that coatings can be grown with residual compressive stresses in them, but thicknesses are limited to about 5 μm. Coatings made by PVD are much smoother than by CVD and can be deposited on to sharp edges. Experience has shown that they are more suitable for milling operations (because of their compressive stresses) and finishing operations (because of the possibility of using sharp edged tools (down to 10 μm to 20 μm edge radius). The range of coating types is not as wide as with CVD. TiN was the first coating type successfully to be developed by PVD. This was followed by TiC and Ti(C,N); and (Ti,Al)N has also been developed. There is great difficulty in generating Al_2O_3 coatings with a strong, coherent microstructure. Cermets as well as cemented carbides are being coated by PVD.

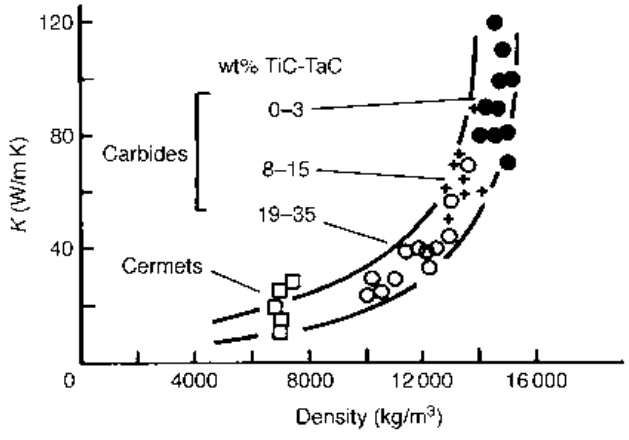


Fig. 3.29 The relation between thermal conductivity and density for coated cemented carbide and cermet substrate materials

Coating developments and summary

Coating technologies continue to develop. For example, there are intermediate processes between CVD and PVD in which coatings are formed with the chemical variability of CVD but in which the substrate needs to be heated only to, say, 800˚C. Today there is a wide variety of choice in the purchase of coated tools and production engineers rely heavily on the advice of tool manufacturers and their own practical trials. Tool manufacturers are rather secretive about their manufacturing processes; and even about what the substrate material is beneath a coating. When an engineer buys a coated tool he or she rarely knows what is beneath the coating. Short of cutting up a tool and examining it, the next best way of satisfying curiosity as to what is a tool's substrate, is to weigh it. There is a strong relation between density and carbide composition – and between that and tool thermal conductivity – as shown in Figure 3.29.

This section has concentrated on TiC, TiN and Al_2O_3 coatings on cemented carbides. At the time of writing, there is much activity in trying to develop PCD-coated tools. There are also many instances in which high speed steel tools are coated with PVD TiN, TiCN or TiAlN. Chromium nitride, boron nitride and boron carbide coatings are also under investigation. TiN and TiC coatings have also been found to be useful on silicon nitride ceramic tools. However, as far as this chapter is concerned, the main lesson is that Surface Engineering has enabled the substrates of cutting tools to be designed for hardness and toughness, separately from considerations of wear resistance. As far as CVD-coated tools are concerned, the depth over which material composition and properties change is significant relative to the distances over which stresses and temperatures penetrate the tool. For PVD-coated tools, the variations of composition and properties are much more superficial.

3.2.8 Tool insert geometries

At the start of Section 3.2, the stresses in a tool were considered, assuming the tool to have a plane rake face. This later led to a conclusion of the minimum wedge angle that a tool should have to avoid failure by yielding or fracture (Figure 3.27). In practice, many tools

do not have plane rake faces. This is particularly true of indexable inserts, manufactured by sintering and first mentioned in Chapter 1.

There are three main reasons for modifying cutting edge geometry: to strengthen the edge, to reduce cutting forces and to control chip flow. The basic ways of achieving these are illustrated, in two dimensions, in Figure 3.30.

Edge strengthening involves changing the edge shape over distances of the same order as the feed length. Figure 3.30(a) shows an edge region chamfered at an angle α_c over a length T and honed to an edge radius *R*. Recommendations in the early 1990s for edge preparation of ceramic cutting tools were typically chamfers for a length *T* between 0.5*f* and 0.75*f* for turning operations and 1.2*f* to 1.5*f* for milling; with α_c from 15° to 30° depending on the severity of the machining operation; and edge radii ranging from 0.013 mm to 0.076 mm for finishing operations, up to 0.13 mm in more severe conditions (Adams *et al.*, 1991). Today, with improved grinding procedures (and perhaps better ceramic tool toughness too), chamfer lengths for general machining are reduced to 0.1 to 0.4*f* for turning and 0.5*f* for milling; and edge radii in general machining are 0.02 to 0.03 mm, with no radiusing – only chamfering – for finishing operations.

Changes to reduce cutting forces involve altering the rake face over lengths of several times the feed (Figure 3.30(b)). The rake face beyond a land *h* of length between 1*f* and 2*f* is cut-away to a depth *d* typically also between 1*f* and 2*f*, established over a length *L* from 3*f* to 6*f*. The land restriction causes a reduced chip thickness.

A disadvantage of cutting away the rake as just described is that, generally, the chips become straighter, and in a continuous process (such as turning) this can lead to long unbroken chips that are difficult to dispose of. In order to control the flow, the cut-away region is usually ended in a back wall (Figure 3.30(c)), so that the cut-away forms some groove shape. When a chip hits the back wall, it is deflected and has a good chance of

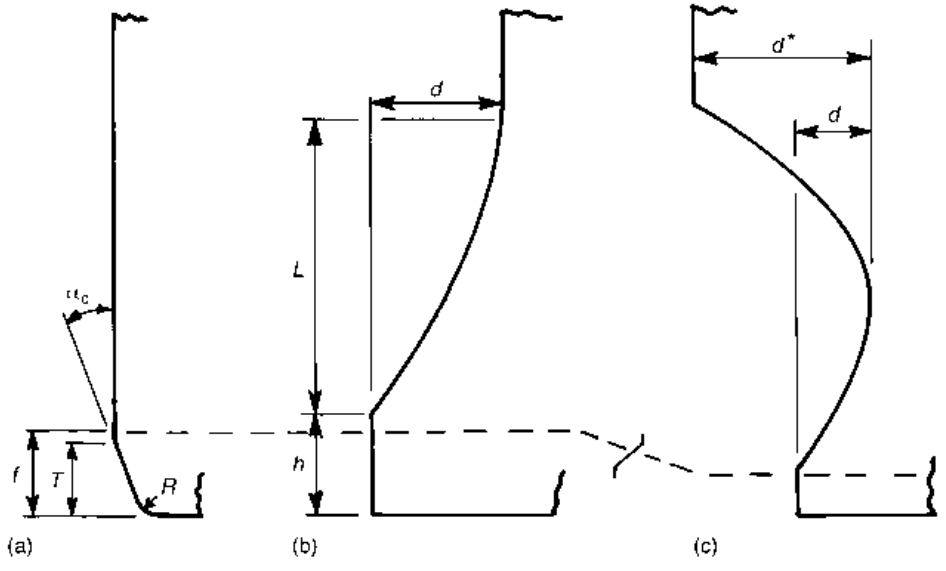


Fig. 3.30 Modifications to a square cutting edge for (a) edge strengthening, (b) cutting force reduction and (c) chip control

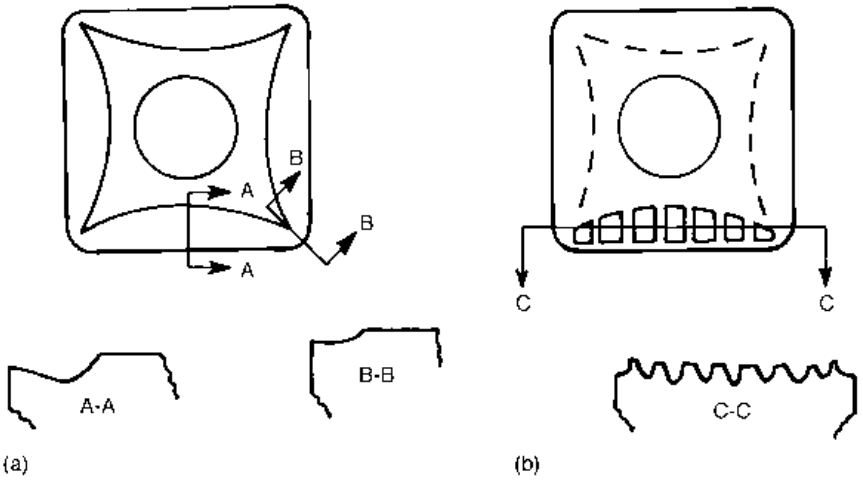


Fig. 3.31 Three-dimensional opportunities for edge strengthening and chip control, not to scale: rake profile varying (a) smoothly along a cutting edge and (b) in a 'rib and pocket' manner

breaking when its tip hits either the tool holder or the work. There is a wide variety of practical groove shapes. They can be curved or triangular, symmetrical or unsymmetrical. The height d^* of the back wall can be greater or less than the groove depth d. In some cases, the back wall is formed without a groove at all. Inserts can be designed for use over a wide range of feeds by creating the groove features as a series of terraces, so that the smallest feeds involve chip contact only with the terrace nearest the cutting edge and larger feeds result in contact over several terraces. Of course, the larger feed features of Figures 3.30(b) and (c) can be combined with the sub-feed strengthening features of Figure 3.30(a).

Figure 3.30 takes a two-dimensional view of a cutting edge. Real inserts are three dimensional – and this gives further opportunity for ingenuity in tool design. Sections as in Figure 3.30 can be varied along the cutting edge. This possibility is shown in Figure 3.31(a). The rake face groove at the corner of an insert can be shaped differently from that along the edge, either to ensure that the corner is strong enough or to help guide the chip away from the corner region, or both.

A different type of modification is shown in Figure 3.31(b). A curling chip is more likely to break, when it hits an obstruction, the larger is its second moment of area, I. Chips formed over plane or smoothly varying rake faces are approximately rectangular in section – and have a relatively small I-value. If they can be corrugated, their I value is raised. The rib and pocket form of the rake face in Figure 3.31(b) can cause such corrugation, if it is designed correctly. As an alternative to the rib and pocket style, the whole cutting edge may be made wavy, or bumps instead of pockets can be formed on the rake. Every manufacturer has a different way in which to achieve the same effect. Interested readers should look at manufacturers' catalogues or refer to a recent handbook (Anon, 1994).

Improved tools, combining new shapes with better surface engineering, continue to be developed. Finite element modelling, introduced in Chapter 6, is starting to contribute to these better designs. However, amongst its required inputs is material property information of the sort collected in this chapter.

References

Adams, J. H., Anschuetz, B. and Whitfield, G. (1991) Ceramic cutting tools. In: *Engineered Materials Handbook* Vol. 4 (*Ceramics and Glasses*). Metals Park, Ohio: ASM.

Anon (1994) *Modern Metal Cutting – A Practical Handbook*. Sandviken: AB Sandvik Coromant.

Brookes, K. J. A. (1992) *World Directory and Handbook of Hardmetals and Hard Materials* 5th edn. East Barnet, UK: International Carbide Data.

Hoyle, G. (1988) *High Speed Steels*. London: Butterworths.

Kobayashi, S. and Thomsen, E. G. (1959) Some observations on the shearing process in metal cutting. *Trans. ASME* **81B**, 251–262; and Eggleston, D. M., Herzog, R. and Thomsen, E. G. *Trans. ASME* **81B**, 263–279.

Komanduri, R. and Samanta, S. K. (1989) Ceramics. In: *Metals Handbook* 9th edn. Vol. 16 (*Machining*). Metals Park, Ohio: ASM.

Sata, T. (1968) Machinability of calcium-deoxidised steels. *Bull. Jap. Soc. Prec. Eng*. **3(1)**, 1–8.

Santhanam, A. T. and Quinto, D. T. (1994) Surface engineering of carbide, cermet and ceramic cutting tools. In: *Metals Handbook*, 10th edn, Vol. 5 (*Surface Engineering*). Metals Park, Ohio: ASM.

Santhanam, A. T., Tierney, P. and Hunt, J. L. (1990) Cemented carbides. In: *Metals Handbook* 10th edn, Vol. 2 (*Properties and Selection: Nonferrous alloys and Special Purpose Materials*). Metals Park, Ohio: ASM.

Trent, E. M. (1991) *Metal Cutting*, 3rd edn. Oxford: Butterworth Heinemann.

4

Tool damage

Chapter 3 considered cutting tool minimum property requirements (both mechanical and thermal) to avoid immediate failure. By failure is meant damage so large that the tool has no useful ability to remove work material. Attention is turned, in this chapter, to the mechanisms and characteristics of lesser damages that accumulate with use, and which eventually cause a tool to be replaced. In reality, there is a continuous spectrum of damage severities, such that there is no sharp boundary between what is to be considered here and what might in practice be described as immediate failure. There is some overlap between this chapter and the previous one.

Chapters 2 and 3 have demonstrated that cutting tools must withstand much higher friction and normal stresses – and usually higher temperatures too – than normal machine tool bearing surfaces. There is, in most cases, no question of avoiding tool damage, but only of asking how rapidly it occurs. The damages of a cutting tool are influenced by the stress and temperature at the tool surface, which in turn depend on the cutting mode – for example turning, milling or drilling; and the cutting conditions of tool and work material, cutting speed, feed rate, depth of cut and the presence or not of cutting fluid and its type. In Chapter 2, it was described in general that wear is very sensitive to small changes in sliding conditions. In machining, the tool damage mode and the rate of damage are similarly very sensitive to changes in the cutting operation and the cutting conditions. While tool damage cannot be avoided, it can often be reduced if its mode and what controls it is understood. Section 4.1 describes the main modes of tool damage.

The economics of machining were introduced in Chapter 1. To minimize machining cost, it is necessary not only to find the most suitable tool and work materials for an operation, but also to have a prediction of tool life. At the end of a tool's life, the tool must be replaced or reground, to maintain workpiece accuracy, surface roughness or integrity. Section 4.2 considers tool life criteria and life prediction.

4.1 Tool damage and its classification

4.1.1 Types of tool damage

Tool damage can be classified into two groups, wear and fracture, by means of its scale and how it progresses. Wear (as discussed in Chapter 2) is loss of material on an asperity

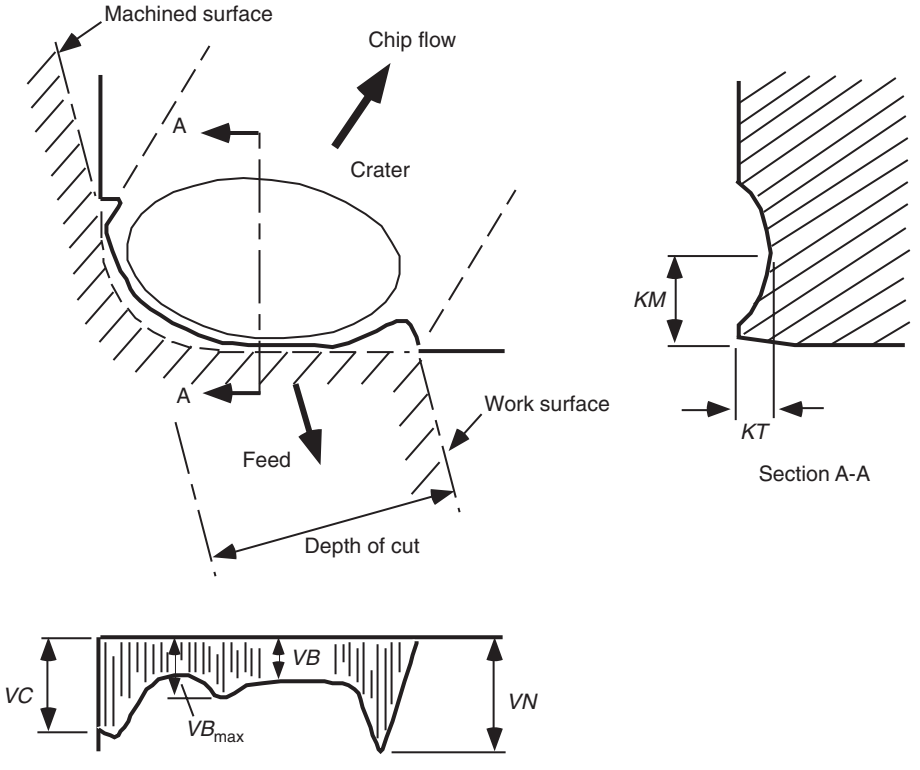
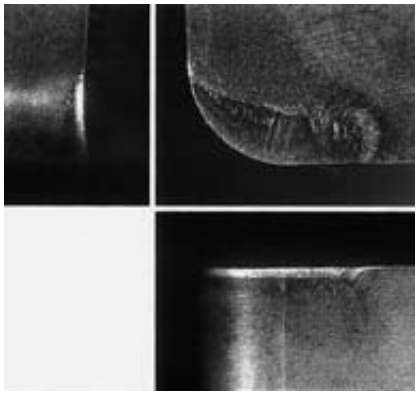


Fig. 4.1 Typical wear pattern of a carbide tool

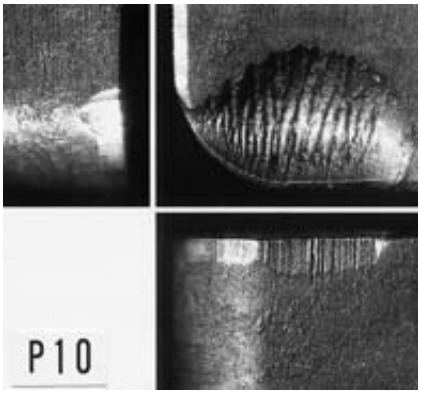
or micro-contact, or smaller scale, down to molecular or atomic removal mechanisms. It usually progresses continuously. Fracture, on the other hand, is damage at a larger scale than wear; and it occurs suddenly. As written above, there is a continuous spectrum of damage scales from micro-wear to gross fracture.

Figure 4.1 shows a typical damage pattern – in this case wear – of a carbide tool, cutting steel at a relatively high speed. Crater wear on the rake face, flank wear on the flank faces and notch wear at the depth of cut (DOC) extremities are the typical wear modes. Wear measures, such as *VB*, *KT* are returned to in Section 4.2.

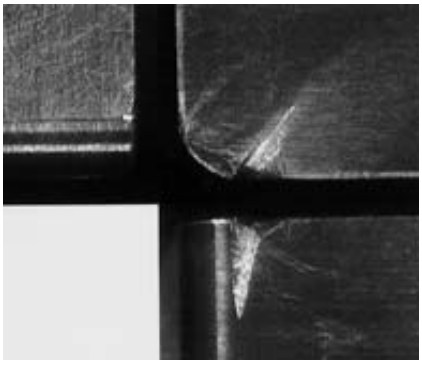
Damage changes, however, with change of materials, cutting mode and cutting conditions, as shown in Figure 4.2. Figure 4.2(a) shows crater and flank wear, with negligible notch wear, after turning a medium carbon steel with a carbide tool at high cutting speed. If the process is changed to milling, a large crater wear with a number of cracks becomes the distinctive feature of damage (Figure 4.2(b)). When turning Ni-based super alloys with ceramic tools (Figure 4.2(c)) notch wear at the DOC line is the dominant damage mode while crater and flank wear are almost negligible. Figure 4.2(d) shows the result of turning a carbon steel with a silicon nitride ceramic tool (not to be recommended!). Large crater and flank wear develop in a very short time. In the case of turning β-phase Ti-alloys with a K-grade carbide tool, large amounts of work material are observed adhered to the tool, and part of the cutting edge is damaged by fracture or chipping (Figure 4.2(e)).



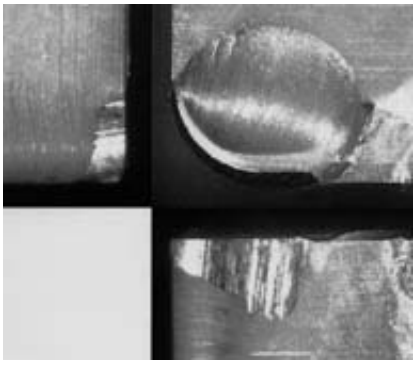
(a) Turning a 0.45% carbon steel



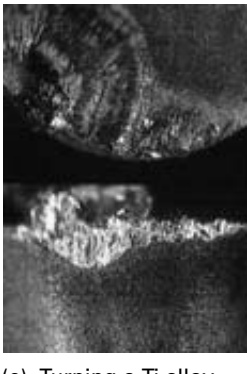
(b) Face milling a 0.45% carbon steel



(c) Turning Inconel 718



(d) Turning a 0.45% carbon steel



(e) Turning a Ti alloy

Fig. 4.2 Typical tool damage observations – both wear and fracture: (a) Tool: cemented carbide P10, $v = 150$ m min^{-1}, $d = 1.0$ mm, $f = 0.19$ mm rev^{-1}, $t = 5$ min; (b) tool: cemented carbide P10, $v = 400$ m min^{-1}, $d = 1.0$ mm, $f = 0.19$ mm tooth^{-1}, $t = 5$ min; (c) tool: Al_2O_3/TiC ceramic tool, $v = 100$ m min^{-1}, $d = 0.5$ mm, $f = 0.19$ mm rev^{-1}, $t = 0.5$ min; (d) tool: Si_3N_4 ceramic tool, $v = 300$ m min^{-1}, $d = 1.0$ mm, $f = 0.19$ mm rev^{-1}, t = 1 min; (e) tool: cemented carbide P10, $v = 150$ m min^{-1}, $d = 0.5$ mm, $f = 0.1$ mm rev^{-1}, $t = 2$ min.

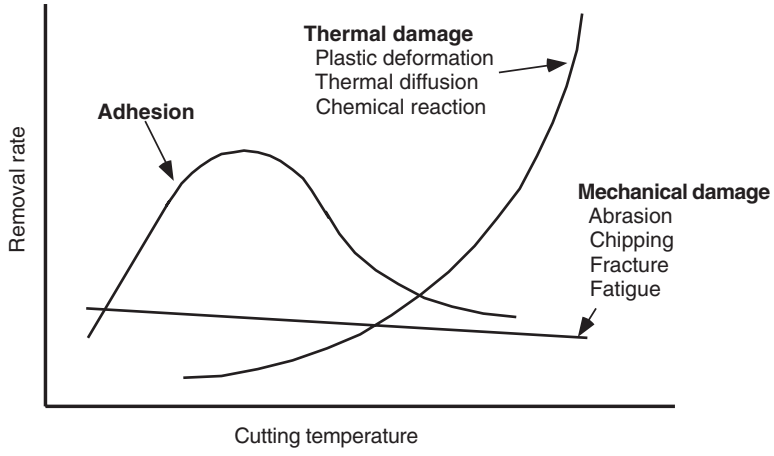


Fig. 4.3 Tool damage mechanisms and cutting temperature

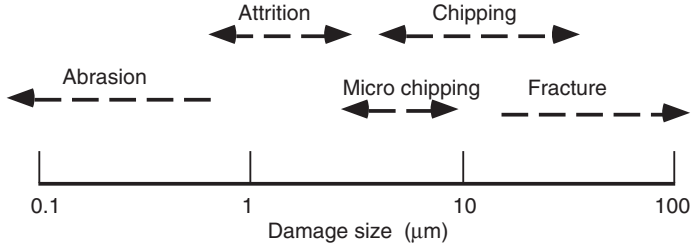


Fig. 4.4 Classification of mechanical damages

4.1.2 Causes of tool damage

Chapter 2.4 outlined the general conditions leading to abrasive, adhesive and chemical wear mechanisms. In the context of cutting tool damage, the importance and occurrence of these mechanisms can be classified by cutting temperature, as shown in Figure 4.3. Three causes of damage are qualitatively identified in the figure: mechanical, thermal and adhesive. Mechanical damage, which includes abrasion, chipping, early fracture and fatigue, is basically independent of temperature. Thermal damage, with plastic deformation, thermal diffusion and chemical reaction as its typical forms, increases drastically with increasing temperature. (It should be noted that thermal diffusion and chemical reaction are not the direct cause of damage. Rather, they cause the tool surface to be weakened so that abrasion, mechanical shock or adhesion can then more easily cause material removal.) Damage based on adhesion is observed to have a local maximum in a certain temperature range.

Mechanical damage

Whether mechanical damage is classified as wear or fracture depends on its scale. Figure 4.4 illustrates the different modes, from a scale of less than 0.1 μm to around 100 μm (much greater than 100 μm becomes failure).

Abrasive wear (illustrated schematically in Figure 2.29) is typically caused by sliding

hard particles against the cutting tool. The hard particles come from either the work material's microstructure, or are broken away from the cutting edge. Abrasive wear reduces the harder is the tool relative to the particles and generally depends on the distance cut (see Section 4.2.2).

Attrition wear occurs on a scale larger than abrasion. Particles or grains of the tool material are mechanically weakened by micro-fracture as a result of sliding interaction with the work, before being removed by wear.

Next in size comes chipping (sometimes called micro-chipping at its small-scale limit). This is caused by mechanical shock loading on a scale that leads to large fluctuations in cutting force, as opposed to the inherent local stress fluctuations that cause attrition.

Finally, fracture is larger than chipping, and is classified into three types: early stage, unpredictable and final stage. The early stage occurs immediately after beginning a cut if the tool shape or cutting condition is improper; or if there is some kind of defect in the cutting tool or in its edge preparation. Unpredictable fracture can occur at any time if the stress on the cutting edge changes suddenly, for example caused by chattering or an irregularity in the workpiece hardness. Final stage fracture can be observed frequently at the end of a tool's life in milling: then fatigue due to mechanical or thermal stresses on the cutting edge is the main cause of damage.

Thermal damage – plastic deformation

The plastic deformation type of thermal damage referred to in Figure 4.3 is observed when a cutting tool at high cutting temperature cannot withstand the compressive stress on its cutting edge. It therefore occurs with tools having a high temperature sensitivity of their hardness as their weakest characteristic. Examples are high speed steel tools in general; and high cobalt content cemented carbide tools, or cermet tools, used in severe conditions, particularly at a high feed rate. Deformation of the edge leads to generation of an improper shape and rapid material removal.

Thermal damage – diffusion

Wear as a result of thermal diffusion occurs at high cutting temperatures if cutting tool and work material elements diffuse mutually into each other's structure. This is well known with cemented carbide tools and has been studied over many years, by Dawihl (1941), Trent (1952), Trigger and Chao (1956), Takeyama and Murata (1963), Gregory (1965), Cook (1973), Uehara (1976), Narutaki and Yamane (1976), Usui *et al.* (1978) and others.

The rates of processes controlled by diffusion are exponentially proportional to the inverse of the absolute temperature θ. In the case of wear, different researchers have proposed different pre-exponential factors: Cook (1973) suggested depth wear h should increase with time t (equation 4.1(a)); earlier, Takeyama and Murata (1963) also suggested this and the further possibility of sliding distance s being a more fundamental variable (equation 4.1(b)); later Usui *et al.* (1978), following the ideas of contact mechanics and wear considered in Chapter 2.4, proposed wear should also increase with normal contact stress σ_n (equation 4.1(c)). In all these cases, a plot of ln(wear rate) against $1/\theta$ gives a straight line, the slope of which is $-C_2$.

$$\frac{dh}{dt} = C_1 \exp\left[-\frac{C_2}{\theta}\right] \tag{4.1a}$$

$$\frac{dh}{ds} = C_1 \exp\left[-\frac{C_2}{\theta}\right] \tag{4.1b}$$

$$\frac{dh}{ds} = C_1 \sigma_n \exp\left[-\frac{C_2}{\theta}\right] \tag{4.1c}$$

Figure 4.5 shows experimental results for both the crater and flank depth wear rates of a 0.25%C and a 0.46%C steel turned by a P20 grade carbide tool, plotted after the manner of equation (4.1c). Two linear regions are seen: in this case the boundary is at $1/\theta \approx 8.5 \times 10^{-4}$ K^{-1} (or $\theta \approx 1175$ K). The slope of the higher temperature data ($\theta > 1175$ K) is typical of diffusion processes between steels and cemented carbides (Cook, 1973). The smaller slope at lower temperatures is typical of a temperature dependent mechanical wear process, for example abrasion.

Diffusion can be directly demonstrated at high temperatures in static conditions. Figure 4.6 shows a typical result of a static diffusion test in which a P-grade cemented carbide tool was loaded against a 0.15% carbon steel for 30 min at 1200˚C. A metallographic section through the interface between the carbide tool and the steel, etched in 4% Nital (nitric acid

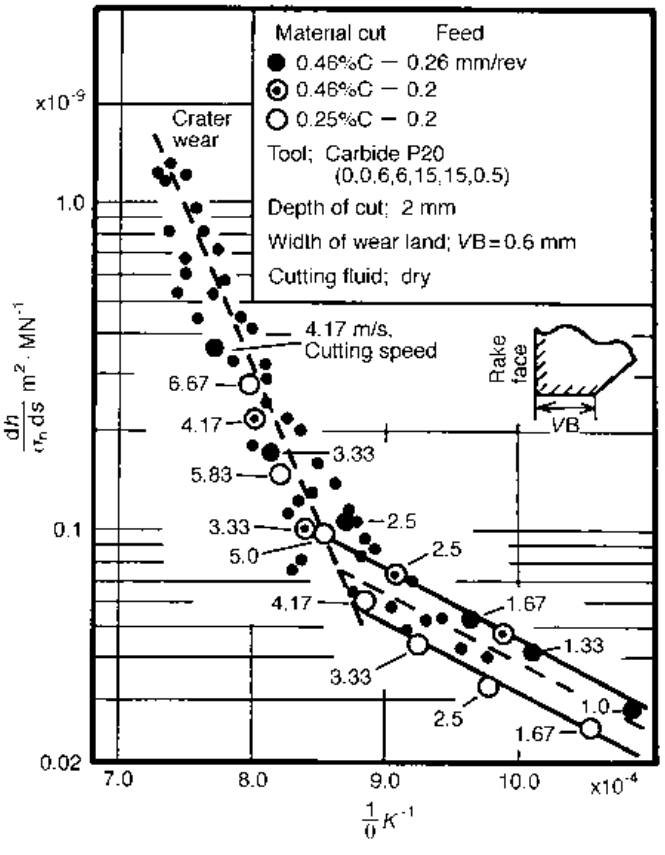


Fig. 4.5 Crater and flank depth wear rates for carbon steels turned by a P20 carbide, from Kitagawa *et al.* (1988)

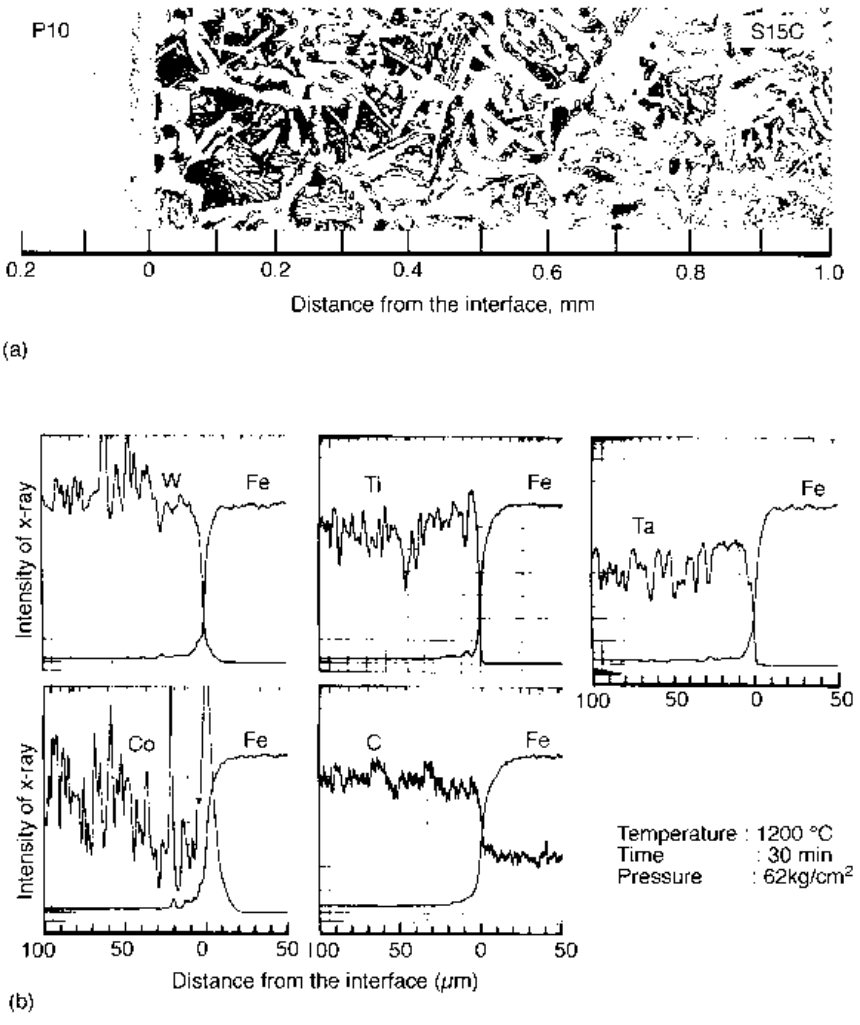
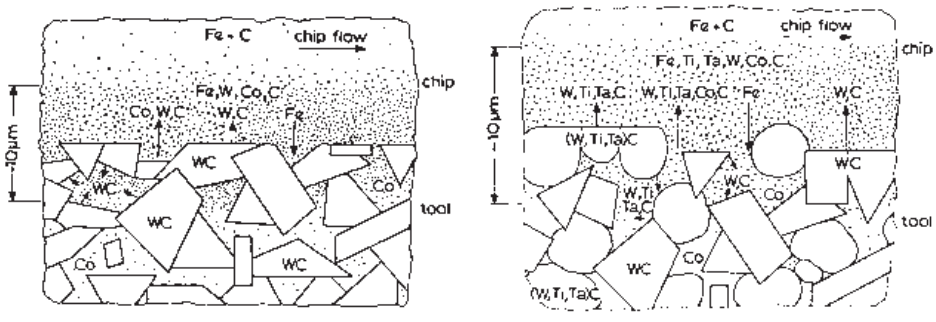


Fig. 4.6 Typical static diffusion test results, for P10 coupled to 0.15% C steel (Narutaki and Yamane, 1976): (a) Section through the interface etched by Nital; (b) Diffusion of elements analysed by EPMA

and alcohol) shows that the pearlite in the steel has increased from its original level. This means that carbon from the cemented carbide has diffused into the steel. Furthermore, electron probe micro-analysis (EPMA) shows that Co and W from the tool material also diffuse into the steel; and iron from the steel diffuses into the tool material. Many researchers agree that mutual diffusion is the cause of carbide tool diffusion wear, but there is not agreement in detail as to the mechanism that then results in material removal.

Naerheim and Trent (1977) have proposed that the wear rates of both WC-Co (K-grade) and WC-(Ti,Ta,W)C-Co (P-grade) cemented carbides are controlled by the rate of diffusion of tungsten (and Ti and Ta) and carbon atoms together into the work material, as indicated in Figure 4.7. This view is based on transmission electron microscope (TEM) observations on crater wear that show no structural changes in the tool's carbide grains within a distance



WC-Co cemented carbide tools

WC-(Ti, Ta, W)C-Co cemented carbide tools

Fig. 4.7 Model of diffusion wear, after Naerheim and Trent (1977)

of 0.01 μm of the tool–chip interface. The slower wear of P-grade than K-grade materials is explained by slower diffusion in the former than the latter case. Naerheim and Trent state that, in their cutting tests, pulled-out carbide grains were not observed adhering to the underside of chips. This was not Uehara's (1976) experience. He collected chips after turning a 0.47% C steel with a K-grade or a P-grade tool, dissolved the chips in acid to extract adhered carbides and finally passed the solution through a 0.1 μm filter, to classify the carbide sizes. With K-grade tools, he only observed carbides less than 0.1 μm in size, in accord with Trent. However, with P-grade tools he observed carbides greater than 0.1 μm in size. This suggests a different wear mechanism for K- and P-type materials.

Other examples of diffusion wear are the severe wear of diamond cutting tools, silicon nitride ceramic tools and SiC whisker reinforced alumina ceramic tools when machining steel. Carbon, silicon and nitrogen all diffuse easily in iron at elevated temperatures; and silicon nitride and silicon carbide dissolve readily in hot iron.

Thermal diffusion wear of carbide tools can be decreased if a layer acting as a barrier to diffusion is deposited on the tool. There are two types of layer in practice: one is as provided by coated tools; the other is a protective oxide layer deposited on the wear surfaces during cutting special deoxidized steels (for example Ca-deoxidized steels), commonly known as a 'belag' layer.

Thermal damage – chemical reaction

Chemical reaction wear occurs when chemical compounds are formed by reaction of the tool with the work material (or with other materials, such as oxygen in air or sulphur and chlorine in a cutting fluid) and when the compounds are then carried away by the chip (from the rake face) or work (from the flank faces).

Oxidation wear is best known. The cutting tool and/or work material are oxidized; and the tool surface, either directly weakened by oxidation or by reaction with oxidized work material, is then carried away by the chip. An example of oxidation wear occurs on the rake face of cemented carbide tools in high speed milling of steel. In milling, crater wear increases drastically with an increase of cutting speed, more so than in turning. Figure 4.8 shows the increase of both flank and crater wear with cutting edge engagement time for two different cutting speeds when turning and milling a 0.45% plain carbon steel under the same feed and depth of cut conditions. The increase of wear rate with cutting speed is

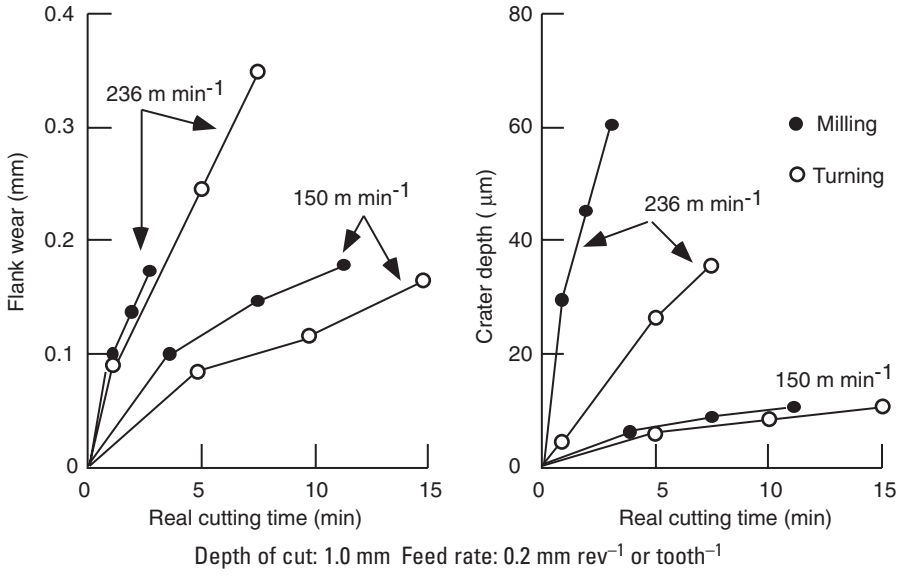


Fig. 4.8 Comparisons of wear when milling and turning a 0.45% carbon steel with a P10 cemented carbide tool, after Yamane and Narutaki (1983)

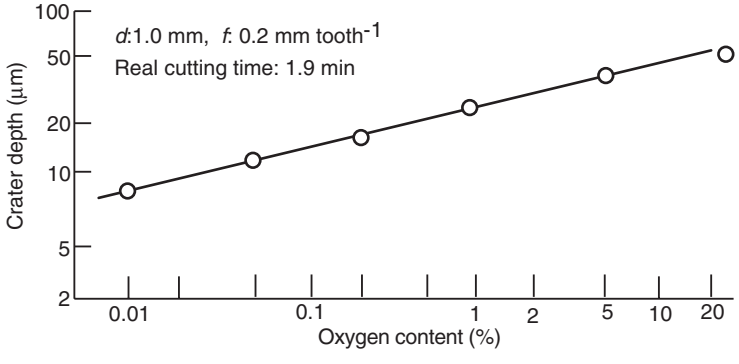


Fig. 4.9 Influence of oxygen on crater wear when milling a 0.45% carbon steel with a P10 carbide tool at 239 m/min (Yamane and Narutaki, 1983)

much less on the flank face than on the rake face, and the difference between milling and turning is much less on the flank too, probably because of the lower temperature on the flank. On the rake face, crater wear is almost the same in turning and milling at the relatively low cutting speed of 150 m/min, but at 236 m/min crater wear in milling becomes much more rapid than in turning. The wear mechanism is the oxidation of chips adhered to the rake face during the out-of-cut time in milling, to form FeO, followed by reaction of the FeO with the cemented carbide tool to weaken it.

Direct evidence of the influence of oxygen on carbide tool wear in milling comes from machining in a controlled atmosphere environment (Figure 4.9). In the same high speed

cutting conditions as in Figure 4.8, crater wear rate is reduced from its milling towards its turning level by reducing the amount of oxygen in the atmosphere.

Oxidation wear is also found with alumina ceramic tools in high speed milling of steels. The same oxidation of steel adhering to the tools' rake face occurs as with cemented carbide tools. FeO reacts with alumina at high temperatures to form weak, easily removed, mixed oxides.

Thermal damage – electro-motive force (EMF) wear

A further thermal wear mechanism, not listed in Figure 4.3, is considered to occur by several researchers. When cutting with electrically conducting tools, such as high speed steels and cemented carbides, the tool and work materials generally have different chemical compositions. A thermal EMF, based on the difference between the cutting temperature and room temperature is generated (the Seebeck effect). An electric current then flows around the closed circuit of work, tool and machine tool. Reports about its effect on wear do not always agree with one another and further investigation is required to establish if it is important.

Adhesion

The third damage mechanism in Figure 4.3 is adhesive wear. It occurs when work or chip material pressure welds (adheres) to the tool – which generally requires high temperature – and has high strength in that condition. Stainless steels, Ni-based super alloys and Ti alloys show this behaviour well. If the adhesive shear strength between the tool and welded chip or work is larger than a failure strength away from the interface, adhesive transfer between work, chip and tool will occur. Transfer will be from the chip or work to the tool if the weakest point is in the chip or work – this can lead to one type of built-up edge formation. Transfer from the tool to the work or chip occurs if the weakest point is in the tool. Adhesive wear is the repeated adhesion of material to the tool, followed by failure within the tool. At low cutting temperatures, it is reduced because of a low adhesion tendency. At high cutting temperatures it is reduced because thermal softening changes failure from within the tool back to the interface or to within the work or chip. Thus, it peaks at some intermediate temperature. Its peak magnitude increases as the tool's resistance to shear failure reduces.

Examples of adhesive wear are shown in Figures 4.10 and 4.11, after turning a Ni-based super alloy, Inconel 718, with an alumina/TiC ceramic tool. Figure 4.10 shows scanning electron microscopy (SEM) views of the DOC notch wear region after dissolving adhered chip material in nitric acid. The damaged surface of region A looks fractured rather than worn. Figure 4.11 shows the characteristic peaking of the notch wear at some temperature, in this case near to 1000˚C.

4.1.3 Tool damage and cutting conditions

Figure 4.3 is qualitative. The temperatures at which thermal and adhesion damages occur vary with the tool material, as well as with what work the tool is machining. Figure 4.12 shows some detail of this – there is some overlap with Chapter 3 (Tables 3.6 and 3.7).

Diamond, the hardest cutting tool material, starts to carbonize in air over 600˚C. It diffuses into iron or steel at higher temperatures, causing diffusion wear problems with these work materials.

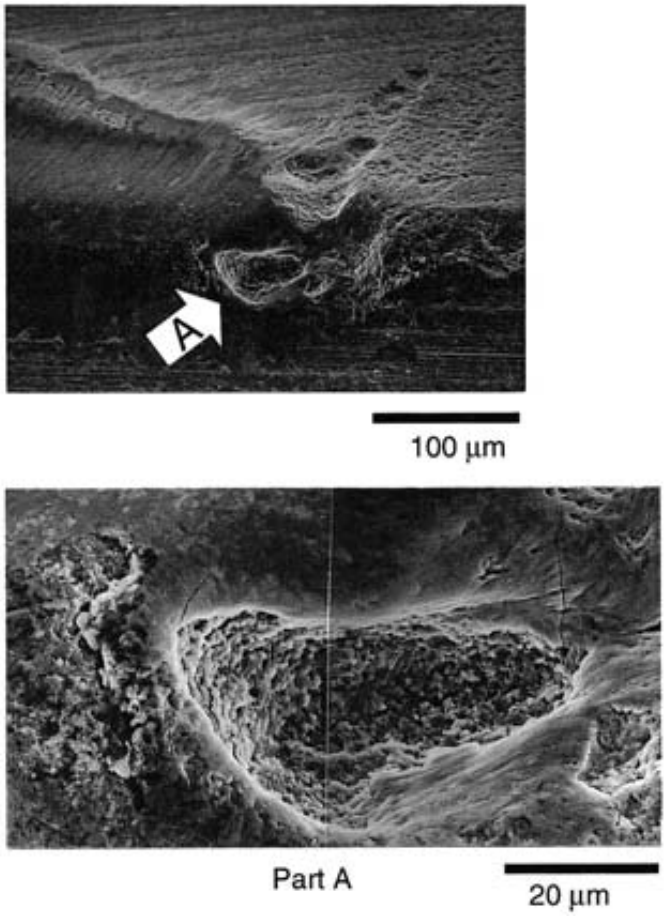


Fig. 4.10 SEM views of DOC region of an Al_2O_3/TiC tool after turning Inconel 718

Cubic boron nitride (cBN), a material that does not exist naturally but is synthesized under ultrahigh pressure and temperature, is stable against diffusion for practically all metals. Its weakness is its failure under shear, so adhesive wear occurs when cutting materials with a tendency to adhere to it.

Among ceramic tools, alumina, silicon nitride and whisker reinforced ceramics, alumina tools are the most stable. Silicon nitride reacts readily with iron or steel at high temperature, leading to diffusion wear. Diffusion wear also leads to tool weakening of whisker reinforced ceramics, if the whisker component reacts with the work material – SiC whiskers reacting with steel is an example of this. In general, all ceramic tools, because of their low toughness, are susceptible to adhesive wear when machining materials that adhere to them.

For cemented carbide tools, thermal diffusion wear increases remarkably at temperatures over 1000˚C, while the main thermal damage of HSS tools is plastic deformation, caused by hardness reduction over the tempering temperature range.

cBN, ceramic and cemented carbide tools also tend to react at high temperature with

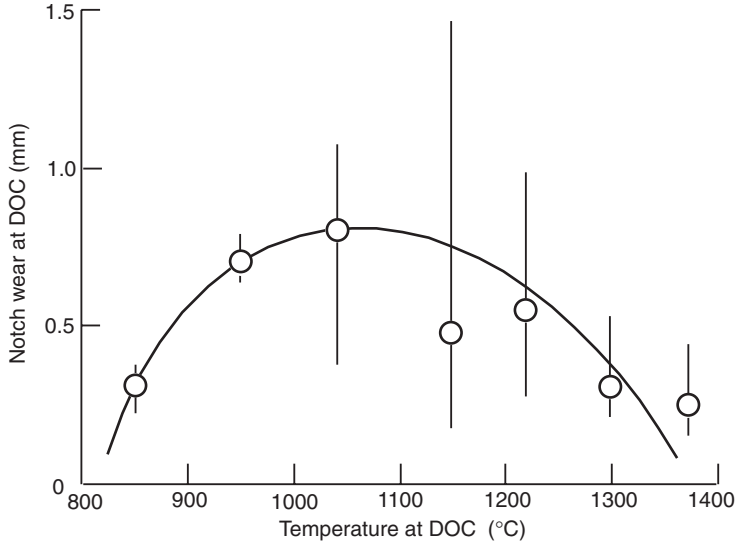


Fig. 4.11 Al_2O_3/TiC DOC notch wear variation with temperature, turning Inconel 718 for 30 s at various speeds at a feed of 0.19 mm and a depth of cut of 0.5 mm, after Narutaki and Yamane (1993)

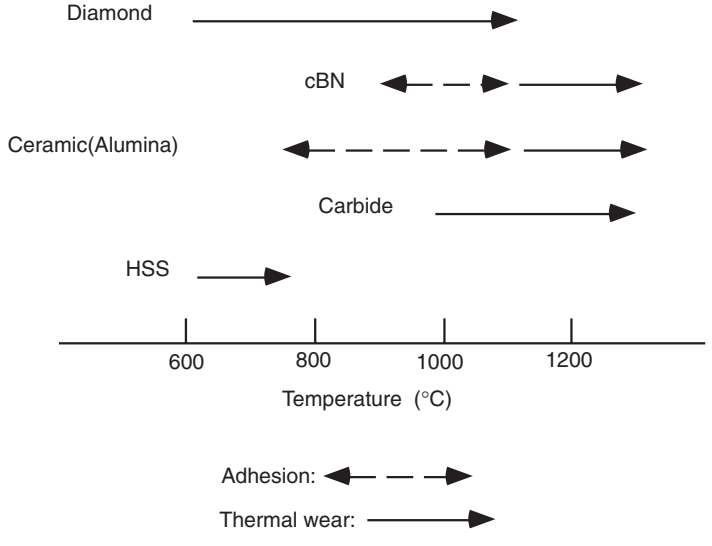


Fig. 4.12 Tool damage and cutting temperature

unstable oxides, such as FeO, or sulphides, such as MnS. If such oxides or sulphides are present at the tool/chip or work interface, thermal wear based on reactions with them must be considered.

In the case of coated tools, wear of the tool depends on the coated material, until it is worn away. TiC, TiN, (TiAl)N and Al_2O_3 are the most popular as they have high thermal stability and low reactivity with steels, compared with WC. Therefore, coated tools can

withstand thermal wear at higher cutting temperature. After the layer is worn away, the tool's wear rate, in general, returns to that of the substrate.

4.2 Tool life

As tool damage, by wear or fracture, increases, the surface roughness and accuracy of the machined surface deteriorates. Eventually the tool must be changed. Some criteria must be developed to decide when to do this. In factories there is a tendency to adopt flexible criteria according to the needs of a particular operation, while in laboratories inflexible criteria are adopted to evaluate tool and work material machining capabilities.

4.2.1 Tool life criteria

In factories, where the concern is acceptable accuracy or surface roughness, the most suitable way to judge tool life is by measuring the size or roughness of machined parts. Life is determined when the measured levels exceed a limit. In this approach, it is the parts that must always be measured; however, this is not always easy or cost-effective. An alternative is to monitor the wear of the tool and to judge the state of the part from a tool wear criterion. Unfortunately, there is not always a unique relation between tool wear and work material surface roughness or accuracy; and measuring tool wear increases non-productive time. Since chip shape sometimes changes with tool wear, change of chip shape is occasionally used as a criterion; but commonly in mass production industries, as a consequence of difficulties of measurement and interpretation, the simple approach is taken to specify the number of machined parts expected from a tool. The number is determined from a primary test, but it must be set with some safety factor – so this leads to machining costs being greater than they could be.

In laboratories, tool wear is almost always used as a life criterion because it is easy to determine quantitatively. The amount of flank wear is often used as the criterion because it is flank wear that influences work material surface roughness and accuracy. When abrasion is the main cause of flank wear, the wear pattern is relatively uniform and easy to measure. A standard measure of tool life is the time to develop a flank wear land (VB, see Figure 4.1) length of 300 μm, although this is more related to a life limited by failure than by surface finish or accuracy. However, when chipping generated by adhesion or thermal cracking is the main cause of flank wear, the wear pattern is usually irregular, the more so the more difficult-to-cut is the material. The mean flank wear width does not then determine life. It is recommended that the wear width at both the mid-point of the depth of cut and at its maximum position are checked. The maximum and where it occurs give useful information of the wear mechanism and hence how the wear might be reduced.

Unlike flank wear, crater wear does not influence surface roughness and accuracy. It can, in thermal damage conditions and for some work materials, result in tool failure. It is therefore also a useful measure of wear, able to distinguish between the machinability of different tool and work combinations. A crater depth (KT) from 0.05 mm to 0.1 mm is generally used as an end of life criterion.

Complete failure of a cutting edge is hardly ever used as a tool life test criterion for turning and milling operations because it causes cutting to stop and can seriously damage the work material and even sometimes the tool holder or machine tool. However, it is used

in drilling, particularly with HSS drills, because drill breakage is a common practical failure mode and it is usually possible to recover from it.

4.2.2 Tool life equations

The ability to predict tool life is obviously important for tool management. However, the previous sections' considerations of what is tool life, together with life dependence on the tool (material, geometry, surface integrity and internal defects), the work (hardness, strength, chemical composition, etc), the cutting conditions (speed, feed, depth of cut, cutting fluids), the cutting mode (turning, milling, drilling and others) and the machine tool (for example stiffness, state of maintenance), make a universal life criterion an impossibility.

It is, however, possible to develop tool life equations if the cutting mode and machine tool are limited and the tool wear is simple and dominated by a single mechanism. At the end of the 19th century, F. W. Taylor machined 30 000 tons of work material to collect tool life data to establish the first tool management system. These data, although promoting machining technology (as introduced in Chapter 1), are almost useless today because tool and work materials have changed. However, his tool life equation (Taylor's equation, also introduced in Chapter 1) remains well known and is still widely used.

Taylor's equation – influence of cutting speed on tool life

Over limited ranges of cutting speed V, tool life T is often observed to vary linearly with V on a log-log scale, as shown in Figure 4.13. An example has already been given in Chapter 1 (Figure 1.24). Taylor's tool life equation for such a limited speed range has also been given (Equation (1.3)), but is repeated here for convenience:

$$VT^n = C \tag{4.2}$$

C is the cutting speed that gives a lifetime of one (in whatever units are being used). n is an exponent that shows the sensitivity of life to change of speed. In practice, it usually

Fig. 4.13 Relation between tool life and cutting speed over a wide speed range

takes a value in the range 0.2 to 1.0 although, in Figure 4.13, the dashed line covers a range of speeds for which n is negative.

If a wear mechanism is independent of temperature, it is expected that, after some running-in period, wear increases in proportion to cut distance (Archard's wear law – Chapter 2.4). A tool subject to that wear mechanism would have a life determined by cut distance, independent of cutting speed (unless the amount of running-in varied with speed). It would thus have a lifetime inversely proportional to cutting speed. n would equal 1.0. Simple mechanical wear modes are characterized by n values close to 1. When thermal damage occurs, because temperature increases with cutting speed and thermal damage increases with temperature, n values reduce, the more so the more temperature sensitive is the wear.

The break points shown in Figure 4.13 indicate changes of wear mechanism, such as the increase of thermal wear at higher cutting speeds, or increasing chipping or fracture at lower speeds. It is always necessary to exercise care when relying for tool management on tool life curves based on small amounts of data (and, for economic reasons, they are often in practice based on small amounts of data). It is sensible to support measurements of the amount of wear by observations of wear patterns and mechanisms as well.

Figure 4.14 shows some examples of tool life observations. Figure 4.14(a), for turning

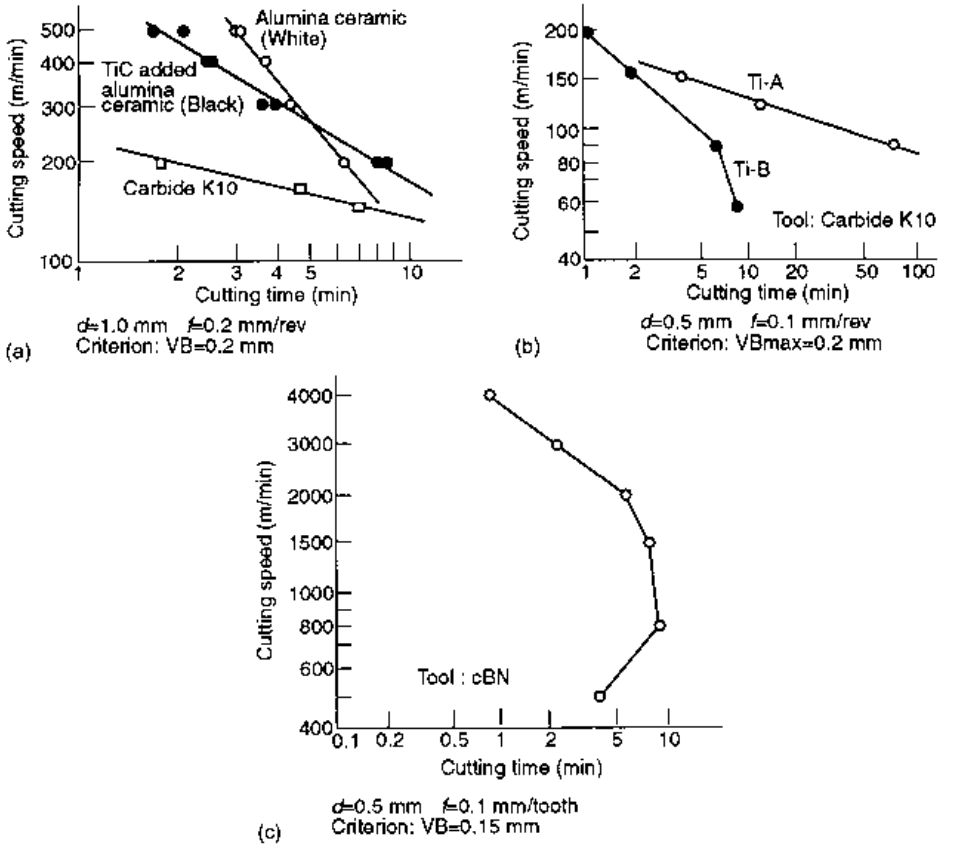


Fig. 4.14 Examples of tool life curves: (a) Turning of ductile cast iron; (b) turning of different types of Ti-alloys; (c) face milling of grey cast iron

a ductile cast iron, shows ceramic tools having a larger n-value than a carbide tool – for the white alumina, $n \approx 1.2$. This very high value indicates a mechanical wear mechanism reducing in intensity with increasing cutting speed. Figure 4.14(b), for turning two different Ti-alloys with a carbide tool, is an example of where a break point falls within the practical cutting speed range. Figure 4.14(c), for face milling a grey cast iron, shows a condition in which tool life decreases with reducing cutting speed.

Taylor's equation – influence of feed and depth of cut

Tool life is influenced by feed and depth of cut, as well as by cutting speed. Additional life equations are

$$f T^{n}_{2} = C_2; \qquad dT^{n}_{3} = C_3 \tag{4.3}$$

and these may be combined with equation (4.2) (replacing n there by n_1) to give

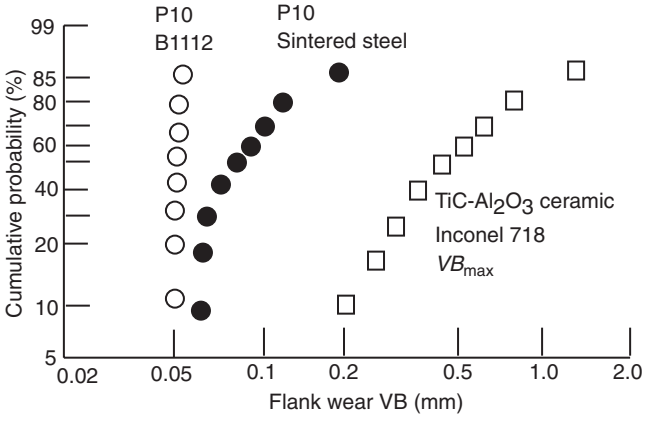
$$V^{1/n_1} f^{1/n_2} d^{1/n_3} \mathrm{T} = C' \tag{4.4}$$

When tool life is limited by thermal damage mechanisms, $n_1 < n_2 < n_3$: i.e. cutting speed has a larger influence on life than does feed than does depth of cut, reflecting the influences of these variables on cutting temperature. If, however, tool life is determined by chipping and fracture failures, n_2 and n_3 can become smaller.

4.2.3 Tool life fluctuations

It is almost impossible to keep cutting conditions exactly constant in practical machining. Even if it were possible, it would be found that tool life and failure are phenomena based on probability. Fluctuations cannot be avoided in these. However, the range of fluctuations is influenced by the damage mechanism. It is easy to imagine larger fluctuations when chipping, or fracture rather than abrasion is the main mechanism.

Figure 4.15 shows the cumulative probability of flank wear development after 1 min of



B112 - P10, V = 200m/min, d = 0.5mm, f = 0.1mm/rev
Sintered steel - P10, V = 200m/min, d = 0.5mm, f = 0.1mm/rev
Inconel 718 - Al_2O_3-TiC ceramic, V = 200m/min, d = 0.5mm, f = 0.19mm/rev

Fig. 4.15 Distributions of flank wear after turning free-cutting steel B1112 and difficult-to-cut sintered steel and Inconel 718

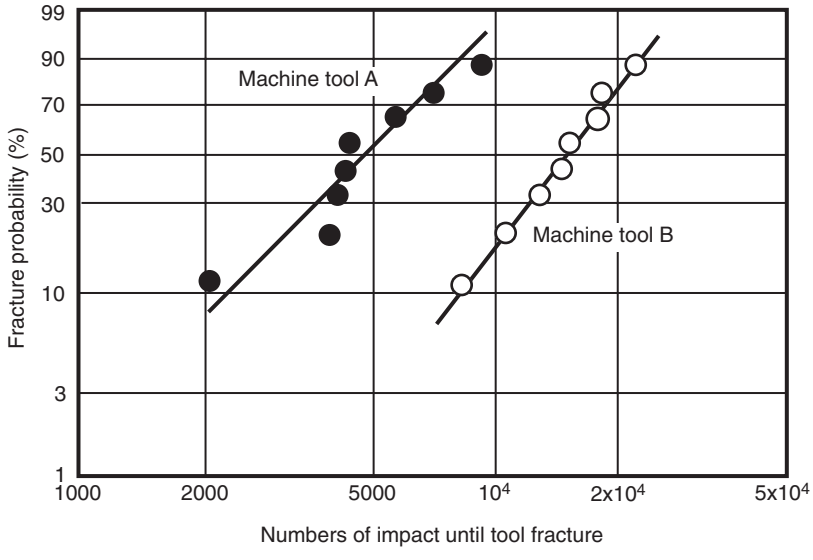


Fig. 4.16 Distributions of tool life limited by fracture when milling a quenched die steel (HRC60) with an Al_2O_3/TiC ceramic tool, on two different milling machines

turning a resulphurized free machining steel and a sintered steel with a P10 carbide tool (plotted on a Weibull chart). Abrasion was the main cause of tool wear with the free machining steel, while edge chipping was the mechanism with the sintered steel. The different slopes of the Weibull plots are clear. The figure also shows the distribution for turning Inconel 718 with an Al_2O_3/TiC ceramic tool. As well as the greater amount of wear, the similarity of slope between this and the sintered steel observations is striking.

Figure 4.16 is an example of tool wear and wear distribution influenced by the machine tool. It gives the results of face milling a quenched die steel with an Al_2O_3/TiC ceramic tool, in the same conditions apart from the machine tool used. Tool wear was by edge chipping or fracture. Machine B obviously provides better resistance against this type of damage. This is due to a better stiffness, maybe a better dynamic stiffness.

4.3 Summary

This chapter complements Chapter 3 on tool bulk properties, by focusing on the mechanisms of cutting edge damage and their characteristic developments with time. Cutting edges experience much higher normal and shear stresses than almost any other type of bearing surface and, at high cutting speeds, high temperatures are also generated. It is not surprising that tool lives are measured in minutes rather than in hours, and certainly not in days.

Abrasion occurs with all tools if the work material has hard enough phases, and self-abrasion follows from other mechanical causes of damage. Mechanical damages, of

increasing size – from attrition, to chipping, to fracture – increase the more brittle is the tool material and they are relatively insensitive to temperature changes. Thermal damages follow diffusion and chemical reactions. They are very sensitive to temperature and are particularly variable from one tool and work combination to another. Adhesive wear depends on both mechanical and thermal factors, and passes through a maximum rate as temperature increases.

For all these reasons of complexity and further influences of mode of cutting, and of the machine tools themselves, on tool life, it has not been attempted to provide comprehensive guidance on tool damage rates. Rather, the goal has been to emphasize what phenomena can occur, and what their effects look like, so mechanisms limiting life in different circumstances may be recognized and sensible directions for improved performance may then be investigated.

References

Cook, N. H. (1973) Tool wear and tool life. *Trans. ASME J. Eng. Ind.* **95B**, 931–938.

Dawihl, W. (1941) Die Vorgange beim Verschleiss von Hartmetallegierungen. *Stahl und Essen* **61**, 210–213.

Gregory, B. (1965) Surface interaction of cemented carbide tool material and Armco iron. *Brit. J. Appl. Phys.* **16**, 689–695.

Kitagawa, T., Maekawa, K., Shirakashi, T. and Usui, E. (1988) Analytical prediction of flank wear of carbide tools in turning plain carbon steels (Part 1). *Bull. Jap. Soc. Prec. Eng.* **22**(4), 263–269.

Naerheim, Y. and Trent, E. M. (1977) Diffusion wear of cemented carbide tools when cutting steel at high speeds. *Metals Technology* **4**, 548–556.

Narutaki, N. and Yamane, Y. (1976) Wear mechanism of carbide tool based on the reaction between tool and work material (Part 1 – reaction test). *Bull. Jap. Soc. Prec. Eng.* **10**(3), 95–100.

Narutaki, N. and Yamane, Y. (1993) High-speed machining of Inconel 718 with ceramic tools. *Annals CIRP* **42**(1), 103–106.

Takeyama, H. and Murata, R. (1963) Basic investigation of tool wear. *Trans ASME J. Eng. Ind.* **85**, 33–38.

Trent, E. M. (1952) Some factors affecting wear on cemented carbide tools. *Proc. I. Mech. E. Lond.* **166**, 64–74.

Trigger, K. J. and Chao, B. T. (1956) The mechanism of crater wear of cemented carbide tools. *Trans ASME* **78**,1119–1126.

Uehara, K. (1976) On the generating mechanism of wear particles on carbide cutting tools. *J. Japan Soc. Prec. Eng.* **42**(6), 445–452.

Usui, E., Shirakashi, T. and Kitagawa, T. (1978) Analytical prediction of three dimensional cutting process Pt. 3. *Trans ASME J. Eng. Ind.* **100**, 236–243.

Yamane, Y. and Narutaki, N. (1983) The effect of atmosphere on tool failure in face milling (1st report). *J. Jap. Soc. Prec. Eng.* **49**(8), 521–527.

5

Experimental methods

Previous chapters have presented optical and electron microscope pictures of chip sections and worn tools, and the results of cutting force and temperature measurements. In addition to cutting force measurements, acoustic emission is also used to study the health of a cutting process. This chapter explains a number of these experimental methods.

5.1 Microscopic examination methods

5.1.1 The quick-stop technique

Direct observations as well as theoretical analyses are needed to clarify chip formation mechanisms. Ideally, such observations would be during cutting, to follow dynamic

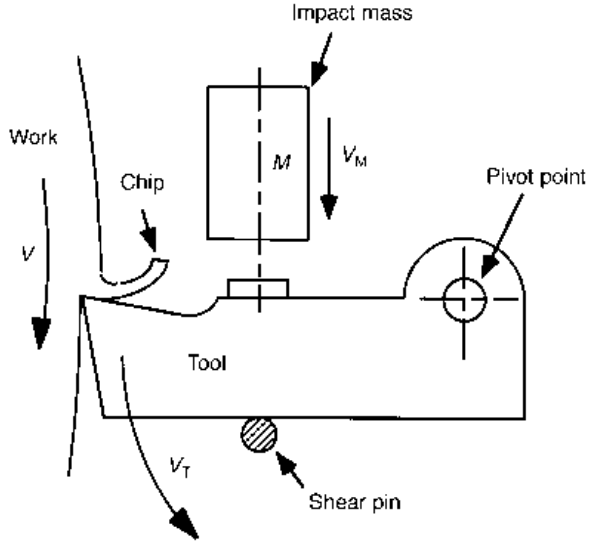


Fig. 5.1 The principle of a quick-stop device for use in turning

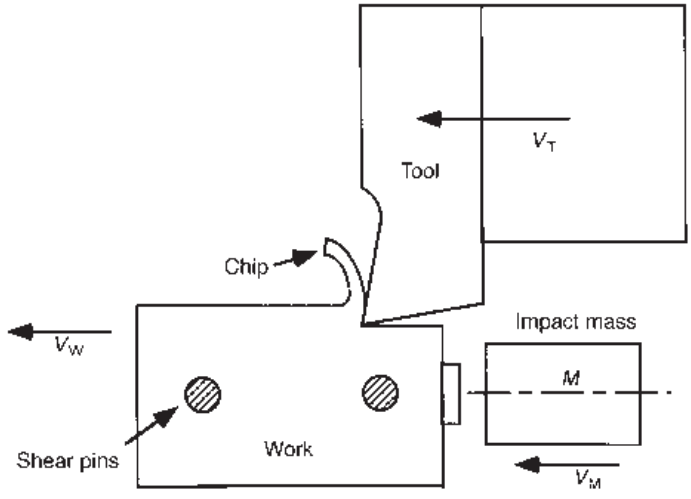


Fig. 5.2 The principle of a quick-stop device for use in milling

variations of chip flow. Although video cameras have been used to gain an external overview of dynamic chip motions, and it is possible to look through transparent tools (for example made from diamond or sapphire) directly at the chip contact, it is difficult, in general, to resolve much because of the small scale of the deforming region and usually the high cutting velocities. Experimentalists are prepared to lose dynamic information to gain microscopic detail, by freezing the motion, for later study. The quick-stop technique is a popular method for achieving this. The machining process is stopped quickly, by moving the tool and work material apart at a speed greater – preferably much greater – than the cutting speed. The chip is left attached to the work (sometimes with a fragment of the cutting edge attached as well). The photographs in Figure 2.4 are polished and etched sections of quick-stopped chips.

Figure 5.1 is a schematic view of a quick-stop device for use with a stationary tool and a moving workpiece, such as in turning, while Figure 5.2 shows a device that could be used for a stationary work and moving tool, as in milling. In Figure 5.1, the tool is supported at a pivot point and by a shear pin. A mass M is made to strike the tool holder with a speed V_M. If the impact force is much greater than needed to break the shear pin, the mass will then cause the tool holder to swing quickly away from the chip. The tool holder's velocity V_T does not instantaneously reach the cutting velocity V that is necessary for cutting to stop, because of its inertia. However, to minimize the retraction time, M and V_M should be made large and the inertia of the holder should be made small.

In practice, V_M is frequently made large by firing the mass M from a gun (although for low cutting speed turning tests, hitting the tool holder with a hammer can be sufficient). A device that uses a humane killer gun (normally used for stunning animals prior to slaughter) with its captive bolt as the mass M was reported to achieve a tool displacement of 2.5 mm in 1.2×10^{-4} s (Williams *et al.*, 1970). If this is assumed to have occurred at approximately constant acceleration, and it is supposed that, for a successful quick-stop, V_T must reach V in a cut distance less than $f/10$, then this device can be used successfully, provided

$$V[\text{m/min}] \leq 354\sqrt{f[\text{mm}]} \tag{5.1}$$

For a feed of 0.13 mm, the largest allowable cutting speed is 128 m/min, while for f = 0.5 mm, the largest speed is 250 m/min. These speeds are larger than those represented in Figure 2.4, but are not large compared with what can be of interest in modern high speed machining. The acceleration required of the tool increases as the square of the cutting velocity, so successful quick-stops become rapidly more difficult as the cutting speed increases.

A similar discussion could be developed in terms of the device of Figure 5.2. However, in milling, it is more difficult to guide the work material away from the cutting edges, and the work and its holder have higher inertia than the tool and its holder in turning. The quick-stop must be synchronized with the intermittent cutting action. There must be a very special reason to pursue a quick-stop in milling, to make the difficulties worthwhile.

Quick-stops can show different results, depending on the adhesion between the chip and the tool (Figure 5.3). If there is low adhesion, a clean separation between the two will occur, as shown in Figure 5.3(a). Coated tools usually show this behaviour. If there is high adhesion relative to the strength of the chip or tool, any of the results of Figure 5.3(b) to (d) can occur. If it is particularly desired to preserve the chip/tool interface, a result like Figure 5.3(d) can be engineered by artificially weakening the tool with a notch or crack on its rake face.

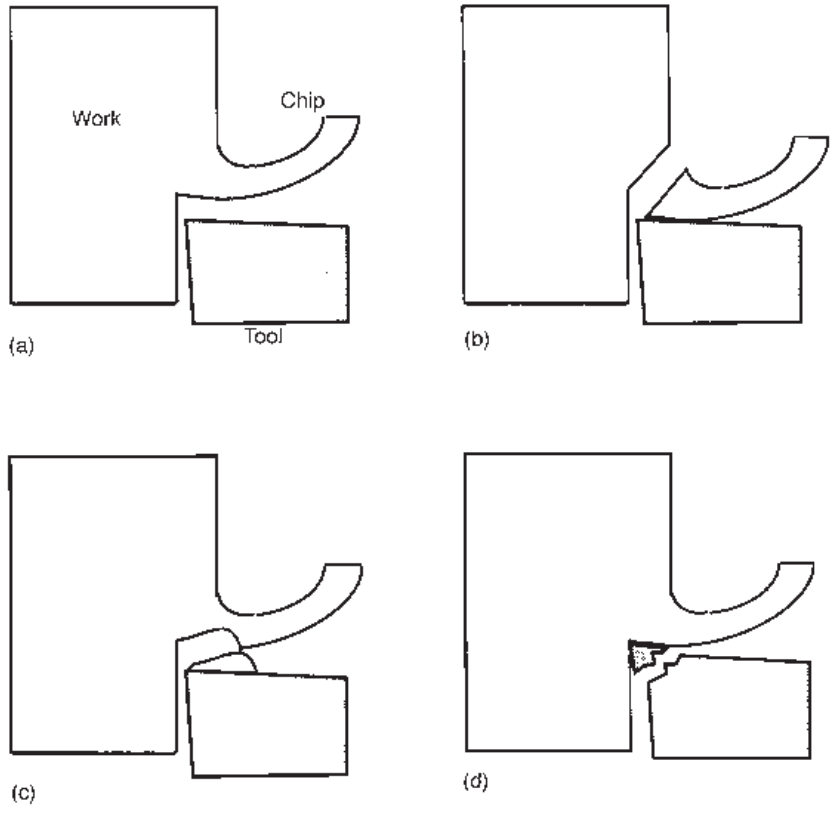
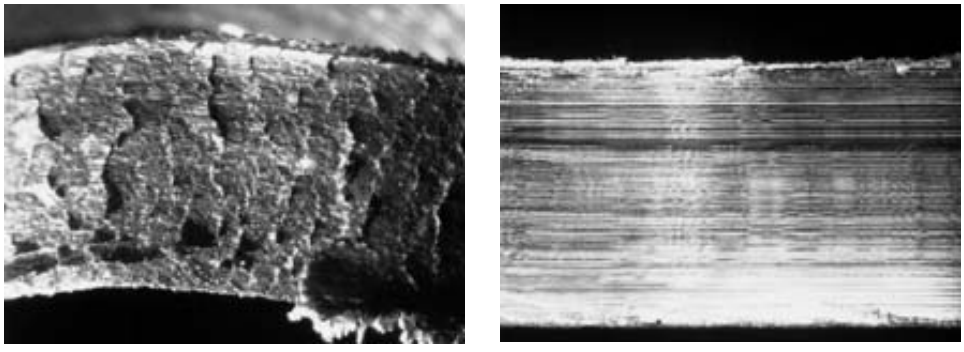


Fig. 5.3 Modes of quick-stop separation



(a) (b)

Fig. 5.4 The back surface of chips formed from 0.15% C steel by P20 carbide tools: (a) with built-up edge, v = 40 m min^{-1}, d = 2.0 mm, f = 0.08 mm rev^{-1}; (b) without built-up edge, v = 100 m min^{-1}, d = 2.0 mm, f = 0.12 mm rev^{-1}.

5.1.2 Other chip form and wear observations

Careful observation of tools and chips after machining can often reveal useful information, without the need for quick-stops. For example, the built-up edge (BUE) formed in machining is usually unstable. It is carried away on the back surface of chips, so observation of the chips (Figure 5.4) can give information as to whether BUE is formed or not. It is obvious that information about wear is obtained by looking at the cutting tools at any time after cutting.

Chapter 4 has shown examples of SEM and EPMA used to study wear and contact conditions in great detail. The magnifications of these techniques are not always necessary. In many cases, a low magnification optical microscope, × 10 or × 20, is enough. Such a microscope on an *X–Y* measurement stage is commonly used in laboratories or machine shops to record wear images and their sizes. Wet photography and printing paper used to be used for archiving information for many years. Now, a high quality CCD camera and a personal computer with a large memory can do the job.

5.2 Forces in machining

5.2.1 Resultant forces

Forces in machining can be measured in two main ways: directly or indirectly. Direct measurements involve mounting a tool (in turning) or the tool or workpiece (in milling) on a dynamometer, which responds to the forces by creating electrical signals in proportion to them. These measurements are used when the forces need to be known accurately both in magnitude and direction, for example if thrust, feed and the main cutting forces in turning are required (Figure 5.5), or the torque and thrust force in drilling are needed.

Indirect measurements involve deductions from the machine tool behaviour. For example, the power used by the main spindle motor increases with the main cutting force or torque; and that used by the feed motions can be related to the feed force. Particularly with

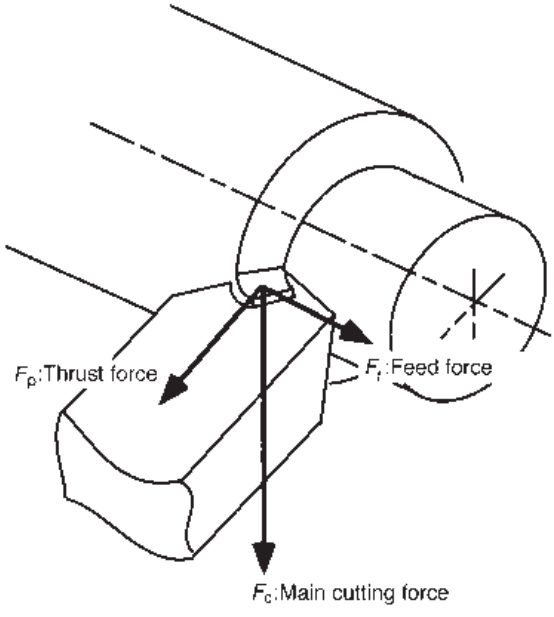


Fig. 5.5 Forces acting on cutting tool in turning

NC machines, which are fitted with high sensitivity and response main and feed drive motors, indirect methods can be used to determine the active forces. Indirect methods are less accurate than direct methods, but can be sufficient for monitoring purposes. The main consideration here will be direct methods.

Tool dynamometers – general points

A tool dynamometer should have high sensitivity, high rigidity, high frequency response, high linearity and low drift. Sensitivity is expressed as electrical output per unit force input. Useful dynamometers must be able to discriminate at least 1% of full scale output. Rigidity depends strongly on the dynamometer's construction. The force sensing transducer is usually the least rigid element of a dynamometer's structure: different types of element are considered in the following subsections.

Frequency response depends on a dynamometer's natural frequency and damping characteristics. In line with elementary dynamics, these may be described in terms of the response of a viscously damped elastic system subjected to a harmonic forcing system:

$$m\ddot{x} + c\dot{x} + kx = P_{\mathrm{m}} \sin \omega t \tag{5.2}$$

Figure 5.6 shows how the amplitude ratio (the response relative to the response in static conditions) of such a system varies with frequency ratio (the frequency relative to the system's undamped natural frequency of $\sqrt{k/m}$) and damping factor c/c_{c}, where c_{c} is the critical damping coefficient. The figure shows that for a linear response between amplitude and force (and hence a linear response between a dynamometer's output and force), a damping factor slightly less than 1, around 0.7, is desirable and then a dynamometer could be used at frequency ratios up to 0.2 to 0.3.

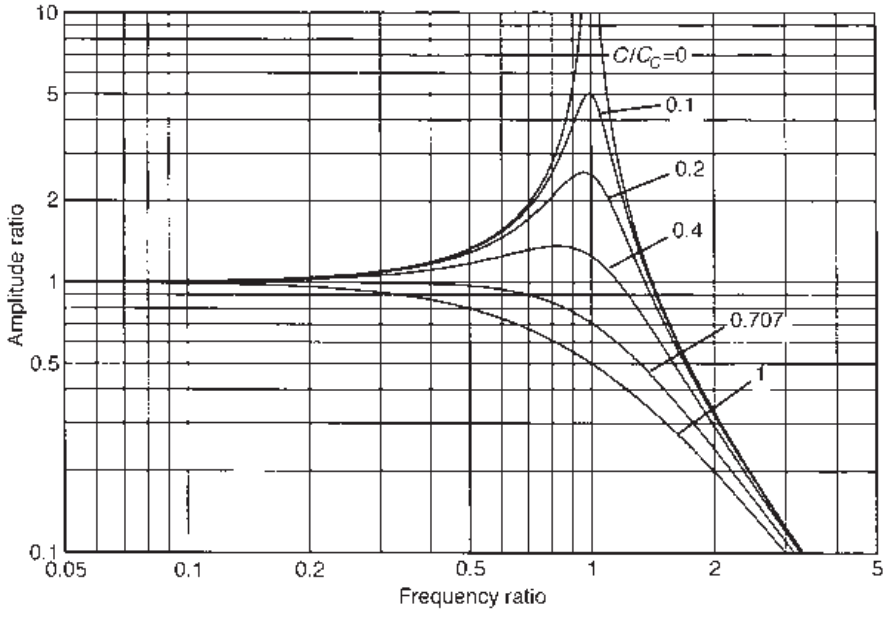


Fig. 5.6 The frequency response of a damped forced vibration system

Linearity and drift are usually more influenced by the electrical elements (including signal amplification) than by the mechanical elements of a dynamometer. Systems with linearity better than 0.5% of full scale output are required. Drift, which describes the stability of output (both from the dynamometer transducer and amplification system) over time, can be a problem with cutting force dynamometers because of the sensitivity of electrical elements to temperature changes and the tendency of machining to heat its surroundings.

Strain gauge dynamometers

A common type of dynamometer uses strain gauges to sense elastic strains caused by cutting forces. Figure 5.7 shows a basic elastic beam type dynamometer with gauges bonded to its surface. It also shows an example of a wire-type gauge and a Wheatstone bridge and amplifier system usually used to measure strain changes in the gauges. The main cutting force F_C will cause the beam to bend, so that the gauge on the top surface will be placed in tension, that on the bottom surface will be placed in compression, and those gauges on the side surfaces (at the neutral axis) will experience no strain. Likewise, a feed force will strain the side-face gauges but not those at the top or bottom. The arrangement shown in Figure 5.7 is not sensitive to force along the axis of the beam as this causes equal strain changes in all gauges.

The fractional resistance change of a strain gauge ($\Delta R/R$) is related to its fractional length change or direct strain ($\Delta L/L$) by its gauge factor K_s:

$$K_s = (\Delta R/R)/(\Delta L/L) \tag{5.3}$$

For wire strain gauges, K_s is typically from 1.75 to 3.5. Strains down to 10^{-6} may be detected with a bridge circuit. The upper limit of strain is around 2×10^{-3}, determined by the elastic limit of the beam.

A disadvantage of the simple cantilever dynamometer is that the gauges' strains depend

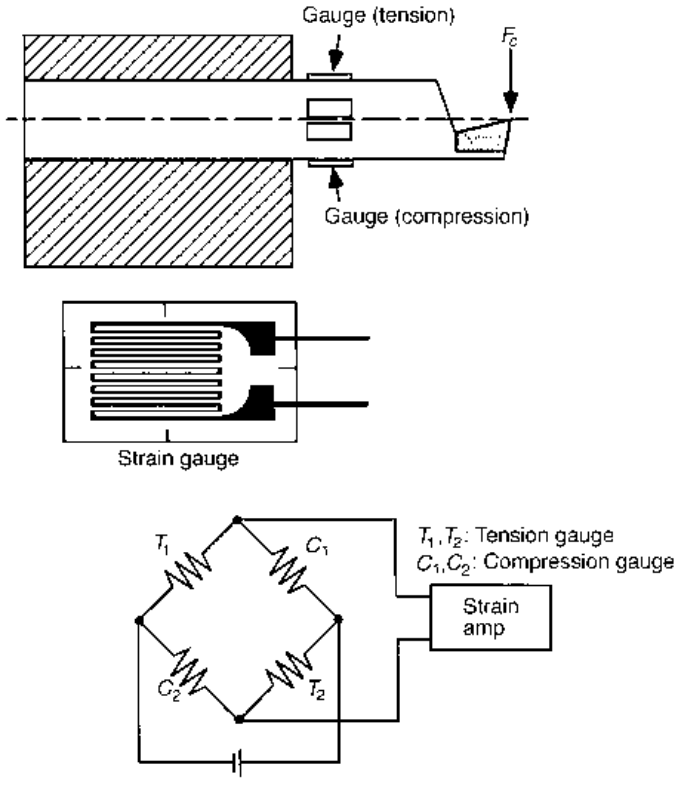


Fig. 5.7 A strain gauged cantilever dynamometer with its bridge circuit

basically on the moment applied to the section at which they are positioned. They therefore depend on the gauges' distance from where the load is applied, as well as on the size of the load. Better designs, less sensitive to where the load is applied, are the octagonal ring and parallel beam designs shown in Figure 5.8. Supporting the load on well-separated thin sections results in the sum of the strains in the gauges being unchanged when the point of application of the load is changed, even though the strains are redistributed between the sections. It is possible to connect the strain gauges in a bridge circuit so that the output is not sensitive to where the force is applied.

The choice of parallel beams or octagonal rings is a matter of manufacturing choice. For both, it is important, as a matter of convenience, to minimize cross-sensitivity between the different orthogonal components of electrical output and mechanical input. For the parallel beam design, this is achieved by manufacturing the two sets of beams perpendicular to each other. For the octagonal ring design, it is important to choose a particular shape of octagon. When a circular ring (Figure 5.9) is loaded radially there is zero strain at the positions B and B′, ± 39.9° from the point of application of the radial load; likewise when the ring is loaded tangentially, there is zero strain at A and A′, ± 90° from the load. Gauges placed at A and A′ will respond only to radial loads; and at B and B′ only to tangential loads. The strains will depend on the loads and the ring dimensions (radius R, thickness t and width b) and Young's modulus E as

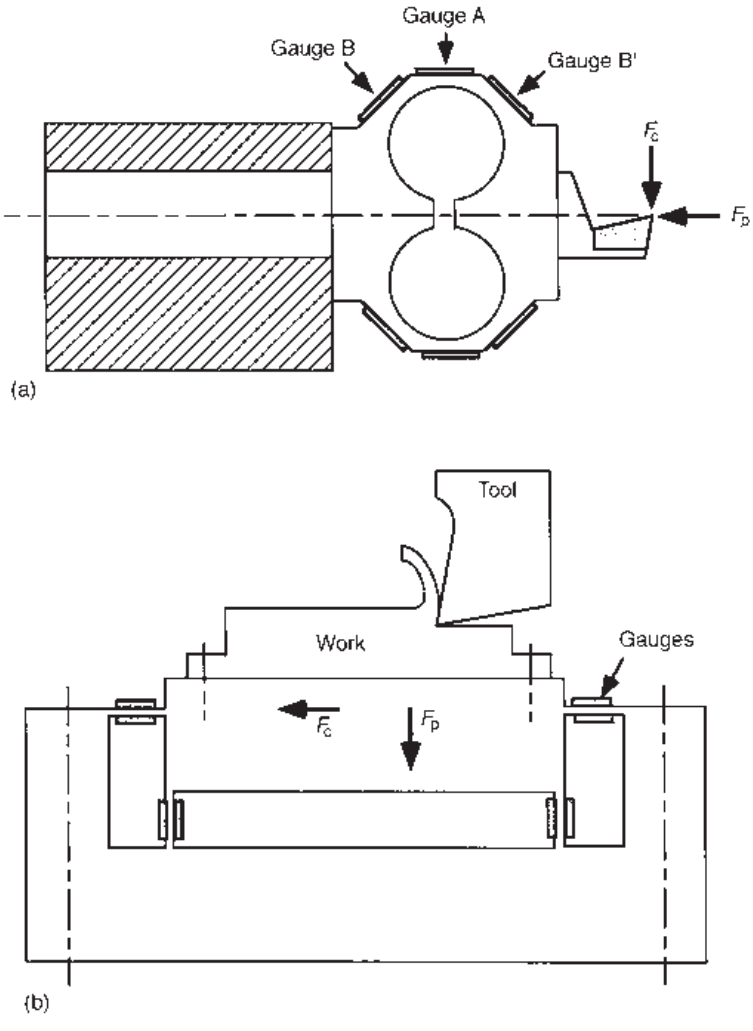


Fig. 5.8 Octagonal ring and parallel beam dynamometer designs: (a) Octagonal ring type tool dynanometer; (b) parallel beam type tool dynanometer

$$\left.\begin{aligned} \varepsilon_{A,A'} &= \pm \frac{1.09F_P}{Ebt^2} \\ \varepsilon_{B,B'} &= \pm \frac{2.18F_c}{Ebt^2} \end{aligned}\right\} \tag{5.4}$$

The manufacture of the ring outer surface as an octagon rather than a cylinder is just a practical matter.

The need to generate detectable strain imposes a maximum allowable stiffness on a dynamometer. This, in turn, with the mass of the dynamometer depending on its size or on the mass supported on it, imposes a maximum natural frequency. Simple beam

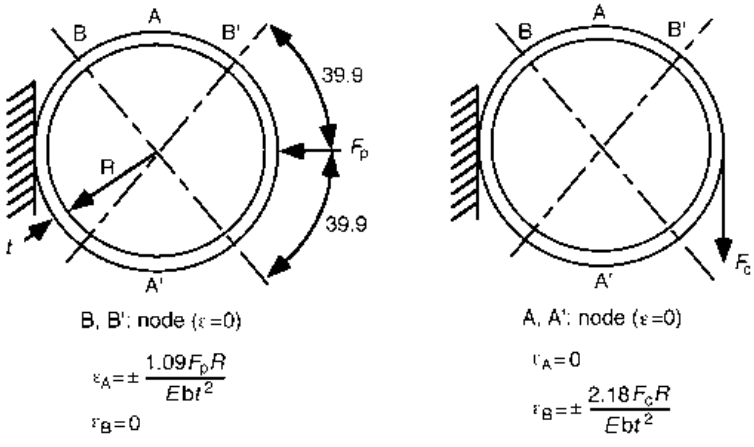


Fig. 5.9 The loading of a ring by radial and tangential forces

dynamometers, suitable for measuring forces in turning from 10 N to 10 kN, can be designed with natural frequencies of a few kHz. The ring and the strut types of dynamometer tend to have lower values, of several hundred Hz (Shaw, 1984, Chapter 7). These frequencies can be increased tenfold if semiconductor strain gauges (K_s from 100 to 200) are used instead of wire gauges. However, semiconductor gauges have much larger drift problems than wire gauges. They are used only in very special cases (an example will be given in Section 5.2.2). An alternative is to use piezoelectric force sensors.

Piezoelectric dynamometers

For certain materials, such as single crystals of quartz, Rochelle salt and barium titanate, a separation of charge takes place when they are subjected to mechanical force. This is the piezoelectric effect. Figure 5.10 shows the principle of how it is used to create a three-axis force dynamometer. Each force component is detected by a separate crystal oriented relative to the force in its piezoelectric sensitive direction. Quartz is usually chosen as the piezoelectric material because of its good dynamic (low loss) mechanical properties. Its piezoelectric constant is only $\approx 2 \times 10^{-12}$ coulombs per Newton. A charge amplifier is therefore necessary to create a useful output. Because the electrical impedance of quartz is high, the amplifier must itself have high input impedance: 10^5 MΩ is not unusual.

Figure 5.11 shows the piezoelectric equivalent of the dynamometers of Figure 5.8. The stiffness is basically that of the crystals themselves. Commercial machining dynamometers are available with natural frequencies from 2 kHz to 5 kHz, depending on size.

5.2.2 Rake face stress distributions

In addition to overall force measurements, the stresses acting on cutting tools are important, as has been indicated in earlier chapters. Too large stresses cause tool failure, and friction stresses strongly influence chip formation. The possibility of using photoelastic studies as well as split-tool methods to determine tool stresses has already been introduced in Chapter 2 (Section 2.4). The main method for measuring the chip/tool contact stresses

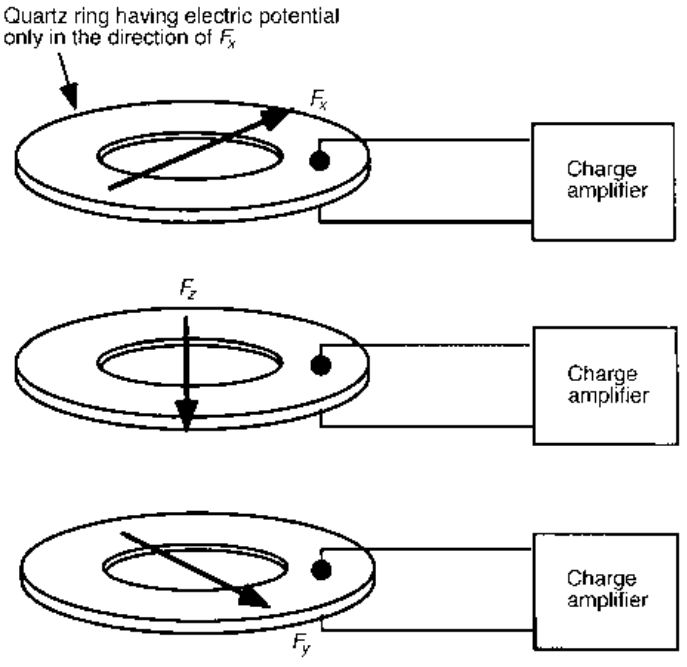


Fig. 5.10 The principle of piezoelectric dynamometry

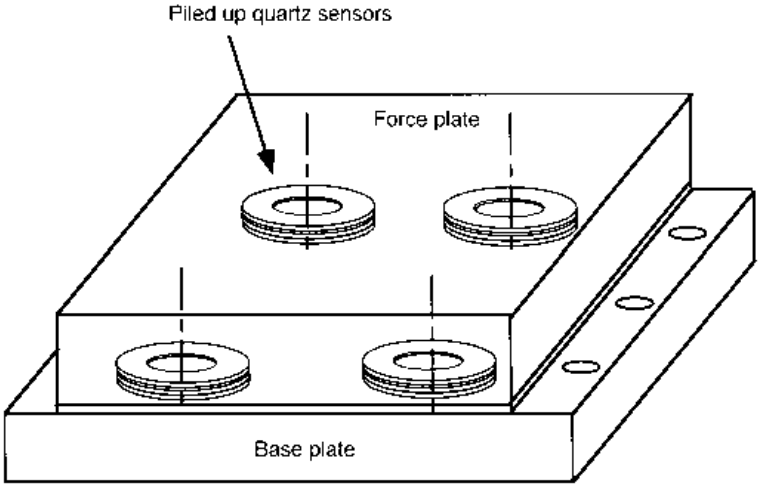


Fig. 5.11 A piezoelectric tool dynamometer

is the split-tool method (Figure 2.21), although even this is limited – by tool failure – to studying not-too-hard work materials cut by not-too-brittle tools.

Figure 5.12 shows a practical arrangement of a strain-gauged split-tool dynamometer. The part B of the tool (tool 1 in Figure 2.21) has its contact length varied by grinding away its rake face. It is necessary to measure the forces on both parts B and A, to check that the

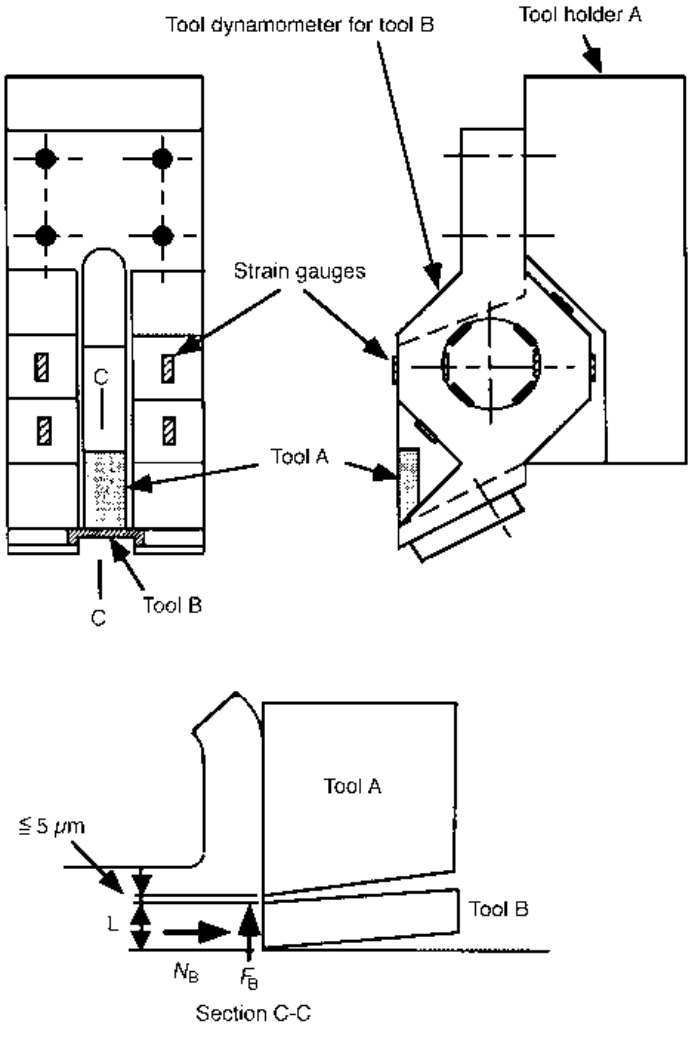


Fig. 5.12 A split-tool dynamometer arrangement

sum of the forces is no different from machining with an unsplit tool. It is found that if extrusion into the gap between the two tool elements (g, in Figure 2.21) is to be prevented, with the surfaces of tools A and B (1 and 2 in Figure 2.21) at the same level, the gap should be less than 5 μm wide (although other designs have used values up to 20 μm and a downward step from 'tool 1' to 'tool 2'). The greatest dynamometer stiffness is required. This is an instance when semiconductor strain gauges are used. Piezoelectric designs also exist.

Split-tool dynamometry is one of the most difficult machining experiments to attempt and should not be entered into lightly. The limitation of the method – tool failure, which prevents measurements in many practical conditions that could be used to verify finite element predicted contact stresses and also to measure friction stresses directly – leaves a major gap in machining experimental methods.

5.3 Temperatures in machining

There are two goals of temperature measurement in machining. The more ambitious is quantitatively to measure the temperature distribution throughout the cutting region. However, it is very difficult, because of the high temperature, commonly over 700˚C even for cutting a plain carbon steel at cutting speeds of 100 m/min, and the small volume over which the temperature is high. The less ambitious goal is to measure the average temperature at the chip/tool contact. Thermocouple methods can be used for both (the next section concentrates on these); but thermal radiation detection methods can also be used (Section 5.3.2 summarizes these). (It is possible in special cases to deduce temperature fields from the microstructural changes they cause in tools – see Trent, 1991 – but this will not be covered here.)

5.3.1 Thermocouple methods

Figure 5.13 shows an elementary thermocouple circuit. Two materials A and B are connected at two junctions at different temperatures T_1 and T_2. The electro-motive force (EMF) generated in the circuit depends on A and B and the difference in the temperatures T_1 and T_2. A third material, C, inserted at one of the junctions in such a way that there is no temperature difference across it, does not alter the EMF (this is the law of intermediate metals).

In common thermocouple instrument applications, A and B are standard materials, with a well characterized EMF dependence on temperature difference. One junction, usually the colder one, is held at a known temperature and the other is placed in a region where the

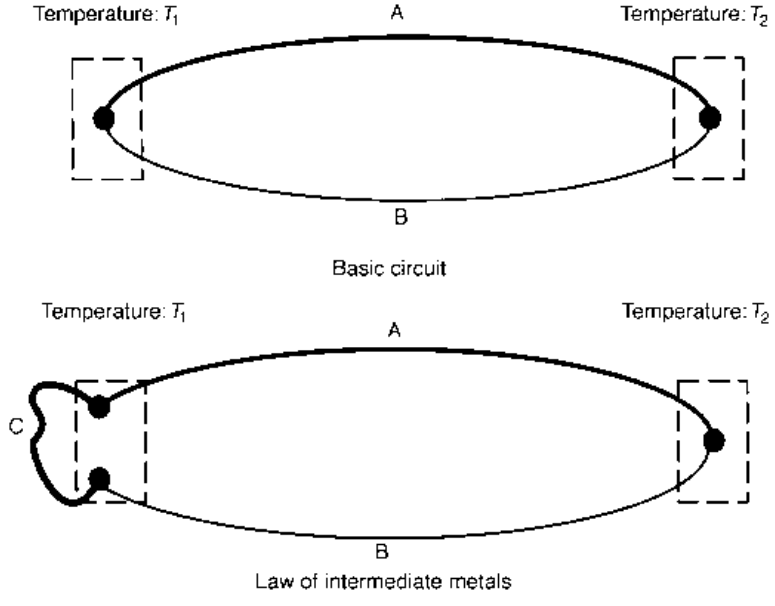


Fig. 5.13 An elementary thermocouple circuit (above) with an intermediate metal variant (below)

temperature is to be deduced from measurement of the EMF generated. Standard material combinations are copper-constantan (60%Cu–40%Ni), chromel (10%Cr–90%Ni)–alumel (2%Al–90%Ni-Si-Mn) and platinum–rhodium. In metal machining applications, it is possible to embed such a standard thermocouple combination in a tool but it is difficult to make it small enough not to disturb the temperature distribution to be measured. One alternative is to embed a single standard material, such as a wire, in the tool, to make a junction with the tool material or with the chip material at the tool/chip interface. By moving the junction from place to place, a view of the temperature distribution can be built up. Another alternative is to use the tool and work materials as A and B, with their junction at the chip/tool interface. By this means, the average contact temperature can be deduced. This application is considered first, with its difficulties stemming from the presence of intermediate metals across which there may be some temperature drop.

Tool–work thermocouple measurements

Figure 5.14 shows a tool–work thermocouple circuit for the turning process. The hot junction is the chip/tool interface. To make a complete circuit, also including an EMF recorder, requires wires to be connected between the recorder and the tool and the recorder and the work. In the latter case, because the work is rotating, the wire must pass through some slip-ring device. If the junctions A, B and C, between the work and slip ring, the slip ring and recorder wire and the tool and the recorder wire, are all at the same (cold junction) temperature, the circuit from A to C is all intermediate and has no effect on the EMF. But this is often not the case.

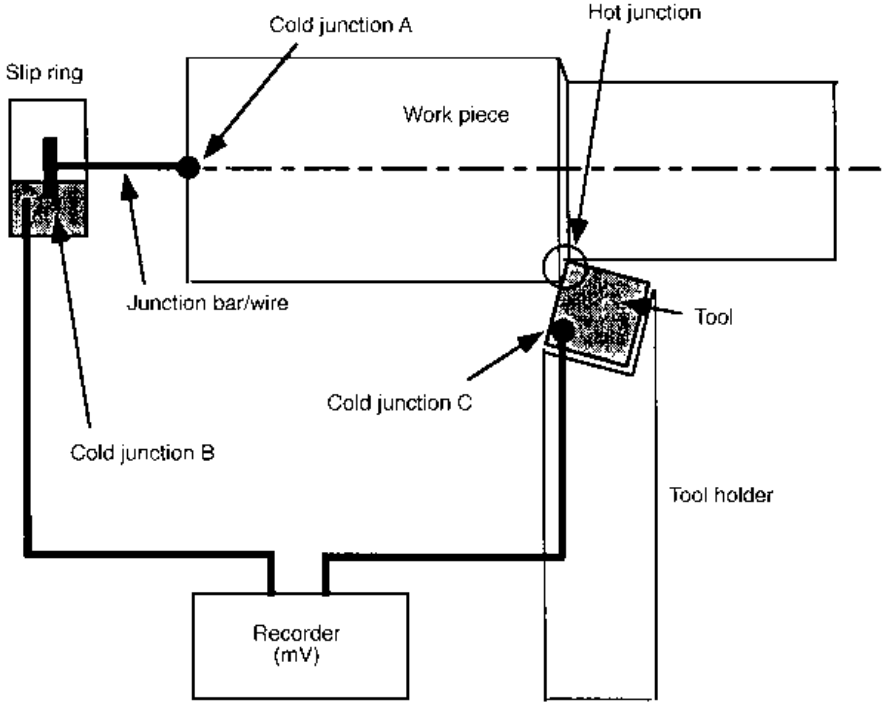


Fig. 5.14 A tool–work thermocouple circuit

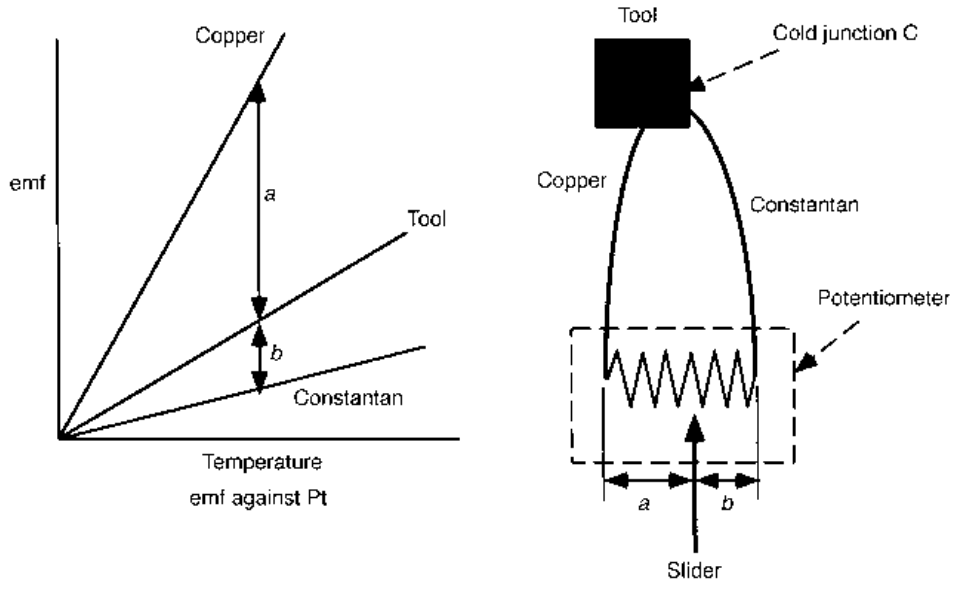


Fig. 5.15 A circuit for compensating the cold junction C

Dry slip rings, with their rubbing interface, frequently create an EMF. The solution is to use a liquid mercury contact. If an indexable insert is used as the cutting edge, the distance from the hot junction to the cold junction C may be only 10 mm. In this case, to eliminate error due to C heating up, either the measurement time must be kept very short, or the insert must be extended in some way – for example by making the connection at C from the same material as the insert (but this is often not practical) – or the heating must be compensated. Figure 5.15 shows a cold junction compensation circuit and its principle. The single wire connection at C is replaced by a standard thermocouple pair of wires which are terminated across a potentiometer in a region where the temperature is not affected by the cutting. The connection to the EMF recorder is then taken from the potentiometer slider. The thermocouple wire materials are chosen so that the tool material has an intermediate EMF potential between them, relative to some third material, for example platinum. The slider is set at the point of interior division of the potentiometer, at the same ratio *a*/*b* as the tool material potential is intermediate between the two thermocouple materials. Copper and constantan are found suitable to span the potentials of most tool materials.

Tool–work thermocouple calibration

The EMF measured in cutting must be converted to temperature. Generally, the EMF–temperature relation for tool–work thermocouples is non-linear. It can even vary between nominally the same tool and work materials. It is essential to calibrate the tool–work thermocouple using the same materials as in the cutting test. Figure 5.16 shows one calibration arrangement and Figure 5.17 shows its associated measurement circuit. In this arrangement, the tool–work thermocouple EMF is not measured directly. Instead, the EMF between the tool and a chromel wire is measured at the same time as that of a

chromel–alumel thermocouple at the same temperature. Thus, the tool–chromel EMF versus temperature characteristic is calibrated against the chromel–alumel standard. This is repeated for the work–chromel combination. The tool–work EMF versus temperature relation is the difference between the tool–chromel and work–chromel relations.

Figure 5.16 shows an overview of the tool or work in contact with the chromel–alumel thermocouple (detail in Figure 5.17). The contact is made at one focus of an infrared heating furnace with reflecting walls, shaped as an ellipsoid of revolution, with a 1 kW halogen lamp at the other focus. The chromel–alumel thermocouple is fixed to the furnace body and the tool or work is pressed on to it by a spring. It is necessary to prepare the tool and work materials as rods in this method, but it is possible to heat the hot junction to 1000˚C in about 10 s: the lengths of the rods, to avoid the need for cold junction compensation circuitry, need only be sufficient to be beyond the heat diffusion distance over this time. Example results, for a P10 carbide tool and a 0.45% plain carbon steel work, are given in Figure 5.18. Even at 1000˚C the EMF is only 10 mV, so a high sensitivity recorder is needed.

Inserted thermocouple measurements

Figure 5.19 shows two further possibilities of tool temperature measurement. In Figure 5.19(a), a small diameter hole has been bored in the tool and a fine standard thermocouple

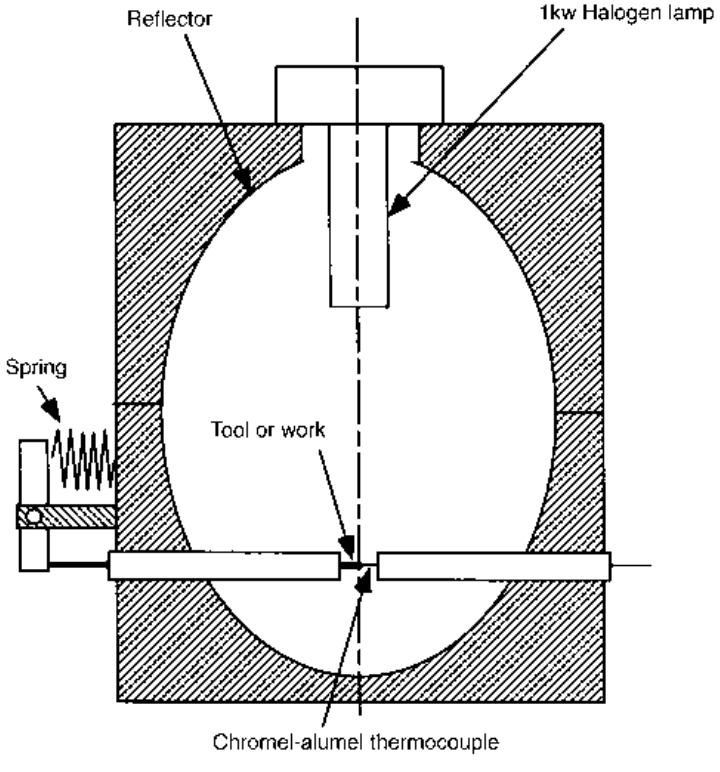


Fig. 5.16 A tool–work thermocouple calibration set-up

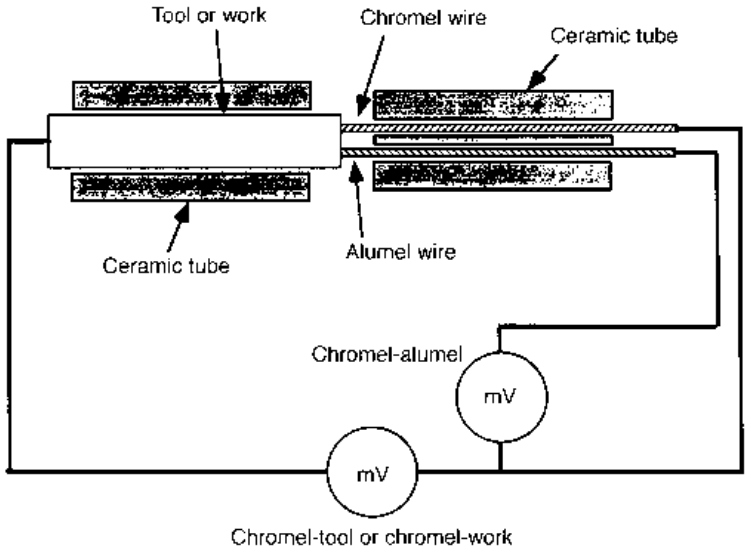


Fig. 5.17 A detail of the hot junction and the associated measurement circuit

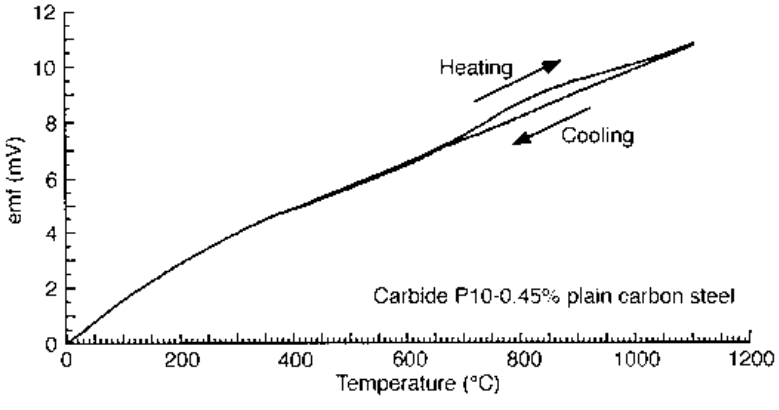


Fig. 5.18 Calibration test results for P10 carbide and a 0.45% plain carbon steel

has been inserted in it. It has the advantage that a precise measurement of temperature at the bottom of the hole can be made, relying on the standard thermocouple, but a disadvantage that the hole may disturb the temperature gradients in the tool. If it is desired to measure the temperature distribution in the tool, while only boring one hole, the rake and clearance faces of the tool may be progressively ground away, to vary the position of the hole relative to the cutting edge.

A finer hole may be bored if only one wire is to be inserted in it. Figure 5.19(b) shows a single wire, for example chromel, or in this case platinum, making contact with the work at the chip–tool interface. In this way, the temperature at a specified point can be measured,

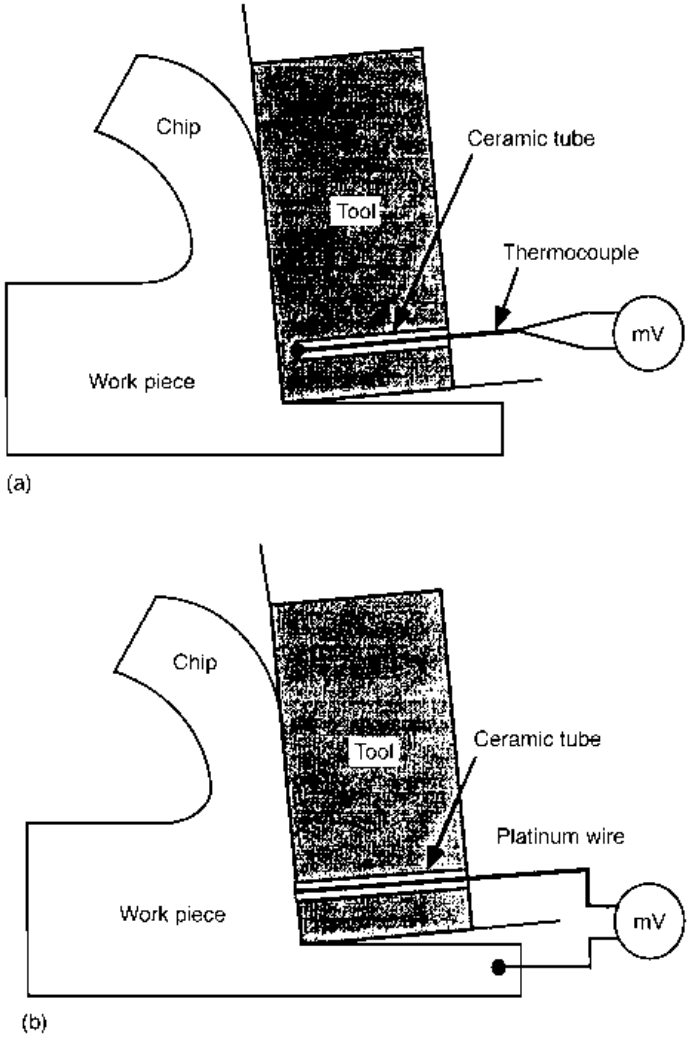


Fig.5.19 (a) Inserted thermocouple or (b) thermocouple wire

but it is necessary to calibrate the thermocouple, as was done with the tool–work thermocouple.

5.3.2 Radiation methods

Inserted thermocouple methods require special modifications to the cutting tools. The tool–work thermocouple method only determines average contact temperatures; and cannot be used if the tool is an insulator. Thermal imaging methods, measuring the radiation from a surface, have a number of attractions, if surface temperatures are of interest.

The laws of electromagnetic energy radiation from a black body are well known. The power radiated per unit area per unit wavelength W_λ depends on the absolute temperature T and wavelength λ according to Planck's law:

$$W_\lambda = \frac{2\pi hc^2}{\lambda^5} \frac{1}{\left(e^{\frac{ch}{\lambda kT}} - 1 \right)} \tag{5.5}$$

where h (Planck's constant) = 6.626×10^{-34} J s, c (speed of light) = 2.998×10^8 m s^{-1} and k (Boltmann's constant) = 1.380×10^{-23} J K^{-1}.

Equation (5.5) can be differentiated to find at what wavelength λ_{max} the peak power is radiated (or absorbed), or integrated to find the total power W. Wien's displacement law and the Stefan–Boltzmann law result:

$$\begin{aligned} \lambda_{max} T &= 2897.8\ \mu\text{m K} \\ W[\text{W m}^{-2}] &= 5.67 \times 10^{-8} T^4 \end{aligned} \tag{5.6}$$

Figure 5.20 shows the characteristic radiation in accordance with these laws. Temperatures measured in industry are usually 2000 K or less. Most energy is radiated in the infrared range (0.75 μm to 50 μm). Therefore, infrared measurement techniques are needed. Much care, however, must be taken, as real materials like cutting tools and work materials are not black bodies. The radiation from these materials is some fraction α of the black body value. α varies with surface roughness, state of oxidation and other factors. Calibration under the same conditions as cutting is necessary.

One of the earliest measurements of radiation from a cutting process was by Schwerd (1933). Since then, methods have followed the development of new infrared sensors. Point measurements, using collimated beams illuminating a PbS cell sensor, have been used to measure temperatures on the primary shear plane (Reichenbach, 1958), on the tool flank

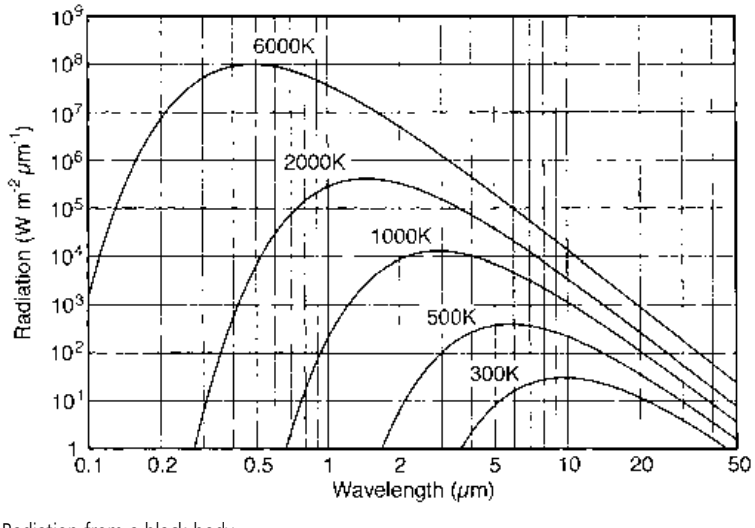


Fig. 5.20 Radiation from a black body

(Chao *et al*., 1961) and on the chip surface (Friedman and Lenz, 1970). With the development of infrared sensitive photographic film, temperature fields on the side face of a chip and tool have been recorded (Boothroyd, 1961) and television-type infrared sensitive video equipment has been used by Harris *et al*. (1980).

Infrared sensors have continued to develop, based on both heat sensing and semiconductor quantum absorption principles. The sensitivity of the second of these is greater than the first, and its time constant is quite small too – in the range of μs to ms. Figure 5.21 shows a recent example of the use of the second type. Two sensors, an InSb type sensitive in the 1 μm to 5 μm wavelength range and a HgCdTe type, sensitive from 6 μm to 13 μm, were used: more sensitive temperature measurements may be made by comparing the signals from two different detectors.

Most investigations of temperature in metal cutting have been carried out to understand the process better. In principle, temperature measurement might be used for condition monitoring, for example to warn if tool flank wear is leading to too hot cutting conditions. However, particularly for radiant energy measurements and in production conditions, calibration issues and the difficulty of ensuring the radiant energy path from the cutting zone to the detector is not interrupted, make temperature measurement for such a purpose not reliable enough. Monitoring the acoustic emissions from cutting is

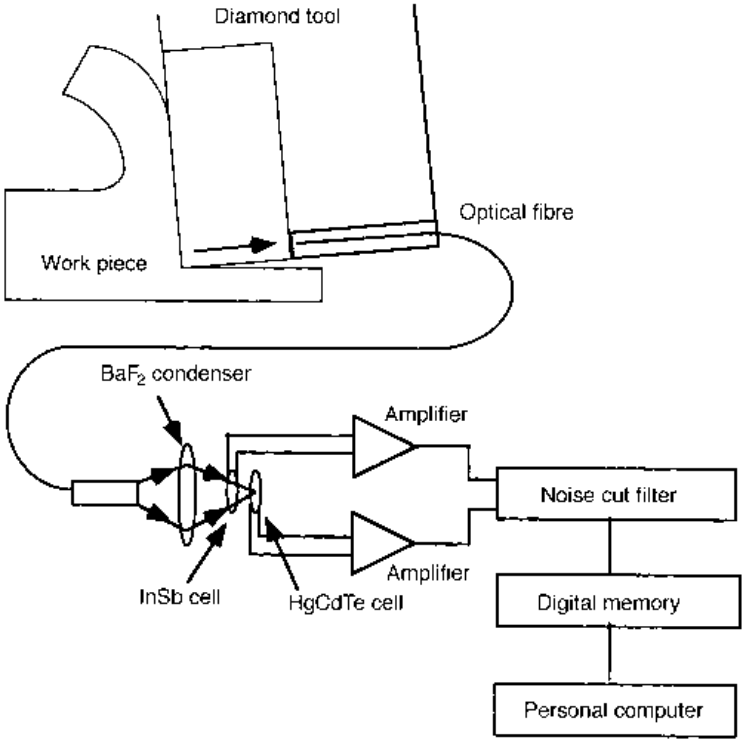


Fig. 5.21 Experimental set-up for measuring the temperature of a chip's back surface at the cutting point, using a diamond tool and infrared light, after Ueda *et al.* (1998)

another way, albeit an indirect method, to study the state of the process, and this is considered next.

5.4 Acoustic emission

The dynamic deformation of materials – for example the growth of cracks, the deformation of inclusions, rapid plastic shear, even grain boundary and dislocation movements – is accompanied by the emission of elastic stress waves. This is acoustic emission (AE). Emissions occur over a wide frequency range but typically from 100 kHz to 1 MHz. Although the waves are of very small amplitude, they can be detected by sensors made from strongly piezoelectric materials, such as $BaTiO_3$ or PZT ($Pb(Zr_xTi_{1-x})O_3$; $x = 0.5$ to 0.6).

Figure 5.22 shows the structure of a sensor. An acoustic wave transmitted into the sensor causes a direct stress $E(\Delta L/L)$ where E is the sensor's Young's modulus, L is it length and ΔL is its change in length. The stress creates an electric field

$$T = g_{33}E(\Delta L/L) \tag{5.7a}$$

where g_{33} is the sensor material's piezoelectric stress coefficient. The voltage across the sensor, TL, is then

$$V = g_{33}E\Delta L \tag{5.7b}$$

Typical values of g_{33} and E for PZT are 24.4×10^{-3} V m/N and 58.5 GPa. It is possible, with amplification, to detect voltages as small as 0.01 mV. These values substituted into equation (5.7b) lead to the possibility of detecting length changes ΔL as small as 7×10^{-15} m: for a sensor with $L = 10$ mm, that is equivalent to a minimum strain of 7×10^{-13}. AE

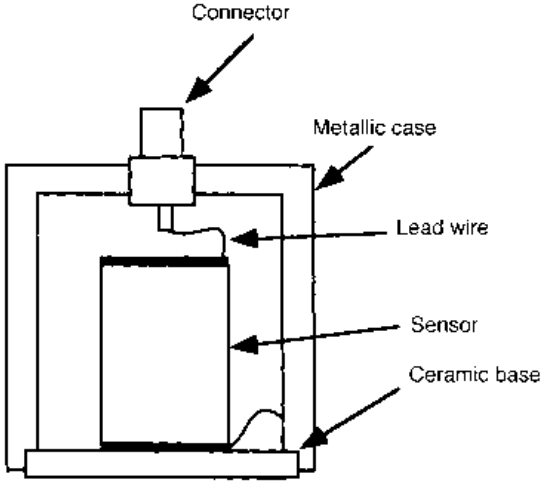


Fig. 5.22 Structure of an AE sensor

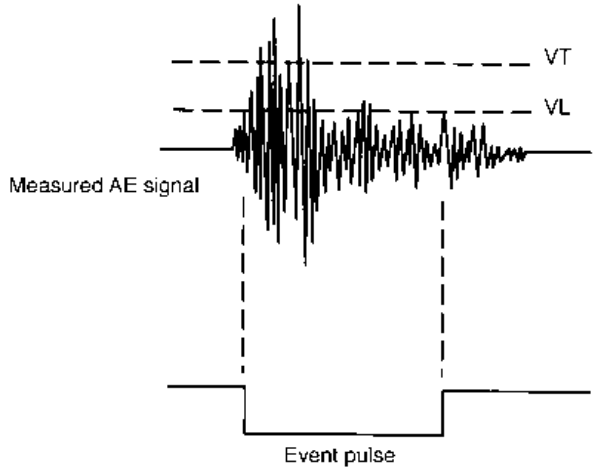


Fig. 5.23 An example of an AE signal and signal processing

strain sensing is much more sensitive than using wire strain gauges, for which the minimum detectable strain is around 10^{-6}.

The electrical signal from an AE sensor is processed in two stages. It is first passed through a low noise pre-amplifier and a band-pass filter (≈100 kHz to 1 MHz). The resulting signal typically has a complicated form, based on events, such as in Figure 5.23. In the second stage of processing, the main features of the signal are extracted, such as the number of events, the frequency of pulses with a voltage exceeding some threshold value *VL*, the maximum voltage *VT*, or the signal energy.

The use of acoustic emission for condition monitoring has a number of advantages. A small number of sensors, strategically placed, can survey the whole of a mechanical system. The source of an emission can be located from the different times the emission takes to reach different sensors. Its high sensitivity has already been mentioned. It is also easy to record; and acoustic emission measuring instruments are lightweight and small. However, it also has some disadvantages. The sensors must be attached directly to the system being monitored: this leads to long term reliability problems. In noisy conditions it can become impossible to isolate events. Acoustic emission is easily influenced by the state of the material being monitored, its heat treatment, pre-strain and temperature. In addition, because it is not obvious what is the relationship between the characteristics of acoustic emission events and the state of the system being monitored, there is even more need to calibrate or train the measuring system than there is with thermal radiation measurements.

In machining, the main sources of AE signals are the primary shear zone, the chip–tool and tool–work contact areas, the breaking and collision of chips, and the chipping and fracture of the tool. AE signals of large power are generally observed in the range 100 kHz to 300 kHz. Investigations of their basic properties and uses in detecting tool wear and chipping have been the subject of numerous investigations, for example Iwata and Moriwaki (1977), Kakino (1984) and Diei and Dornfeld (1987). The potential of using AE is seen in Figure 5.24. It shows a relation between flank wear *VB* and the amplitude level

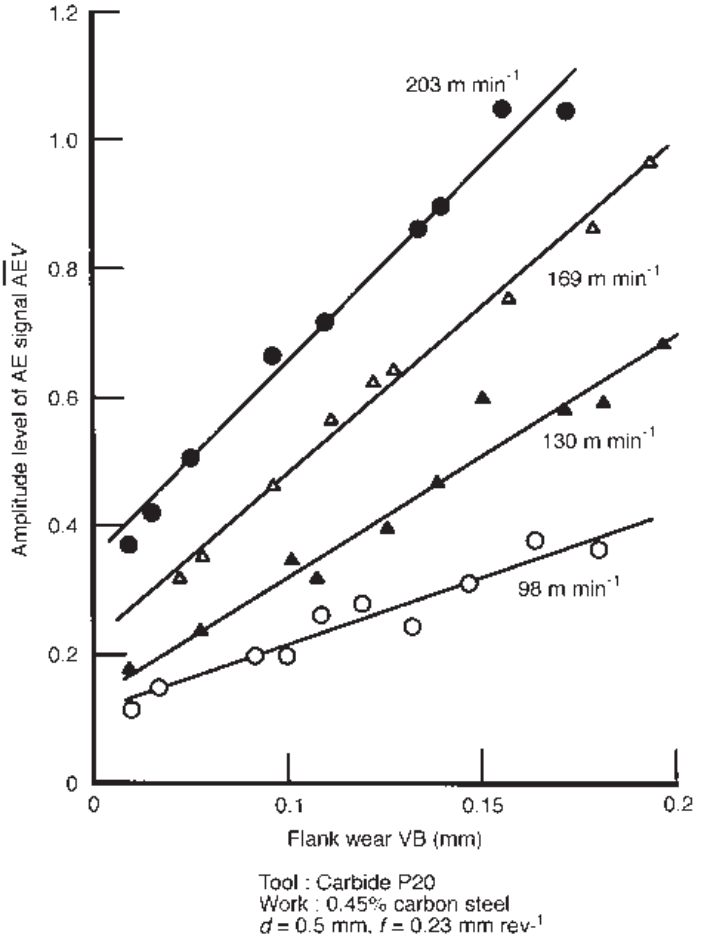


Fig. 5.24 Relation between flank wear VB and amplitude of AE signal, after Miwa *et al.* (1981)

of an AE signal in turning a 0.45% plain carbon steel (Miwa, 1981). The larger the flank wear, the larger the AE signal, while the rate of change of signal with wear changes with the cutting conditions, such as cutting speed.

References

Boothroyd, G. (1961) Photographic technique for the determination of metal cutting temperatures. *British J. Appl. Phys.* **12**, 238–242.

Chao, B. T., Li, H. L. and Trigger, K. J. (1961) An experimental investigation of temperature distribution at tool flank surface. *Trans. ASME J. Eng. Ind.* **83**, 496–503.

Diei, E. N. and Dornfeld, D. A. (1987) Acoustic emission from the face milling process – the effects of process variables. *Trans ASME J. Eng. Ind.* **109**, 92–99.

Friedman, M. Y. and Lenz, E. (1970) Determination of temperature field on upper chip face. *Annals CIRP* **19**(1), 395–398.

Harris, A., Hastings, W. F. and Mathew, P. (1980) The experimental measurement of cutting temperature. In: *Proc. Int. Conf. on Manufacturing Engineering*, Melbourne, 25–27 August, pp. 30–35.

Iwata, I. and Moriwaki, T. (1977) An application of acoustic emission to in-process sensing of tool wear. *Annals CIRP* **26**(1), 21–26.

Kakino, K. (1984) Monitoring of metal cutting and grinding processes by acoustic emission. *J. Acoustic Emission* **3**, 108–116.

Miwa, Y., Inasaki, I. and Yonetsu, S. (1981) In-process detection of tool failure by acoustic emission signal. *Trans JSME* **47**, 1680–1689.

Reichenbach, G. S. (1958) Experimental measurement of metal cutting temperature distribution. *Trans ASME* **80**, 525–540.

Schwerd, F. (1933) Uber die bestimmung des temperaturfeldes beim spanablauf. *Zeitschrift VDI* **77**, 211–216.

Shaw, M. C. (1984) *Metal Cutting Principles*. Oxford: Clarendon Press.

Trent, E. M. (1991) *Metal Cutting*, 3rd edn. Oxford: Butterworth Heinemann.

Ueda, T., Sato, M. and Nakayama, K. (1998) The temperature of a single crystal diamond tool in turning. *Annals CIRP* **47**(1), 41–44.

Williams, J. E, Smart, E. F. and Milner, D. (1970) The metallurgy of machining, Part 1. *Metallurgia* **81**, 3–10.

6

Advances in mechanics

6.1 Introduction

Chapter 2 presented initial mechanical, thermal and tribological considerations of the machining process. It reported on experimental studies that demonstrate that there is no unique relation between shear plane angle, friction angle and rake angle; on evidence that part of this may be the influence of workhardening in the primary shear zone; on high temperature generation at high cutting speeds; and on the high stress conditions on the rake face that make a friction angle an inadequate descriptor of friction conditions there. Chapters 3 to 5 concentrated on describing the properties of work and tool materials, the nature of tool wear and failure and on experimental methods of following the machining process. This sets the background against which advances in mechanics may be described, leading to the ability to predict machining behaviours from the mechanical and physical properties of the work and tool.

This chapter is arranged in three sections in addition to this introduction: an account of slip-line field modelling, which gives much insight into continuous chip formation but which is ultimately frustrating as it offers no way to remove the non-uniqueness referred to above; an account of the introduction of work flow stress variation effects into modelling that removes the non-uniqueness, even though only in an approximate manner in the first instance; and an extension of modelling from orthogonal chip formation to more general three-dimensional (non-orthogonal) conditions. It is a bridging chapter, between the classical material of Chapter 2 and modern numerical (finite element) modelling in Chapter 7.

6.2 Slip-line field modelling

Chapter 2 presented two early theories of the dependence of the shear plane angle on the friction and rake angles. According to Merchant (1945) (equation (2.9)) chip formation occurs at a minimum energy for a given friction condition. According to Lee and Shaffer (1951) (equation (2.10)) the shear plane angle is related to the friction angle by plastic flow rules in the secondary shear zone. Lee and Shaffer's contribution was the first of the slip-line field models of chip formation.

6.2.1 Slip-line field theory

Slip-line field theory applies to plane strain (two-dimensional) plastic flows. A material's mechanical properties are simplified to rigid, perfectly plastic. That is to say, its elastic moduli are assumed to be infinite (rigid) and its plastic flow occurs when the applied maximum shear stress reaches some critical value, k, which does not vary with conditions of the flow such as strain, strain-rate or temperature. For such an idealized material, in a plane strain plastic state, slip-line field theory develops rules for how stress and velocity can vary from place to place. These are considered in detail in Appendix 1. A brief and partial summary is given here, sufficient to enable the application of the theory to machining to be understood.

First of all: what are a slip-line and a slip-line field; and how are they useful? The analysis of stress in a plane strain loaded material concludes that at any point there are two orthogonal directions in which the shear stresses are maximum. Further, the direct stresses are equal (and equal to the hydrostatic pressure) in those directions. However, those directions can vary from point to point. If the material is loaded plastically, the state of stress is completely described by the constant value k of maximum shear stress, and how its direction and the hydrostatic pressure vary from point to point. A line, generally curved, which is tangential all along its length to directions of maximum shear stress is known as a slip-line. A slip-line field is the complete set of orthogonal curvilinear slip-lines existing in a plastic region. Slip-line field theory provides rules for constructing the slip-line field in particular cases (such as machining) and for calculating how hydrostatic pressure varies within the field.

One of the rules is that if one part of a material is plastically loaded and another is not, the boundary between the parts is a slip-line. Thus, in machining, the boundaries between the primary shear zone and the work and chip and between the secondary shear zone and the chip are slip-lines. Figure 6.1 sketches slip-lines OA, A′D and DB that might be such boundaries. It also shows two slip-lines inside the plastic region, intersecting at the point 2 and labelled α and β, and an element of the slip-line field mesh labelled EFGH (with the shear stress k and hydrostatic pressure p acting on it); and it draws attention to two regions labelled 1 and 3, at the free surface and on the rake face of the tool. The theory is developed in the context of this figure.

As a matter of fact, Figure 6.1 breaks some of the rules. Some correct detail has been sacrificed to simplify the drawing – as will be explained. Correct machining slip-line fields are introduced in Section 6.2.2.

The variation of hydrostatic pressure with position along a slip-line is determined by force equilibrium requirements. If the directions of the slip-lines at a point are defined by the anticlockwise rotation ϕ of one of the lines from some fixed direction (as shown for example at the centre of the region EFGH); and if the two families of lines that make up the field are labelled α and β (also as shown) so that, if α and β are regarded as a right-handed coordinate system, the largest principal stress lies in the first quadrant (this is explained more in Appendix 1), then

$$\left.\begin{array}{l} p + 2k\phi = \text{constant, along an } \alpha\text{-line} \\ p - 2k\phi = \text{constant, along a } \beta\text{-line} \end{array}\right\} \tag{6.1}$$

Force equilibrium also determines the slip-line directions at free surfaces and friction surfaces (1 and 3 in the figure) – and at a free surface it also controls the size of the hydrostatic pressure. By definition, a free surface has no force acting on it. From this, slip-lines

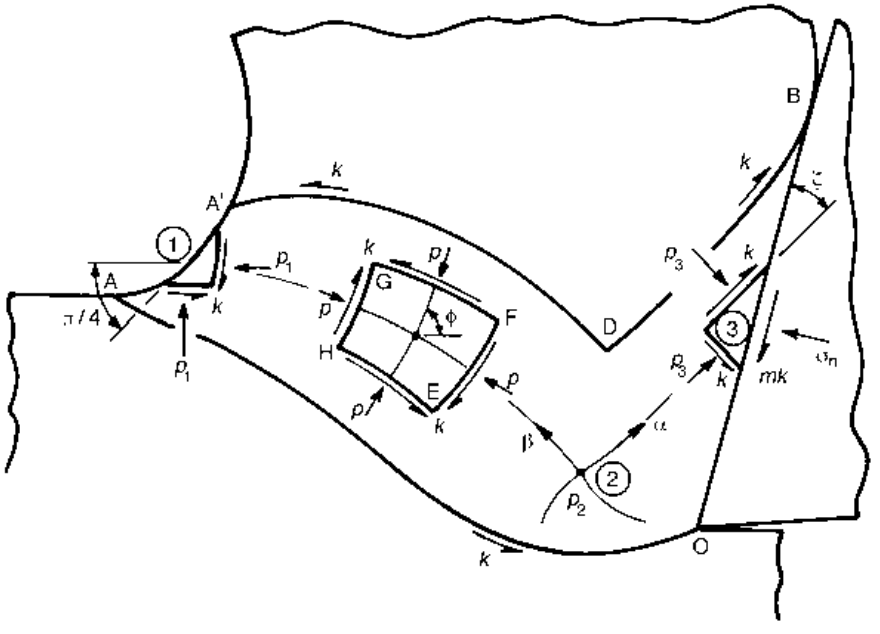


Fig. 6.1 A wrong guess of a chip plastic flow zone shape, to illustrate some rules of slip-line field theory

intersect a free surface at 45° and the hydrostatic pressure is either $+k$ or $-k$ (depending respectively on whether the free surface normal lies in the first or second quadrant of the coordinate system). At a friction surface, where the friction stress is defined as mk (as introduced in Chapter 2), the slip lines must intersect the surface at an angle ζ (defined at 3 in the figure) given by

$$\cos 2\zeta = m \tag{6.2}$$

As an example of the rules so far, equation (6.1) can be used to calculate the hydrostatic pressure p_3 at 3 if the hydrostatic pressure p_1 is known ($p_1 = +k$ in this case) and if the directions of the slip-lines ϕ_1, ϕ_2 and ϕ_3 at points 1, 2 and 3 are known (point 2 is the intersection of the α and β lines connecting points 1 and 3). Then, the normal contact stress, σ_n, at 3 can be calculated from the force equilibrium of region 3:

$$\left.\begin{aligned} p_3 &= k - 2k[(\phi_1 - \phi_2) - (\phi_2 - \phi_3)] \\ \sigma_n &= p_3 + k \sin 2\zeta \end{aligned}\right\} \tag{6.3}$$

Rules are needed for how ϕ varies along a slip-line. It can be shown that the rotations of adjacent slip-lines depend on one another. For an element such as EFGH

$$\left.\begin{aligned} \phi_F - \phi_G &= \phi_E - \phi_H \\ \text{or} \qquad \phi_H - \phi_G &= \phi_E - \phi_F \end{aligned}\right\} \tag{6.4}$$

From this, the shapes of EF and GF are determined by HG and HE. By extension, in this example, the complete shape of the primary shear zone can be determined if the shape of the boundary AO and the surface region AA′ is known.

One way in which Figure 6.1 is in error is that it violates the second of equations (6.4). The curvatures of the α-lines change sign as the β-line from region 1 to region 2 is traversed. Another way relates to the velocities in the field that are not yet considered. A discontinuous change in tangential velocity is allowed on crossing a slip-line, but if that happens the discontinuity must be the same all along the slip-line. In Figure 6.1, a discontinuity must occur across OA at O, because the slip-line there separates chip flow up the tool rake face from work flow under the clearance face. However, no discontinuity of slope is shown at A on the free surface, as would occur if there were a velocity discontinuity there.

6.2.2 Machining slip-line fields and their characteristics

A major conclusion of slip-line field modelling is that specification of the rake angle α and friction factor m does not uniquely determine the shape of a chip. More than one field can be constructed, each with a different chip thickness and contact length with the tool. The possibilities are fully described in Appendix 1. Figure 6.2 sketches three of them, for $\alpha = 5^\circ$ and $m = 0.9$, typical for machining a carbon steel with a cemented carbide tool. The estimated variations along the rake face of σ_n/k and of the rake face sliding velocity as a fraction of the chip velocity, U_{rake}/U_{chip}, are added to the figures, and so is the final

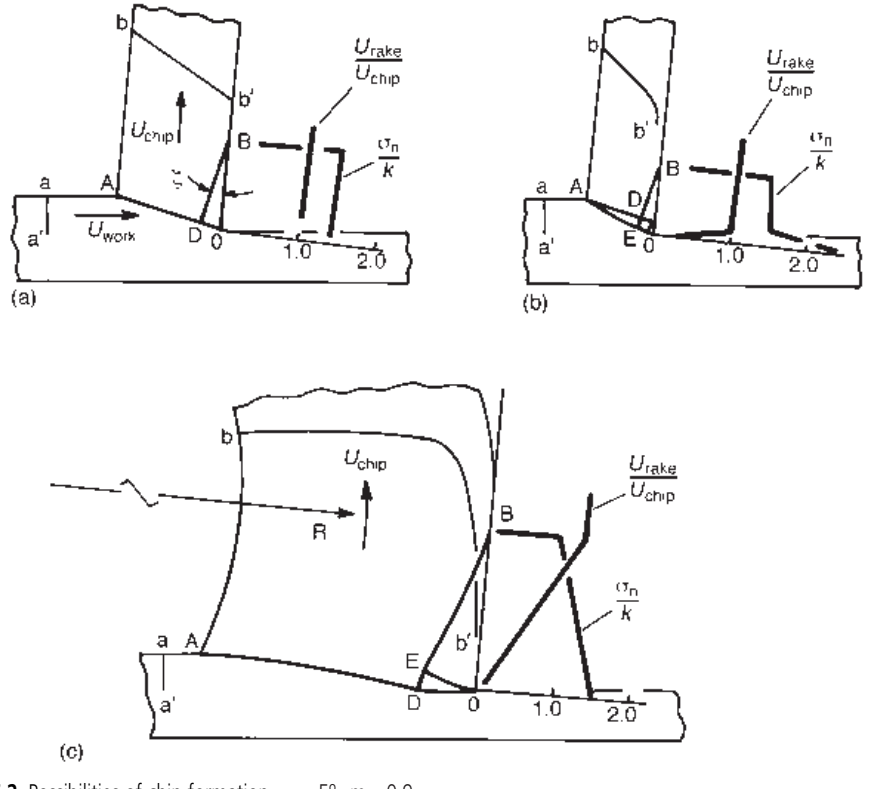


Fig. 6.2 Possibilities of chip formation, $\alpha = 5°$, $m = 0.9$

shape bb′ of an originally straight line aa′, which has passed through the chip formation zone.

Figure 6.2(a) is the Lee and Shaffer field. The slip-lines OA and DB are straight. Consequently, the hydrostatic stress is constant in the field: its value is not determined by a free surface condition at A (the plastic zone at A has no thickness) but from the condition that the chip is free – there is no resultant force across ADB. The straightness of the slip-lines results in a constant normal stress along the chip/tool contact, and a sliding velocity U_{rake} everywhere equal to the chip velocity. The line bb′ is also straight, its orientation determined by the difference between the chip and work velocities.

Figure 6.2(b) shows a field introduced by Kudo (1965). The shear plane AD of Lee and Shaffer's field is replaced by a straight-sided fan shaped region ADE, centred on A. The result is that it describes thinner chips with shorter contact lengths. The rake face normal contact stress is calculated to increase and the rake face sliding velocity to reduce close to the cutting edge. The chip is formed straight, but its reduced velocity near the cutting edge causes the line bb′ to become curved. Such curved markings are frequently observed in real chips (Figure 2.4).

Figure 6. 2(c) shows a field introduced by Dewhurst (1978). Its boundaries OA and DB are curved; and a fan shaped region ODE is centred on O. The result is the formation of a curled chip, with some radius R, thicker and with a longer contact length than the Lee and Shaffer field. The hydrostatic pressure and the velocity vary continuously from place to place. The normal contact stress and the rake face sliding velocity vary over the entire chip/tool contact length; and bb′ is grossly curved.

The normal contact stress variations reproduce the range of observations made experimentally (Figure 2.22), except of course they do not show the elastically stressed tail of the experimental data.

The Kudo and Dewhurst fields that are illustrated are, in each case, just one of a family of possibilities, each with a different fan angle DAE (the Kudo field) or different rotation from A to D (the Dewhurst field). All that is required is that the hydrostatic pressure at A, calculated for each field from the free chip boundary condition, is able to be contained by the surrounding work or chip (which is supposed to be rigid). For each possibility that satisfies this, the average friction and normal rake face contact stress can be calculated, to obtain the effective friction angle at the contact. The chip thickness to feed ratio can also be determined to obtain the effective shear plane angle. Equation (2.5b) can then be used to determine the dimensionless specific cutting and thrust forces. Figure 6.3 plots results from such an exercise, for two values of rake angle. The observed non-uniqueness found experimentally, shown here and also in Figure 2.15, fits well within the bounds of slip-line field theory.

Unfortunately, slip-line field theory cannot explain why any one expermental condition leads to a particular data point in Figure 6.3. It does conclude though, that the increased shear plane angle at constant friction angle is associated with a reduced chip/tool contact length. Factors that lead to a reduced contact length, perhaps such as increased friction heating with increased cutting speed, leading to reduced rake face shear stresses, are beyond the simplifying assumptions of the theory of constant shear flow stress.

Figure 6.3(b) supports the view that if cutting could be carried out with 30° rake angle tools, the spread of allowable specific forces would be very small and it would not matter much that slip-line field theory cannot explain where in the range a particular result will

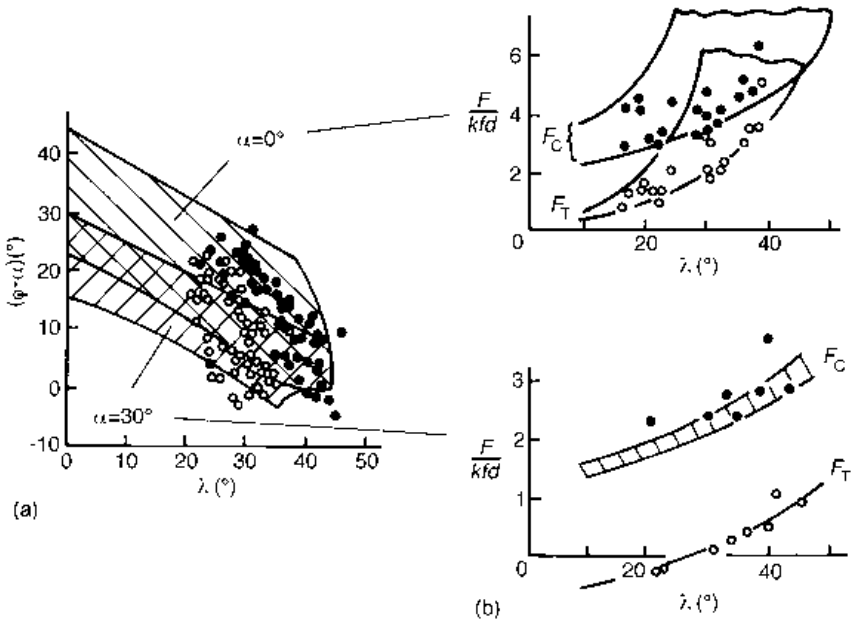


Fig. 6.3 Slip-line field allowed ranges of (a) (ϕ–α) and (b) specific forces and λ, for tools of rake angle 0° and 30°: experimental data for carbon steels (Childs, 1980)

lie. Unfortunately, to avoid tool breakage, rake angles closer to 0˚ are more common. The ranges of allowable specific forces at a particular friction angle are then large.

6.2.3 Further considerations

In addition to directly estimating machining parameters, slip-line field theory may be used to stimulate thought about the machining process and its modelling.

In Chapter 2, around Figure 2.11, it was discussed how work-hardening might change the mean level of hydrostatic stress on the shear plane, and hence the angle ($\phi + \lambda - \alpha$) between the resultant force and the shear plane. The mean level of hydrostatic stress can now be seen to be variable even in the absence of work-hardening, depending on the choice of slip-line field. Figure 6.4 shows the range of values of ($\phi + \lambda - \alpha$), as a function of ϕ, allowed by the Kudo and Dewhurst fields. Values are found from 0.5 to 2.0. These compare with 1.2 to 1.4 deduced experimentally for fully work-hardened materials in Figure 2.11(b). It is arguable that some of the further variation of ($\phi + \lambda - \alpha$) observed in Figure 2.11(b), attributed to work-hardening induced pressure variation along the primary shear plane, could be due to a free surface hydrostatic pressure changed for other reasons. The line $\tan(\phi + \lambda - \alpha) = [1 + 2(\pi/4 - \phi)]$ added to Figure 6.4 relates to this and is returned to in Section 6.3.

In Figure 6.3, rake face friction is described by the friction angle λ, even though the friction factor m is believed to be a physically more realistic way to describe the conditions. This is a practical consideration: λ is easier to measure. It is interesting therefore to look in a little more detail at the relation between m and λ. Figure 6.5 shows, as the hatched

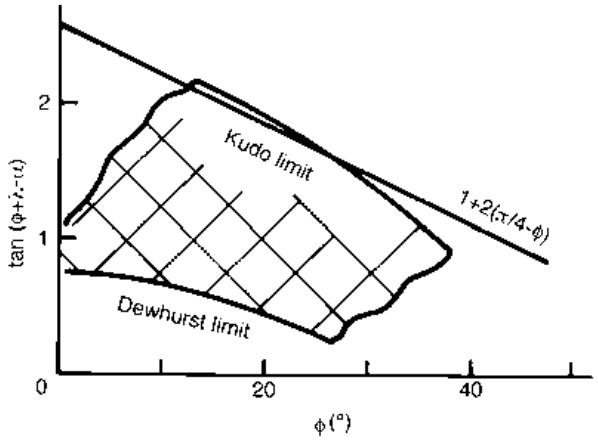


Fig. 6.4 Slip-line field predicted ranges of tan(ϕ+λ–α), dependent on ϕ, for α = 0°

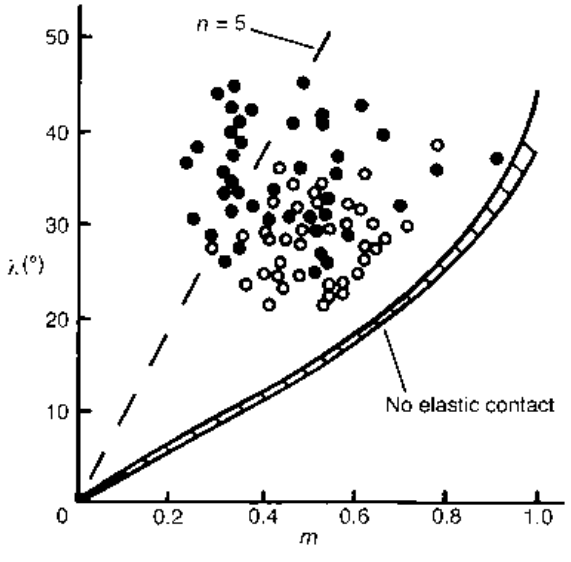


Fig. 6.5 Effects of elastic contact on relations between λ and m. Experimental data for carbon steels, (after Childs, 1980)

region, the slip-line field predicted relationship between λ and m for $\alpha = 0^{\circ}$ (in fact the relationship is almost independent of α). There is almost a one-to-one relationship between the two. It also shows experimental observations for carbon steels – the m values were deduced by dividing the measured rake face friction force per unit depth of cut by the total chip/tool contact length – and experiment and theory do not agree. The reason is that the measured contact lengths include an elastic part, less loaded than the plastic part. The deduced m values are averages over a plastic and an elastic regime. This was considered in a paper by Childs (1980). In that paper, an empirical modification to slip-line field theory was made,

considering elastic contact forces as external forces on an otherwise free chip. The line $n = 5$ in Figure 6.5 was deduced for an elastic contact length five times the plastic length. The elastic contact should not be ignored in machining analyses.

Slip-line field modelling may also be applied to machining with restricted contact tools (Usui *et al*., 1964), with chip breaker geometry tools (Dewhurst, 1979), with negative rake tools (Petryk, 1987), as well as with flank-worn tools (Shi and Ramalingham, 1991), to give an insight into how machining may be changed by non-planar rake face and cutting edge modified tools. Figures 6.6 and 6.7 give examples.

Figure 6.6 is concerned with modifications to chip flow caused by non-planar rake-faced tools. As the chip/tool contact length is reduced below its natural value by cutting away the rake face (Figure 6.6(a)), the sliding velocity on the remaining rake face is reduced, with the creation of a stagnant zone, and the chip streams into the space created by cutting away the tool. If a chip breaker obstruction, of slope δ, is added some distance l_B from the cutting edge of a plane tool (Figure 6.6(b)), its effect on chip curvature and cutting forces can be estimated. The combination of these effects can give some guidance on the geometrical design of practical chip-breaker geometry tools.

The slip-line fields of Figure 6.7 show how, with increasingly negative rake angle, a stagnant zone may develop, eventually (Figure 6.7(c)) allowing a split in the flow, with material in the region of the cutting edge passing under the tool rather than up the rake face. The fields in this figure, at first sight, are for tools of an impractically large negative rake angle. However, real tools have a finite edge radius, can be worn or can be manufactured with a negative rake chamfer. The possibility of stagnation that these fields signal, needs to be accomodated by numerical modelling procedures.

6.2.4 Summary

In summary, the slip-line field method gives a powerful insight into the variety of possible chip flows. A lack of uniqueness between machining parameters and the friction stress

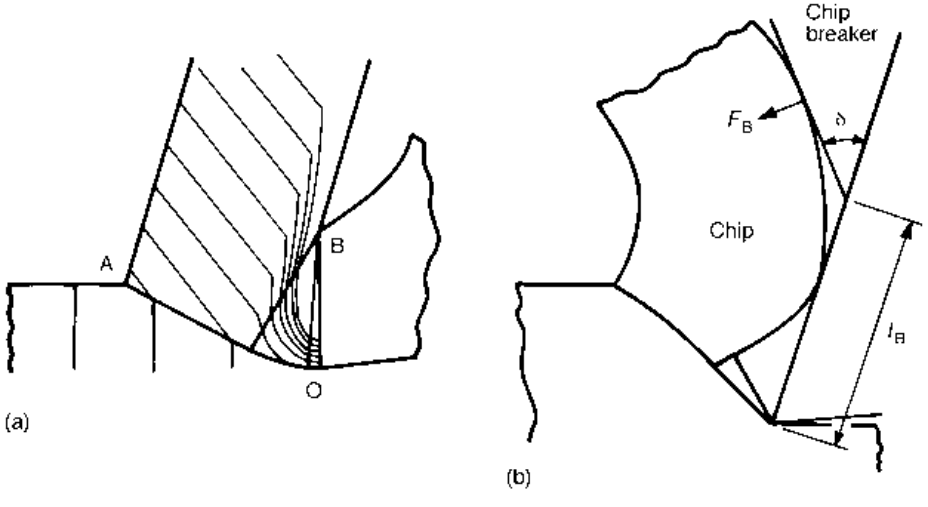


Fig. 6.6 Slip-line field models of cutting with (a) zero rake restricted contact and (b) chip breaker geometry tools, after Usui *et al.* (1964) and Dewhurst (1979)

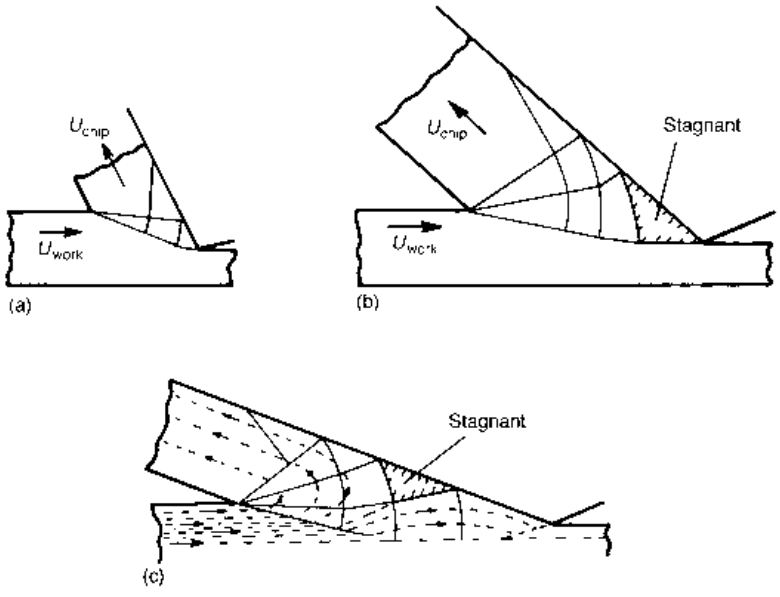


Fig. 6.7 Chip flows with tools, from (a) to (c) of increasingly negative rake (after Petryk, 1987)

between the chip and tool is explained by the freedom of the chip, at any given friction stress level, to take up a range of contact lengths with the tool. Chip equilibrium is maintained for different contact lengths by allowing the level of hydrostatic stress in the field to vary. The velocity fields indicate where there are regions of intense shear, which should be taken into account later in numerical modelling. They also illustrate how velocities might vary in the secondary shear zone, a topic that will be returned to later. They also show a range of variations of normal contact stress on the rake face that is observed in practice. However, a frustrating weakness of the slip-line field approach is that it offers no way, within the limitations of the rigid perfectly plastic work material model, of removing the non-uniqueness: what does control the chip/tool contact length in a given situation? Additionally, it can offer no way of taking into account variable flow stress properties of real materials, demonstrated experimentally to have an influence. An alternative modelling, concentrating on material property variation effects, is introduced in the next section.

6.3 Introducing variable flow stress behaviour

Slip-line field modelling investigates the variety of chip formation allowed by equilibrium and flow conditions while grossly simplifying a metal's yield behaviour. A complementary approach is to concentrate on the effects of yield stress varying with strain (strain hardening) and in many cases with strain rate and temperature too, while simplifying the modelling of equilibrium and flow. Pioneering work in this area is associated with the name of Oxley. The remainder of this section relies heavily on his work, which is summarized in *Mechanics of Machining* (Oxley, 1989). Developments may be considered in four phases: firstly experimental and numerical studies of actual chip flows, by the method of visioplasticity; secondly, simplifications allowing analytical relations to be developed between stress variations in the

primary shear zone and material flow properties, dependent on strain, strain rate and temperature; thirdly, a consideration of stress conditions in the secondary shear zone; and finally, a synthesis of these, allowing the prediction of chip flow from work material properties.

6.3.1 Observations of chip flows

Visioplasticity is the study of experimentally observed plastic flow patterns. In its most complete form, strain rates throughout the flow are deduced from variations of velocity with position, and strains are calculated by integrating strain rates with respect to time along the streamlines of the flow. The temperatures associated with the plastic work are calculated from heat conduction theory. Then, from independent knowledge of the variation of flow stress with the strain, strain rate and temperature, it can be attempted to deduce what the stress variations are throughout the flow and what resultant forces are needed to create the flow. Alternatively, measured values of the forces can be used to deduce how the flow stress varied. Frequently, however, the accuracy of flow measurement is not good enough to support this entire scheme. Nonetheless, useful insights come from only partial success.

In the case of plane strain flows, the first step is usually to determine the maximum shear strain rate trajectories of the flow, and from these to construct the slip-line field. Departures of the field's shape from the rules established for perfectly plastic solids (Section 6.2) are commonly observed. Figure 6.8(a) shows an early example of a chip primary shear zone investigated in this way (Palmer and Oxley, 1959). In addition to flow calculations in deriving this field, Palmer and Oxley also applied the force equilibrium constraint, that the slip-lines should intersect the free surface AA′ at 45°. The field is for a mild steel machined at the low cutting speed of 12 mm/min and a feed of 0.17 mm. At the low strain rates and temperatures generated in this case, departures from perfect plasticity are expected to be due only to strain hardening. The strain hardening behaviour of the material was measured in a simple compression test.

Two conclusions arise from Figure 6.8 (and from other examples that could have been chosen). First, and most obviously, the entry and exit slip lines OA and OA′ are of opposing curvature. The field violates equation (6.4). This is a direct effect of work-hardening.

Secondly, and less obviously, there is a problem with the constraint placed on the field that the slip-lines should meet the free surface at 45°. By revisiting the derivation of equations (6.1) (Appendix 1, Section 1.2.2), and removing the constraint of no strain hardening, it is easy to show that

$$\left.\begin{aligned} \frac{\partial p}{\partial s_1} + 2k\frac{\partial \phi}{\partial s_1} - \frac{\partial k}{\partial s_2} &= 0 \qquad \text{along an } \alpha - \text{line} \\ \frac{\partial p}{\partial s_2} - 2k\frac{\partial \phi}{\partial s_2} - \frac{\partial k}{\partial s_1} &= 0 \qquad \text{along a } \beta - \text{line} \end{aligned}\right\} \tag{6.5}$$

where s_1 and s_2 are distances along an α and a β slipline respectively. In Figure 6.8(a), as in Figure 6.1, AC is a β line and CA′ an α line. After estimating the variations of k, $\partial k/\partial s_1$ and $\partial k/\partial s_2$ in the region of AA′C, Palmer and Oxley concluded, from the application of equation (6.5), that the hydrostatic pressure at A′ could not equal the shear yield stress of the work hardened material at A′, as it should according to the further constraint imposed

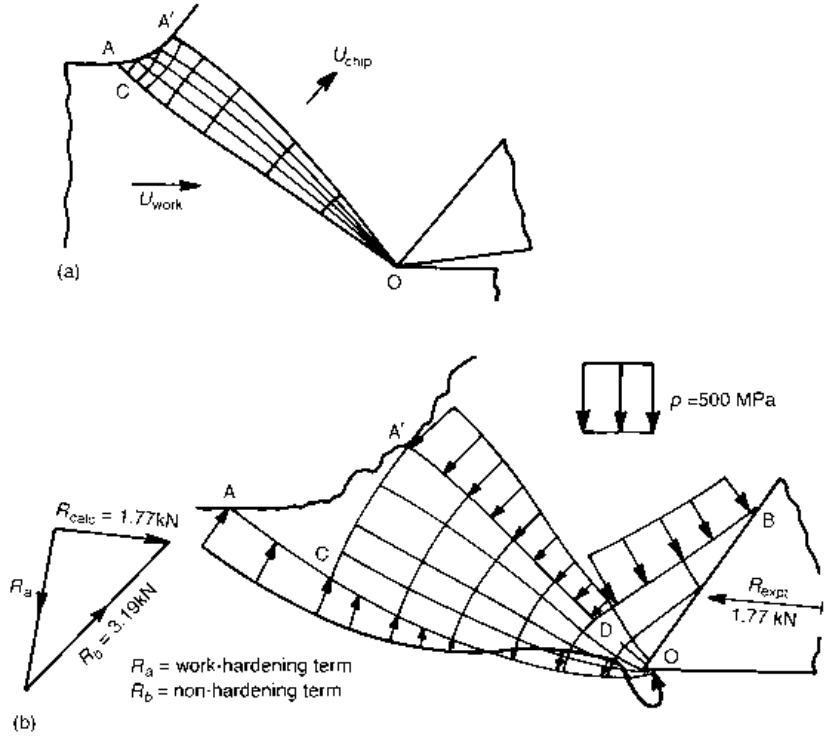


Fig. 6.8 Experimentally derived slip-line fields for slow speed machining of mild steels, after (a) Palmer and Oxley (1959), and (b) Roth and Oxley (1972)

by the free surface boundary condition there. Palmer and Oxley resolved the contradiction by suggesting that plastic flow was not steady at the free surface. The smoothed free surface in Figure 6.8(a) is, in reality, corrugated and therefore the slip-lines should not be constrained to intesect the smoothed profile at 45°.

The result of a later study (Roth and Oxley, 1972), still at low cutting speed to exclude the effects of strain rate and temperature on flow stress – now also including an estimate of the secondary shear zone shape – is shown in Figure 6.8(b). At A, the entry boundary OA is still made to intersect the free surface at 45°: there, continuity of flow ensures that the free surface slope is known (velocity discontinuities cannot exist in a hardening material – discontinuities that would occur in a non-hardening material are broadened into narrow zones). However, a free surface constraint has not been placed on the exit boundary direction at A′; and no attempt has been made to detail the field within the near-surface region AA′C.

Roth and Oxley applied equations (6.5) to the calculation of hydrostatic stress along all the field boundaries, assuming that at A its value was that of the shear yield stress there. These are shown in the figure. Along the entry boundary OA, hydrostatic stress variations are dominated by the effect of work hardening. Integration of the hydrostatic and shear stresses with respect to distance along OA gives the force acting across it. Inclusion of work hardening gives a value of 1.77 kN (in line with experiment), while omitting it gives 3.19 kN, in a grossly different direction.

Over the exit boundaries BD and DA′, where strain hardening has reduced the rate of change of shear flow stress across the slip lines, the variations approach those expected of a non-hardening material. They depend on the direction changes along the lines. The exit region OBDA′ is visually similar in this example to the non-hardening slip-line field proposed by Dewhurst (Figure 6.2(c)). The whole field is this, with the primary shear plane replaced by a work hardening zone of finite width.

In a parallel series of experiments, Stevenson and Oxley (1969–70, 1970–71) extended the direct observations of chip flows to higher cutting speeds, but with a changed focus, to assess how large might be the strain rate and temperature variations in the primary shear zone. Figure 6.9(a) is a sketch of the streamlines that they observed when machining a 0.13%C free-machining steel at a cutting speed of 105 m/min and a feed of 0.26 mm. Figure 6.10 shows, for a range of cutting speeds, the derived variations of maximum shear strain rate along a central streamline, such as aa′ in Figure 6.9(a). The peak of maximum shear strain rate is observed to occur close to the line OA″ that would be described as the shear plane in a shear plane model of the machining process. The peak maximum shear strain rate was measured to vary in proportion to the notional primary shear plane velocity (from equation (2.3)) and inversely as the length s of the shear plane (assumed to be $f/\sin\phi$):

$$\dot{\gamma}_{OA''} = C\,\frac{U_{\text{primary}}}{s} \equiv C\,\frac{U_{\text{work}}}{f}\;\frac{\cos\alpha\,\sin\phi}{\cos(\phi-\alpha)} \tag{6.6}$$

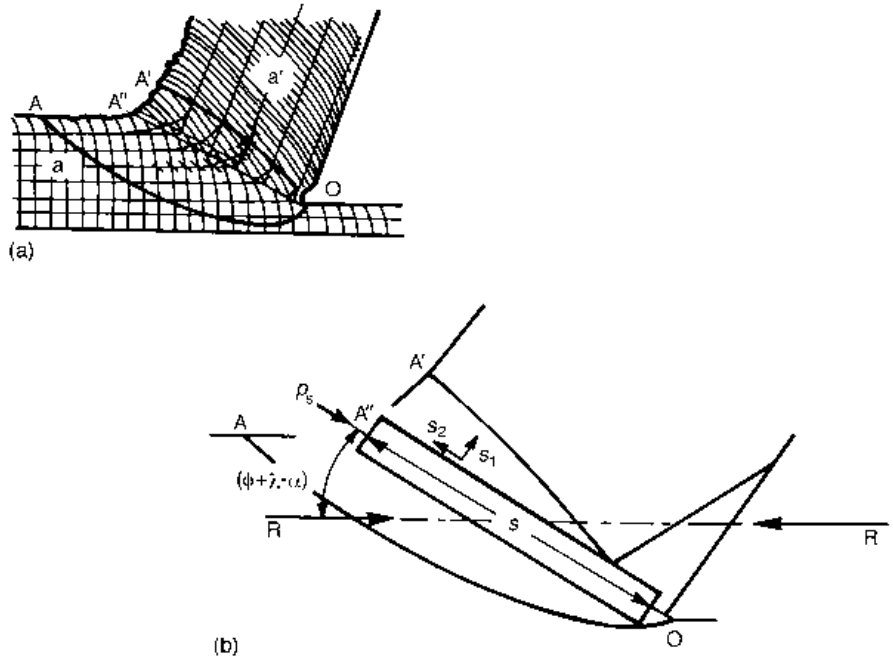


Fig. 6.9 (a) Stream-lines of the flow of a 0.13%C free-cutting steel and (b) a simplification for later analysis (Section 6.3.2)

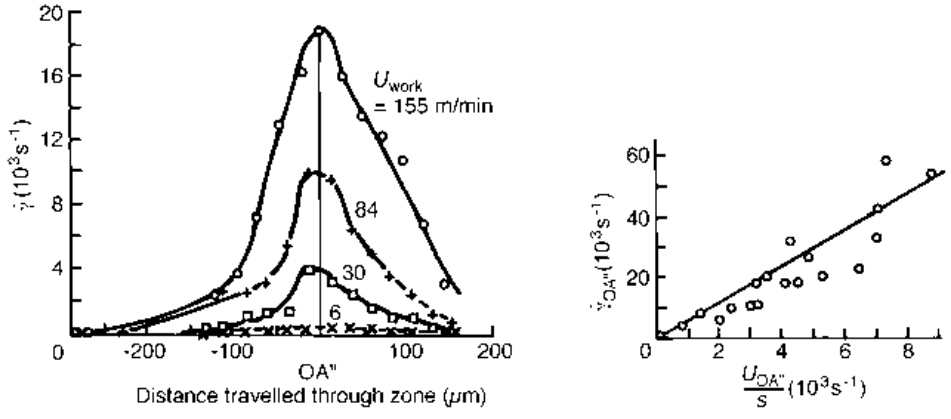


Fig. 6.10 Shear strain rate variations along a central stream-line, and peak shear strain rate changes with cutting speed and feed, as described in the text

In this case, the best-fit constant of proportionality C is 5.9. In many practical machining operations, peak shear strain rates are of the order of 10^4/s.

It is interesting to consider the value of $C = 5.9$ in the light of the length-to-width ratio of the primary shear zone, equal to 2, derived in Chapter 2 from Figure 2.10 and equation (2.7). The average shear strain rate may be roughly half the peak rate. It is also the total shear strain divided by the time for material to pass through the primary zone. This time is the width of the zone divided by the work velocity normal to the plane, namely $U_{work}\sin\phi$. An easy manipulation equates the length-to-width ratio to $C/2$, or about 3 in this case. A consistent view emerges of a primary shear region in which the strain rates do in fact peak along a plane OA″ but which in its totality may not be as narrow compared with its length as is commonly believed.

Temperature rises in the primary zone have already been considered in Chapter 2. Stevenson and Oxley used the same approach described there to obtain the total temperature rise from the measured cutting forces resolved on to the shear plane. In the notation of this book, combining equations (2.4a), (2.5c) and (2.14), and remembering that only a fraction $(1 - \beta)$ of generated heat flows into the chip

$$\Delta T_1 = \frac{(1-\beta)}{\rho C} \frac{F_C \cos\phi - F_T \sin\phi}{fd} \frac{\cos\alpha}{\cos(\phi - \alpha)} \tag{6.7}$$

However, as will be seen in the next section, there is a particular interest in the temperature rise in the plane OA″ where the strain rate is largest. Stevenson and Oxley took the temperature along OA″ to be

$$T_{OA''} = T_0 + \eta \Delta T_1 \tag{6.8}$$

where η can range from 0 to 1. Usually, they took it to equal 1, but this is not consistent with OA″ being upstream of the exit boundary of the primary zone. They commented that lower values (0.7 to 0.95) might be better (Oxley, 1989).

6.3.2 Approximate analysis in the primary shear zone

Although a complete analysis of hydrostatic stress variations in the primary shear zone, as in Figure 6.8(b), might be useful in considering the possible fracture of chips during their formation, it might not be necessary if the objective is only to predict the force transmission (the magnitude and direction) across the shear zone. If, for example, along the plane surface OA″ in Figure 6.9(a), variations of hydrostatic stress are dominated by flow stress variations rather than by rotations in the slip-line field, an approximate analysis of stress along OA″, neglecting rotations, might be sufficient. This is the approach developed by Oxley.

Figure 6.9(b) combines aspects of Figures 6.8(b) and 6.9(a), showing the boundaries of a typical flow field but emphasizing a narrow rectangular region around the plane OA″. The hydrostatic stress at A″ is supposed to have some value p_s. Then, by analogy with the derivation of equation (2.7) (Chapter 2), and after assuming pressure variations along OA″ (of length s) are dominated by $\partial k/\partial s_1$, the direction of the resultant force R across OA″ is given by

$$\tan(\phi + \lambda - \alpha) = \frac{p_s}{k_{OA''}} - \frac{1}{2}\,\frac{s}{k_{OA''}}\,\frac{\partial k}{\partial s_1} \tag{6.9a}$$

The size of R (with d, the depth of cut) is found from

$$R\cos(\phi + \lambda - \alpha) = sd.k_{OA''} \tag{6.9b}$$

Oxley showed how to relate the second term on the right-hand side of equation (6.9a) to the work-hardening behaviour of the material, expressed as

$$\bar{\sigma} = \sigma_0 \bar{\varepsilon}^n \tag{6.10}$$

and to the shear strain-rate on OA″, from equation (6.6), in order to replace equation (6.9a) by

$$\tan(\phi + \lambda - \alpha) = \frac{p_s}{k_{OA''}} - Cn \tag{6.11}$$

The term Cn may be thought of as a correction to the value $p_s/k_{OA''}$ that $\tan(\phi + \lambda - \alpha)$ would have in the absence of any strain hardening effects. The non-uniqueness of the non-hardening circumstance has already been considered in section 6.2. There, Figure 6.4 gives a range for the variation of $\tan(\phi + \lambda - \alpha)$ with ϕ, for the example of a zero rake angle tool. In his work, Oxley constrained the range of allowable non-hardening relations, to propose that

$$\frac{P_s}{k_{OA''}} = 1 + 2\left(\frac{\pi}{4} - \phi\right) \tag{6.12}$$

This can be seen in Figure 6.4 to be close to the upper boundary of the allowable range. Then, finally,

$$\tan(\phi + \lambda - \alpha) = 1 + 2(\pi/4 - \phi) - Cn \tag{6.13}$$

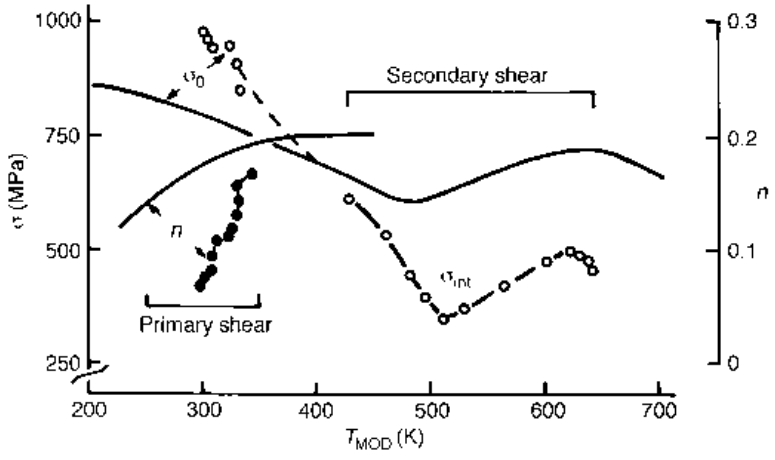


Fig. 6.11 Variations of σ_0 (o) and n (•) for a low carbon steel, derived from machining tests, compared with compression test data (—)

To the extent that constraining the variations of $p_s/k_{OA''}$ is valid, equations (6.9b) and (6.13) may be used to investigate the strain, strain-rate and temperature dependence of flow in the primary shear zone. Stevenson and Oxley (1969–70, 1970–71) carried out turning tests on a 0.13%C steel at cutting speeds up to around 300 m/min, measuring tool forces and shear plane angles. They calculated n from equation (6.13), assuming $C = 5.9$. They calculated $k_{OA''}$ from equation (6.9b), and multiplied it by $\sqrt{3}$ to obtain the equivalent flow stress on OA″; they calculated the equivalent strain on OA″, assuming it to be half the total strain; and finally derived σ_0 (equation (6.10)). They also calculated the strain rate and temperature on OA″. Figure 6.11 shows the variations with strain rate and temperature they derived for σ_0 and n. Strain rate and temperature are combined into a single function, known as the velocity modified temperature, T_{MOD} (K):

$$T_{MOD} = T\{1 - \nu \log(\dot{\bar{\varepsilon}}/\dot{\bar{\varepsilon}}_0)\} \tag{6.14}$$

There are materials science reasons (Chapter 7) why strain rate and temperature might be combined in this way. ν is a material property constant that was taken to be 0.09, and $\dot{\bar{\varepsilon}}_0$ is a reference strain rate that was taken to be 1.

The figure also shows data derived from compression tests on a similar carbon steel and further data (σ_{int}) determined from the analysis of secondary shear flow, which will be discussed in Section 6.3.3. The data for machining and compression tests are not in quantitative agreement, but there is a qualitative similarity in their variations with velocity modified temperature that supports the view that at least some part of the variation of machining forces and shear plane angles with cutting speed is due to the variation of flow stress with strain, strain rate and temperature.

There are clearly a number of assumptions in the procedures just described: that all the variation in $(\phi + \lambda - \alpha)$ is due to variation in n; that the parallel-sided shear zone model is adequate (strain rates in practice will vary from the cutting edge to the free surface, as the actual shear zone width varies); and that C really is a constant of the machining process. In later work, Oxley investigated the sensitivity of his modelling to variations of C. A

change to C causes a change to the hydrostatic stress gradient along the primary shear plane and hence to the normal contact stress on the tool at the cutting edge, $\sigma_{n,O}$. Adding the constraint that $\sigma_{n,O}$ derived from the primary shear plane modelling should be the same as that from secondary shear modelling (Section 6.3.3), he concluded – for the same steel for which he had initially given the value $C = 5.9$, but over a wider range of feed, speed and rake angle cutting conditions – that C might vary between 3.3 and 7.1. The interested reader is referred to Oxley (1989).

6.3.3 Flow in the secondary shear zone

With the partial exception of slow speed cutting tests like those of Roth and Oxley (Figure 6.8), visioplasticity studies have never been accurate enough to give information on strain rate and strain distributions in the secondary shear zone on a par with the level of detail revealed in the primary shear zone. Certainly at high cutting speeds, grids or other internal markers necessary for following the flow are completely destroyed. Nor is there any way, equivalent to applying equation (6.13) in the primary zone, of deducing the strain hardening exponent n for flow in the secondary shear zone. So, even if a flow stress could be deduced for material there, the extraction of a σ_0 value (equation (6.10)) and the estimation of a T_{MOD} value for it might be thought to be impractical. Yet Figure 6.11 contains, in the variation of σ_{int} with T_{MOD}, such plastic flow stress information. The insights and assumptions that enabled this data to be presented are worth considering.

Oxley explicitly suggested that in the secondary shear zone strain-hardening would be negligible above a strain of 1.0. This allowed him, from equation (6.10) with $\bar{\varepsilon} = 1$, to identify σ_0 with $\bar{\sigma}$. It is a major issue in materials' modelling for machining – and is returned to in Chapter 7.4 – to determine how in fact flow stress does vary with strain at the high strains generated in secondary shear. Oxley then suggested that $\bar{\sigma}$ is the same as σ_{int}, or $\sqrt{3}\tau_{av}$, where τ_{av} is the average friction stress over the chip/tool contact area (obtained by dividing the friction force by the measured contact area). This is reasonable, from considerations of the friction conditions in machining (Chapter 2), provided there is a negligible elastic contact region. Oxley argued that this was the case, on the basis of his (Roth and Oxley, 1972) low speed observations, but the observations of Figure 6.5 do not support that.

To determine a T_{MOD} value, he estimated representative temperatures and strain rates in the secondary shear zone. For the strain rate $\dot{\bar{\varepsilon}}_{int}$ he supposed the secondary shear zone to have an average width δt_2, and that the chip velocity varied from zero at the rake face to its bulk value U_{chip} across this width. Then

$$\dot{\bar{\varepsilon}}_{int} \equiv \frac{\dot{\gamma}_{int}}{\sqrt{3}} = \frac{U_{chip}}{\sqrt{3}\,\delta t_2} \tag{6.15}$$

He took the representative temperature to be the average at the rake face, calculated in a manner similar to equation (2.18), but allowing for the variation of work thermal properties with temperature and for the fact that heat generated in secondary shear is not entirely planar but is distributed through the secondary shear zone (Hastings *et al.*, 1980). In the notation of this book, equation (2.18) is modified by a factor c

$$(T - T_0)_{\text{secondary shear}} = (1 - \beta)\frac{k\gamma}{(\rho C)_{\text{work}}} + 0.75c\frac{\tau_{\text{av}} U_{\text{chip}} l}{K_{\text{work}}}\left(\frac{\kappa_{\text{work}}}{U_{\text{chip}} l}\right)^{1/2} \tag{6.16}$$

with

$$c = 1 \text{ if } \delta t_2 \left(\frac{U_{\text{chip}}}{\kappa_{\text{work}} l}\right)^{1/2} < 0.3; \qquad c = 10^{0.06-0.2\delta t}\left(\frac{U_{\text{chip}}}{\kappa_{\text{work}} l}\right)^{1/2} \text{ if } \delta t_2 \left(\frac{U_{\text{chip}}}{\kappa_{\text{work}} l}\right)^{1/2} \geq 0.3$$

The calculated (σ_{int}, T_{MOD}) data in Figure 6.11 result from these assumptions. That they follow the variations expected from independent mechanical testing gives some support to these insights. There is one assumption to which it is particularly interesting to return: that is, that the sliding velocity at the chip/tool interface is zero. This strongly influences both the calculated strain rate and the need for the correction, c, to the temperature calculation. The slip-line field modelling does not support such a severe reduction of chip movement. Figure 6.2, for example, shows sliding velocities reduced to zero only in some circumstances and then only near to the cutting edge. Resolving the conflict between these variable flow stress and slip-line field views of rake face sliding velocities leads to insight into conditions at the rake face during high speed (temperature affected) machining.

In his work, Oxley identified two zones of secondary shear, a broader one and a narrower one within it, closest to the rake face. This narrower zone has also been identified by Trent who describes it as the flow-zone and, when it occurs, as a zone in which seizure occurs between the chip and tool (Trent, 1991). Figure 6.12(a) shows Oxley's measurements of the narrower zone's thickness, for a range of cutting speeds and feeds, for the example of a 0.2%C steel turned with a –5° rake angle tool (other results, for a 0.38%C steel and a +5° rake tool, could also have been shown). The flow-zone is thinner the larger the cutting speed and the lower the feed. In Figure 6.12(b), the observations are replotted against $t(\kappa_{\text{work}}/(U_{\text{work}} f))^{1/2}$. This is the same as $(\kappa_{\text{work}} l/U_{\text{chip}})^{1/2}$, which occurs in equation (6.16), if it is assumed that the contact length l is equal to the chip thickness t. The experimental results lie within a linear band of mean slope 0.2. The flow-zone lies

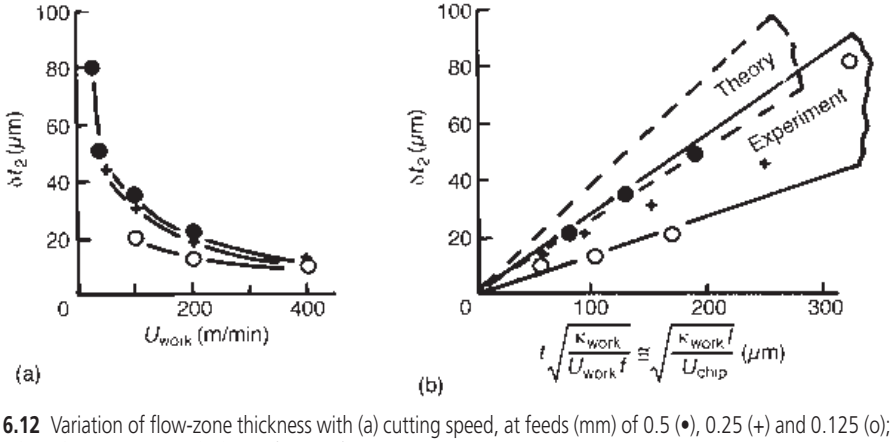


Fig. 6.12 Variation of flow-zone thickness with (a) cutting speed, at feeds (mm) of 0.5 (•), 0.25 (+) and 0.125 (o); and (b) replotted to compare with theory (see text)

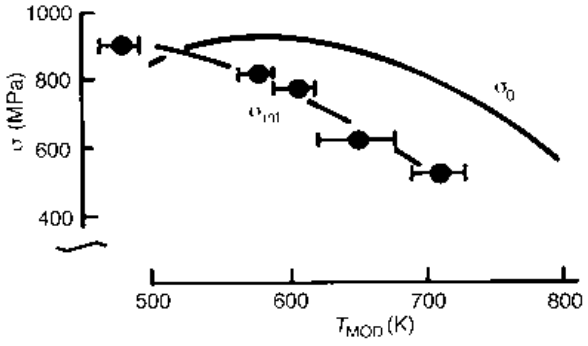


Fig. 6.13 0.45%C steel data from Figure 2.22(a), replotted (•) as σ_{int} versus T_{MOD} and compared with σ_o for a similar steel taken from Oxley (1989)

within, and is proportional to the thickness of, the chip layer heated by sliding over the tool.

Oxley pointed out that the temperature of the flow zone would reduce the thicker it was, through the factor c (equation (6.16)); and that its strain rate would increase the thinner it was (equation (6.15)). These influences of thickness on strain rate and temperature would result in there being a thickness for which the velocity modified temperature would be a maximum, and the shear flow stress a minimum (provided T_{MOD} was above about 620 K for the example in Figure 6.11). He proposed that the thickness would take the value that would maximize T_{MOD}. This gives the band of values labelled 'Theory' in Figure 6.12(b). The predicted band lies about 50% above the observed one, sufficiently close to give validity to the proposal.

In Chapter 2 (Figure 2.22(a)), direct measurements of the variation of friction factor m with rake face temperature were presented, for turning a 0.45%C steel. Flow-zone thickness was not measured in those tests. However, if the experimental relationship shown in Figure 6.12(b) is assumed to hold, the data of Figure 2.22(a) can be converted to a dependence of $\sqrt{3}mk$ (or σ_{int}) on T_{MOD}. Figure 6.13 shows the result and compares it with the value of σ_o for a 0.45%C steel used by Oxley. The agreement between the two sets of data is better than in Figure 6.11, but not perfect. It could be made perfect by supposing the strain rate to be only one tenth of the assumed value (as could be the case if the chip velocity was not reduced to zero at the rake face). Or maybe it should not be perfect: it has been argued that the tests from which σ_o values are derived are not close enough to machining conditions and that equation (6.10) has not the proper form to model flow behaviour over large ranges of extrapolation (Chapter 7.4). These are points of detail still to be resolved. However, it is close enough to reinforce the proposition that the plateau friction stress in machining is the shear flow stress of the chip material at the strain, strain rate and temperature that prevails in the flow-zone; and that that is governed by the localization of shear caused by minimization of the flow stress in the flow-zone. This wording is preferred, rather than maximization of T_{MOD}, as possibly applying more generally to materials whatever is their exact functional dependence of flow stress on strain, strain rate and temperature.

Dealing with average values of strain rate and temperature at the rake face avoids the question of how these vary along the rake face. It is still an open question as to why there

is a plateau value of friction stress, considering the large variation of strain, strain rate and temperature from one end of the flow-zone to the other. However, one thing is certain for the development of numerical (such as finite element) methods that may answer that question: the finite element mesh must be sufficiently fine next to the rake face to be able to resolve details of the flow zone. Figure 6.12(b) gives, at least for carbon steels, guidance of how fine that is: less than one fifth of $(\kappa_{work} l/U_{chip})^{1/2}$, or down to a few micrometres at high cutting speeds and low feeds.

6.3.4 Summary

Oxley developed his primary and secondary shear modelling into an iterative scheme for predicting cutting forces and shear plane angles from variations of work material flow stress with strain, strain rate and temperature. It is fully described by Hastings *et al*. (1980) and in Oxley (1989).

His work has shown that, in the primary shear zone, flow stress variations can significantly alter resultant forces in both magnitude and direction (Figure 6.8). Additionally, in the secondary shear zone, it suggests that, at least for high speed machining of metals without free-machining additives, the plateau friction stress is closely linked to the way in which shear localization occurs in a narrow flow-zone next to the rake face.

To develop those observations in to a predictive scheme, he found it necessary to restrict the possibilities of free surface hydrostatic stress variation that slip-line field theory has shown to be possible (Figure 6.4). He then observed that the non-uniqueness of slip-line field modelling was removed. Oxley's scheme involves two restrictive assumptions: that the hydrostatic stress at the free surface of the primary shear zone is given by equation (6.17) and that the normal contact stress is uniform over the chip/tool contact area (the latter also implies a negligible elastic part of the contact length). The first ignores the variety allowed by slip-line field modelling (Figure 6.5(b)). Many experiments (and slip-line field modelling) show exceptions to the second assumption. However, the main importance of his work, not affected by this detail, is the removal of the non-uniqueness predicted by slip-line modelling. Only one of the range of allowed results of a slip-line model (for example Figure 6.3) will create the rake face temperatures and strain rates that result in the assumed rake face shear stress.

The challenge for machining mechanics is to combine these materials-led ideas with the insights given by slip-line field modelling, in order to remove the restrictive assumptions relating to hydrostatic stress variations. The complexity of the geometrical and materials interactions is such that fundamental (as opposed to empirical) studies of the machining process require numerical, finite element, tools.

6.4 Non-orthogonal (three-dimensional) machining

Sections 6.2 and 6.3 have considered mechanics and materials issues in modelling the machining process, in orthogonal (two-dimensional or plane strain) conditions. This is sufficient for understanding the basic processes and physical phenomena that are involved. However, most practical machining is non-orthogonal (or three-dimensional): a comprehensive extension to this condition is necessary for the full benefits of modelling to be realized. Many published accounts of three-dimensional effects have considered special cases, using elementary geometry as their tools (Shaw *et al*., 1952; Zorev, 1966; Usui *et al*., 1978; Usui and Hirota, 1978; Arsecularatne *et al*., 1995). This section introduces the further

complexity of three-dimensional geometry in a more general manner than before, based on linear algebra.

Three-dimensional aspects of machining were briefly mentioned in Chapter 2 (Section 2.2.1 and Figure 2.2). Some basic terms like cutting edge approach angle, inclination angle and tool nose radius were introduced. The difference between feed and depth of cut (set by machine tool movements) and uncut chip thickness and cutting edge engagement length (related parameters, from the point of view of chip formation) was also explained. In this book, the term *feed* is generally used for both feed and uncut chip thickness, and *depth of cut* is used for both depth of cut and cutting edge engagement length. This section is the main part in which feed and depth of cut are used, properly, only to describe the parameters set by the machine tool.

6.4.1 An overview

The main feature (introduced in Section 2.2.1) of non-orthogonal machining is that the chip's direction of flow over the rake face is not normal to the cutting edge, but at some angle η_c to the normal, measured in the plane of the rake face. A second feature (not considered in Section 2.2.1) is that usually the uncut chip thickness varies along the cutting edge; and then the chip cross-section is not rectangular. This occurs whenever the nose radius of the cutting tool is engaged in cutting. Figure 6.14(a) shows both these features, as well as defining the chip flow direction η_c as positive when rotated clockwise from the normal to the cutting edge. It also shows the cutting, feed and depth of cut force components, F_c, F_f and F_d, of the work on the tool. If it is assumed that all parts of the chip are travelling with the same velocity, U_{chip}, (i.e. that there is no straining or twisting in the chip) then all material planes containing U_{chip} and the cutting velocity U_{work} are parallel to each other. The figure shows two such planes (hatched). The area of the planes decreases from D to A, where A and D lie at the extremities of the cutting edge engaged with the work.

Figure 6.14(b) shows any one of the hatched planes, simplified to a shear plane model of the machining process. The particular value of the uncut chip thickness is t_{1e} and the accompanying chip thickness is t_{2e}. The subscript e stands for effective and emphasizes that the plane of the figure is the U_{work}–U_{chip} plane. The rake angle in this plane, α_e, differs from that in the plane normal to the cutting edge. However, the condition that η_c is the same for every plane determines that so is α_e; and the condition that U_{chip} is the same on every plane requires that the effective shear plane angle ϕ_e is also the same on every plane. Equations (2.2) to (2.4) for orthogonal machining, in Chapter 2, can be extended to the circumstance of Figure 6.14(b) to give, after slight manipulation,

$$\tan \phi_e = \frac{\cos \alpha_e}{(t_{2e}/t_{1e}) - \sin \alpha_e} \tag{6.17a}$$

$$U_{chip} = \frac{\sin \phi_e}{\cos(\phi_e - \alpha_e)} U_{work} \tag{6.17b}$$

$$U_{primary} = \frac{\cos \alpha_e}{\cos(\phi_e - \alpha_e)} U_{work} \tag{6.17c}$$

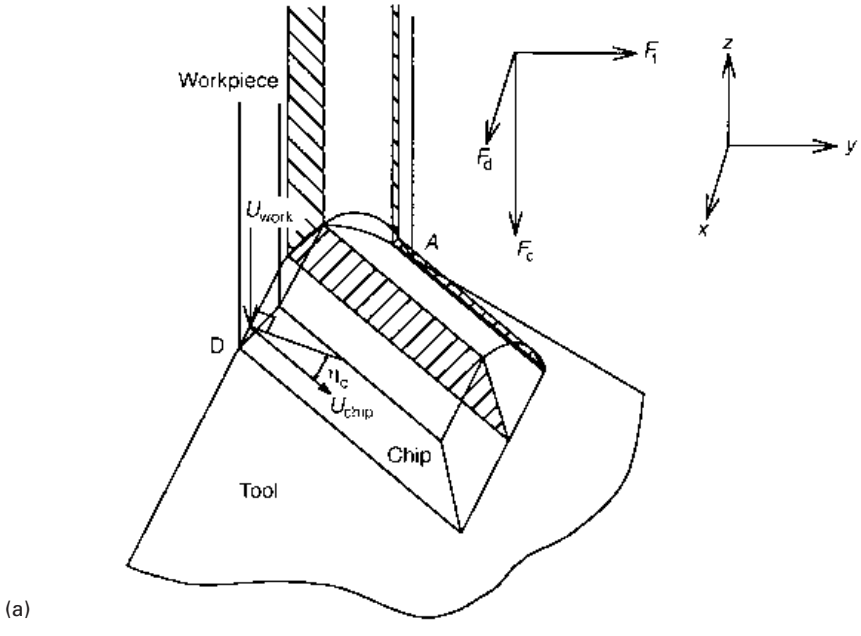


Fig. 6.14 (a) A three-dimensional cutting model showing (hatched) planes containing the cutting and chip velocity and (b) a shear plane model on one of those planes

$$\gamma = \cot \phi_e + \tan(\phi_e - \alpha_e) \tag{6.17d}$$

Furthermore, resolution of the three force components F_c, F_f and F_d in the direction of primary shear, and division by the primary shear surface area, gives the primary shear stress, as in orthogonal machining. However, the direction of primary shear depends not only on ϕ_e but also on η_c and the tool geometry. When, in addition, t_{1e} varies along the cutting edge, the primary shear surface is curved: consequently, its area can be difficult to

calculate. The description of machining, after the manner of Chapter 2, is inherently more complicated in the three-dimensional case.

The main independent variables, for a given tool geometry are η_c and ϕ_e. There are two basic ways to determine them, either from experiment (the descriptive manner of Chapter 2) or by prediction, both described in principle as follows.

Experimental analysis of three-dimensional machining

If the three force components F_c, F_f and F_d are measured, and resolved into components in the plane of, and normal to, the rake face of the tool, η_c can be obtained from the condition that the chip flow direction is opposed to the direction of the resultant (friction) force in the plane of the rake face. α_e can then be determined from η_c and the tool geometry. Equation (6.17a) can then be used to determine ϕ_e from the measurement of chip thickness. When the chip thickness varies along the cutting edge, a modification of the equation must be used

$$\tan \phi_e = \frac{\cos \alpha_e}{A_{fc}/A_{uc} - \sin \alpha_e} \tag{6.18}$$

where A_{fc} and A_{uc} are, respectively, the cross-sectional areas of the formed and uncut chip; and A_{fc} must be measured (for example by weighing a length of chip and dividing by the length and the chip material's density). Once η_c and ϕ_e are known, they may be used, with the tool geometry and the set feed and depth of cut, to estimate the primary shear plane area A_{sh}; the shear force F_{sh} on the shear plane may be calculated from the measured force components; and the shear stress τ_{sh} obtained from F_{sh}/A_{sh}. Other quantities may then be derived; for example, the work per unit volume on material flowing through the primary shear plane, for estimating the primary shear temperature rise, is $\tau_{sh}\gamma$.

Prediction in three-dimensional machining

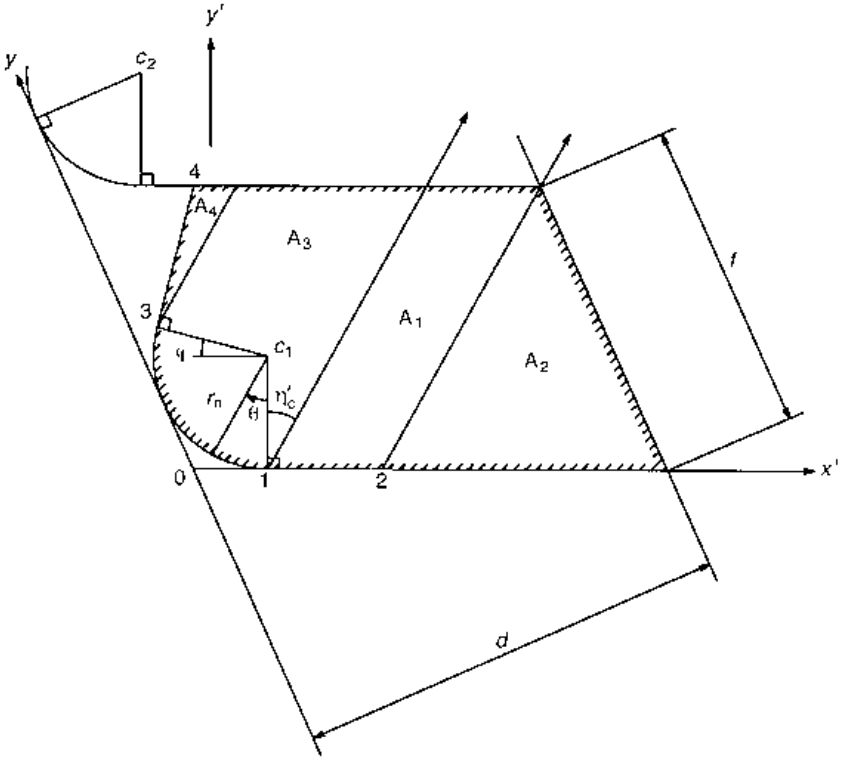
The earliest attempts at prediction in three-dimensional machining concentrated on η_c. Stabler (1951) suggested that η_c should equal the cutting edge inclination angle λ_s (defined in Figure 2.2 and more rigorously in Section 6.4.2); this is a first approximation. As seen later, it is not well supported by experiment. A better idea, based on geometry and due to Colwell (1954), is that, in a view normal to U_{work}, the chip will flow at right angles to the line AD joining the extremities of the cutting edge engagement (Figure 6.14(a)).

The best agreement with experiment, short of complete three-dimensional analyses *ab initio* (which hardly exist yet), is obtained by regarding the three-dimensional circumstance as a perturbation of orthogonal machining at the same feed, depth of cut and cutting speed (for example Usui *et al*., 1978; Usui and Hirota, 1978). In such an approach, the effective rake angle (α_e) is recognized to change with η_c. It is supposed that the friction angle λ, ϕ_e and τ_{sh} (and, in Usui's case, the rake face friction force per uncut chip area projected on to the rake face, F_{fric}/A_{uf}) are the same functions of α_e in three-dimensional machining as they are of α in orthogonal machining. These functions are determined either by orthogonal machining experiments or simulations. Finally, η_c is obtained as the value that minimizes the energy of chip formation under the constraints of the just described dependencies of λ, ϕ_e, τ_{sh} and F_{fric}/A_{uf} on α_e. This approach, in which both η_c and ϕ_e are obtained – although empirical in its minimum energy assumption – is a practical way to extend orthogonal modelling to three-dimensions.

A range of cases

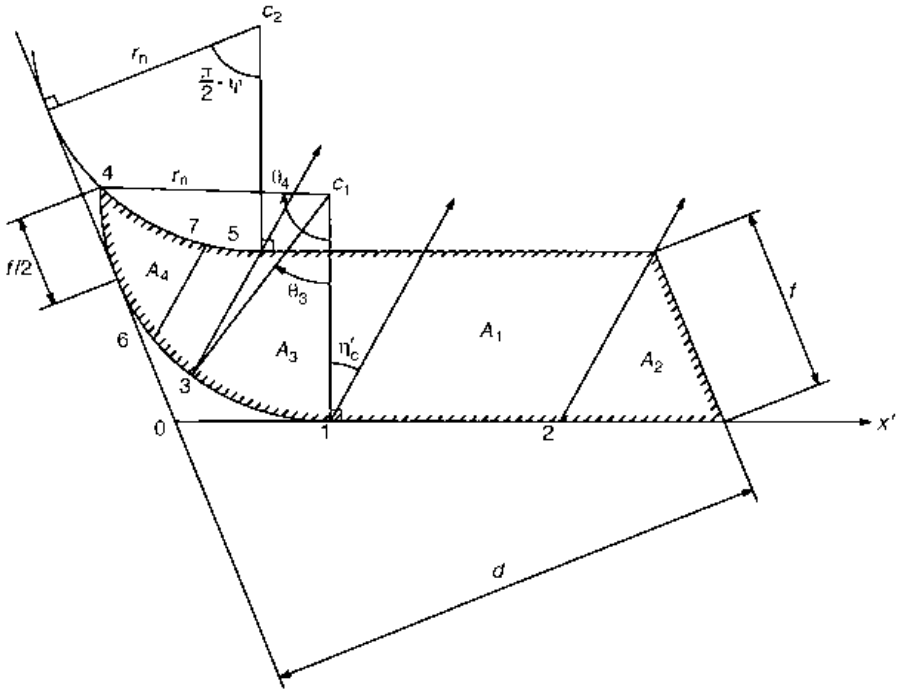
As the relative sizes of the feed, depth of cut and tool nose radius change, the shape of the uncut chip cross-section changes. Figure 6.15 shows four examples for the turning process, with which many engineers and certainly all tool engineers are familiar, but which could represent any process, as discussed in Chapter 2. The hatched areas are the uncut chip areas projected onto a plane normal to the cutting velocity. The directions and size of the feed and depth of cut are marked. Points such as 1 and 2 lie on the major cutting edge; and 3 and 4 on the tool nose radius or the minor cutting edge. Figure 6.15(a) is a case in which both the feed and depth of cut are large compared with the tool's nose radius; in Figure 6.15(b), the feed is becoming small compared with the nose radius, but the depth of cut remains large; in Figure 6.15(c), the depth of cut is reducing; and in Figure 6.15(d), machining is confined entirely to the nose radius region. The different cross-section shapes in these cases lead to different detail in estimating the shear plane and other areas. The further detail in the figures is concerned with this and is returned to later.

Different combinations of tool cutting edge approach and inclination angles, and rake face rake angles, lead to further variety in considering special cases. Formulae for use in three-dimensional analyses, for handling this wide range of variety, both in tool angular values and linear dimensions of the uncut chip, are derived in Sections 6.4.2 to 6.4.7, before their applications are considered in Section 6.4.8.

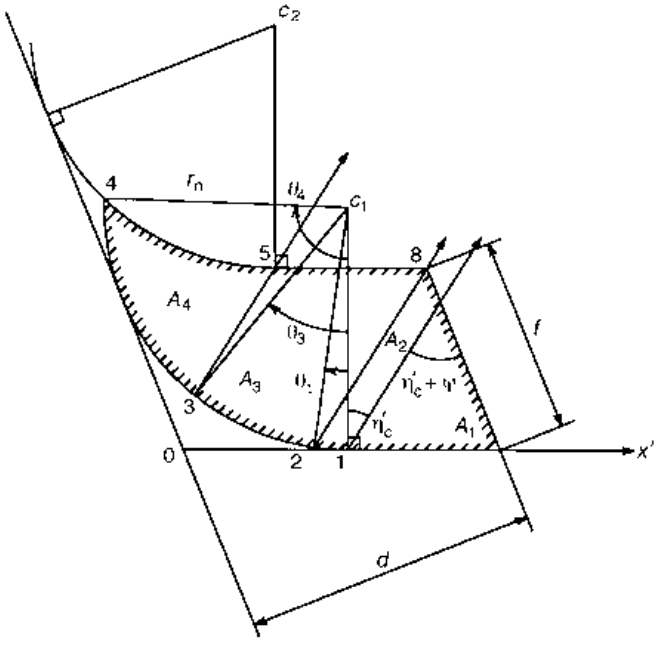


(a)

Fig. 6.15 Uncut chip cross-sections in single point turning: (a) case 1, (b) case 2, (c) case 3 and (d) case 4

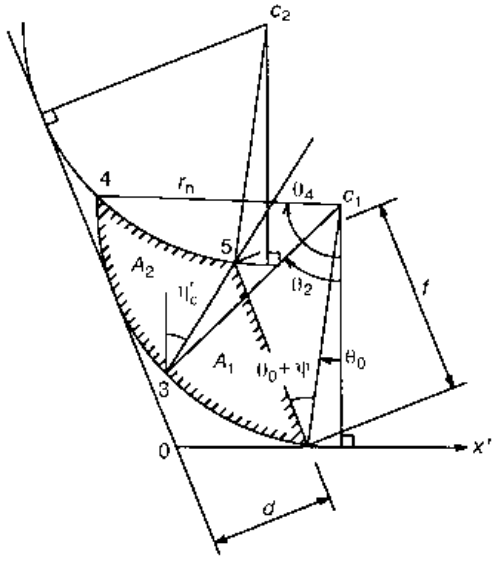


(b)



(c)

Fig. 6.15 *continued*



(d)

Fig. 6.15 *continued*

6.4.2 Tool geometry

Figure 6.16 shows plans and elevations, and defines tool angles, of a plane rake face turning tool oriented in a lathe. The treatment here is in terms of that, but (as has just been written) the results may be applied to any other machining process. O*A is parallel to the depth of cut direction and O*B to the feed direction of the machine tool (the cutting velocity direction O*C is normal to both O*A and O*B). The cutting tool has major (or side) and minor (or end) cutting edges which, in projection onto the O*AB plane, are inclined at the approach angles ψ and κ'_r to O*A and O*B, as shown (ψ here is $\pi/2$ minus the major cutting edge angle κ_r introduced in Chapter 2). In addition, the tool has a nose radius r_n, also measured in the O*AB plane. The slope of the rake face is determined by the angles, α_f (side rake), that the intersection of the rake face with the plane through O* normal to O*A makes with O*B and, α_p (back rake), that the intersection of the rake face with the plane through O* normal to O*B makes with O*A. Clearance angles γ_f and γ_p and the sign conventions for the angles, + or – as indicated, are also defined.

The figure also shows other views, defining other commonly described tool angles with their sign conventions. The major cutting edge inclination angle λ_s (already introduced but included here for completeness) is the direction between the major cutting edge and the normal to the cutting velocity in the plane containing the major cutting edge and the cutting velocity. The normal rake angle α_n is the angle, in the plane normal to the cutting edge, between the intersection with that plane of the rake face and the normal to the plane containing the cutting edge and the cutting velocity. Finally, the orthogonal rake angle α_o is similarly defined to α_n, but in the plane normal to the projection of the major cutting edge in the plane O*AB.

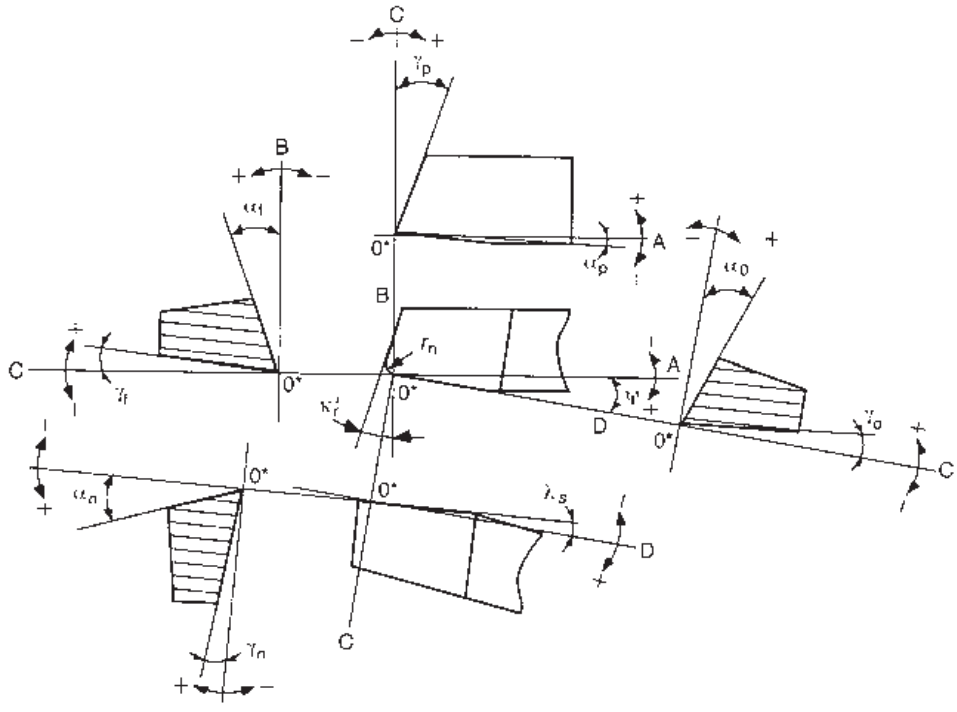


Fig. 6.16 Single point cutting tool geometry

6.4.3 Coordinate systems

The analysis of three-dimensional machining is aided by the introduction of six Cartesian coordinate systems, all with the same origin O (O is not the same as O*) at the intersection of the major and minor cutting edges. These systems may be written (x, y, z), (x', y', z'), (X, Y, Z), (X', Y', Z'), (ξ, η, ζ) and (ξ', η', ζ') and are defined in Figures 6.17. Transformations between the first four aid the analysis of cutting geometry and all of them are useful for force analysis.

The (x, y, z) system (Figure 6.17(a))

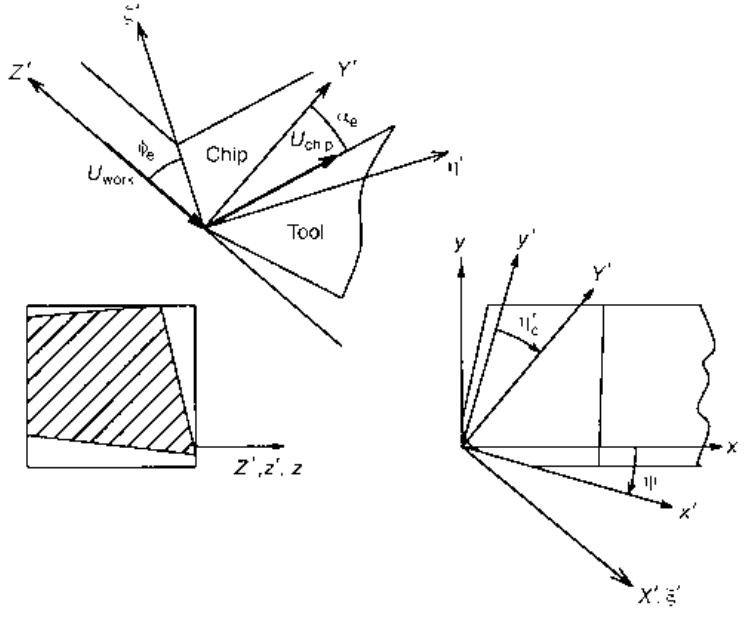
This system is aligned to major directions in the machine tool, with x directed opposite to the depth of cut, y opposite to the feed and z in the direction of the cutting velocity (the workpiece is supposed to move towards the stationary tool, with cutting velocity U_{work} in the $-z$ direction).

The (x′, y′, z′) system (Figure 6.17(a))

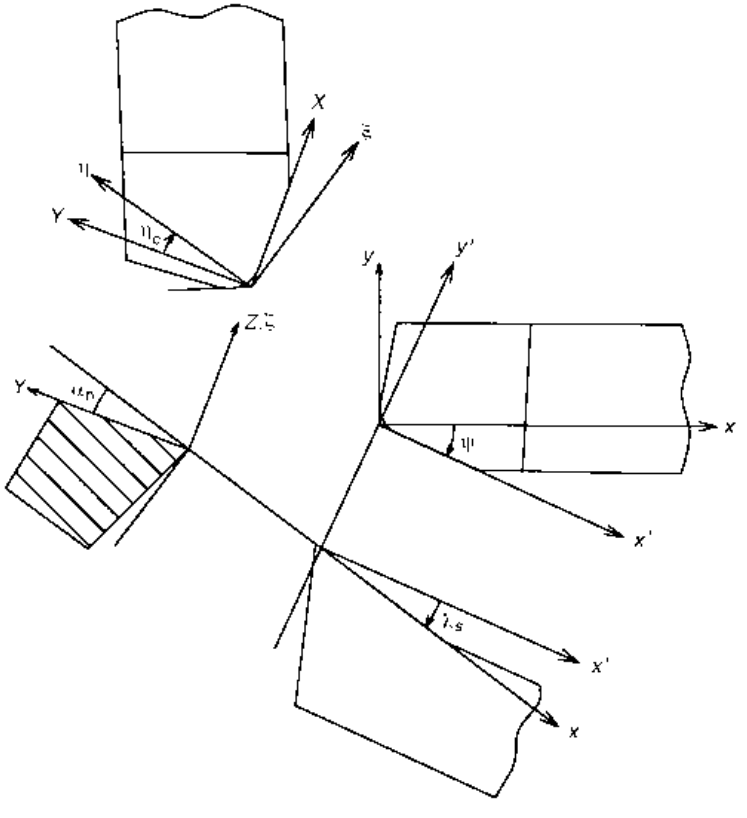
This is obtained by a clockwise rotation of (x, y, z) about z, by the amount ψ. It serves to link cutting tool and machine tool centred points of view. The coordinate transformation from (x', y', z') to (x, y, z) may be written as

$$\mathbf{x} = \mathbf{L}_1 \mathbf{x}' \tag{6.19a}$$

where $\mathbf{x}$ and $\mathbf{x}'$ are position vectors in the (x, y, z) and (x', y', z') systems and



(a)



(b)

Fig. 6.17 (a) (x, y, z), (x', y', z'), (X', Y', Z') and (ξ', η', ζ') and (b) (x', y', z'), (X, Y, Z) and (ξ, η, ζ) coordinate systems

$$\mathbf{L}_1 = \begin{bmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{6.19b}$$

In terms of the inverse or transposed matrices $\mathbf{L}_1^{-1}$ or $\mathbf{L}_1^{\mathbf{T}}$ respectively, the inverse transform is

$$\mathbf{x}' = \mathbf{L}_1^{-1}\,\mathbf{x} = \mathbf{L}_1^{\mathbf{T}}\mathbf{x} \tag{6.19c}$$

The (X, Y, Z) system (Figure 6.17(b))

In this cutting tool centred system, X lies along the major cutting edge, Y is in the plane of, and Z is normal to, the rake face. The transformation from (X, Y, Z) to (x', y', z') is accomplished in two stages, first by rotating (X, Y, Z) about the X-axis by the amount α_n, then about the y' axis by the amount λ_s:

$$\mathbf{x}' = \mathbf{L}_2\mathbf{X} \tag{6.20a}$$

where

$$\left.\begin{aligned} \mathbf{L}_2 = \mathbf{L}_{21}\mathbf{L}_{22} &= \begin{bmatrix} \cos\lambda_s & 0 & \sin\lambda_s \\ 0 & 1 & 0 \\ -\sin\lambda_s & 0 & \cos\lambda_s \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha_n & \sin\alpha_n \\ 0 & -\sin\alpha_n & \cos\alpha_n \end{bmatrix} \\ &\equiv \begin{bmatrix} \cos\lambda_s & -\sin\lambda_s\sin\alpha_n & \sin\lambda_s\cos\alpha_n \\ 0 & \cos\alpha_n & \sin\alpha_n \\ -\sin\lambda_s & -\cos\lambda_s\sin\alpha_n & \cos\lambda_s\cos\alpha_n \end{bmatrix} \end{aligned}\right\} \tag{6.20b}$$

The (X′, Y′, Z′) system (Figure 6.17(a))

This system is the first to be introduced here from the point of view of chip formation. Z' is parallel to z and z', still in the cutting direction, but X' is normal to, and Y' is in, the plane containing the cutting and chip velocities. In terms of the chip flow direction projected in the x'–y' plane, defined by η_c' (different from but related to η_c) the transformation from (X', Y', Z') to (x', y', z') is

$$\mathbf{x}' = \mathbf{L}_3\mathbf{X}' \tag{6.21a}$$

where

$$\mathbf{L}_3 = \begin{bmatrix} \cos\eta_c' & \sin\eta_c' & 0 \\ -\sin\eta_c' & \cos\eta_c' & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{6.21b}$$

The (ξ, η, ζ) system (Figure 6.17(b))

This system is also concerned with chip flow. It is obtained by clockwise rotation of the (X, Y, Z) frame about Z by the amount of the chip flow direction η_c. The η direction is then parallel to the chip flow direction, in the plane of the rake face. Transformation from (ξ, η, ζ) to (X, Y, Z) is

$$\mathbf{X} = \mathbf{L}_4\mathbf{X}_\xi \tag{6.22a}$$

where

$$\mathbf{L}_4 = \begin{bmatrix} \cos\eta_c & \sin\eta_c & 0 \\ -\sin\eta_c & \cos\eta_c & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{6.22b}$$

The (ξ', η', ζ') system (Figure 6.17(a))

Finally, clockwise rotation of the (X', Y', Z') frame about X' by the amount of the effective shear angle ϕ_e gives a system in which ξ' remains normal to the plane containing the cutting and chip velocities and ζ' lies in the shear plane. To transform from (ξ', η', ζ') to (X', Y', Z'),

$$\mathbf{X}' = \mathbf{L}_5 \mathbf{X}_{\xi'} \tag{6.23a}$$

where

$$\mathbf{L}_5 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi_e & \sin\phi_e \\ 0 & -\sin\phi_e & \cos\phi_e \end{bmatrix} \tag{6.23b}$$

6.4.4 Relations between tool and chip flow angles

In the three-dimensional cutting model described in Section 6.4.1, the chip flow direction η_c and the effective shear angle ϕ_e are the basic independent variables for a given tool geometry and cutting conditions. Key dependent parameters used in their determination are the effective rake angle α_e and the chip flow direction η'_c in the x'–y' plane. In this section the dependence of α_e and η'_c on η_c and tool geometry, characterized by the normal rake α_n and the cutting edge inclination angle λ_s is first derived. Conversions between α_n and other measures of tool rake are then developed.

Dependence of α_e and η'_c on η_c, α_n and λ_s

The unit vector in the chip flow direction may be expressed in two different ways in the $\mathbf{x}'$ coordinate system to obtain the required relationships. In a notation $\mathbf{a}(\mathbf{b})$, which expresses a vector $\mathbf{a}$ in the $\mathbf{b}$ coordinate system, the unit vector in the chip flow direction $\mathbf{e}_\eta$ may be expressed in the $\mathbf{X}$ and $\mathbf{X}'$ systems as

$$\left.\begin{aligned} \mathbf{e}_\eta(\mathbf{X}) &= \begin{Bmatrix} \sin\eta_c \\ \cos\eta_c \\ 0 \end{Bmatrix} \\ \mathbf{e}_\eta(\mathbf{X}') &= \begin{Bmatrix} 0 \\ \cos\alpha_e \\ -\sin\alpha_e \end{Bmatrix} \end{aligned}\right\} \tag{6.24}$$

and these may be transformed to $\mathbf{e}_\eta(\mathbf{x}')$ respectively as

$$\mathbf{e}_\eta(\mathbf{x}') = \mathbf{L}_2 \mathbf{e}_\eta(\mathbf{X}) = \begin{Bmatrix} \cos\lambda_s \sin\eta_c - \sin\lambda_s \sin\alpha_n \cos\eta_c \\ \cos\alpha_n \cos\eta_c \\ -\sin\lambda_s \sin\eta_c - \cos\lambda_s \sin\alpha_n \cos\eta_c \end{Bmatrix} \tag{6.25a}$$

$$\mathbf{e}_\eta(\mathbf{x}') = \mathbf{L}_3\mathbf{e}_\eta(\mathbf{X}') = \begin{Bmatrix} \sin\eta_c' \cos\alpha_e \\ \cos\eta_c' \cos\alpha_e \\ -\sin\alpha_e \end{Bmatrix} \tag{6.25b}$$

Equating the components of equations 6.25(a) and (b) leads to the required results:

$$\sin\alpha_e = \sin\lambda_s \sin\eta_c + \cos\lambda_s \sin\alpha_n \cos\eta_c \tag{6.26a}$$

$$\tan\eta_c' = (\cos\lambda_s \sin\eta_c - \sin\lambda_s \sin\alpha_n \cos\eta_c)/(\cos\alpha_n \cos\eta_c) \tag{6.26b}$$

Relations between tool rake and cutting edge angles

Transformation of the unit vector normal to the rake face, originally expressed in the **X** coordinate system, into its **x** and **x′** forms, leads to the required relations (using $\mathbf{x} = \mathbf{L}_1\mathbf{L}_2\mathbf{X}$ for the first case):

$$\mathbf{e}_\mathbf{Z}(\mathbf{x}) = \mathbf{L}_1\mathbf{L}_2\mathbf{e}_\mathbf{Z}(\mathbf{X}) = \mathbf{L}_1\mathbf{L}_2 \begin{Bmatrix} 0 \\ 0 \\ 1 \end{Bmatrix} = \begin{Bmatrix} \cos\psi \sin\lambda_s \cos\alpha_n + \sin\psi \sin\alpha_n \\ -\sin\psi \sin\lambda_s \cos\alpha_n + \cos\psi \sin\alpha_n \\ \cos\lambda_s \cos\alpha_n \end{Bmatrix} \tag{6.27a}$$

$$\mathbf{e}_\mathbf{Z}(\mathbf{x}') = \mathbf{L}_2\mathbf{e}_\mathbf{Z}(\mathbf{X}) = \mathbf{L}_2 \begin{Bmatrix} 0 \\ 0 \\ 1 \end{Bmatrix} = \begin{Bmatrix} \sin\lambda_s \cos\alpha_n \\ \sin\alpha_n \\ \cos\lambda_s \cos\alpha_n \end{Bmatrix} \tag{6.27b}$$

Because $\tan\alpha_p$ and $\tan\alpha_f$ are, respectively, the ratios of the x to z and y to z components of $\mathbf{e}_\mathbf{Z}(\mathbf{x})$ and $\tan\alpha_o$ is the ratio of the y' to z' components of $\mathbf{e}_\mathbf{Z}(\mathbf{x}')$, from equations 6.27(a) and (b):

$$\tan\alpha_p = \cos\psi \tan\lambda_s + \sin\psi \tan\alpha_n/\cos\lambda_s \tag{6.28a}$$

$$\tan\alpha_f = -\sin\psi \tan\lambda_s + \cos\psi \tan\alpha_n/\cos\lambda_s \tag{6.28b}$$

$$\tan\alpha_o = \tan\alpha_n/\cos\lambda_s \tag{6.28c}$$

while further inversion and substitution results in

$$\tan\lambda_s = \cos\psi \tan\alpha_p - \sin\psi \tan\alpha_f \tag{6.28d}$$

$$\tan\alpha_n/\cos\lambda_s = \sin\psi \tan\alpha_p + \cos\psi \tan\alpha_f \tag{6.28e}$$

6.4.5 Force relationships

The resultant cutting force **F** is most commonly reported by its cutting force F_c, feed force F_f and thrust force F_d components in the (x, y, z) coordinate system, as shown in Figure 6.14 (and also earlier in Figure 5.7). However, the components in the (ξ, η, ζ) and (ξ', η', ζ') frames are more fundamental to chip formation. In the former, the rake face friction force F_{fric} is F_η, the rake face normal force F_N is $-F_\zeta$ and, because friction force is parallel to the direction of motion, $F_\xi = 0$; in the latter, the shear force F_{sh} on the primary shear surface is $-F_{\zeta'}$.

The coordinate transformation

$$\mathbf{F}(\mathbf{x}) = \mathbf{L}_1\mathbf{L}_2\mathbf{L}_4\mathbf{F}(\mathbf{X}_\xi) \tag{6.29}$$

with $\mathbf{F}^T(\mathbf{x}) = \{F_d, F_f, -F_c\}$ and $\mathbf{F}^T(\mathbf{X}_\xi) = (0, F_{fric}, -F_N)$, may be used to derive (after substitution of equation (6.26b) to eliminate η_c in favour of η_c')

$$\begin{Bmatrix} F_d \\ F_f \\ F_c \end{Bmatrix} = \mathbf{L}_6 \begin{Bmatrix} F_{fric} \\ F_N \end{Bmatrix} \tag{6.30a}$$

where

$$\mathbf{L}_6 = \begin{bmatrix} (\cos\psi \tan\eta'_c + \sin\psi)\cos\alpha_n \cos\eta_c & -\cos\psi \sin\lambda_s \cos\alpha_n - \sin\psi \sin\alpha_n \\ (\cos\psi - \sin\psi \tan\eta'_c)\cos\alpha_n \cos\eta_c & -\cos\psi \sin\alpha_n + \sin\psi \sin\lambda_s \cos\alpha_n \\ \sin\alpha_e & \cos\lambda_s \cos\alpha_n \end{bmatrix} \tag{6.30b}$$

The inverse transformation of equation (6.29) with $F_\xi = 0$ enables the chip flow direction and then the rake face friction and normal forces to be obtained from F_c, F_f, F_d and the tool rake and cutting edge angles:

$$\tan\eta_c = \frac{(F_d \cos\psi - F_f \sin\psi)\cos\lambda_s + F_c \sin\lambda_s}{(F_d \sin\psi = F_f \cos\psi)\cos\alpha_n + \{(F_f \sin\psi - F_d \cos\psi)\sin\lambda_s + F_c \cos\lambda_s\}\sin\alpha_n} \tag{6.31a}$$

$$\begin{Bmatrix} F_{fric} \\ F_N \end{Bmatrix} = \mathbf{L}^T{}_6 \begin{Bmatrix} F_d \\ F_f \\ F_c \end{Bmatrix} \tag{6.31b}$$

Transformation of the cutting force components between the (x, y, z) and (ξ', η', ζ') frames is achieved by

$$\begin{Bmatrix} F_{\xi'} \\ F_{\eta'} \\ F_{\zeta'} \end{Bmatrix} = \mathbf{L}^T{}_5 \mathbf{L}^T{}_3 \mathbf{L}^T{}_1 \begin{Bmatrix} F_d \\ F_f \\ -F_c \end{Bmatrix} = \mathbf{L}^T{}_7 \begin{Bmatrix} F_d \\ F_f \\ -F_c \end{Bmatrix} \tag{6.32a}$$

where

$$\mathbf{L}^T{}_7 = \begin{bmatrix} \cos(\eta'_c + \psi) & -\sin(\eta'_c + \psi) & 0 \\ \cos\phi_e \sin(\eta'_c + \psi) & \cos\phi_e \cos(\eta'_c + \psi) & -\sin\phi_e \\ \sin\phi_e \sin(\eta'_c + \psi) & \sin\phi_e \cos(\eta'_c + \psi) & \cos\phi_e \end{bmatrix} \tag{6.32b}$$

Then, since $F_{sh} = -F_{\zeta'}$ and the shear stress τ_{sh} on the shear plane is this divided by the shear surface area A_{sh}

$$F_{sh} = -\{F_d \sin(\eta'_c + \psi) + F_f \cos(\eta'_c + \psi)\}\sin\phi_e + F_c \cos\phi_e \tag{6.33a}$$

$$\tau_{sh} = F_{sh}/A_{sh} \tag{6.33b}$$

6.4.6 Shear surface area relations

In terms of l_ϕ, the length of the shear surface along the shear direction on a cutting velocity–chip velocity plane, and of $\mathbf{e}_\phi$ and $\mathbf{e}_S$, the unit vectors along the shear direction and cutting edge respectively, the area A_{sh} of the curved shear surface is obtained by integration along the cutting edge S:

$$A_{sh} = \int_S \| \mathbf{e}_\phi \times \mathbf{e}_S \| \, l_\phi \mathrm{d}S \tag{6.34a}$$

where ‖ ‖ is the norm of the vector. However, the cutting edge shape, through ψ, κ'_r and r_n, is defined on the x–y or x'–y' plane: amongst these, x' is parallel to the major cutting edge direction. It is therefore convenient to carry out the integration in the (x', y', z') coordinate system, along the projected infinitesimal cutting edge length dS' in the x'–y' plane:

$$A_{sh} = \int_{S'} \| \mathbf{e}_\phi \times \mathbf{e}_S \| \, l_\phi \frac{dS}{dS'} \, dS' \tag{6.34b}$$

Evaluating A_{sh}: l_ϕ *and* $\mathbf{e}_\phi(\mathbf{x}')$

Expressions for l_ϕ and $\mathbf{e}_\phi(\mathbf{x}')$ are simplified by the assumption that the effective shear plane angle ϕ_e is a constant, the same on every cutting velocity–chip velocity plane, for given cutting conditions. In terms of the effective uncut chip thickness t_{1e} (Figure 6.14(b)),

$$l_\phi = \frac{t_{1e}}{\sin \phi_e} \tag{6.35}$$

$\mathbf{e}_\phi$ in the $\mathbf{x}'$ coordinate system is obtained from the fact that, in the $\mathbf{X}'$ coordinate system, it has components $(0, \sin\phi_e, \cos\phi_e)$. Thus

$$\mathbf{e}_\phi(\mathbf{x}') = \mathbf{L}_3 \mathbf{e}_\phi(\mathbf{X}') = \mathbf{L}_3 \begin{Bmatrix} 0 \\ \sin \phi_e \\ \cos \phi_e \end{Bmatrix} = \begin{Bmatrix} \sin \eta'_c \sin \phi_e \\ \cos \eta'_c \sin \phi_e \\ \cos \phi_e \end{Bmatrix} \tag{6.36}$$

Evaluating A_{sh}: $\mathbf{e}_S(\mathbf{x}')$ (dS/dS')

The unit vector along the x'–y' plane projection of the cutting edge, in general makes some angle θ with x' and may be written $\mathbf{e}_\theta(\mathbf{x}')$, with (x', y', z') components $(\cos\theta, -\sin\theta, 0)$. Along the major cutting edge, $\theta = 0$; along the minor cutting edge, $\theta = \pi/2 + (\kappa'_r - \psi)$; and round the nose radius region θ varies from one to the other. Because $\mathbf{e}_\theta dS'$ is the projection on the x'–y' plane of $\mathbf{e}_S dS$,

$$\mathbf{e}_S(\mathbf{x}') \frac{dS}{dS'} = \mathbf{e}_\theta(\mathbf{x}') + c\mathbf{e}_{z'}(\mathbf{x}') = \begin{Bmatrix} \cos \theta \\ -\sin \theta \\ c \end{Bmatrix} \tag{6.37}$$

where $\mathbf{e}_{z'}$ is the unit vector in the z' direction and c is a constant that may be found from the condition that, because $\mathbf{e}_S$ lies in the X–Y plane, the Z component of $\mathbf{e}_S$ is zero. In the (X, Y, Z) coordinate system

$$\frac{dS}{dS'} \mathbf{e}_S(\mathbf{X}) = \frac{dS}{dS'} \mathbf{L}_2^T \mathbf{e}_S(\mathbf{x}') =$$

$$\begin{Bmatrix} \cos \lambda_s \cos \theta - c \sin \lambda_s \\ -\sin \lambda_s \sin \alpha_n \cos \theta - \cos \alpha_n \sin \theta - c \cos \lambda_s \sin \alpha_n \\ \sin \lambda_s \cos \alpha_n \cos \theta - \sin \alpha_n \sin \theta + c \cos \lambda_s \cos \alpha_n \end{Bmatrix} \tag{6.38}$$

Hence

$$c = \frac{\tan \alpha_n \sin \theta - \sin \lambda_s \cos \theta}{\cos \lambda_s} \tag{6.39}$$

Evaluating A_{sh}: the result

Substitution of equations (6.35), (6.36) and (6.37) with (6.39) into equation (6.34b) and performing the vector multiplication leads to

$$A_{\text{sh}} = \int_{S'} A_\theta \frac{t_{1\text{e}}}{\sin \phi_{\text{e}}} \, \text{d}S' \tag{6.40a}$$

with

$$A_\theta^2 = \cos^2(\eta'_{\text{c}} - \theta) + \{c \sin \theta_{\text{e}} - \cos \theta_{\text{e}} \sin(\eta'_{\text{c}} - \phi)\}^2 \tag{6.40b}$$

Along the major cutting edge ($\theta = 0$), A_θ takes a constant value that may be written A_{major}; along the minor cutting edge ($\theta = \pi/2 + \varphi$, where $\varphi = \kappa'_{\text{r}} - \psi$) it takes another constant value that may be written A_{minor}.

Example calculations of A_{sh}

The functional dependence of A_θ and $t_{1\text{e}}$ on S' commonly changes along S'. The evaluation of A_{sh} is then accomplished by dividing the range of S' into as many intervals as there are different variations. If there are N such,

$$A_{\text{sh}} = \sum_{i=1}^{N} \int_{\lim_{1,i}}^{\lim_{2,i}} A_{\theta,i} \frac{t_{1\text{e},i}}{\sin \phi_{\text{e}}} \, \text{d}S'_i \tag{6.41}$$

A number of special cases have been introduced in Figure 6.15. In Figure 6.15(a), for example, there are four intervals indicated by A_1, A_2, A_3, A_4. In Table 6.1, the values of $\lim_{1,i}$, $\lim_{2,i}$, $A_{\theta,i}$, $t_{1\text{e},i}$ and the appropriate form of $\text{d}S'_i$ are listed for this example (case 1) as well as for the other three cases, 2 to 4, of Figures 6.15(b) to (d) respectively ($N = 4$ for cases 2 and 3, and $N = 2$ for case 4). In the table, the subscripts 1, 2, 3, etc indicate the positions 1, 2, 3, etc in Figure 6.15 at which the quantities are calculated, or else the subscripted quantities are defined in the figure. Further details of the values of the quantities are listed in Table 6.2; and how they are obtained follows next.

The values of x'_1, x'_2 and y'_3 in Table 6.2 are obtained from Figure 6.15(a), θ_4 from Figure 6.15(b) and θ_0 from Figure 6.15(d), by inspection. More explanation is needed of the values of θ_1, θ_2 and θ_3 and of $t_{1\text{e},\theta 3\text{–}4}$. In Figure 6.15(c), θ_1 is defined as the direction between the major cutting edge and the tangent to the cutting edge at point 2. Point 2 lies on the same chip flow line that passes through point 8; and the direction of all flow lines is given in the x'–y' plane by η'_{c}. Thus, θ_1 in Table 6.2 is found from

$$\tan \eta'_{\text{c}} = \frac{x'_8 - x'_2}{y'_8 - y'_2} = \frac{(d - r_{\text{n}} + r_{\text{n}} \sin \psi)/\cos \psi - f \sin \psi + r_{\text{n}} \sin \theta_1}{r_{\text{n}} \cos \theta_1 - r_{\text{n}} + f \cos \psi} \tag{6.42}$$

θ_2 and θ_3 are obtained by a similar argument. They are obtained from the conditions that points 3 and 5 in Figure 6.15(d) and Figure 6.15(c) respectively lie on the same chip flow line.

$t_{1\text{e},\theta 3\text{–}4}$, the effective uncut chip thickness between points 3 and 4 in Figure 6.15(b), is found from the condition that point 7 lies on the circle of radius r_{n} centred on C_2 and point 7 is displaced from point 6 by $t_{1\text{e}}$ in the chip flow direction. Thus

Table 6.1 Particular values of coefficients and variables in equation (6.41)

Case	i	dS'_i	$\lim_{1,i}$	$\lim_{2,i}$	$A_{\theta,i}$	$t_{1e,i}$
1	1	dx'	x'_1	x'_2	A_{major}	$\frac{f\cos\psi}{\cos\eta'_c}$
	2	dx'	x'_2	$d/\cos\psi$	A_{major}	$\left(\frac{f\cos\psi}{\cos\eta'_c}\right)\left(\frac{d/\cos\psi - x'}{d/\cos\psi - x'_2}\right)$
	3	$r_n d\theta$	0	$\pi/2+\varphi$	A_θ	$\frac{f\cos\psi - r_n + r_n\cos\theta}{\cos\eta'_c}$
	4	$\frac{dy'}{\cos\varphi}$	y'_3	$f\cos\psi$	A_{minor}	$\frac{f\cos\psi - y'}{\cos\eta'_c}$
2	1	dx'	x'_1	x'_2	A_{major}	$\frac{f\cos\psi}{\cos\eta'_c}$
	2	dx'	x'_2	$d/\cos\psi$	A_{major}	$\left(\frac{f\cos\psi}{\cos\eta'_c}\right)\left(\frac{d/\cos\psi - x'}{d/\cos\psi - x'_2}\right)$
	3	$r_n d\theta$	0	θ_3	A_θ	$\frac{f\cos\psi - r_n + r_n\cos\theta}{\cos\eta'_c}$
	4	$r_n d\theta$	θ_3	θ_4	A_θ	$t_{1e,\theta 3-4}$
3	1	dx'	x'_1	$d/\cos\psi$	A_{major}	$\left(\frac{f\cos\psi}{\cos\eta'_c}\right)\left(\frac{d/\cos\psi - x'}{d/\cos\psi - x'_2}\right)$
	2	$r_n d\theta$	0	θ_1	A_θ	$\frac{d - r_n + r_n\sin(\theta+\psi)}{\sin(\eta'_c+\psi)}$
	3	$r_n d\theta$	θ_1	θ_3	A_θ	$\frac{f\cos\psi - r_n + r_n\cos\theta}{\cos\eta'_c}$
	4	$r_n d\theta$	θ_3	θ_4	A_θ	$t_{1e,\theta 3-4}$
4	1	$r_n d\theta$	θ_0	θ_2	A_θ	$\frac{d - r_n + r_n\sin(\theta+\psi)}{\sin(\eta'_c+\psi)}$
	2	$r_n d\theta$	θ_2	θ_4	A_θ	$t_{1e,\theta 3-4}$

$$(x'_6 - x'_{C2} + t_{1e,\theta 3-4}\sin\eta'_c)^2 + (y'_6 - y'_{C2} + t_{1e,\theta 3-4}\cos\eta'_c)^2 = r_n^2 \qquad (6.43)$$

$t_{1e\theta 3-4}$ is obtained as the solution to the quadratic equation (6.43) after substituting

$$(x'_6 - x'_{C2}) = f\sin\psi - r_n\sin\theta \quad \text{and} \quad (y'_6 - y'_{C2}) = -f\cos\psi - r_n\cos\theta$$

6.4.7 Uncut chip cross-section areas A_{uc} and A_{uf}

The uncut chip cross-section area A_{uc} in the x'–y' plane, also required in the theory of three-dimensional machining, is

$$A_{uc} = \int_{S'} t_{1e}\cos(\eta'_c - \theta)\,dS' \qquad (6.44)$$

For the particular cases of Figure 6.15, it can be obtained geometrically, without integration. For case 1

Table 6.2 Values of the coefficients x'_1, x'_2, y'_3, θ_i (i = 0 to 4) and $t_{1e,\theta 3-4}$

x'_1	x'_2	y'_3
$\dfrac{r_n(1-\sin\psi)}{\cos\psi}$	$\dfrac{d}{\cos\psi} - f(\sin\psi + \cos\psi\tan\eta'_c)$	$r_n(1+\sin\varphi)$
θ_0	θ_1	
$\sin^{-1}\left\{\dfrac{r_n - d}{r_n}\right\} - \psi$	$\eta'_c + \sin^{-1}\left\{\dfrac{f}{r_n}\sin(\psi+\eta'_c) - \dfrac{(d-r_n)\cos\eta'_c + r_n\sin(\psi+\eta'_c)}{r_n\cos\psi}\right\}$	
θ_2	θ_3	
$\eta'_c + \sin^{-1}\left\{\dfrac{f}{r_n}\sin(\psi+\eta'_c) - \sin(\theta_0 - \eta'_c)\right\}$	$\eta'_c + \sin^{-1}\left\{\dfrac{f}{r_n}\sin(\psi+\eta'_c) - \sin\eta'_c\right\}$	
θ_4	$t_{1e,\theta 3-4}$	
$\dfrac{\pi}{2} - \psi + \sin^{-1}\dfrac{f}{2r_n}$	$r_n\cos(\eta'_c - \theta) + f\cos(\eta'_c + \psi)$ $- \left\{\{r_n\cos(\eta'_c - \theta) + f\cos(\eta'_c + \psi)\}^2 - f^2 - 2r_n f\cos(\psi+\theta)\right\}^{1/2}$	

$$A_{uc} = fd + r_n^2\left(\frac{\pi}{4} + \frac{\varphi}{2} - \frac{1+\sin\varphi}{\cos\varphi}\right) + r_n f\left(\frac{\cos\psi + \sin\kappa'_r}{\cos\varphi} - 1\right) - \frac{f^2\sin\kappa'_r\cos\psi}{2\cos\varphi} \tag{6.45a}$$

while for cases 2, 3 and 4

$$A_{uc} = (d - r_n)f + r_n^2\sin^{-1}\frac{f}{2r_n} + \frac{1}{4}f\left(4r_n^2 - f^2\right)^{1/2} \tag{6.45b}$$

A_{uf}, the projection onto the rake face, along the cutting direction, of the uncut chip cross-section area is readily shown to be the division of A_{uc} by the z' component of $\mathbf{e}_Z(\mathbf{x}')$ in equation (6.27b):

$$A_{uf} = \frac{A_{uc}}{\cos\lambda_s\cos\alpha_n} \tag{6.46}$$

6.4.8 Predictions from three-dimensional models

The relations from the previous sections may finally be used in the prediction of chip flow direction and cutting force components. Colwell's (1954) approach and the energy approach initiated by Usui (Usui *et al*. 1978; Usui and Hirota, 1978; also Usui, 1990), will particularly be developed and compared with experiments.

Colwell's geometrical model

The chip flow direction η'_c in the $x' - y'$ plane is assumed to be perpendicular to the projected chord AD joining the extremities of the cutting edge engagement (Figure 6.14) in the $x' - y'$ plane. It is readily found by trigonometry, for the four cases shown in Figure 6.15. For Case 1 (Figure 6.15(a))

$$\tan \eta'_c = \frac{f \cos^2 \psi \cos \varphi}{(d - r_n)\cos \varphi + r_n (\sin \kappa'_r + \cos \psi) - f \cos^2 \psi \sin \varphi} \tag{6.47a}$$

while for Cases 2, 3 and 4 (in terms of θ_4 given in Table 6.2)

$$\tan \eta'_c = \frac{r_n(1 - \cos \theta_4)\cos \psi}{d + r_n (\sin \psi + \sin \theta_4 \cos \psi - 1)} \tag{6.47b}$$

Then, from equation (6.26b),

$$\tan \eta_c = \frac{\cos \alpha_n \tan \eta'_c + \sin \lambda_s \sin \alpha_n}{\cos \lambda_s} \tag{6.48}$$

This result alone is not sufficient for predicting machining forces: shear plane prediction is required as well.

Usui's energy model

As introduced in Section 6.4.1, it is assumed that ϕ_e, τ_{sh}, λ and F_{fric}/A_{uf} are the same functions of α_e in three-dimensional machining as they are of α in orthogonal machining. From Chapter 2 (Section 2.2), in orthogonal conditions

$$\frac{F_{fric}}{A_{uf}} = \frac{\tau_{sh} \sin \lambda}{\cos(\phi + \lambda - \alpha)} \frac{\cos \alpha}{\sin \phi} \tag{6.49a}$$

Then, in three-dimensional conditions

$$F_{fric} = \frac{\tau_{sh} A_{uf} \sin \lambda}{\cos(\phi_e + \lambda - \alpha_e)} \frac{\cos \alpha_e}{\sin \phi_e} \tag{6.49b}$$

The friction work rate is $F_{fric}U_{chip}$ and the primary shear work rate is $\tau_{sh}A_{sh}U_{primary}$. After applying equations (6.17b) and (6.17c), the total work rate is

$$E_{cutting} = \left\{ A_{sh} \frac{\cos \alpha_e}{\cos(\phi_e - \alpha_e)} + A_{uf} \frac{\sin \lambda \cos \alpha_e}{\cos(\phi_e + \lambda - \alpha_e)\cos(\phi_e - \alpha_e)} \right\} \tau_{sh} U_{work} \tag{6.50}$$

For given tool angles, equations (6.26a) and (6.26b) are used to obtain α_e and η'_c in terms of η_c; A_{sh} is then obtained from η'_c, tool geometry and feed and depth of cut, from equation (6.41), using Tables 6.1 and 6.2 as appropriate; A_{uf} is determined from tool geometry, feed and depth of cut by equations (6.46) and (6.45). Thus, equation (6.50), with ϕ_e, τ_{sh} and

λ as functions of α_e, is converted to a function of η_c, tool geometry, feed and depth of cut and can be minimized with respect to η_c.

Once the energy is minimized, the cutting force component F_c is obtained from that energy divided by the cutting speed; and F_{fric} is found from equation (6.49b). The normal force on the rake face, F_N, is then found by manipulation of equation (6.30): from the relation between F_c, F_{fric} and F_N

$$F_N = \frac{F_c - F_{fric} \sin \alpha_e}{\cos \lambda_s \cos \alpha_n} \tag{6.51}$$

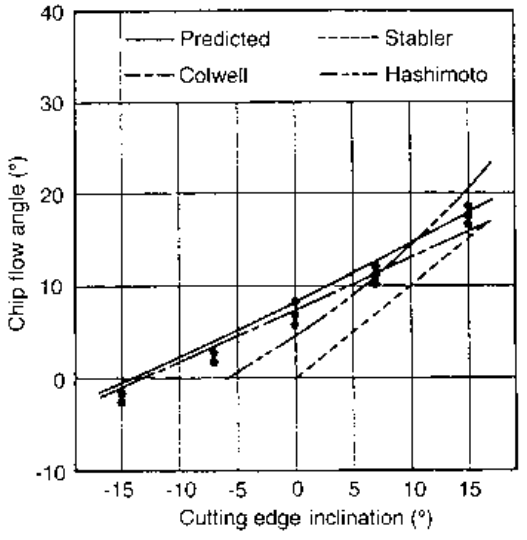
Equation (6.30) can also be used to obtain the feed and depth of cut force components. (It is not correct to determine F_N directly from F_{fric} and the friction angle, as the friction angle is defined, for the purposes of the energy minimization, in the cutting velocity–chip velocity plane; and this does not contain the normal to the rake face.)

Comparison with experiments

The predictions of the various models have been compared by Usui and Hirota (1978), for machining a medium (0.45%C) carbon steel with a P20 grade carbide tool. The orthogonal cutting data for this were established by experiment as (with angles in rad and τ_{sh} in MPa)

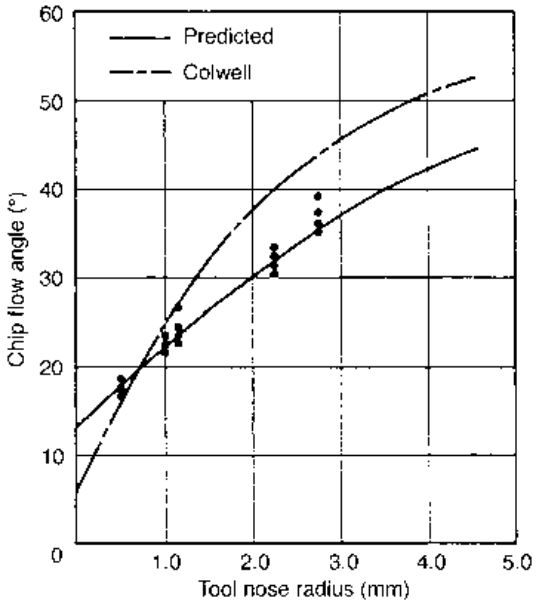
$$\left.\begin{aligned} \phi &= \exp(0.581\alpha - 1.139) \\ \tau_{sh} &= 517.4 - 19.89\alpha \\ \lambda &= \exp(0.848\alpha - 0.416) \end{aligned}\right\} \tag{6.52}$$

Figure 6.18 compares the measured and predicted dependencies of chip flow angle on cutting edge inclination and tool nose radius. The energy method gives closer agreement



(a)

Fig. 6.18 The dependence of η_c on (a) λ_s and (b) r_n, for machining a carbon steel (after Usui and Hirota, 1978)

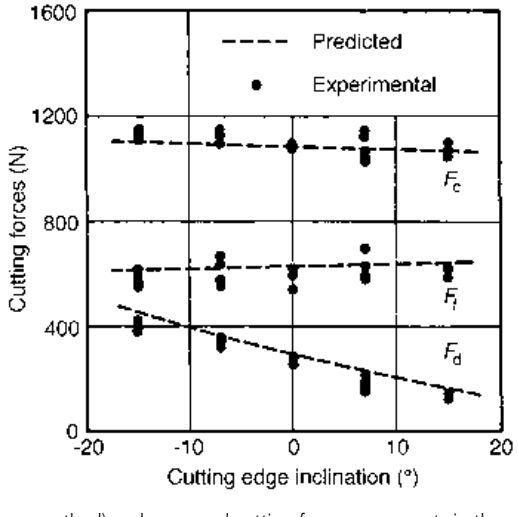


(b)

Fig. 6.18 *continued*

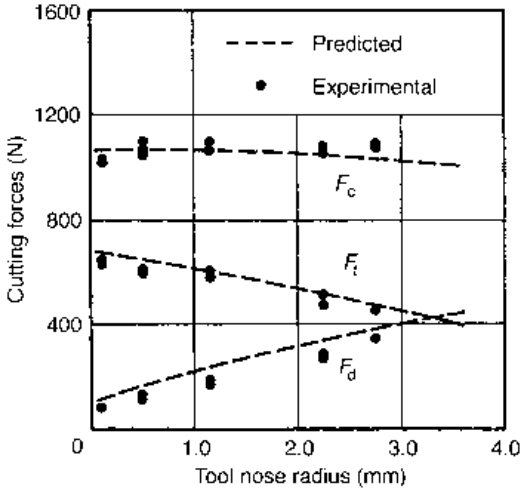
with experiment than Stabler's or Colwell's prediction or a third prediction due to Hashimoto and Kuise (1966). Figure 6.19, for the same conditions, shows that the energy method also predicts the force components well.

The good results with the energy method come despite its approximations, that ϕ_e is the same on every cutting velocity–chip velocity plane and that ϕ_e, τ_{sh}, λ and F_{fric}/A_{uf} depend



(a)

Fig. 6.19 Predicted (energy method) and measured cutting force components in the same conditions as Figure 6.18



(b)

Fig. 6.19 *continued*

only on α_e for a given tool geometry, cutting speed and feed. In reality, chips do curl and twist, so ϕ_e can vary from plane to plane (although, from Chapter 2, the extra deformation from this is small compared with the main primary shear). In addition, around the tool nose radius, the uncut chip thickness varies: it could be imagined that ϕ_e, τ_{sh}, λ and F_{fric}/A_{uf} should be allowed to vary with t_{1e} as well as with α_e. Whether there are conditions in which this extra refinement is necessary is unknown.

In the example just considered, the orthogonal cutting data were obtained by experiment. The main interest today is that such data can be obtained by simulation, by the finite element methods that are the subject of the following chapters.

References

Arsecularatne, J. A., Mathew, P. and Oxley, P. L. B. (1995) Prediction of chip flow direction and cutting forces in oblique machining with nose radius tools. *Proc. I. Mech. E. Lond.* **209Pt.B**, 305–315.

Childs, T. H. C. (1980) Elastic effects in metal cutting chip formation. *Int. J. Mech. Sci.* **22**, 457–466.

Colwell, L. V. (1954) Predicting the angle of chip flow for single-point cutting tools. *Trans. ASME* **76**, 199–204.

Dewhurst, P. (1978) On the non-uniqueness of the machining process. *Proc. Roy. Soc. Lond.* **A360**, 587–610.

Dewhurst, P. (1979) The effect of chip breaker constraints on the mechanics of the machining process. *Annals CIRP* **28 Part 1**, 1–5.

Hashimoto, F. and Kuise. H. (1966) The mechanism of three-dimensional cutting operations. *J. Japan Soc. Prec. Eng.* **32**, 225–232.

Hastings, W. F., Mathew, P. and Oxley, P. L. B. (1980) A machining theory for predicting chip geometry, cutting forces, etc, from work material properties and cutting conditions. *Proc. Roy. Soc. Lond.* **A371**, 569–587.

Kudo, H. (1965) Some new slip-line solutions for two-dimensional steady-state machining. *Int. J. Mech. Sci.* **7**, 43–55.

Lee, E. H. and Shaffer, B. W. (1951) The theory of plasticity applied to a problem of machining. *Trans. ASME J. Appl. Mech.* **18,** 405–413.

Merchant, M. E. (1945) Mechanics of the metal cutting process. *J. Appl. Phys.* **16,** 318–324.

Oxley, P. L. B. (1989) *Mechanics of Machining*. Chichester: Ellis Horwood.

Palmer, W. B. and Oxley, P. L. B. (1959) Mechanics of metal cutting. *Proc. I. Mech. E. Lond.* **173**, 623–654.

Petryk, H. (1987) Slip-line field solutions for sliding contact. In *Proc. Int. Conf. Tribology – Friction, Lubrication and Wear Fifty years On*, London, 1–3 July, pp. 987–994 (IMechE Conference 1987–5).

Roth, R. N. and Oxley, P. L. B. (1972) A slip-line field analysis for orthogonal machining based on experimental flow fields. *J. Mech. Eng. Sci.* **14**, 85–97.

Shaw, M. C., Cook, N. H. and Smith, P. A. (1952) The mechanics of three dimensional cutting operations. *Trans. ASME* **74**, 1055–1064.

Shi, T. and Ramalingam, S. (1991) Slip-line solution for orthogonal cutting with a chip breaker and flank wear. *Int. J. Mech. Sci.* **33**, 689–704.

Stabler, G. V. (1951) The fundamental geometry of cutting tools. *Proc. I. Mech. E. Lond.* **165**, 14–26.

Stevenson, M. G. and Oxley, P. L. B. (1969–70) An experimental investigation of the influence of speed and scale on the strain-rates in a zone of intense plastic deformation. *Proc. I. Mech. E. Lond.* **184**, 561–576.

Stevenson, M. G. and Oxley, P. L. B. (1970–71) An experimental investigation of the influence of strain-rate and temperature on the flow stress properties of a low carbon steel using a machining test. *Proc. I. Mech. E. Lond.* **185**, 741–754.

Trent, E. M. (1991) *Metal Cutting,* 3rd edn. Oxford: Butterworth Heinemann.

Usui, E., Kikuchi, K. and Hoshi K. (1964) The theory of plasticity applied to machining with cutaway tools. *Trans ASME, J. Eng. Ind* **B86**, 95–104.

Usui, E., Hirota, A. and Masuko, M. (1978) Analytical prediction of three dimensional cutting process (Part 1). *Trans. ASME J. Eng. Ind.* **100**, 222–228.

Usui, E. and Hirota, A. (1978) Analytical prediction of three dimensional cutting process (Part 2). *Trans. ASME J. Eng. Ind.* **100**, 229–235.

Usui, E. (1990) *Modern Machining Theory*. Tokyo: Kyoritu-shuppan (in Japanese).

Zorev, N. N. (1966) *Metal Cutting Mechanics*. Oxford: Pergamon Press.

7

Finite element methods

In the previous chapter, Sections 6.2 and 6.3 established some of the difficulties and issues in analysing even steady-state and plane strain chip formation. The finite element method is a natural tool for handling the non-linearities involved. Section 6.4 suggested how orthogonal (plane strain) results could be extended to three-dimensional conditions. An eventual goal, particularly for non-plane rake-faced tools, must be the direct analysis of three-dimensional machining; and the finite element method would appear to be the best candidate for this. Chip formation is a difficult process to analyse, even by the finite element method. This chapter is mainly concerned with introducing the method and reviewing the learning process – from the 1970s to the present – of how to use it. Its applications are the subject of Chapter 8.

There are, in fact, several finite element methods, not just one. There is a coupling of thermal and mechanical analysis methods. In the mechanical domain, different approaches have been tried and are still in use. The differences cover how material stress–strain relations are described (modelling elasticity as well as plasticity, or neglecting elastic components of stress and strain); how flow variations are described (relative to fixed axes, or convecting with material elements – the Eulerian and Lagrangian views of fluid and solid mechanics); how the elements are constructed (uniform, or structured according to physical intuition, or allowed to remesh adaptively in response to the results of the calculations); and how some factors more specific to metal machining (for example the separation of the chip from the work) are dealt with. A general background to these (to raise awareness of issues more than to support use in detail) is given in Section 7.1. Section 7.2 surveys developments of the finite element approach (applied to chip formation), from the 1970s to the 1990s. Section 7.3 gives some additional background information to prepare for the more detailed material of Chapter 8. To obtain accurate answers from finite element methods (as much as for any other tool) it is necessary to supply accurate information to these methods. Section 7.4 considers the plastic flow behaviour of materials at the high strains, strain rates and temperatures that occur in machining, a topic introduced in Chapter 6.3.

7.1 Finite element background

Fundamental to all finite element analysis is the replacement of a continuum, in which problem variables may be determined exactly, by an assembly of finite elements in which

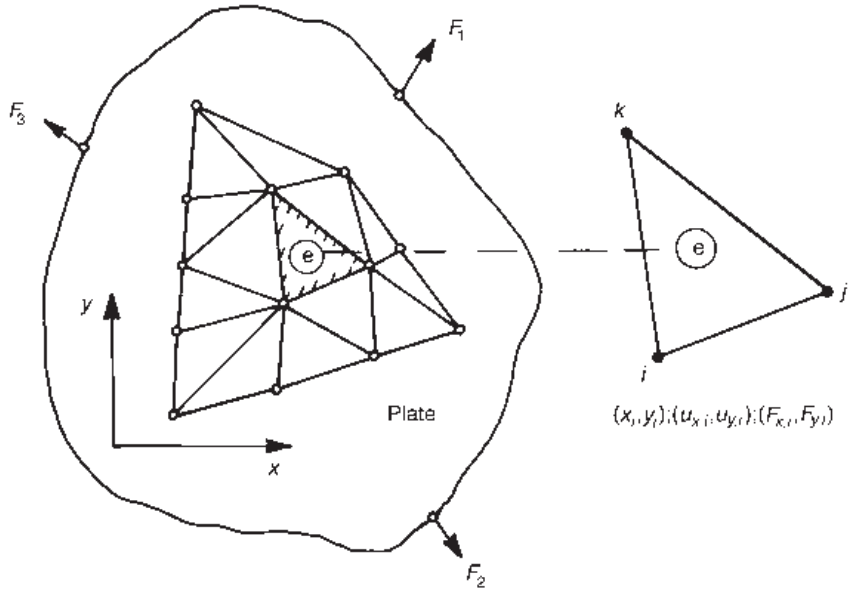


Fig. 7.1 To illustrate the finite element method for a mechanics problem

the problem variables are only determined at a set number of points: the nodes of the elements. Between the nodes, the values of the variables, or quantities derived from them, are determined by interpolation.

A simple example may be given to demonstrate the method: calculating the stresses and strains in a thin plate (thickness t_h) loaded elastically in its plane by three forces F_1, F_2 and F_3. The plate is divided into triangular elements – the most simple type possible. Some of them are shown in Figure 7.1.

The nodes of the problem are the vertices of the elements. Each element, such as that identified by 'e', is defined by the position of its three nodes, (x_i,y_i) for node i and similarly for j and k. The external loadings cause x and y displacements of the nodes, $(u_{x,i},u_{y,i})$ at i and similarly at j and k. The adjacent elements transmit external forces to the sides of the element, equivalent to forces $(F_{x,i},F_{y,i})$, $(F_{x,j},F_{y,j})$ and $(F_{x,k},F_{y,k})$ at the nodes.

Strain – displacement relations

Displacements within the element are, by linear interpolation

$$u_x = a_1 + a_2x + a_3y; \qquad u_y = a_4 + a_5x + a_6y \tag{7.1}$$

From the definition of strain as the rate of change of displacement with position, and choosing the coefficients a_1 to a_6 so that, at the nodes, equation (7.1) gives the nodal displacements,

$$\varepsilon_{xx} = \frac{\partial u_x}{\partial x} = a_2 = \frac{(y_j - y_k)u_{x,i} + (y_k - y_i)u_{x,j} + (y_i - y_j)u_{x,k}}{2\Delta} \tag{7.2}$$

where Δ is the area of the element; and similarly for the other strains ε_{yy} and γ_{xy}. Matrix algebra allows a compact way of writing these results:

$$\begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \gamma_{xy} \end{Bmatrix} = \frac{1}{2\Delta} \begin{bmatrix} y_j - y_k & 0 & y_k - y_i & 0 & y_i - y_j & 0 \\ 0 & x_k - x_j & 0 & x_i - x_k & 0 & x_j - x_i \\ x_k - x_j & y_j - y_k & x_i - x_k & y_k - y_i & x_j - x_i & y_i - y_j \end{bmatrix} \begin{Bmatrix} u_{x,i} \\ u_{y,i} \\ u_{x,j} \\ u_{y,j} \\ u_{x,k} \\ u_{y,k} \end{Bmatrix} \tag{7.3a}$$

or, more compactly still

$$\{\varepsilon\}_{\text{element}} = [B]_{\text{element}} \{u\}_{\text{element}} \tag{7.3b}$$

where $[B]_{\text{element}}$, known as the B-matrix, has the contents of equation (7.3a).

Elastic stress – strain relations

In plane stress conditions, as exist in this thin plate example, Hooke's Law is

$$\begin{Bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{xy} \end{Bmatrix} = \frac{E}{1-\nu^2} \begin{bmatrix} 1 & \nu & 0 \\ \nu & 1 & 0 \\ 0 & 0 & \dfrac{1-\nu}{2} \end{bmatrix} \begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \gamma_{xy} \end{Bmatrix}; \text{ or } \qquad \{\sigma\} = [D]\{\varepsilon\} \tag{7.4}$$

Combining equations (7.3b) and (7.4)

$$\{\sigma\}_{\text{element}} = [D][B]_{\text{element}} \{u\}_{\text{element}} \tag{7.5}$$

Nodal force equations, their global assembly and solution

Finally, the stresses in the element can be related to the external nodal forces, either by force equilibrium or by applying the principle of virtual work. Standard finite element texts (see Appendix 1.5) show

$$\{F\}_{\text{element}} = t_{\text{h}} \Delta_{\text{element}} [B]^{\text{T}}_{\text{element}} [D][B]_{\text{element}} \{u\}_{\text{element}} \tag{7.6}$$

Equations (7.6) for every element are added together to create a global relation between the forces and displacements of all the nodes:

$$\{F\}_{\text{global}} = [K]\{u\}_{\text{global}} \qquad \text{or, more simply} \qquad \{F\} = [K]\{u\} \tag{7.7}$$

where $[K]$, the global stiffness matrix, is the assembly of $t_{\text{h}} \Delta_{\text{element}} [B]^{\text{T}}_{\text{element}} [D][B]_{\text{element}}$.

For the assembled elements, the resultant external force on every node is zero, except for where, in this example, the forces F_1, F_2 and F_3 are applied. The column vector $\{F\}$ is a known quantity: equations (7.7) are a set of linear equations for the unknown displacements $\{u\}$. After solving these equations, the strains in the elements and then the stresses can be found from equations (7.3) and (7.4).

These steps of a finite element mechanics calculation are for the circumstances of small strain elasticity. Plasticity introduces some changes and large deformations require more care in the detail.

Rigid–plastic or elastic–plastic modelling

In plastic flow conditions, such as occur in machining and forming processes, it is natural to consider nodal velocities $\dot{u}$ instead of displacements u as the unknowns. Strain rates in

an element are derived from rates of change of velocity with position, in the same way that strains are derived from rates of change of displacement with position. Over some period of time, the strain rates generate increased strains in an element. In a time dt strain increments are:

$$\{\mathrm{d}\varepsilon\} = [B]\{\dot{u}\}\mathrm{d}t \tag{7.8}$$

The strain increment components have both elastic and plastic parts. The plastic parts are in proportion to the total stress components but the elastic parts are in proportion to the stress increment components (as described in Appendix 1). If elastic parts of a flow are ignored, plastic flow rules lead to relations between the total stresses and the strain increments. These lead, in turn, (Appendix 1.5 gives better detail) to finite element equations of the form

$$\{F\} = [K]\{\dot{u}\}\mathrm{d}t \tag{7.9a}$$

Ignoring the elastic strains is the rigid-plastic material approximation. Equation (7.9a) is commonly solved directly for the velocity of a flow, by iteration on an initial guess.

If the elastic strain parts of a plastic flow are not ignored, the flow rules lead to relations between element stress increment and strain increment components. The finite element equations become

$$\{\mathrm{d}F\} = [K]\{\dot{u}\}\mathrm{d}t \qquad \text{or} \qquad \{\dot{F}\} = [K]\{\dot{u}\} \tag{7.9b}$$

In order to predict the state of an element, it is necessary to integrate the solution of equation (7.9b) along an element's loading path, from its initially unloaded to its current position.

The above descriptions are highly simplified. Appendix 1.5 gives more detail, particularly of the non-linearities of the finite element equations that enter through the rigid–plastic or elastic–plastic $[D]$ matrix within the $[K]$ matrix. The main point to take forward is that elastic–plastic analysis gives a more complete description of process stresses and strains but, because it is necessary to follow the development of a flow from its transient start to whatever is its final state, and because of its high degree of non-linearity, it is computationally very intensive. Rigid–plastic finite element modelling requires less computing power because it is not necessary to follow the path of a flow so closely, and the equations are less non-linear; but it ignores elastic components of strain. Particularly in machining, when thin regions of plastic distortion (the primary and secondary shear zones) are sandwiched between elastic work, chip and tool, this is a disadvantage. Nonetheless, both rigid–plastic and elastic–plastic finite element analysis are commonly applied to machining problems.

Eulerian or Lagrangian flow representation

There is a choice, in dividing the region of a flow problem into elements, whether to fix the elements in space and allow the material to flow through them (the Eulerian view), or to fix the elements to the flowing material, so that they convect with the material (the Lagrangian view). Figure 7.2 illustrates these options. In the Eulerian case, attention is drawn to how velocities vary from element to element (for example elements 1 and 2) at the same time. In the Lagrangian case, attention is focused on how the velocity of a particular element varies with time. Each view has its advantages and disadvantages.

The advantage of the Eulerian view is that the shapes of the elements do not change

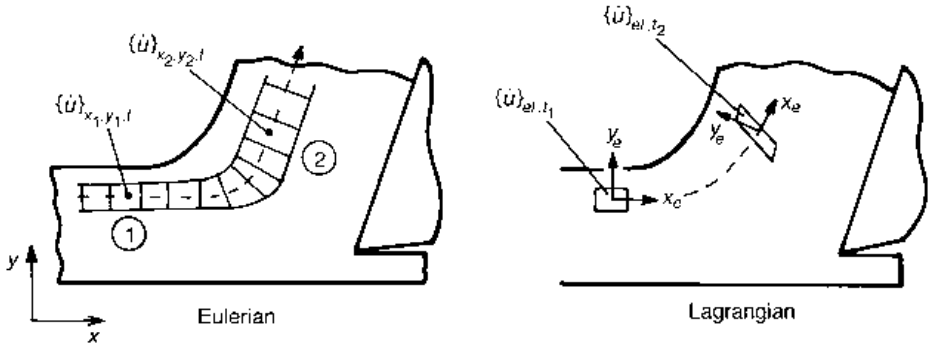


Fig. 7.2 Eulerian and Lagrangian views of a plastic flow

with time, so the coefficients of the [*B*] matrix, which depend on element shape (for example equation (7.3), for a triangular element), need only be computed once. However, in a problem such as machining, in which determining the location of the free surface of the chip is part of the problem to be solved, it is not clear where the elements should be drawn. It is necessary to develop the free surface boundaries of the element mesh by iteration. A more general problem is how to describe the convection of material property changes, like strain hardening, from element to element. (Eulerian analyses are more common in fluid mechanics than in solid mechanics because fluid properties vary less with deformation than do those of solids.) In steady flow problems, it is assumed that material properties convect along the streamlines.

The Lagrangian view has no problem with convection of material properties. The state of a material is fixed in an element. However, the element changes shape during a flow: the [*B*] matrix requires continued updating. This leads to geometrical non-linearities in addition to material non-linearities in the finite element equations. In extreme cases it may become necessary to simplify a distorted element shape by remeshing (see the next section). There is a further complication. An element most likely rotates as well as distorts as it passes through a flow (as shown in Figure 7.2). After a while, its local *x* and *y* directions will differ from those of other elements. However, a common set of axes is required for the transformation of individual element equations to a global assembly. Counter-rotating the local element coordinate system, as well as updating the [*B*] matrix, is repeatedly required.

Structured or adaptive meshing – and other matters

It is common sense that a finer mesh is needed where problem variables (velocity, temperature) vary strongly with position than where they do not. In metal machining, fine detail is needed to model the primary and secondary shear zones. This poses no problem for Eulerian meshes: a choice is made where to refine the mesh and by how much. However, for computing efficiency with a Lagrangian mesh, there is a need to refine and then coarsen how the material is divided into elements as it flows into and out of plastic shear zones.

The need to refine Lagrangian meshes is particularly accute near the cutting edge of a tool, where the work material flow splits into flow under the cutting edge and flow into the chip. A range of approaches to separation at the cutting edge has been developed, from introducing an artificial crack in the work, to highly adaptive remeshing, to developing special elements with singularities in them. These are not needed in Eulerian analyses.

In addition to the choice of finite element method based on computational criteria, particular softwares for metal machining should be able to model the variation of flow stress with strain, strain rate and temperature (Section 6.3) and the variation of rake face friction conditions from high load to low load conditions (Chapter 2, Section 2.4)

Summary

The choice of finite element methods for machining problems involves rigid-plastic or elastic–plastic material models; Eulerian or updated Lagrangian flow treatments; structured or adaptive meshes; chip/work separation criteria needed or not needed; and coupling to thermal calculation models or not. Some of the achievements of these approaches, and methods of overcoming computational problems, are chronicled in the next section. On balance, the updated Lagrangian analyses' advantage of easily tracking material property changes outweighs the disadvantages of computational complexity. The simplicity of Eulerian computations is not fully realized in the large free surface movement conditions of a chip forming process.

7.2 Historical developments

The 1970s

The earliest finite element chip formation studies (Zienkiewicz, 1971; Kakino, 1971) avoided all the problems of modelling large flows by simulating the loading of a tool against a pre-formed chip (Figure 7.3). A small strain elastic–plastic analysis demonstrated the development of plastic yielding along the primary shear plane as a tool was displaced against the chip. This work has a number of limitations, making it of historical interest only. For

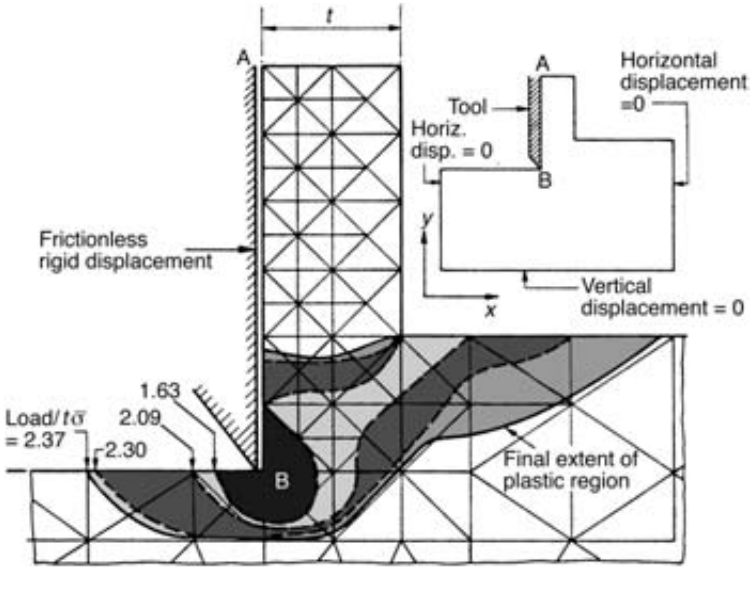


Fig. 7.3 Shear zone development, loading a pre-formed chip (Zienkiewicz, 1971)

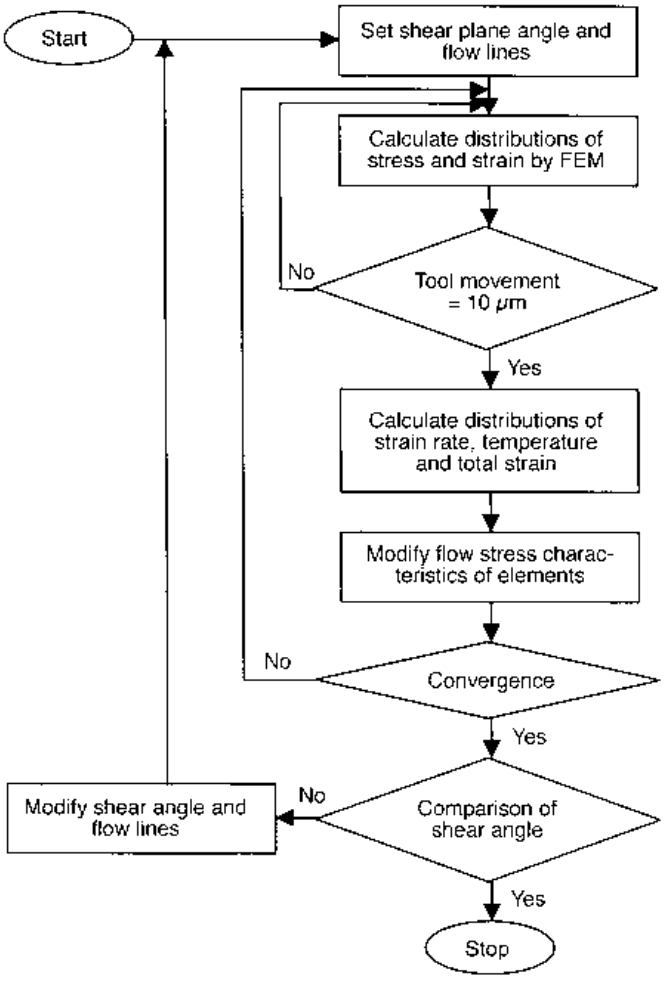


Fig. 7.4 The iterative convergence method (ICM) – Shirakashi and Usui (1976)

example, it neglects friction between the chip and tool, and strain rate and temperature material flow stress variations are not considered either. More fundamentally, it assumes the shape of the chip in the first place: the main purpose of chip forming analyses is to predict the shape.

The limitations of this initial work were removed by Shirakashi and Usui (1976). While keeping the computational advantages of supposing the tool to move into a pre-formed chip, they developed an iterative way of changing the shape of the pre-form until the generated plastic flow was consistent with the assumed shape. They also included realistic chip/tool friction conditions (from split-tool experiments), a temperature as well as a mechanical calculation, and material flow stress variations with strain, strain rate and temperature, measured from high strain rate Hopkinson bar tests (see Section 7.4). Their iterative convergence method (ICM) is shown in Figure 7.4.

The first step of the ICM is to assume a steady state chip shape (similar to Figure 7.3, except for supposing there to be a small crack at the cutting edge to enable the chip to

separate from the work) and (for plane strain modelling) to create a three-node triangular mesh following the streamlines of the flow. In the first iteration, the tool is moved against the chip: the development of nodal velocities is followed with an updated Lagrangian elastic–plastic analysis. When it is judged that the plastic flow is fully developed, the nodal velocity field is used to calculate the element strain rates along the streamlines; strains are obtained by integrating the strain rates with respect to time along the streamlines (as if material had reached its current position by flowing along a streamline). Temperatures are calculated from the internal and friction work rates and the work and tool materials' thermal properties (in the first application of the ICM, temperature was calculated by a finite difference method, but later the finite element method was used). Material flow stress is then set according to its strain, strain rate and temperature, the tool and chip are unloaded and the cycle of moving the tool into the chip repeated. This is continued until converged strain rates and temperatures are achieved. At that stage, the flow field is used to modify the initially assumed streamlines to be closer to the calculated flow. The complete cycle is then repeated, and repeated again until the assumed and calculated flow fields agree. The displacement of the tool needed to establish the flow field is sufficiently small that the need to reform the crack at the cutting edge does not arise. Within limits, the crack size does not influence the predicted chip flow.

Figure 7.5 shows chip shape, equivalent plastic strain rate and temperature fields

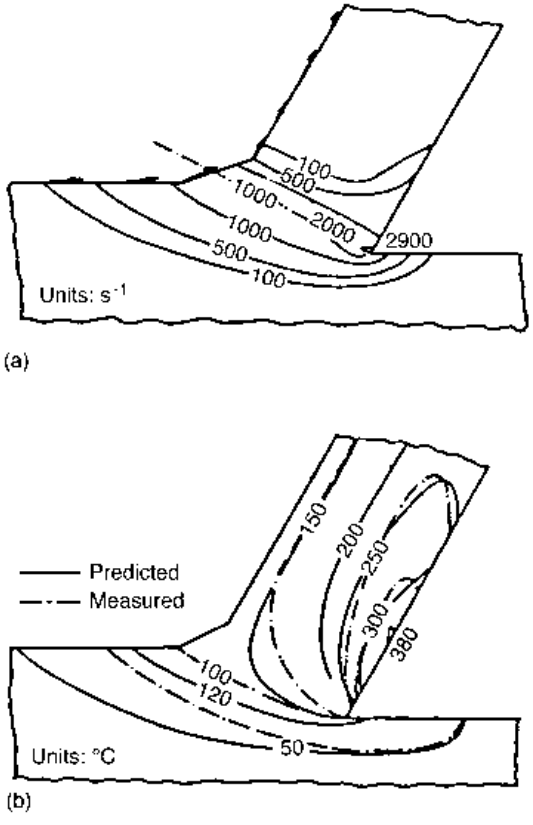


Fig. 7.5 (a) Strain rates and (b) temperatures predicted by the ICM method for dry machining α-brass, cutting speed 48 m/min, rake angle 30°, feed 0.3 mm

calculated by Shirakashi and Usui for machining an α-brass. Chip shape agrees with experiment, as does the temperature field (which was studied experimentally with infrared microscopy).

The procedure of loading a tool against an already formed chip greatly reduces computing capacity requirements and, in the 1970s, made elastic–plastic analysis possible. However, it does not follow the actual path by which a chip is formed and, as outlined in Section 7.1 and Appendix 1, the development of elastic–plastic flows is path dependent. The justification of the method is that it gives good agreement with experiment. The ICM has been developed further, in analyses of cutting fluid action (Usui *et al.*, 1977), built-up-edge formation (Usui *et al.*, 1981) and more recently in studies of low alloy semi free-machining steels (Childs and Maekawa, 1990). It is given further consideration in Section 7.3 and Chapter 8.

The 1980s

Rigid–plastic modelling does not require the actual loading path to be followed (also discussed in Section 7.1 and Appendix 1). Steady state rigid–plastic modelling, within a Eulerian framework, also adjusting an initially assumed flow field to bring it into agreement with the computed field, was first applied to machining by Iwata *et al.* (1984), using software developed from metal forming analyses. They included friction and work hardening and also a consideration of whether the chip would fracture, but not heating (and obviously not elastic effects). Experiments at low cutting speed (0.15 mm/min in a scanning electron microscope) supported their predictions. It was not necessary with the Eulerian frame to introduce a crack at the cutting edge, but it was necessary, to avoid computational difficulties, to give the cutting edge a small radius (about one tenth of the feed).

The mid-1980s, with a growth in available computer power, saw the first non-steady chip formation analyses, following the development of a chip from first contact of a cutting edge with a workpiece, as in practical conditions (Figure 7.6(a)). Updated Lagrangian elastic–plastic analysis was used, and the chip/work separation criterion at the cutting edge

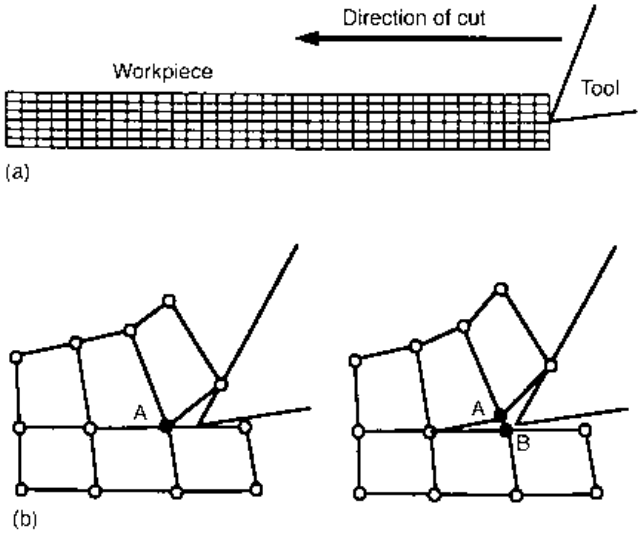


Fig. 7.6 Non-steady state analysis: (a) initial model and (b) separation of nodes at the cutting edge

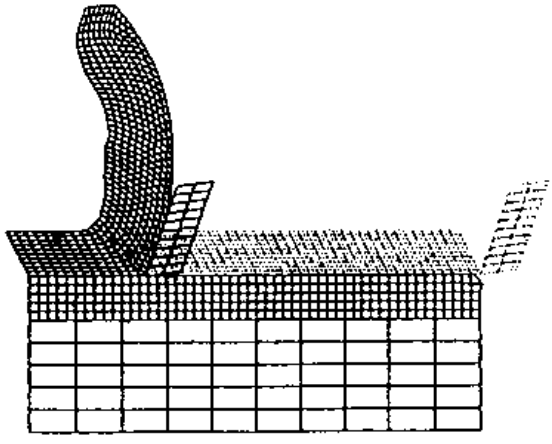


Fig. 7.7 An example of non-steady state analysis (Strenkowski and Carrol, 1985)

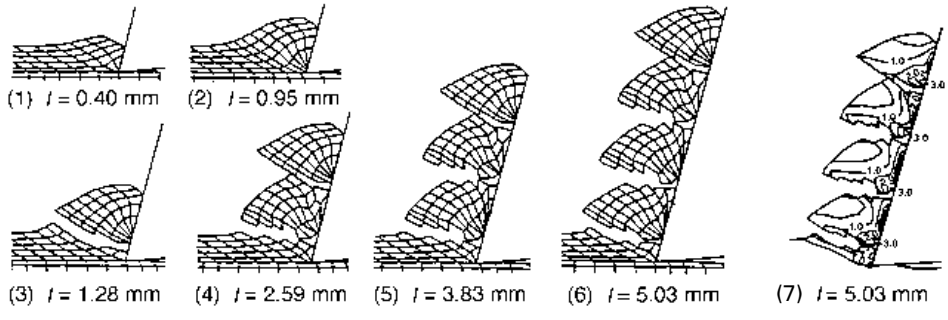


Fig. 7.8 Discontinuous chip formation in β-brass (Obikawa *et al.* 1997): (1–6) element deformation and (7) equivalent plastic strain distribution, at different cut distances *l*

became an issue (Figure 7.6(b)): should the connection between elements be broken by a limiting strain, limiting energy or limiting displacement condition? Figure 7.7 shows the earliest example (Strenkowski and Carrol, 1985), which used a strain-based separation criterion. At that time, neither a realistic friction model nor coupling of the elastic–plastic to thermal analysis (and hence nor a realistic flow stress variation with cutting conditions) was included.

At the same time as plastic flow finite element methods were being developed for metal machining, linear fracture mechanics methods were being developed for the machining of brittle ceramics (Ueda and Sugita, 1983).

The 1990s

The 1990s have seen the development of non-steady analysis, from transient to discontinuous chip formation, the first three-dimensional analyses and the introduction of adaptive meshing techniques particularly to cope with flow around the cutting edge of a tool.

Figure 7.8 shows an updated Lagrangian elastic–plastic simulation of discontinuous chip formation in β-brass at low cutting speed. To obtain this result a geometrical

(displacement controlled) parting criterion at the cutting edge was combined with an empirical crack nucleation and growth criterion, considered further in Section 7.3 and Chapter 8. Other authors have taken different approaches to crack growth during chip formation (Ueda *et al.* 1991).

Figures 7.9 and 7.10 are the earliest examples of elastic–plastic steady and non-steady three-dimensional analyses. The steady state example is an extension of the ICM to three

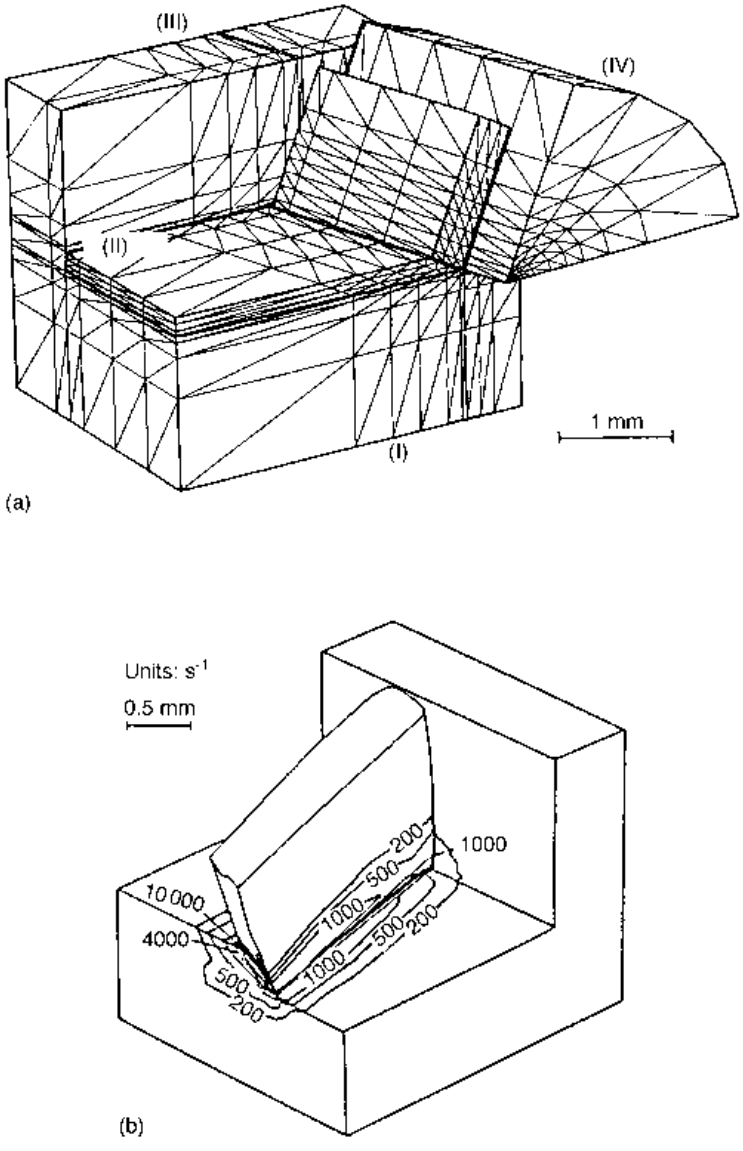


Fig. 7.9 Three-dimensional steady state chip formation by the ICM (Maekawa and Maeda, 1993): (a) initial model and (b) equivalent strain rate distribution

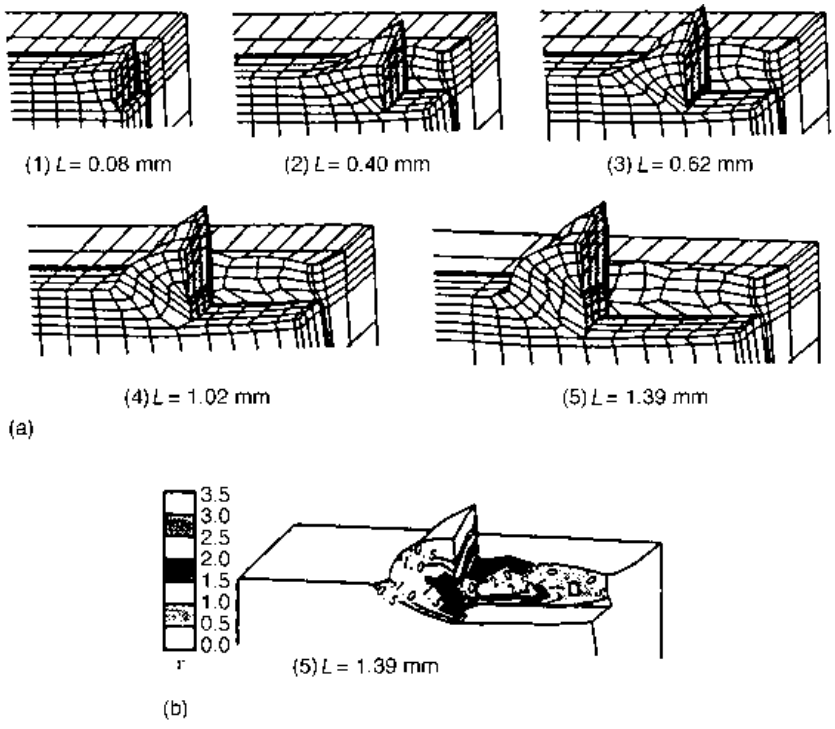


Fig. 7.10 Three-dimensional non-steady chip formation (Sasahara *et al.,* 1994): (a) element deformations and (b) equivalent plastic strain distribution

dimensions. The non-steady example employs a geometrical parting criterion at both the primary and secondary cutting edges. In both these cases, temperature and strain rate effects are ignored, to reduce the computing requirements. This restriction was soon removed: three-dimensional elastic–plastic, thermally coupled, ICM simulation soon became used for cutting tool design, also considered further in Chapter 8 (Maekawa *et al.* 1994).

In parallel with the extension of elastic–plastic methods to non-steady and three-dimensional conditions, the rigid–plastic method (Iwata *et al.*, 1984) was similarly being developed (Ueda and Manabe, 1993; Ueda *et al.*, 1996), with a shift from Eulerian to Lagrangian modelling. Figure 7.11 shows the simulation of spirally curled chip formation during milling with a non-zero axial rake tool. A simple form of remeshing at the cutting edge, instead of a geometrical crack, was introduced to accommodate the separation of the chip from the work.

Adaptive mesh refinement in non-steady flows, whereby during an increment of flow (a time step) the mesh is fixed to the work material in a Lagrangian manner – but between steps the mesh connectivity and size is changed according to rules based on local severities of deformation – offers the advantage over fixed Lagrangian approaches of concentrating the mesh where it is needed most, in the primary shear zone, at the cutting edge and along the rake face. Concentration at the cutting edge provides an alternative to introducing a crack for following the separation of the chip from the work. Both rigid–plastic (Sekhon and Chenot,

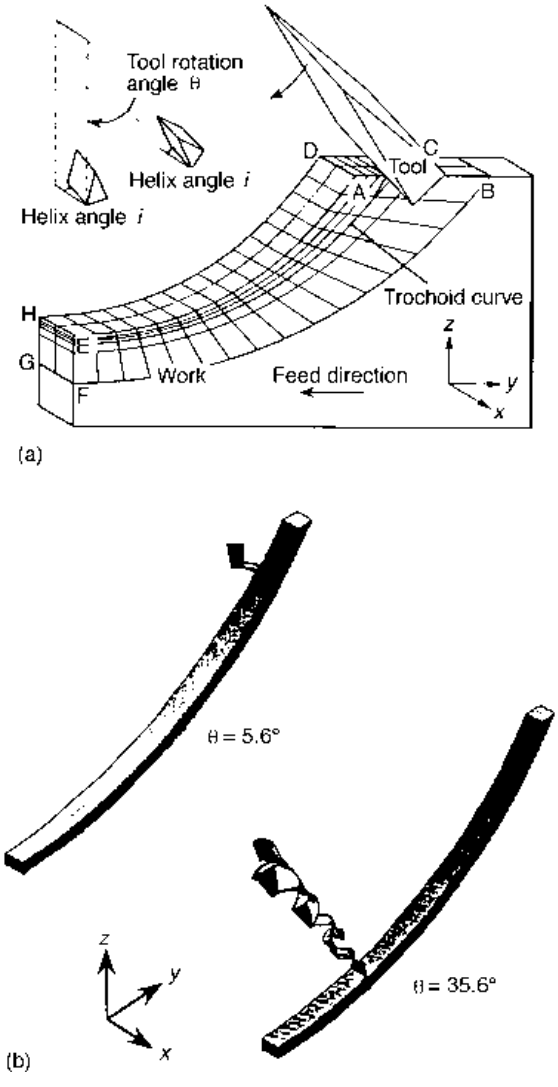


Fig. 7.11 Three-dimensional non-steady chip formation by rigid plastic finite element method (Ueda *et al.*, 1996): (a) initial model and (b) spiral chips

1993; Ceretti *et al.,* 1996) and elastic–plastic (Marusich and Ortiz, 1995) adaptive remeshing softwares have been developed and are being applied to chip formation simulation. They seem more effective than arbitrary Lagrangian–Eulerian (ALE) methods in which the mesh is neither fixed in space nor in the workpiece (for example Rakotomolala *et al.*, 1993).

Summary

The 1970s to the 1990s has seen the development and testing of finite element techniques for chip formation processes. Many of the researches have been more concerned with the development of methods than their immediate application value: the limited availability of

reliable friction and high strain, strain rate and temperature material flow properties did not hold back this work. The ICM approach is the exception: from the start it has been concerned with supporting machining applications. Now that all methods are approaching maturity, attention is shifting to the provision of appropriate friction and material flow property data (see Section 7.4).

In the future there are likely to be three main avenues of finite element modelling of chip formation: (1) the ICM method for steady state processes, because of its extremely high computing efficiency; (2) Lagrangian adaptive mesh refinement methods for unsteady processes, both elastic–plastic as the most complete treatment and rigid–plastic for its fewer computing requirements if elastic effects are not needed; and (3) fixed mesh Lagrangian methods (with chip separation criteria) to support educational studies of unsteady processes in a time effective manner. Chapter 8 will concentrate on the first and the last of these, but a future edition may well include more of the second.

7.3 The Iterative Convergence Method (ICM)

Sections 7.3.1 and 7.3.2 give more details of the ICM method (which was introduced in the previous section), as background to the examples of its use presented in Chapter 8. Section 7.3.3 introduces a treatment of unsteady processes (case (3) above).

7.3.1 Principles and implementation

As has already been described, the ICM method is an updated Lagrangian elastic–plastic finite element analysis for predicting steady state chip flows. Such analyses normally must follow the development of strain along a material's load path and are computationally very intensive. The ICM method replaces the real path by a shorter one: loading the tool onto an already formed chip. It provides a way, by iteration, of finding the formed chip shape that is consistent with the material's flow properties and friction interaction with the tool. A key point is that its finite elements are structured to follow the stream lines of the steady state chip flow (as will be seen in Figure 7.13).

The flow chart of the ICM procedure as it was originally introduced, is shown in Figure 7.4. Figure 7.12 shows its developed form. An initial guess of the chip flow or stream lines (usually of the simple straight shear plane type considered in Chapter 2) is made and the tool is placed so that its rake face just touches the back surface of the chip. Calculation proceeds by incrementally displacing the workpiece towards the tool so that a load develops between the chip and tool. At each increment, it is checked if the plastic flow is fully developed (saturated): if it is not, a further increment is applied (loop I). Once the flow is developed, the initial guess is systematically and automatically reformed to bring it into closer agreement with the calculated flow. The strain rate in each element of the reformed flow is calculated; and the strain distribution is obtained by integrating strain rate along the streamlines. The element flow stress associated with the reformed flow is then estimated; but this requires temperature as well as strain and strain rate to be known. A second loop (loop II), a thermal finite element analysis, is entered to determine the temperature field. Finally, it is checked whether the derived material flow stress, temperature and flow fields have converged: if they have not, the whole iteration is repeated (loop III). The next paragraphs give some details that are special to the calculations.

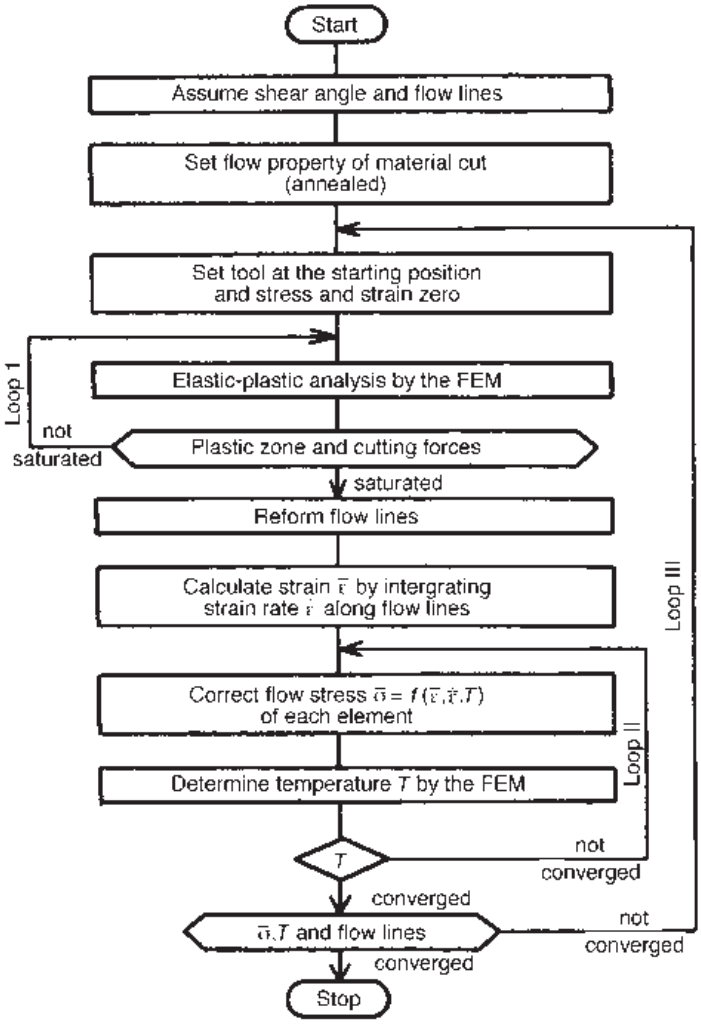


Fig. 7.12 Developed flow-chart of the iterative convergence method

Judgement of saturation of the plastic flow is made either on the basis of the tool load reaching a maximum value or of conservation of volume – i.e. that the computed flow of material out of the plastic zone into the chip balances that of the work into the plastic zone.

Reformation of the flow field supposes that the separation between nodes along a streamline is unchanged by reformation, but that the direction from one node to the next is altered to bring it more closely tangential to the calculated flow. For each flow line consisting of a node sequence $j-1, j, j+1 \ldots$, the updated (x, y) coordinates of node j are given by

$$x_j = x_{j-1} + \frac{\dot{u}_{x,j-1} + \dot{u}_{x,j}}{\ell_j} L_j, \qquad y_j = y_{j-1} + \frac{\dot{u}_{y,j-1} + \dot{u}_{y,j}}{\ell_j} L_j \tag{7.10}$$

where $(\dot{u}_{x,j}, \dot{u}_{y,j})$ are the calculated velocities at node j, ℓ_j is the resultant average velocity of nodes j–1 and j, and L_j is the separation between nodes j–1 and j:

$$\ell_j = \left| (\dot{u}_{x,j-1} + \dot{u}_{x,j})^2 + (\dot{u}_{y,j-1} + \dot{u}_{y,j})^2 \right|^{1/2} \tag{7.11a}$$

$$L_j = \left| (x_{j-1} - x_j)^2 + (y_{j-1} - y_j)^2 \right|^{1/2} \tag{7.11b}$$

The reformation using equations (7.10) and (7.11) is implemented from the beginning to the end of a flow line so that the coordinates (x_{j-1}, y_{j-1}) have already been revised.

The equivalent plastic strain $\bar{\varepsilon}$ in each element is evaluated by the integration of its rate $\dot{\bar{\varepsilon}}$ along the reformed flow lines:

$$\bar{\varepsilon} = \int \dot{\bar{\varepsilon}} \mathrm{d}t = \int \frac{\dot{\bar{\varepsilon}}}{\dot{v}_e} \mathrm{d}\ell \tag{7.12}$$

where $\dot{v}_e$ the element velocity, obtained from the average of an element's nodal velocities. Relations between flow stress, strain, strain rate and temperature are considered in Section 7.4.

Figure 7.13 shows an ICM mesh for two-dimensional machining with a single point tool, in which the x- and y-axes are taken respectively parallel and perpendicular to the cutting direction, in a rectangular Cartesian coordinate system. The tool is assumed to be stationary and rigid, while the workpiece moves towards it at the specified cutting speed.

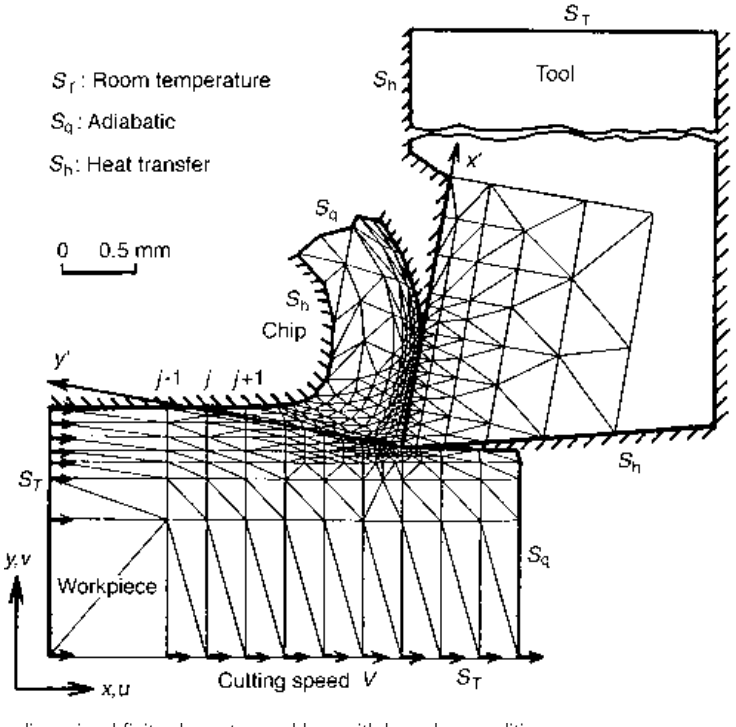


Fig. 7.13 Two-dimensional finite element assemblage with boundary conditions

The mesh is highly refined in the primary and secondary shear zones, in line with the considerations of Chapter 6.

The friction boundary at the tool–chip interface is treated as follows. For the nodes contacting the rake face, the conditions imposed on the finite element equation (equation 7.9(b)) with respect to the nodal force rate $\dot{F}$ and the nodal velocity $\dot{u}$ are:

$$\dot{F}_{x'} = \left(\frac{\mathrm{d}\tau}{\mathrm{d}\sigma_{\mathrm{n}}} \right) \dot{F}_{y'}, \qquad \dot{u}_{y'} = 0 \tag{7.13}$$

where x' and y' are the local coordinate systems parallel and perpendicular to the rake face (as shown in Figure 7.13) and $(\mathrm{d}\tau/\mathrm{d}\sigma_{\mathrm{n}})$ is the local slope of the friction characteristic curve (for example the inset in Figure 2.23) at the value of σ_{n} associated with the nodal force $F_{y'}$.

In the course of the elastic–plastic analysis, loop I of Figure 7.12, the chip contact length may increase or decrease. A chip surface node in contact with the rake face is judged to leave contact if its $F_{y'}$ force becomes tensile; and a node out of contact is judged to come into contact if its reformed y' becomes negative (penetrates the tool). Thus, the ICM method automatically determines the chip-tool contact length as one aspect of determining the chip flow.

The separation of material at the cutting edge is taken into account geometrically. The streamline at the cutting edge bifurcates both onto the rake face and onto the clearance surface of the work. In the ICM calculation, the relative displacement between the work near the cutting edge and the tool is only about 1/20 of the uncut chip thickness. A small crack imposed on the mesh, of that length, is sufficient to cope with separation without additional treatments, such as reconstruction of node and element sequences and special procedures to ensure a force balance at the crack tip. (This is not the case when the actual loading path of an element has to be followed, as in the analysis of unsteady or discontinuous chip flows, to be considered in Section 7.3.3.)

Finally, Figure 7.13 shows the boundary conditions for the temperature analysis (loop II). The forward and bottom surfaces of the work are fixed at room temperature. No heat is conducted across the chip and work exit surfaces (adiabatic condition), although there is of course convection. Heat loss by convection is allowed at those surfaces surrounded by atmosphere. Heat loss by radiation is negligible in the analysis.

7.3.2 ICM simulation examples

The following is an example of the application of the ICM scheme to the two-dimensional machining of an 18%Mn–5%Cr high-hardness steel (Maekawa *et al.*, 1988). The cutting conditions used were a cutting speed of 30 m/min, an uncut chip thickness of 0.3 mm, unit cutting width, a P20 grade carbide tool with a zero rake angle and dry cutting. Figure 7.14 shows the predicted chip shape and nodal displacement vectors. Material separation at the tool tip and chip curl are successfully simulated. Figure 7.15 gives the distribution of equivalent plastic strain rate, showing where severe plastic deformation takes place. The deformation concentrates at the so-called shear plane, but is widely distributed around that plane. The secondary plastic zone is also clearly visible along the rake face, although the deformation is not as severe as in the primary zone.

These features are reflected in the temperature distribution in the chip and workpiece, as shown in Figure 7.16. A maximum temperature of more than 800˚C appears on the rake face at up to two feed distances from the tool tip.

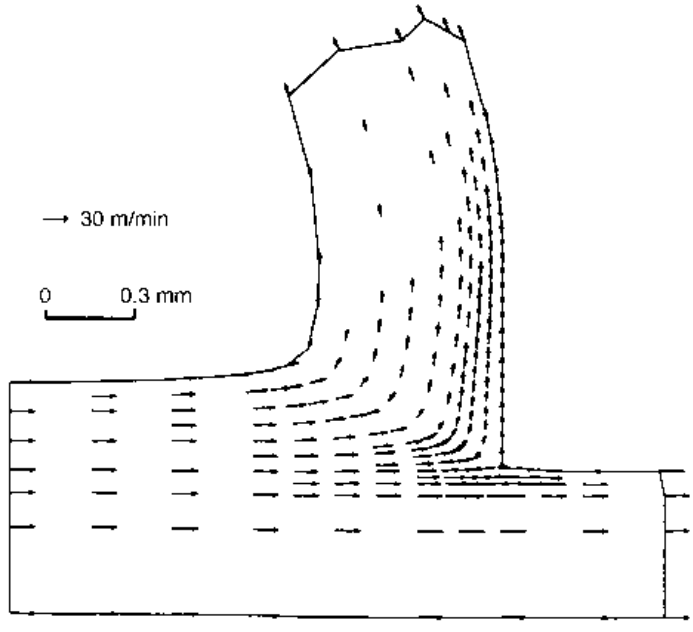


Fig. 7.14 Chip shape and velocity vectors in machining high manganese steel: cutting speed = 30 m/min, undeformed chip thickness = 0.3 mm, width of cut =1 mm, rake angle = 0°, no coolant

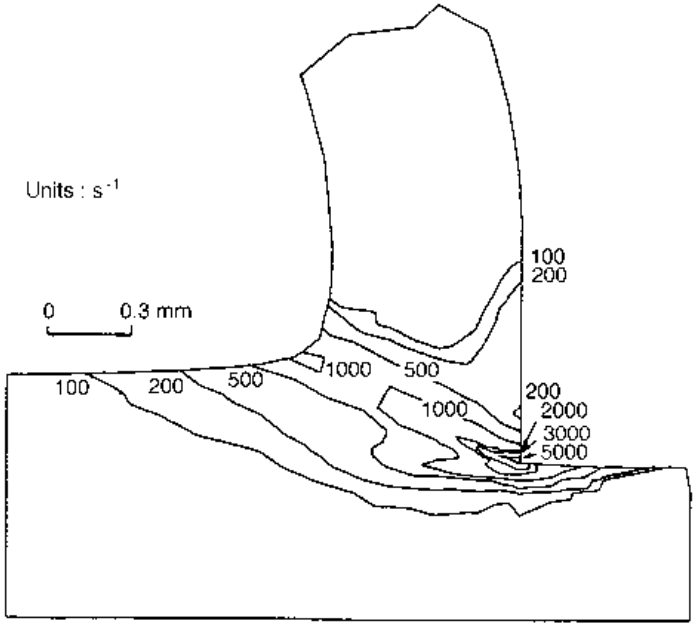


Fig. 7.15 Distribution of equivalent plastic strain rate, showing concentration of plastic deformation: cutting conditions as Figure 7.14

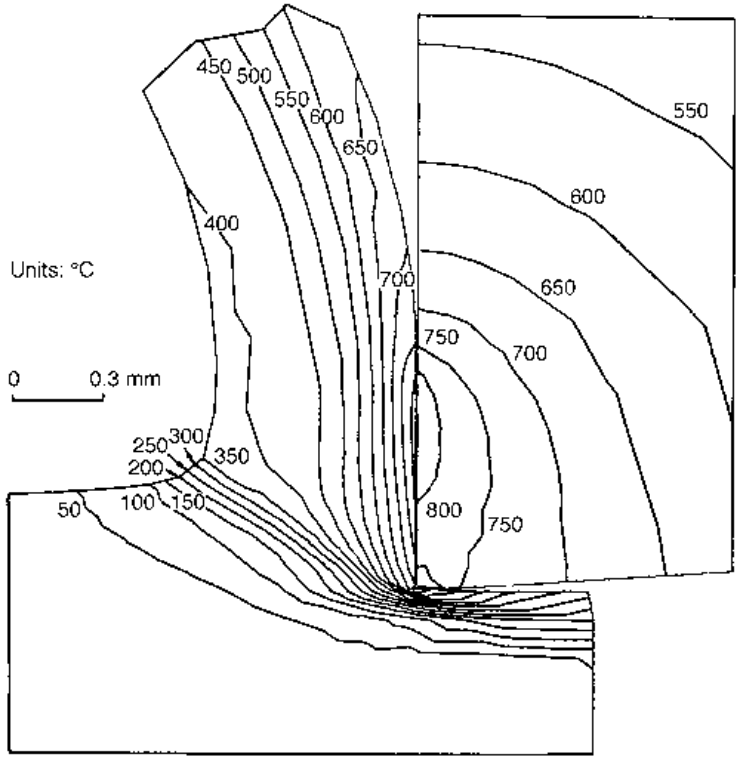


Fig. 7.16 Isotherms near the cutting tip, cutting conditions as Figure 7.14

Experimental verification has also been performed. Figure 7.17 compares the predicted and measured specific cutting forces under the same conditions (but varying speed). The observed force–velocity characteristics are well simulated. Similar agreement was confirmed in other quantities such as chip curl, rake temperature, stresses on the rake face and tool wear. For tool wear, a diffusive wear law as described in equation (4.1) was assumed (Maekawa *et al.,* 1988).

The calculation time for the ICM method depends both on the computer hardware and on the number of finite elements. In the above case, it takes only a few minutes from ICM execution to graphical presentations, using a recent high-specification PC (Pentium II, 400 MHz CPU) and an assemblage of 390 nodes and 780 triangular elements. However, a pre-processor to prepare the finite element assemblage and a post-processor to handle a large amount of data for visualization are required.

Further ICM steady flow examples will be presented in Chapter 8, together with the finite element analysis of unsteady and discontinuous chip formation. The latter requires more consideration of the chip separation criterion.

7.3.3 A treatment of unsteady chip flows

As has been written above, the ICM scheme cannot be applied to the analysis of non-steady metal machining. The iteration around an incremental small strain plastic loading

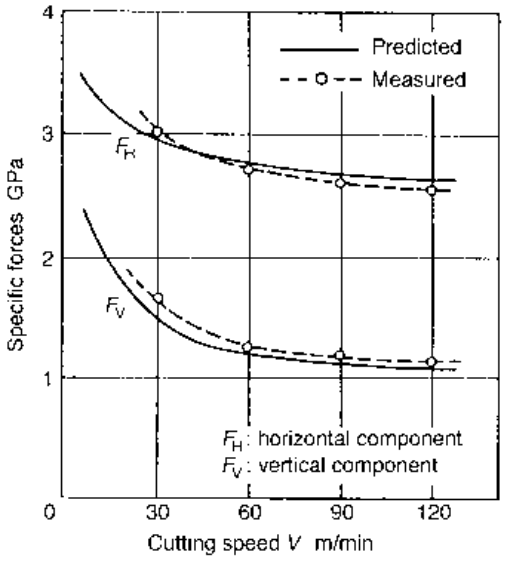


Fig. 7.17 Comparison of predicted specific forces with experiment for the same feed and rake angle as Figure 7.14, but with varying cutting speed

path closely coupled with a steady state temperature calculation (Figure 7.12) must be replaced by an incremental large strain and deformation analysis, coupled with a non-steady state temperature calculation (Appendix 2.4.4.), along the actual material loading path. Movement of the tool relative to the work over distances much greater than the feed, or uncut chip thickness, requires a way of reforming the nodes at the feed depth, as they approach the cutting edge, to form the work clearance surface and the chip surface in contact with the rake face. In addition, if the unsteady flow being treated involves fracture within the primary shear zone, a fracture criterion and a way of handling crack propagation are also needed. All these potentially require more computing power.

The examples of unsteady flow in Chapter 8.2 deal with these complications in the following ways (Obikawa and Usui, 1996; Obikawa *et al.,* 1997). Computational intensity is reduced by using meshes less refined than that shown in Figure 7.13, despite a possible loss of detail in the secondary shear region at high cutting speeds (Figure 6.12). Figure 7.18 shows the four-node quadrilateral finite element meshes used in plane strain conditions, similar to those in Figures 7.6 and 7.7. Hydrostatic pressure variations in large strain elastic–plastic analyses are dealt with easier using four-node quadrilateral than three-node triangular elements (Nagtegaal *et al.*, 1974): more detail of large strain plasticity is summarized in Obikawa and Usui (1996).

Details of node separation at the cutting edge and the propagation of a ductile primary shear fracture are shown respectively in the lower and upper parts of Figure 7.19.

Node separation

A geometrical criterion is used for node separation. A node i reforms to two nodes i and i' once its distance from the cutting edge becomes less than 1/20 of the element's side length

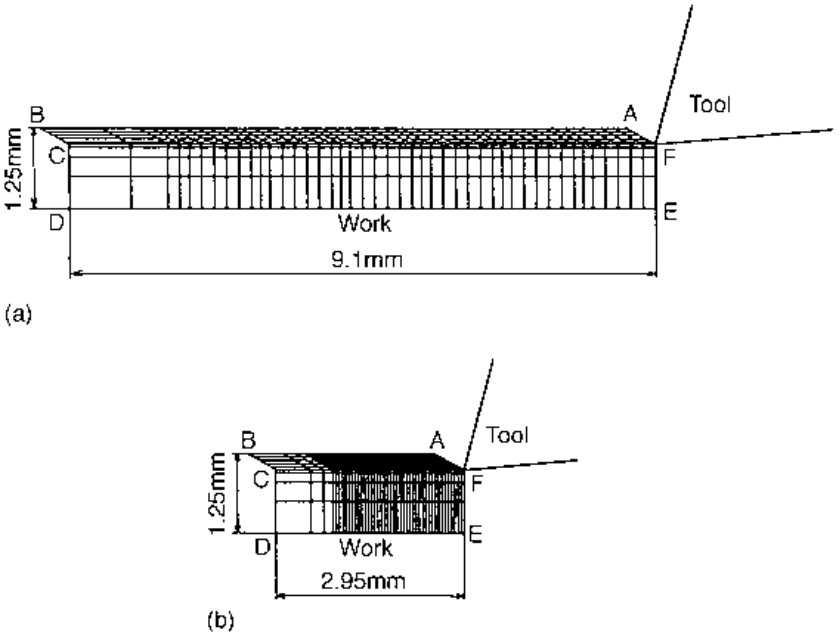


Fig. 7.18 (a) Coarse (b) fine finite element mesh

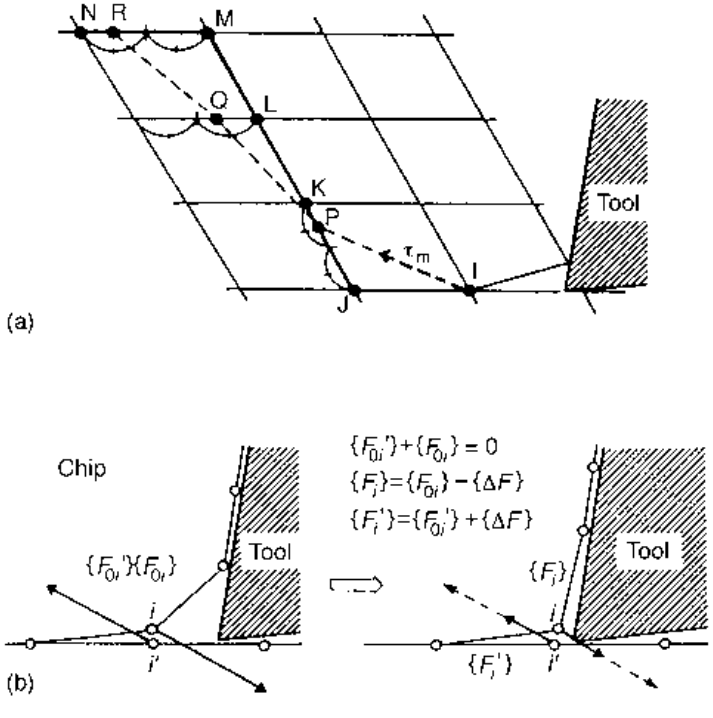


Fig. 7.19 (a) Separation of nodes within a fracturing chip; and (b) release of nodal forces at the cutting edge

(about 5 μm in the examples to be considered) and once the previously separated node has come into contact with the rake face. To avoid a sudden change in nodal forces, which can cause the computation to become unstable, the forces F_i and $F_{i'}$ acting on the separated nodes are not relaxed to zero immediately. Instead their components in the cutting direction are reduced step-by-step, under the constraint that both nodes move parallel to the cutting direction, to reach zero as i reaches the rake face (when the friction boundary condition takes over its movement). As in the ICM method, the small artificial crack at the cutting edge introduced by this procedure does not significantly alter the machining parameters.

Fracture initiation and crack growth

Shear fracture is proposed to occur if the equivalent strain exceeds an amount depending on the size of the hydrostatic pressure p (positive in compression) relative to the equivalent stress $\bar{\sigma}$, and on the absolute temperature T and equivalent strain rate $\dot{\bar{\varepsilon}}$:

$$\bar{\varepsilon} > \bar{\varepsilon}_0 + a\frac{p}{\bar{\sigma}} + f(T, \dot{\bar{\varepsilon}}) \tag{7.14}$$

where $f(T, \dot{\bar{\varepsilon}})$ causes the critical strain to increase with increasing temperature and reducing strain rate, as considered further in Chapter 8.

The upper part of Figure 7.19 shows the method of treating crack propagation, for the case of crack initiation at the cutting edge (a crack may alternatively initiate at the free surface end of the primary shear zone). If the strain at node I exceeds the limit of equation (7.14), an actual crack is assumed to propagate in the direction of the maximum shear stress τ_m to a point P. If point P is closer to node J than to K, a nominal crack is assumed to form along IJ, but if (as shown) P is closer to K, the nominal crack continues along JK to K. If the fracture limit is still exceeded at P, the actual crack continues to propagate in the direction of τ_m there, to Q; and so on to R, until the fracture criterion is no longer satisfied. The nominal crack growth, for the example shown, follows the path IJKLMN.

7.4 Material flow stress modelling for finite element analyses

Flow stress, friction and, as considered in the previous section, fracture behaviour of metals, are all required as inputs to finite element analyses. This final section of this chapter concentrates on the flow stress dependence on strain, strain rate and temperature. The reason is that most of what is known about friction in metal cutting has already been introduced in Chapter 2; and there is insufficient information about the application of ductile shear fracture criteria to machining to enable a sensible review to be made. Only on flow stress behaviour is there more currently to be written.

The topic of flow stress dependence on strain, strain-rate and temperature has been introduced in Section 6.3. There, flow stress was related to strain by a power law, with the constant of proportionality and power law exponent both being functions of strain rate and temperature (equations (6.10) and (6.14)). Comparisons were made between flow stress data deduced from machining tests and high strain-rate compression tests (Figure 6.11). Those compression tests were carried out in a high speed hammer press, driven by

compressed air, on material brought to temperature (up to 1100˚C) by pre-heating in a furnace (Oxley, 1989; from Oyane *et al.,* 1967).

Pre-heating in a furnace allows a material's microstructure to come into thermal equilibrium. This differs from the conditions experienced in metal machining. There, metal is heated and passes through the deforming region in the order of milliseconds. The microstructures of chips, in the hot secondary shear region, appear heavily cold worked and not largely recovered or recrystallized. For steels, traces of austenitization and quenching are hardly ever seen, even though secondary shear temperatures are calculated to be high enough for that to occur for longer heating times. The ideal mechanical testing of metals for machining applications involves high heating rates as well as strain rates.

7.4.1 High heating-rate and strain-rate mechanical testing

Such testing has been developed by Shirakashi *et al*. (1983). A Hopkinson bar creates strain rates up to 2000 s^{-1} in a cylindrical sample of metal (6 mm diameter by 10 mm long). Induction heating and a quench tank heat and cool the sample within a 5 s cycle. A stopping ring limits the strain per cycle to 0.05: multiple cycling allows the effect of strain path (varying strain rate and temperature along the path) on flow stress to be studied. Figure 7.20 shows the principle of the test, with a measured temperature/time result of heating a 0.15%C steel to 600˚C.

Subsidiary tests show that a single sample can be heated for up to 90 s at temperatures up to 680˚C before thermal annealing or age hardening modifies the flow stress generated by straining. Thus, 20 cycles, each taking 5 s, developing a strain up to 1, can be achieved

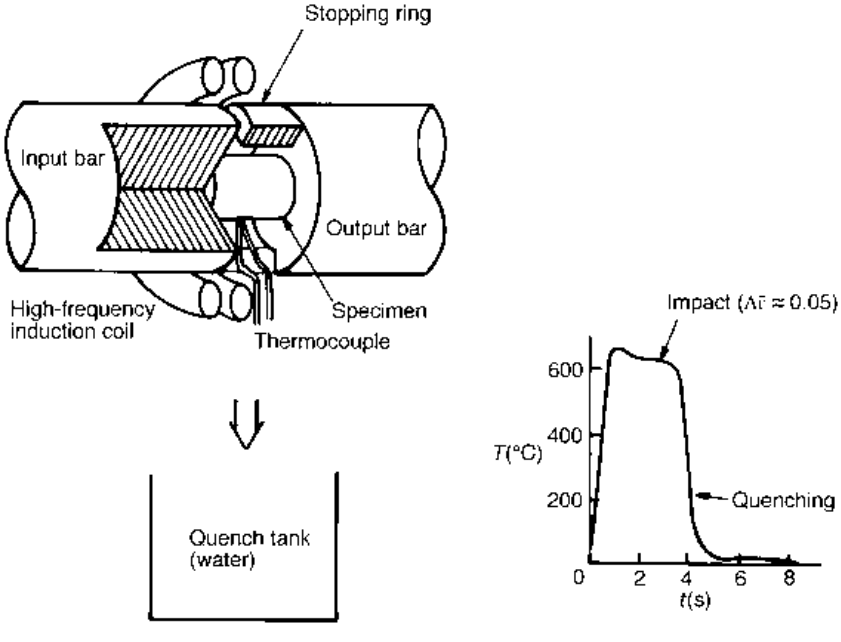


Fig. 7.20 Principle of the rapid heating and quenching high strain rate test (after Shirakashi *et al.*, 1983)

before the time at which temperature degrades the results. Phase transformation prevents useful testing above 720˚C. Even testing at strains up to 1, strain rates up to 2000 and temperatures up to ≈ 700˚C (for steels) does not reach metal cutting secondary shear conditions, but it is the closest yet achieved.

With this equipment, the flow stresses of a range of carbon and low alloy steels have been measured. Varying both strain rate and temperature along a strain path has been observed to influence the stress/strain curve. An empirical equation to represent this has been developed:

$$\bar{\sigma} = A\left(\frac{\dot{\bar{\varepsilon}}}{1000}\right)^{M} \mathrm{e}^{aT}\left(\frac{\dot{\bar{\varepsilon}}}{1000}\right)^{m}\left(\int_{\text{strain path}} \mathrm{e}^{-aT/N}\left(\frac{\dot{\bar{\varepsilon}}}{1000}\right)^{-m/N} \mathrm{d}\bar{\varepsilon}\right)^{N} \tag{7.15a}$$

When straining takes place at constant strain rate and temperature, it reduces to:

$$\bar{\sigma} = A\left(\frac{\dot{\bar{\varepsilon}}}{1000}\right)^{M} \bar{\varepsilon}^{N} \tag{7.15b}$$

where A, M and N may all vary with temperature. Measured values are given in Appendix 4.3. Figure 7.21 gives example results for a low alloy steel (the 0.36C-Cr-Mo-Ni material of Table A4.4).

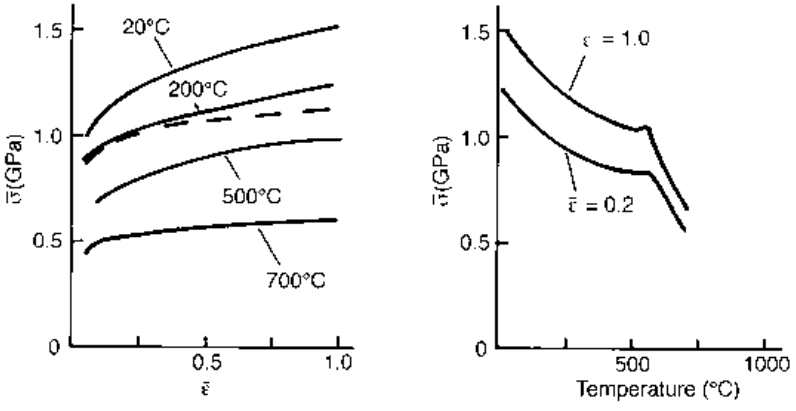


Fig. 7.21 Flow stress behaviour of a low alloy steel: dashed line at 20ºC and a strain rate of 10^{-3} s^{-1}; solid lines at strain rate of 10^{3} s^{-1} and temperatures as marked

The Hopkinson bar equipment has established different laws for non-ferrous face centred cubic metals such as aluminium and α-brass. A much greater strain rate path effect and no temperature path effect has been observed (Usui and Shirakashi, 1982). At temperatures, T˚C, up to about 300˚C (higher temperature data would be useful but is not reported)

$$\bar{\sigma} = A\left(\mathrm{e}^{-\frac{B}{T+273}}\right)\left(\frac{\dot{\bar{\varepsilon}}}{1000}\right)^{M}\left(\int_{\text{strain path}}\left(\frac{\dot{\bar{\varepsilon}}}{1000}\right)^{m} \mathrm{d}\bar{\varepsilon}\right)^{N} \tag{7.16a}$$

which, at constant strain rate, simplifies to the form

$$\bar{\sigma} = A\left(\mathrm{e}^{-\frac{B}{T+273}}\right)\left(\frac{\dot{\bar{\varepsilon}}}{1000}\right)^{M^*}\bar{\varepsilon}^{N} \tag{7.16b}$$

Coefficients in these equations, with data for other alloys too, are also given in Appendix 4.3.

When flow stress data from these Hopkinson bar tests are used in machining simulations in which the predicted temperatures do not rise too far above the ranges to which the data apply, satisfactory agreement with experiments is usually obtained (as will be seen in later chapters). However, with the increasing capabilities of tool materials to withstand high temperatures and the consequent increase of practical cutting speeds, there is a need to extend the range of validity of flow stress equations.

7.4.2 Other approaches to flow stress modelling

A number of the finite element studies reported in Section 7.2 (Rakotomolala *et al*., 1993, Sekhon and Chenot, 1993, Marusich and Ortiz, 1995) have used the empirical flow stress equations first used in dynamic impact applications, combining power law strain hardening, power law or logarithmic strain-rate effects and linear or power law thermal softening. Two examples are

$$\left.\begin{aligned}\bar{\sigma} &= (A + B\bar{\varepsilon}^{\mathrm{n}})(1 + C1\mathrm{n}\dot{\bar{\varepsilon}})(1 - [(T - T_{\mathrm{amb}})/(T_{\mathrm{melt}} - T_{\mathrm{amb}})]^{m}\\ \bar{\sigma} &= \sigma_0(1 + (\bar{\varepsilon}/\varepsilon_0))^{n}(1 + (\dot{\bar{\varepsilon}}/\dot{\bar{\varepsilon}}_0))^{m}(1 - \alpha(T - T_0))\end{aligned}\right\} \tag{7.17}$$

where the coefficients are either constants or change step-wise with strain rate. These equations have been suggested for computational convenience. As each has four or five adjustable coefficients, they may be able to be trained to give realistic simulations over limited ranges of cutting speed and feed, but they are too simple, compared with observations (for example Figure 7.21) with respect to variations of flow stress with temperature. They do not allow (without modification) modelling the strain path effects that are significant, particularly for face centred cubic (f.c.c.) metals (equation 7.16(a)).

Another approach developed for dynamic impact applications, but applied only recently in machining simulations, is to use flow stress equations based on dislocation mechanics fundamentals. Zerilli and Armstrong (1987) have suggested that the flow stress variations of f.c.c. and b.c.c. metals with strain, strain rate and temperature should take the forms (with temperature T Kelvin):

$$\left.\begin{aligned}&\text{for f.c.c. metals} \quad \bar{\sigma} = C_1 + C_2\bar{\varepsilon}^{0.5}\exp[(-\mathrm{C}_3 + \mathrm{C}_4 1\mathrm{n}\dot{\bar{\varepsilon}})T]\\ &\text{for b.c.c. metals} \quad \bar{\sigma} = C_1 + C_5\bar{\varepsilon}^{n} + C_2\exp[(-C_3 + C_4 1\mathrm{n}\dot{\bar{\varepsilon}})T]\end{aligned}\right\} \tag{7.18}$$

Both these combine strain rate and temperature in the velocity modified temperature form (Chapter 6, equation (6.14)). The form for b.c.c. metals suggests that the dependence of flow stress on strain hardening should not depend on temperature. Figure 7.21 shows this to be the case up to about 600°C but not to be true at higher temperatures. Recently, Goldthorpe *et al.* (1994) have suggested a modification for b.c.c. metals that introduces a temperature dependence of strain hardening, through the reduction of a metal's elastic shear modulus, G, with temperature:

$$\bar{\sigma} = (C_1 + C_5\bar{\varepsilon}^n)(G_T/G_{293}) + C_2 \exp[(-C_3 + C_4 1n\dot{\bar{\varepsilon}})T] \tag{7.19}$$

where, for steels $(G_T/G_{293}) \approx 1.13 - 0.000445T$.

A question arises about extrapolation of these, and other power law equations, to strains much greater than 1. (In Section 6.3, Oxley's assumption that strain hardening ceased for strains greater than 1 was mentioned.) Zerilli and Armstrong (1997), in the context of modelling the behaviour of a titanium alloy, suggest that strain could be replaced by a form that saturates at a limiting, or recovery, value ε_r:

$$\bar{\varepsilon} \Rightarrow \varepsilon_r(1 - \exp[-\bar{\varepsilon}/\varepsilon_r]) \tag{7.20}$$

Gradually, experience of the formulation of flow stress equations for broader ranges of strain, strain rates and temperatures is growing (path dependence remains undeveloped). It can be anticipated that useful fundamentally-based equations for metal machining applications will be developed over the coming years. Eventually the goal of relating flow behaviour to a metal's composition and microstructure will be reached. However today, the empirical forms outlined in Section 7.3.1 are the best validated that are available.

References

Ceretti, E., Fallbohmer, P., Wu, W. T. and Altan, T. (1996) Application of 2-D FEM to chip formation in orthogonal cutting. *J. Materials Processing Tech.* **59**, 169–181.

Childs, T. H. C. and Maekawa, K. (1990) Computer aided simulation and experimental studies of chip flow and tool wear in turning low alloy steels by cemented carbide tools. *Wear* **139**, 235–250.

Goldthorpe, B. D., Butler, A. L. and Church, P. (1994) A wide range constitutive equation for medium and high strength steel. *J. de Physique IV* **4**(C8), 471–476.

Iwata, K., Osakada, K. and Terasaka, Y. (1984) Process modeling of orthogonal cutting by the rigid-plastic finite element method. *Trans ASME J. Eng. Mat. Tech.* **106**, 132–138.

Kakino, Y. (1971) Analysis of the mechanism of orthogonal machining by the finite element method. *J. Japan Soc. Prec. Eng.* **37**(7), 503–508.

Maekawa, K. and Maeda, M. (1993) Simulation analysis of three-dimensional continuous chip formation processes (1st. report) – FEM formulation and a few results. *J. Japan Soc. Prec. Eng.* **59**(11), 1827–1833.

Maekawa, K., Kubo, A. and Kitagawa, T. (1988) Simulation analysis of cutting mechanism in plasma hot machining of high manganese steels, *Bull. Japan Soc. Prec. Eng.* **22**(3), 183–189.

Maekawa, K., Ohhata, H. and Kitagawa, T. (1994) Simulation analysis of cutting performance of a three-dimensional cut-away tool. In Usui, E. (ed.*), Advancement of Intelligent Production.* Tokyo: Elsevier, pp. 378–383.

Marusich, T. D. and Ortiz, M. (1995) Modelling and simulation of high speed machining. *Int. J. Num. Methods in Engineering* **38**, 3675–3694.

Nagtegaal, J. C., Parks, D. M. and Rice, J. R. (1974) On numerically accurate finite element solutions in the fully plastic range. *Comp. Meth. Appl. Mech. Eng.* **4**, 153–177.

Obikawa, T. and Usui, E. (1996) Computational machining of titanium alloy – finite element modeling and a few results. *Trans ASME J. Manufacturing Sci. Eng.* **118**, 208–215.

Obikawa, T., Sasahara, H., Shirakashi, T. and Usui, E. (1997) Application of computational machining method to discontinuous chip formation. *Trans ASME J. Manufacturing Sci. Eng.* **119**, 667–674.

Oxley, P. L. B. (1989) *Mechanics of Machining*, Chicester: Ellis Horwood.

Oyane, M., Takashima, F., Osakada, K. and Tanaka, H. (1967) The behaviour of some steels under dynamic compression. In: *10th Japan Congress on Testing Materials*, pp. 72–76.

Rakotomolala, R., Joyot, P. and Touratier, M. (1993) Arbitrary Lagrangian-Eulerian thermomechanical finite-element modelof material cutting. *Comm. in Num. Methods in Engng.* **9**, 975–987.

Sasahara, H., Obikawa, T. and Shirakashi, T. (1994) FEM analysis on three dimensional cutting – analysis on large deformation problem of tool entry. *Int. J. Japan Soc. Prec. Eng.* **28**(2), 123–128.

Sekhon, G. S. and Chenot, S. (1993) Numerical simulation of continuous chip formation during non-steady orthogonal cutting. *Engineering Computations* **10**, 31–48.

Shirakashi, T. and Usui, E. (1976) Simulation analysis of orthogonal metal cutting process. *J. Japan Soc. Prec. Eng.* **42**(5), 340–345.

Shirakashi, T., Maekawa, K. and Usui, E. (1983) Flow stress of low carbon steel at high temperature and strain rate (Part 1) – propriety of incremental strain method in impact compression test with rapid heating and cooling systems. *Bull. Japan Soc. of Prec. Eng.* **17**(3), 161–166.

Strenkowski, J. S. and Carrol III, J. T. (1985) A finite element model of orthogonal metal cutting. *Trans ASME J. Eng. Ind.* **107**, 349–354.

Trent, E. M. (1991) *Metal Cutting*, 3rd edn. Oxford: Butterworth Heinemann.

Ueda, K. and Sugita, T. (1983) Application of fracture mechanics in micro-cutting of engineering ceramics. *Annals CIRP* **32**(1), 83–86.

Ueda, K. and Manabe, K. (1993) Rigid-plastic FEM analysis of three-dimensional deformation field in chip formation process. *Annals CIRP* **42**(1), 35–38.

Ueda, K., Sugita, T. and Hiraga, H. (1991) A J-integral approach to material removal mechanisms in microcutting of ceramics. *Annals CIRP* **40**(1), 61–64.

Ueda, K., Manabe, K. and Nozaki, S. (1996) Rigid-plastic FEM of three-dimensional cutting mechanism (2nd report) – simulation of plain milling process. *J. Japan Soc. Prec. Eng.* **62**(4), 526–531.

Usui, E. and Shirakashi, T. (1982) Mechanics of machining – from descriptive to predictive theory. *ASME Publication* **PED 7**, 13–35.

Usui, E., Shirakashi, T. and Obikawa, T. (1977) Simulation analysis of cutting fluid action. *J. Japan Soc. Prec. Eng.* **43**(9), 1063–1068.

Usui, E., Maekawa, K. and Shirakashi, T. (1981) Simulation analysis of built-up edge formation in machining of low carbon steel. *Bull. Japan Soc. Prec. Eng.* **15**(4), 237–242.

Zienkiewicz, O. C. (1971) *The Finite Eelement Method in Engineering Science* 2nd edn. Ch. 18. London: McGraw-Hill.

Zerilli, F. J. and Armstrong, R. W. (1987) Dislocation-mechanics based constitutive relations for material dynamics calculations. *J. Appl. Phys*. **61**, 1816–1825.

Zerilli, F. J. and Armstrong, R. W. (1997) Dislocation mechanics based analysis of materials dynamics behaviour. *J. de Physique IV* **7**(C8), 637–648.

8

Applications of finite element analysis

In this chapter, a number of special topics are considered as examples of applications to which finite element methods have already contributed. Built-up edge (BUE) and serrated chip flows were introduced in Chapter 2 (Figure 2.4) but have hardly been mentioned since. The unsteady nature of their flows makes their effective analysis by classical methods impractical. BUE formation is considered in Section 8.1, and discontinuous/unsteady chip formation, including serrated flows, in Section 8.2. The development of free-machining steels remains an important application area for manufacturing industry: the correlation of machinability with other materials' properties through finite element studies is the topic of Section 8.3. The reality that many cutting tools do not have plane rake faces was introduced in Chapter 3 (Section 3.2.8) but this too has not been considered since. Section 8.4 introduces finite element analyses of chip control and the effects of cutting edge shape that have a potential to support rational tool design.

8.1 Simulation of BUE formation

Built-up edges occur at some cutting speed or other in machining most metal alloys containing more than one phase, as machining conditions change from low speed, at which no significant heating occurs, to high speed, when the secondary shear zone becomes too hot to support BUEs (Williams *et al.* 1970). Figures 2.4(d), (e) and (f) show the progression from a heavily cracked chip flow, through BUE formation, to steady flow, in the machining of a 0.15%C steel at a feed of 0.15 mm, as the cutting speed increases from 5 m/min to 55 m/min. Figure 3.14 follows the associated changes in cutting forces and shear plane angle. Many researchers have investigated the effects of cutting temperature, work hardening (and work softening) and adhesion between the chip and tool on the formation and disappearance of the BUE (Pekelharing, 1974). All these factors have some influence, by and large. It is clear that the BUE is an unstable formation. It repeatedly nucleates, grows and breaks away in fragments from the tool, with the disadvantages, among others, that the machined surface is degraded and tool wear (certainly by chipping) is increased.

One point arises concerning the mechanism of nucleation. Is it by a steady secondary shear flow, leading to a continuous pile up of laminates on the tool rake face (Trent, 1963)? Or is it by discrete fractures in the secondary shear zone leading to discontinuous separations of the BUE from the main body of the chip (Shaw *et al.*, 1961)? The next sections

address these questions with the aid of the ICM finite element method (Usui *et al.*, 1981). As the ICM method can only follow steady state chip flows (Chapter 7.3), it is more accurate to write that it is used to assess incipient BUE formation: it supports the discrete fracture viewpoint.

8.1.1 The simulation model

Orthogonal dry machining of a 0.18%C plain carbon steel by a P20-grade carbide tool has been simulated, at cutting speeds from 75 m/min (above BUE formation) down to 30 m/min (within the BUE range), at a feed of 0.3 mm, a rake angle of 10° and a clearance angle of 6° (to match experimental machining studies that were also carried out). The finite element assemblage used was that shown in Figure 7.13.

Flow properties of the workpiece were obtained by the Hopkinson-bar method and formulated according to equation 7.15(a) (also see Appendix 4). Friction at the tool–chip interface was also measured (at a cutting speed of 46 m/min) using the split tool method (Chapter 5): the results were fitted to equation (2.24c), taking k to be the local shear flow stress at the rake face; and with $m = 1$ and $\mu = 1.6$.

8.1.2 Orthogonal machining without BUE

Figure 8.1 shows the predicted chip shape and other quantities at the high cutting speed of 75 m/min. Figure 8.1(a) shows the pattern of distorted grid lines calculated from the nodal

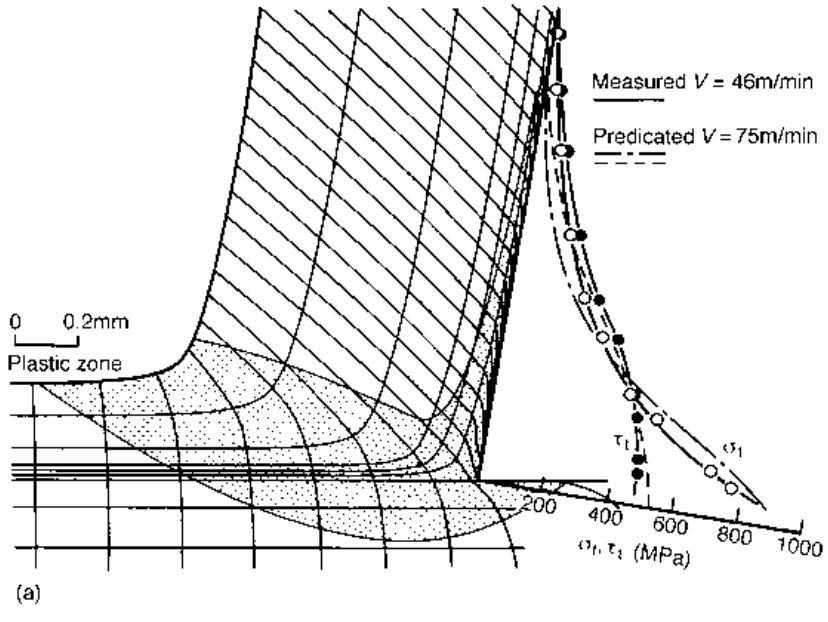


Fig. 8.1 Simulated machining of a 0.18%C steel at a cutting speed of 75 m/min where no BUE appears: (a) distorted grid pattern and normal stress and friction stresses on the tool rake face, (b) distributions of maximum shear strain γ_m and strain rate $\dot{\gamma}_m$ and (c) distributions of shear flow stress k and temperature T (°C)

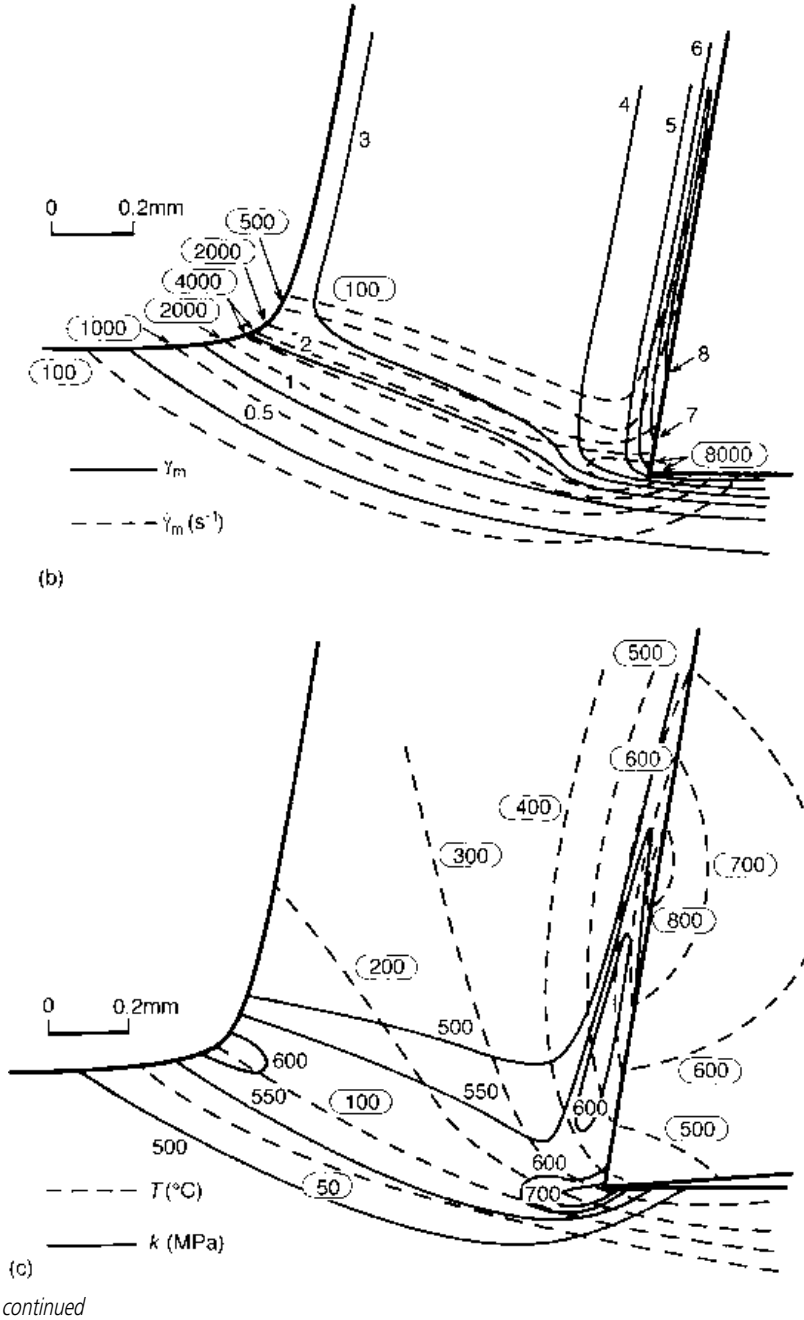


Fig. 8.1 *continued*

velocities along the flow lines (following the method of Johnson and Kudo, 1962). The grey area represents the plastic deformation zone. The predicted normal stress σ_t and friction stress τ_t along the rake face are in reasonable agreement with the measured ones, considering the simulation and experiments were not carried out at exactly the same cutting speeds. (The split-tool measurement was carried out using a planer, the maximum speed of which was 46 m/min.)

Figure 8.1(b) shows the calculated distributions of maximum shear strain γ_m and strain rate $\dot{\gamma}_m$. Deformation is concentrated, as expected, near the shear plane and along the rake face. The strain rate reaches 8000 s^{-1} in front of the cutting edge, and its time integration along the flow lines yields a maximum strain of more than 8 on the rake face.

Figure 8.1(c) shows the distributions of shear flow stress k and temperature T within the chip and tool. The magnitude of the stress rises in the deformation zone because of work hardening, but it is limited to 700 MPa by thermal softening.

8.1.3 Orthogonal machining with BUE

When the cutting speed is reduced to 30 m/min (and a BUE appears in practice), different phenomena appear in the chip. Figure 8.2 first shows an intermediate stage of the ICM iteration, before full convergence of the flow has occurred. In the primary shear region shown in Figure 8.2(a), the directions of the nodal velocities, as indicated by the arrows, are already in reasonable agreement with the reformed streamlines, but this is not the case in the secondary shear zone.

In the secondary shear zone, localized areas of relatively high strain rate are developing. Figure 8.2(b) highlights, by shading, two regions – one attached to the rake face, one separated from it – in which the strain rate $\dot{\gamma}_m$ is 400 s^{-1}, greater than the value of 100 s^{-1}, also shaded, nearer to the cutting edge. This may be contrasted with the strain rate distribution

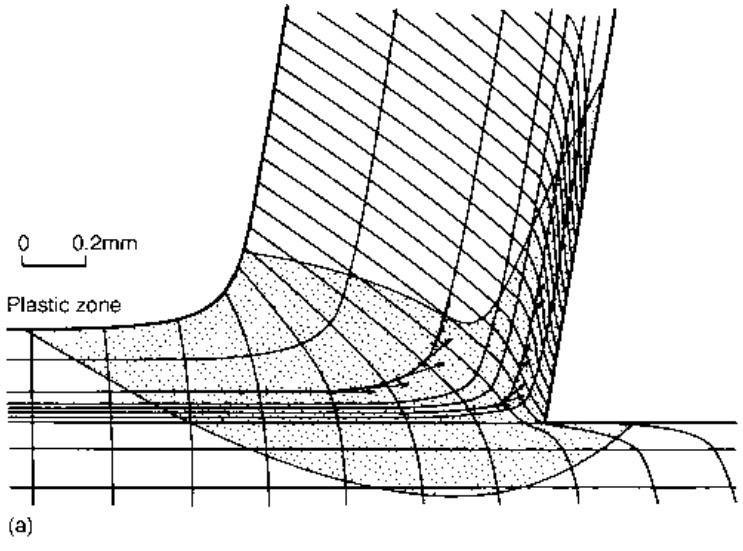


Fig. 8.2 As for Figure 8.1 but at a cutting speed of 30 m/min; and just before convergence of the computation

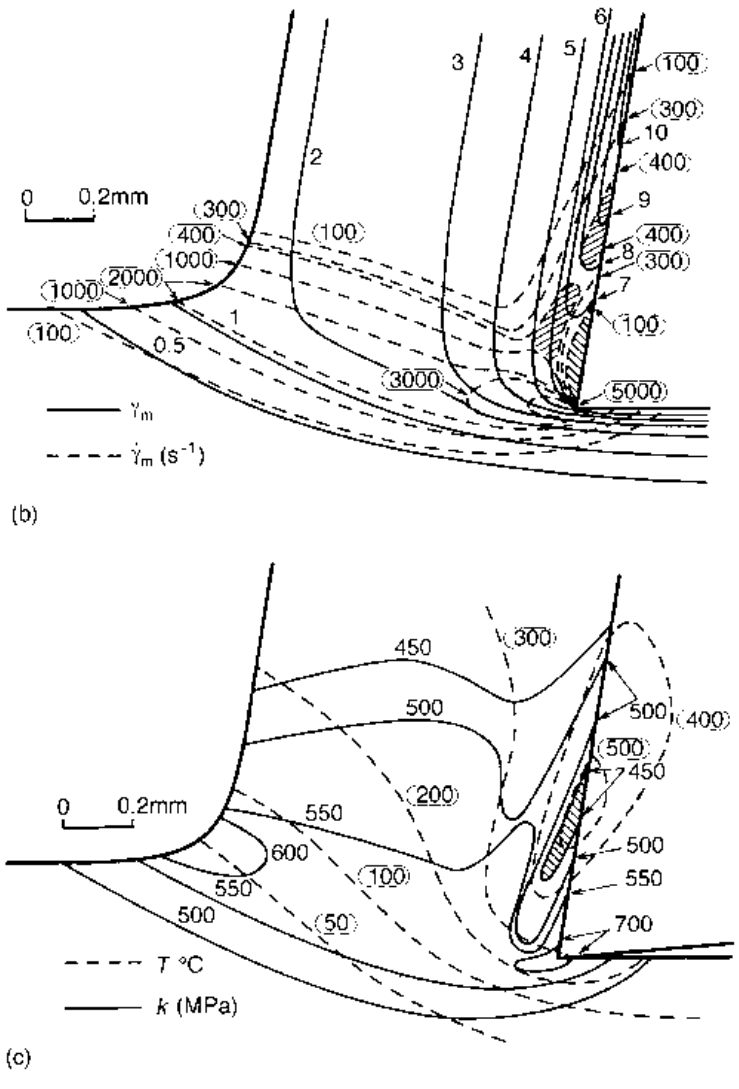


Fig. 8.2 *continued*

at 75 m/min (Figure 8.1(b)) which shows strain rates steadily decreasing with increasing distance from the cutting edge.

The distribution of the shear flow stress (Figure 8.2(c)) follows the uneven distribution of $\dot{\gamma}_m$: a shear flow stress minimum (shaded region) of $k = 450$ MPa appears in the region where $\dot{\gamma}_m = 400\ s^{-1}$.

As the simulation continues to convergence (Figure 8.3), the localization of secondary shear strain rate and shear stress becomes stronger. Although no obvious change in chip thickness or shear plane angle can be seen, the secondary deformation becomes concentrated (Figure 8.3(b)) to form a narrow band, partly separated from the rake face, in which there is a very high shear strain, $\gamma_m = 16$ (or $\bar{\varepsilon} = 9.2$). A relatively low hydrostatic pressure

also exists there (the dashed lines in Figure 8.3(b)). In contrast to the intermediate state (Figure 8.2), the shear flow stress in the band has become increased relative to its surroundings, by work hardening, as shown in Figure 8.3(c).

These changes are all favourable to the separation of the flow by shear fracture, to

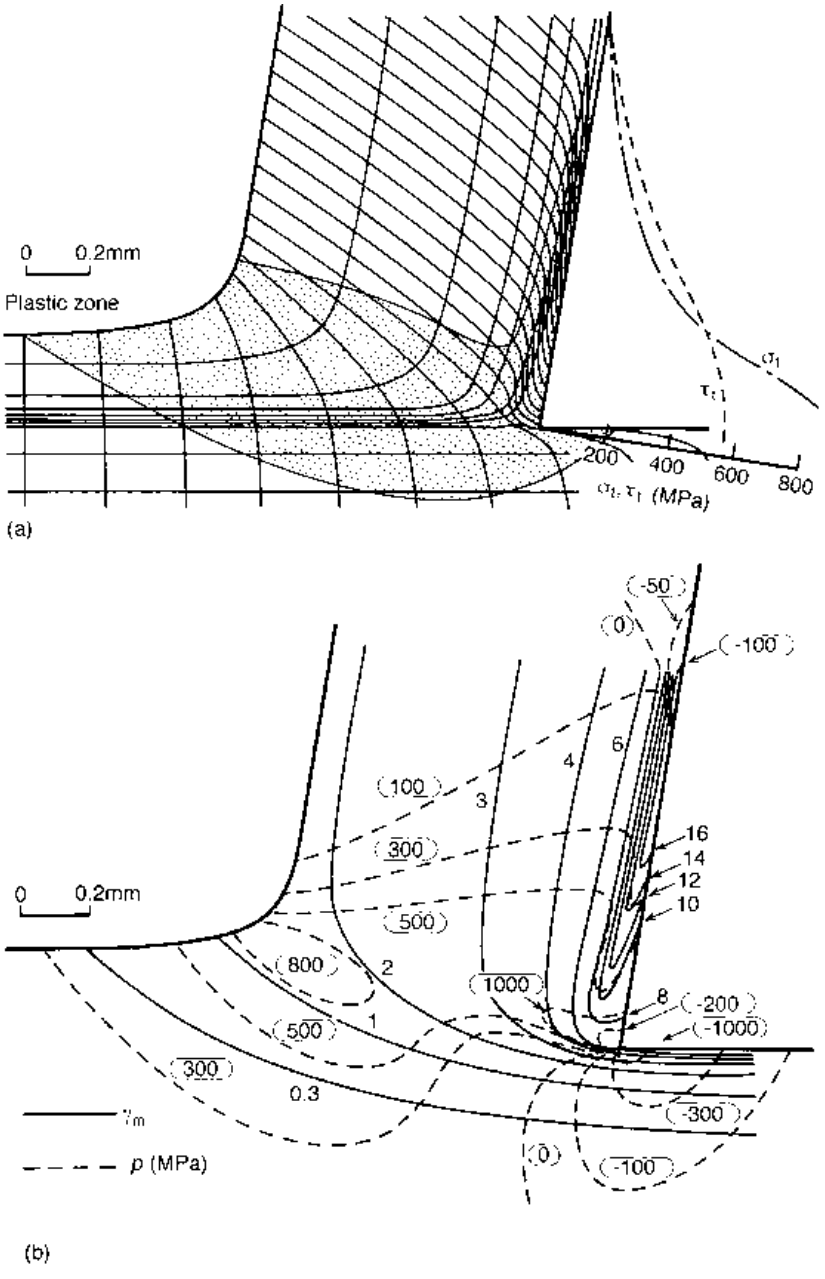


Fig. 8.3 Converged results at the cutting speed of 30 m/min: (a) and (c) as in Figures 8.1 and 8.2 but (b) distribution of γ_m and hydrostatic pressure p

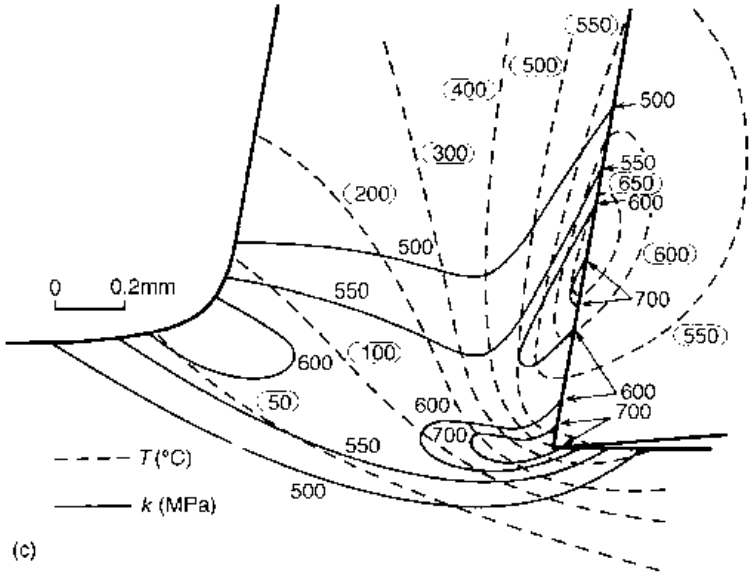


Fig. 8.3 *continued*

generate the nucleus of a BUE. If nucleation occurs, debris stuck to the rake face will have enough hardness to resist loading by the chip body.

Figure 8.4 is a quick-stop observation showing that separation can occur in the chip close to the cutting edge. The ICM simulation described here, apart from only dealing with steady states, has no chip separation criterion within it. If a reliable fracture criterion were available, not only the accumulation of nuclei or the growth of the BUE but also its breakage might be simulated by an extended finite element method.

8.1.4 The role of blue brittleness

Why does the deformation concentrate at $V = 30$ m/min? The primary cause, for steels, can be attributed to blue brittleness. The effect of blue brittleness is expressed in the term

$$A\left(\frac{\dot{\bar{\varepsilon}}}{1000}\right)^{M}$$

in equation (7.15b) (omitting, for simplicity, the path dependence effects in equation (7.15a)). Figure 8.5 shows the relation between flow stress and temperature measured for the 0.18%C steel. Between 400˚C and 600˚C the flow stress increases with temperature (the blue brittle effect). In Figure 8.3(c) the temperature at the boundary between the stagnant secondary zone and the main body of the chip is in the same range. Since carbon steels become brittle near the peak flow stress temperature, fracture is most likely to occur in this condition.

At the cutting speed of 75 m/min, no BUE appears (Figure 8.1). Figure 8.5 indicates that the temperature along the rake face is beyond the blue brittle range. Thermal softening

Fig. 8.4 Experimental distorted grid pattern from a quick stop test at a cutting speed of 25 m/min, f=0.16 mm, d=4 mm, α=10° and without coolant

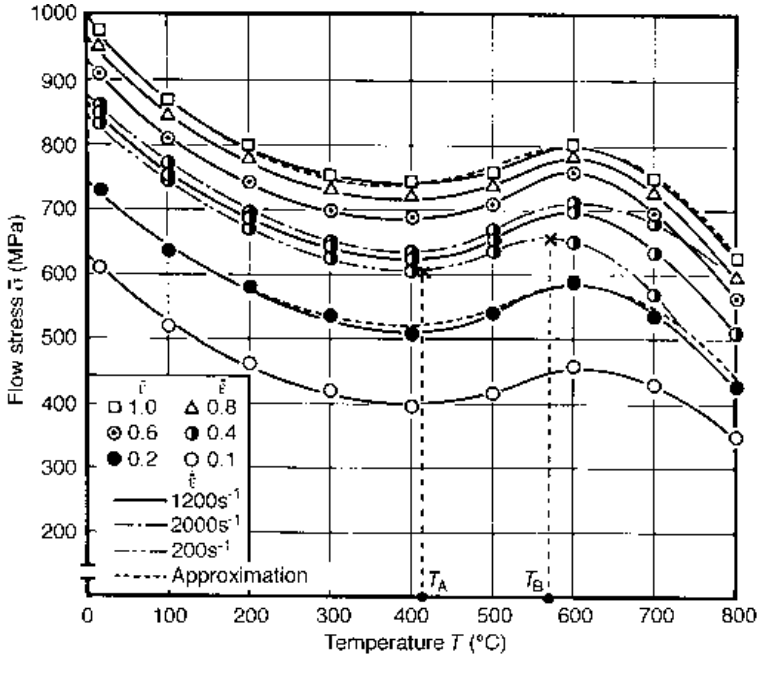


Fig. 8.5 Relation between flow stress and temperature of the 0.18%C steel

above 600˚C is so steep that deformation occurs easily. The secondary flow zone grid lines in Figure 8.1(a), compared with those in Figure 8.3(a), indicate the collapse of the BUE-range stagnant flow. The almost uniform secondary shear flow stress in Figure 8.1(c) can be attributed to compensation between work hardening and thermal softening. It indicates why, despite varying strain, strain rate and temperature along the rake face, split-tool tests show a plateau friction stress almost independent of distance from the cutting edge (although this does, of course, depend on the constitutive law chosen for the simulation, as has been discussed in Chapter 7.4).

In summary, the BUE formation process in steels has successfully been simulated using the finite element method. Under practical cutting conditions where a BUE appears, the chip flow property characterized by blue brittleness assists in developing the secondary shear flow into a stagnant zone. At the boundary between the developed stagnant flow and the main body of the chip, conditions of high strain concentration, low hydrostatic pressure and material brittleness are favourable for the separation of flow to form the nucleus of a BUE. The stagnant flow degenerates at higher cutting speeds because thermal softening prevails over work hardening.

8.2 Simulation of unsteady chip formation

Three examples of unsteady chip formation are described: (1) chip flow, force and residual stress variations in the low speed (13 mm/min) machining of a β-brass (60%Cu–40%Zn), in conditions that lead to discontinuous chip formation (Obikawa *et al.*, 1997); (2) changes in chip formation, and resulting changes in tool fracture probability, during transient chip flow at the end of a cut, for the low speed machining of a different β-brass, in conditions which give continuous chip formation (Usui *et al.,* 1990); and (3) serrated chip formation in machining a Ti-6Al-4V alloy (Obikawa and Usui, 1996). The treatment of unsteady flow is as outlined in Chapter 7.3.3.

Low strain rate mechanical testing showed both brass materials had the same work-hardening behaviour, but that which gave discontinuous chips was less ductile than the other. The low cutting speed of the application means that the effects of strain rate and temperature on flow stress can be neglected. However, it is found that the distribution of strain rate in the primary shear zone influences where a crack initiates – and the dependence of shear fracture on this cannot be neglected. The following expressions for flow stress $\bar{\sigma}$ dependence on strain $\bar{\varepsilon}$ and of fracture criterion on hydrostatic pressure p and strain rate $\dot{\bar{\varepsilon}}$ (relative to cutting speed, to accommodate the distribution effect) are used for positive shear of both brasses in the finite element analysis:

$$\bar{\sigma}\,(\text{MPa}) = 740(\bar{\varepsilon} + 0.01)^{0.27}; \qquad \bar{\varepsilon} \geq a + 0.4\,\frac{p}{\bar{\sigma}} - 0.01\,\frac{\dot{\bar{\varepsilon}}}{V} \tag{8.1a}$$

with $a = 1.57$ for the less, and 10.0 for the more, ductile material, and V the cutting speed in mm/s. Friction between the chip and the tool is modelled according to equation (2.24c), with m and μ both equal to 1.

The fracture due to negative shear at the end of a cut occurs under mixed modes: tensile mode I and shear mode II. The latter is the predominant mode, but the former accelerates crack propagation. Under the conditions that strain rate due to positive shear is less than

that due to negative shear and that a crack nucleates only in the negative shear region, another criterion is applied for the negative shear fracture (Obikawa *et al.*, 1990):

$$\bar{\varepsilon} \geq 1.1 + 0.3\,\frac{p}{\bar{\sigma}} \tag{8.1b}$$

For the Ti-alloy example, strain rate and temperature effects cannot be ignored. The material's flow stress is given in Appendix 4; the shear fracture criterion used is

$$\bar{\varepsilon} \geq \frac{p(\text{MPa})}{12\,600} + 0.09\,\exp\left(\frac{T(\text{K})}{293}\right) - \text{MAX}\left[0.075\log\left(\frac{\dot{\bar{\varepsilon}}}{100}\right),0\right] \tag{8.2}$$

where MAX[,] means the greater of the two choices. Rake face friction is modelled in the same way as for the β-brass, with $m = 1$ but $\mu = 0.6$. (The fracture criterion and that for the β-brass are empirically developed – further developments may be expected in the coming years, in parallel with flow stress modelling improvements as described in Chapter 7.)

8.2.1 Discontinuous chip formation with a β-brass

Figure 8.6 shows the chip formation predicted at different cut distances L for the β-brass, with the material properties of equation (8.1a), machined with a carbide tool of rake angle 15°, at a feed of 0.25 mm. A shear-type discontinuous chip is simulated, with a crack initiating periodically at the tool side of the chip, within the highly deformed workpiece, and propagating towards the free surface side. Figure 8.7 shows the pattern of changing cutting forces. Both horizontal and vertical components increase with cut distance, up to the point where a crack initiates. The crack propagates, accompanied by falling forces. It finally

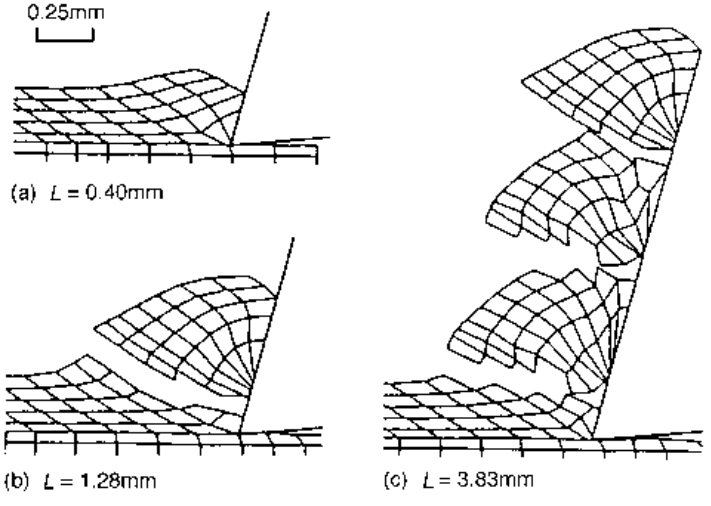


Fig. 8.6 Predicted discontinuous chip in β-brass machining: cutting speed of 13 mm/min, f=0.25 mm, d=1 mm, α=15° and no coolant

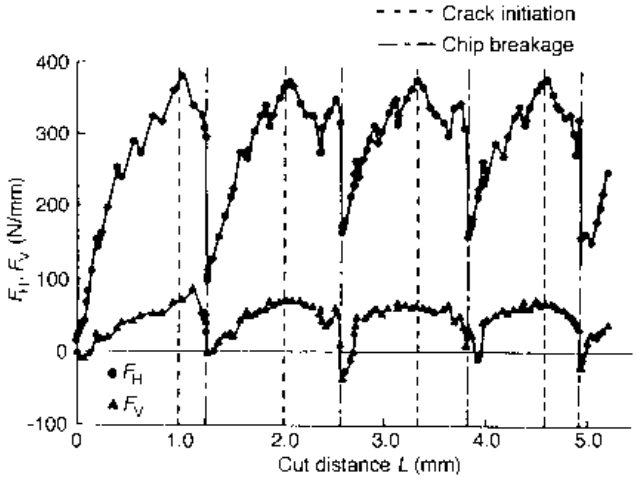


Fig. 8.7 Predicted horizontal and vertical cutting forces for the same conditions as Figure 8.6

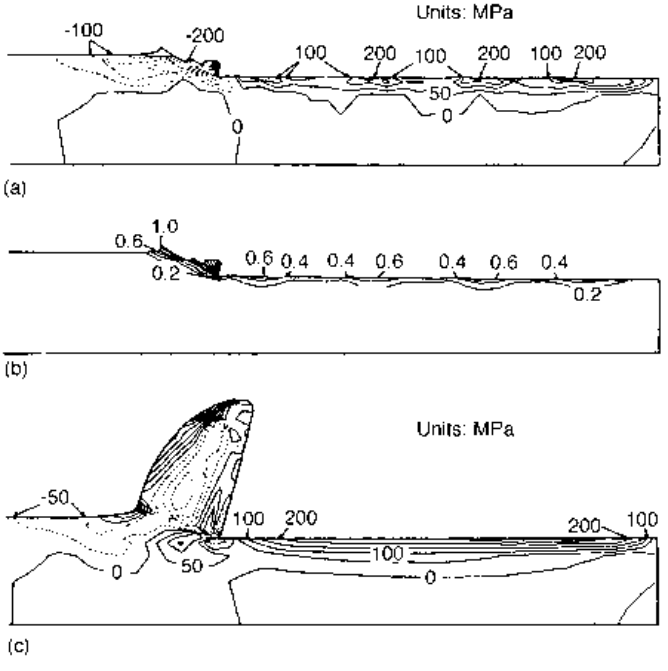


Fig. 8.8 Residual stress and strain in machined layer: (a) direct stress σ_x acting in the cutting direction, (b) equivalent plastic strain; and (c) σ_x in continuous chip formation

penetrates through the chip with a sharp drop in the forces. The force cycle then repeats itself. These tendencies are in accord with experiments (Obikawa *et al.*, 1997).

Residual stress and strain in the machined layer can also be predicted, as shown in Figure 8.8. It shows contours of (a) normal stress σ_x acting in the cutting direction and (b) equivalent plastic strain $\bar{\varepsilon}$, after a cut distance of 5.09 mm and after the cutting forces on

the chip have been relaxed. Periodic variations in σ_x and $\bar{\varepsilon}$ occur synchronously with the cutting force variations (Figure 8.7). For comparison, Figure 8.8(c) shows the continuous chip and the steady residual stress distribution σ_x obtained by removing the possibility of fracture from the simulation.

8.2.2 Tool exit transient chip flow

Figure 8.9 shows changes in chip flow as a cutting tool approaches work-exit conditions (as has been schematically represented in Figure 3.18(b)). Machining with the alumina

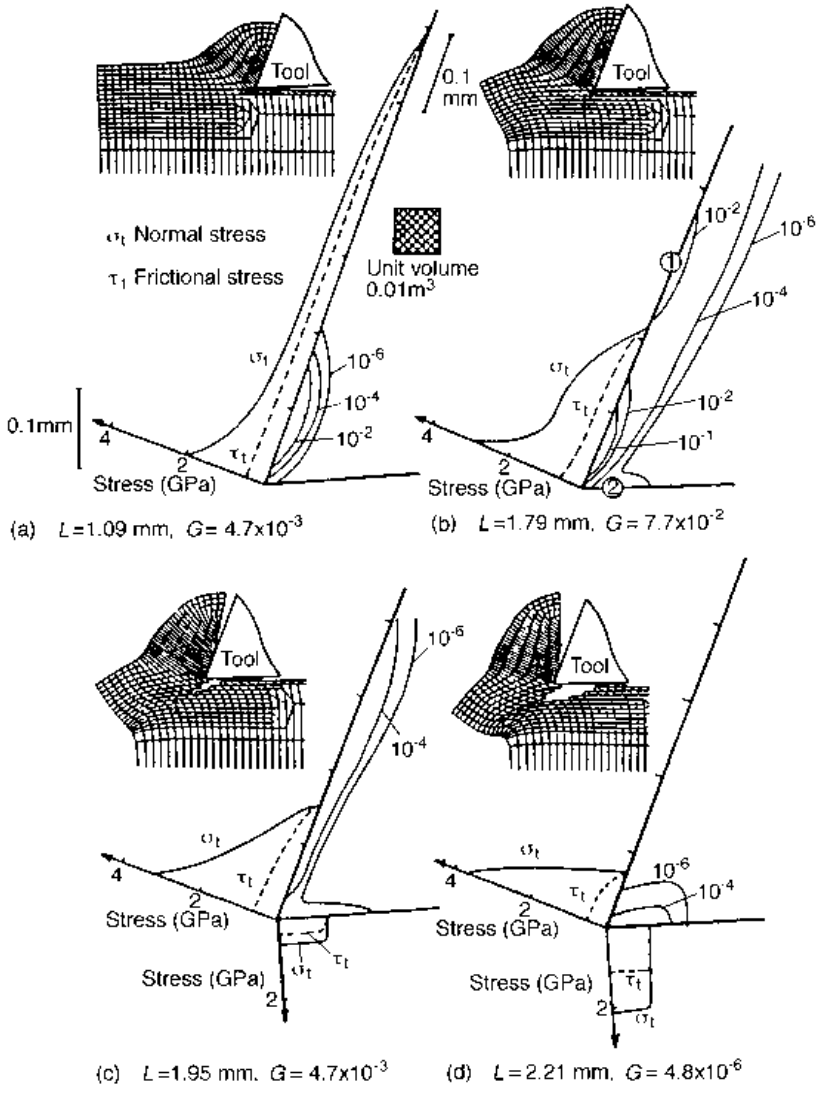


Fig. 8.9 Changes of chip shape and tool edge fracture probability at exit, when machining a β-brass with an alumina ceramic tool at a cutting speed of 13 mm/min, f = 0.25 mm, d = 3 mm, α = 20°, clearance angle γ = 5°, exit angle = 90°, friction coefficient μ = 1.0 and no coolant

tool is begun only 2.5 mm from its end point: in Figure 8.9(a) (L = 1.09 mm) the chip is still in its transient initial formation phase; in Figure 8.9(b) (L = 1.79 mm), material flow into the chip has slowed down as the alternative possibility takes over, of pushing out the end face of the work, by shear at a negative shear plane angle, to form a burr. Eventually (Figure 8.9(c)), a crack forms at the clearance surface and propagates along the negative shear plane towards the end face (Figure 8.9(d)).

The figure also records the changing rake face contact stresses as the end of the cut is approached. The internal stresses have been determined from these by an elastic finite element analysis; and used to assess the probability of tool fracture. The contours within each tool outline are surfaces of constant probability of fracture within a unit volume of 0.01 mm^3, derived from the principal stress distribution in the tool and the tool material's Weibull statistics of failure (Usui *et al.*, 1979, 1982 – see also Chapter 9.2.4). The overall fractional probability of fracture, G, is given by

$$G = 1 - \prod_{i=1}^{n} (1 - G_i) \tag{8.3a}$$

with

$$G_i = \begin{cases} \dfrac{1}{V_0} \displaystyle\int_{V_i} \left(\frac{\sigma^* - \sigma_u}{\sigma_0} \right)^m \mathrm{d}V & (\sigma^* \geq \sigma_u) \\ 0 & (\sigma^* < \sigma_u) \end{cases} \tag{8.3b}$$

where n is the number of finite elements, and G_i is the probability of fracture within one element i, V_0 is a unit volume, V_i is the volume of element i, σ^* is a scalar stress defined in Usui's Weibull statistics model of failure (see Figure 9.8(b)) and σ_u, σ_0 and m are Weibull parameters. In the case of Figure 8.9, G reaches its maximum value of 0.077 at L = 1.79 mm, just before the crack is formed beneath the cutting edge. Once the crack propagates, compressive tool stresses are created, on the tool's clearance face, that reduce the fracture probability. The workpiece fracture relieves the probability of tool fracture; thus, the friction coefficient μ and workpiece brittleness have a strong influence on the tool fracture probability. Reduction of μ from 1.0 to 0.6 increases the shear plane angle to delay negative shear and work crack initiation. This results in an increase in fracture probability, up to a maximum value of 0.293. On the other hand, if a crack initiates early due to workpiece brittleness, as in the machining of a cast iron, a low tool fracture probability is obtained. In cutting experiments, acoustic emission is always detected, when a tool edge fractures, just before the work negative shear band crack forms (Usui *et al.*, 1990).

In Figure 8.9, the exit angle θ, which is the angle between the cutting direction and the face through which the tool exits the work, is 90°. Fracture probability is largest for θ in the range 70° to 100°. Smaller exit angles give rise to safe exit conditions (from the point of view of tool fracture) with little burr formation. Larger angles also give safe exit but large burr formation. Tool exit conditions are of particular interest in milling and drilling. In face milling, the exit angle depends on the ratio of radial depth of cut to cutter diameter (d_R/D, Figure 2.3) and is well-known to affect tool fatigue failures (Pekelharing, 1978). In drilling through-holes, breakthrough occurs at high exit angles (although the three-dimensional nature of the breakthrough makes this statement a simplification of what actually occurs) – and burr formation is a common defect.

8.2.3 Titanium alloy machining

Figure 8.10 shows the pattern of changing chip shape with cut distance L when an $\alpha + \beta$ type Ti-6Al-4V alloy is machined with a carbide tool at a cutting speed of 30 m/min, simulated with the material properties described at the start of Section 8.2. A serrated chip formation is seen. In this case, fractures start at the free surface but never penetrate completely through the chip.

Figure 8.11 shows temperature distributions within the workpiece and tool at the various cut distances corresponding to those in Figure 8.10. Despite a relatively low cutting speed, the temperature in the chip is high, as has been explained in Chapter 2.3. In that chapter, only steady state heat generation was considered. An additional effect of non-steady flow (Figure 8.11(c)) is to bring the maximum temperature rise into the body of the chip, close to the cutting edge.

Many researchers (for example Recht, 1964; Lemaire and Backofen,1972) have attributed serrated chip formation in titanium alloy machining to adiabatic shear or thermal softening in the primary and secondary zones. The results shown in Figures 8.10 and 8.11 contradict this, revealing that the serration arises from the small fracture strain of the alloy, followed by the propagation of a crack and the localization of deformation. However, if the fracture criterion is omitted from a simulation, serrated chip formation can still be observed, but only at higher cutting speeds, for example at 600 m/min (Sandstrom and Hodowany, 1998). It is clear that fracture and adiabatic heating are different mechanisms that can both lead to serrated chip formation. In the case of the titanium alloy, serrated chips occur at cutting speeds too low for adiabatic shear – and then fracture is the cause. However, at higher speeds, the mechanism and form of serration may change, to become adiabatic heating controlled.

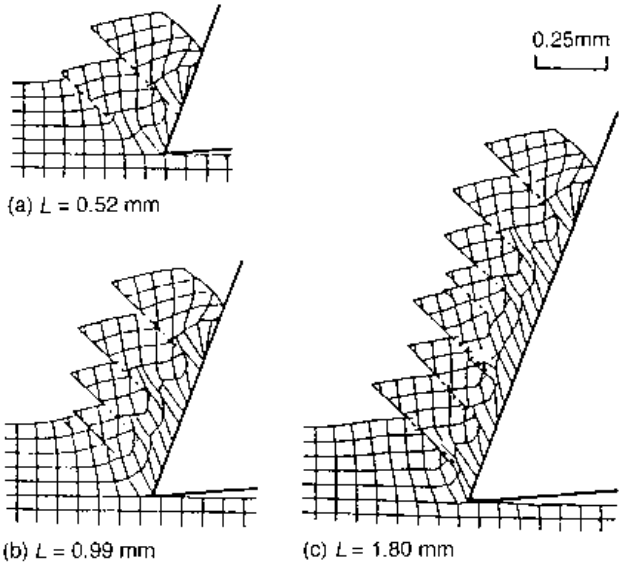


Fig. 8.10 Predicted serrated chip shape in titanium alloy machining by a carbide tool, at a cutting speed of 30 m/min, f = 0.25 mm, d=1 mm, α = 20° and no coolant

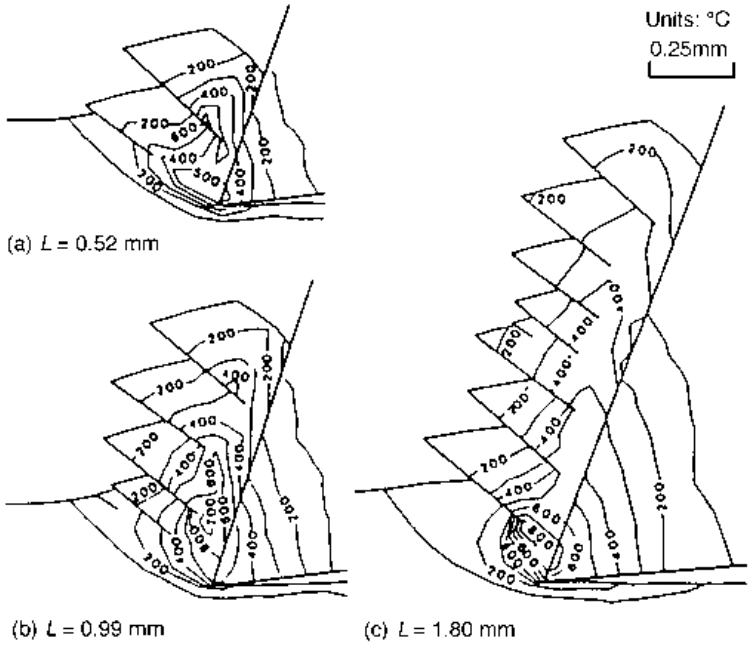


Fig. 8.11 Isotherms near the cutting tip, cutting conditions as Figure 8.10

With other alloy systems, for example some ferrous and aluminium alloys, and with other titanium alloys too, continuous chips may be observed at low cutting speeds, but serrated or segmented chips are seen at high or very high speeds. In some of these cases, serration is almost certainly controlled by adiabatic heating and thermal softening, although in the case of a medium carbon low-alloy steel machining simulation, initial shear fracture has been observed to aid flow localization and facilitate the onset of adiabatic shear (Marusich and Ortiz, 1995; Marusich, 1999); and the importance of fracture in concentrating shear is more strongly argued by some (Vyas and Shaw, 1999). Although the relative importance of fracture and adiabatic shear in individual cases is still a matter for argument, it is certain that an ideally robust finite element simulation software should have the capacity to deal with ductile fracture processes even if, in many applications, the fracture capability remains unused.

8.3 Machinability analysis of free cutting steels

The subject of free cutting steels – steels with more sulphur and manganese than normal (to form manganese sulphide – MnS), and sometimes also with lead additions – was introduced in Chapter 3. Figure 3.16 shows typical force reductions and shear plane angle increases at low cutting speeds of these steels, relative to a steel without additional MnS and Pb. These changes have been attributed to embrittling effects of the MnS inclusions in the primary shear zone (for example Hazra *et al.*, 1974) and a rake face lubricating effect (for example Yamaguchi and Kato, 1980). The lubrication effect has been considered in Chapter 2 (Figure 2.23). The deposition of sulphide and other non-metallic inclusions on

the tool face to reduce wear has also been described (Figure 3.17) and briefly referred to in Chapter 4 – many researchers have studied this (for example Naylor *et al.*, 1976; Yamane *et al.*, 1990). Finite element analysis provides a tool for studying the relations between the cutting conditions (speed, feed, rake angle) and the local stress and temperature conditions in which the lubricating and wear reducing effects must operate. The next sections describe a particular comparative investigation into the machining of four steels: a plain carbon steel, two steels with MnS additions and one steel with MnS and Pb. In this case, the lubrication effects completely explain observed behaviours, with no evidence of embrittlement (Maekawa *et al.*, 1991).

8.3.1 Flow and friction properties of resulphurized steels

The compositions of the four steels are listed in Table 8.1. They are identified as P (plain), X and Y (the steels with MnS added) and L (the steel with MnS and Pb). The steels X and Y differ in the size of their MnS inclusions: Table 8.1 also gives their inclusion cross-section areas.

The flow behaviours of the steels in their as-rolled state were found from Hopkinson-bar compression tests at temperatures T, strain rates $\dot{\bar{\varepsilon}}$ and strains $\bar{\varepsilon}$ from 20 to 700°C, 500 to 2000 s^{-1} and 0 to 1, respectively, as described in Chapter 7.4. Figure 8.12 shows the orientation and size of the specimens: a bar-like test piece of ∅6 mm × 10 mm was cut from the commercial steel bars that were later machined. Figure 8.13 shows example flow stress–temperature curves, at a strain rate of 1000 s^{-1} and two levels of strain, 0.2 and 1.0. The symbols indicate measured values while the solid lines are fitted to equation (7.15a). For the sake of clarity, only the approximated curve for steel P is drawn in the figure. The flow stress is more or less the same for all four steels, although that for steel X, with larger MnS inclusions, is slightly lower than that of the others. The values of A, M, N, a and m (equation (7.15a)) for the steels are listed in Table 8.2.

Table 8.1 Chemical composition of workpiece (wt%)

	C	Mn	P	S	Pb	MnS size (μm^2)
Steel P	0.100	0.400	0.025	0.019	–	–
Steel X	0.070	0.970	0.067	0.339	–	145
Steel Y	0.070	0.910	0.087	0.321	–	124
Steel L	0.080	1.300	0.070	0.323	0.025	–

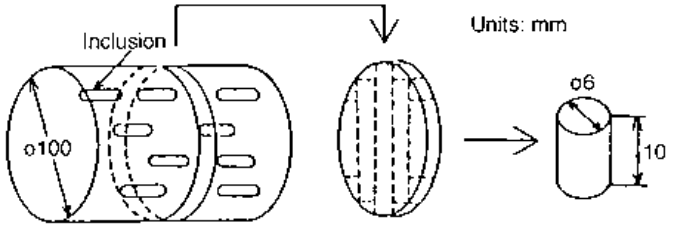


Fig. 8.12 Specimen preparation for high speed compression testing

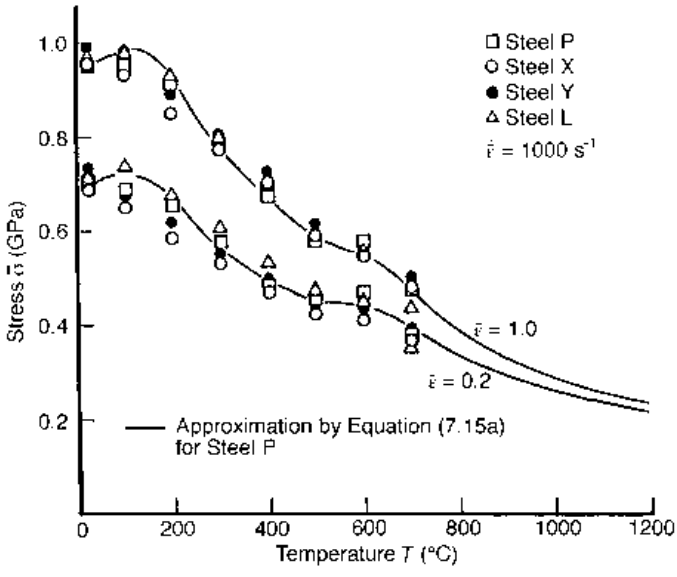


Fig. 8.13 Flow stress–temperature curves at a strain rate of 1000 s^{-1}

Table 8.2 Flow characteristics of steels

	Coefficients of equation (7.10a)
Steel P	$A = 900e^{-0.0011T} + 170e^{-0.00007(T-150)^2} + 110e^{-0.00002(T-350)^2} + 80e^{-0.0001(T-650)^2}$ $M = 0.0323 + 0.000014T$ $N = 0.185e^{-0.0007T} + 0.055e^{-0.000015(T-370)^2}$ $a = 0.00024$, $m = 0.0019$
Steel X	$A = 880e^{-0.0011T} + 120e^{-0.00009(T-150)^2} + 150e^{-0.00002(T-350)^2} + 90e^{-0.0001(T-650)^2}$ $M = 0.0285 + 0.000014T$ $N = 0.18e^{-0.0005T} + 0.1e^{-0.000015(T-430)^2}$ $a = 0.00020$, $m = 0.0052$
Steel Y	$A = 920e^{-0.0011T} + 120e^{-0.00003(T-160)^2} + 110e^{-0.00004(T-340)^2} + 120e^{-0.0001(T-650)^2}$ $M = 0.0315 + 0.000016T$ $N = 0.19e^{-0.0005T} + 0.085e^{-0.000015\ (T-430)^2}$ $a = 0.00032$, $m = 0.0018$
Steel L	$A = 910e^{-0.0011T} + 190e^{-0.00011(T-135)^2} + 150e^{-0.00002(T-330)^2} + 100e^{-0.0001(T-650)^2}$ $M = 0.0325 + 0.000008T$ $N = 0.18e^{-0.0007T} + 0.055e^{-0.000015(T-370)^2}$ $a = 0.00028$, $m = 0.0016$

As for the measurement of friction characteristics at the tool–chip interface, the split-tool method was employed. Figure 8.14 shows the distributions of normal stress σ_t and friction stress τ_t when the steels were turned on a lathe without coolant, by a P20-grade cemented carbide tool at a cutting speed of 100 m/min, a feed of 0.2 mm/rev, a rake angle of 0° and a depth of cut of 2.8 mm. The abscissa is the distance from the cutting edge in the direction of chip flow. The normal stress increases exponentially towards the tool edge, whereas the friction stress has a trapezoidal distribution saturated towards the edge. Steel P shows $\tau_n > \sigma_t$ near the end of contact. The free cutting steels all show $\tau_t < \sigma_t$ there and a shorter contact length than steel P. These tendencies are more evident for steel L and steel Y than steel X.

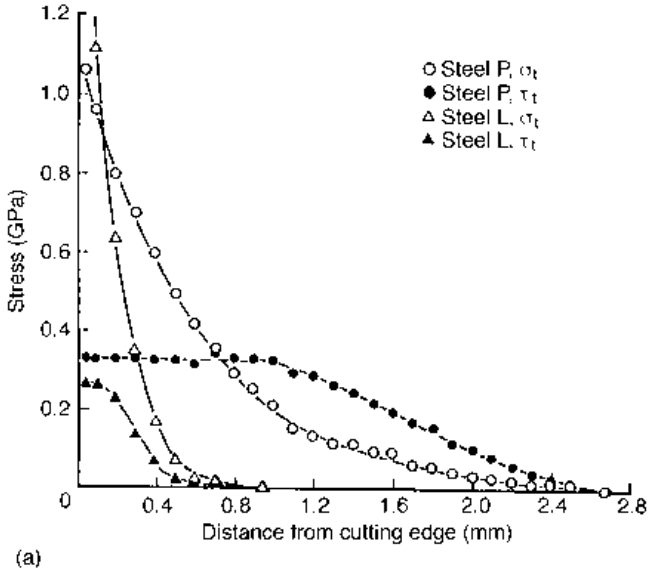


Fig. 8.14 Normal stress σ_t and friction stress τ_t distributions measured on the tool rake face at a cutting speed of 100 m/min: (a) steels P and L; (b) steels X and Y

Rearrangement of Figure 8.14 leads to Figure 8.15 which shows the relationship between τ_t and σ_t (measurements were also made at a cutting speed of 200 m/min). The measured stress distributions can be formulated as equation (2.24d) where the values of m, μ and n^* are listed in Table 8.3. The friction characteristic equation suggests that the lubrication effect of MnS inclusions is evaluated by m and μ, and this is more evident when lead is added to the steel.

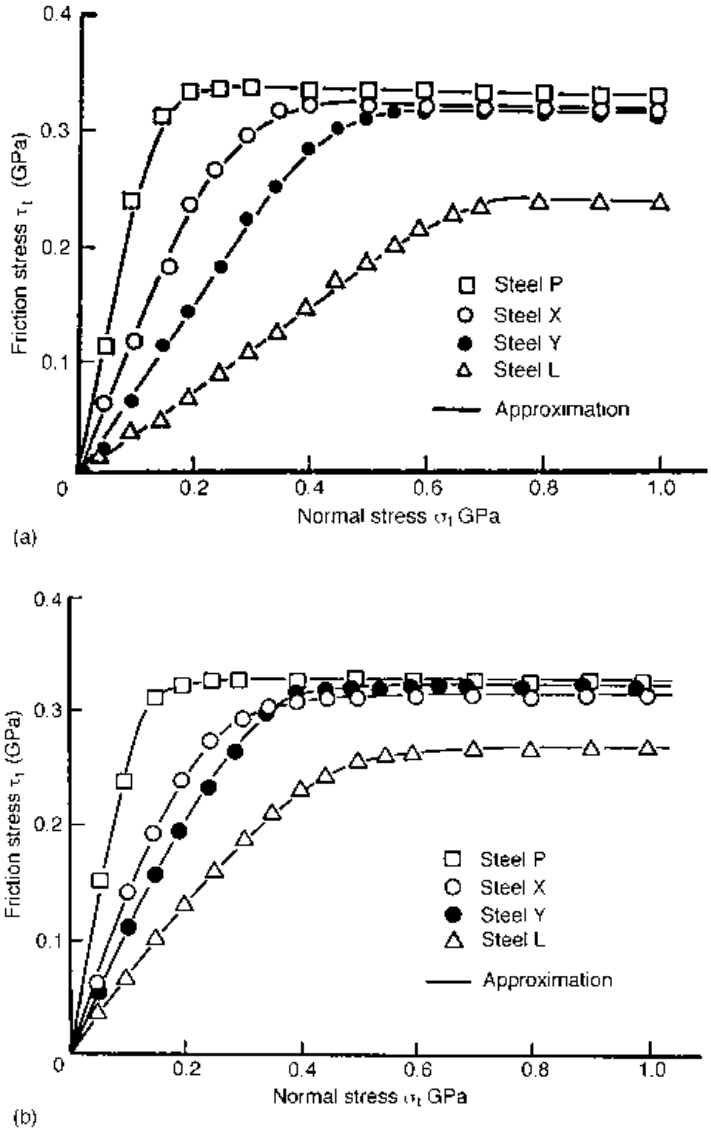


Fig. 8.15 Relations between σ_t and τ_t at cutting speeds of (a)100 m/min and (b) 200 m/min

Table 8.3 Coefficients of friction in characteristic equation (2.24d)

	V=100 m/min			*V*=200 m/min		
	m	μ	n^*	*m*	μ	n^*
Steel P	1.0	2.31	3.89	1.0	2.48	3.22
Steel X	0.99	1.25	3.05	0.96	1.43	2.06
Steel Y	0.97	0.76	5.98	0.99	1.04	3.04
Steel L	0.74	0.38	8.78	0.88	0.72	4.21

8.3.2 Simulated analysis of free cutting actions

Figures 8.16 and 8.17 show contours of equivalent plastic strain rate and isotherms together with chip configurations predicted at the cutting speed of 100 m/min, feed of 0.2

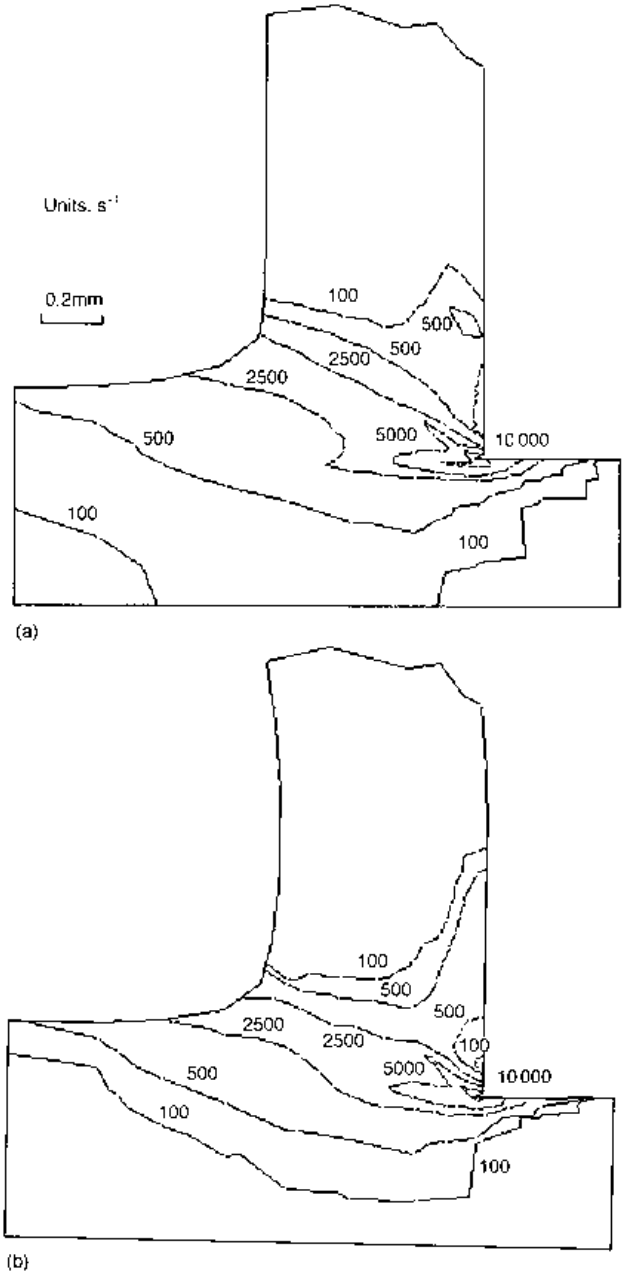
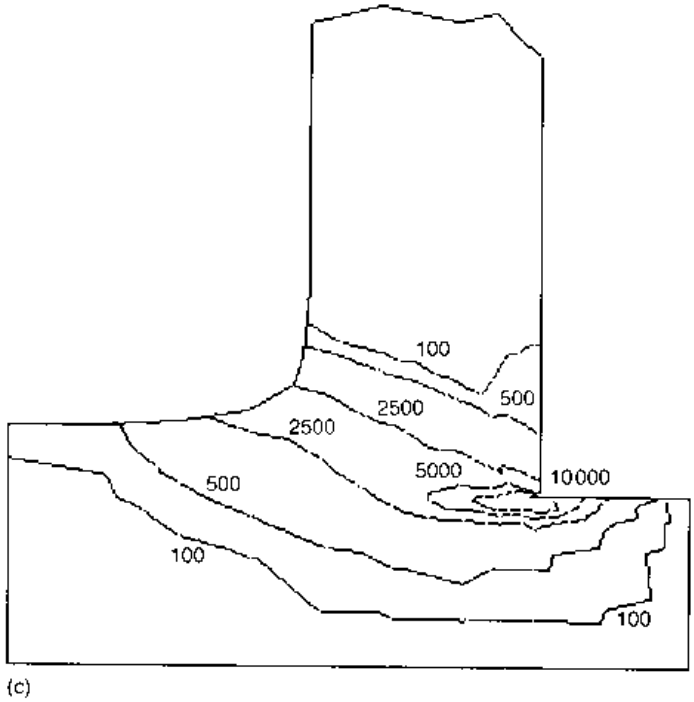
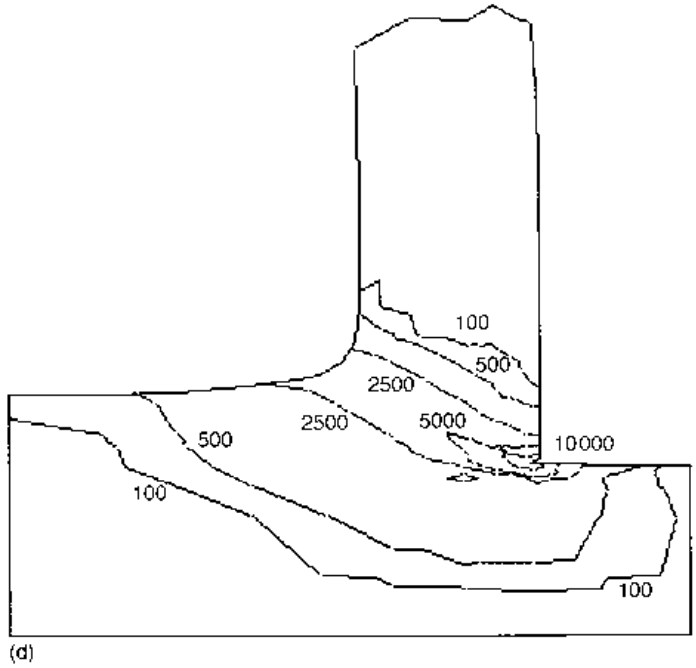


Fig. 8.16 Contours of equivalent plastic strain rate at a cutting speed of 100 m/min, f = 0.2 mm, α = 0° and no coolant: (a) steel P; (b) steel X; (c) steel Y and (d) steel L



(c)



(d)

Fig. 8.16 *continued*

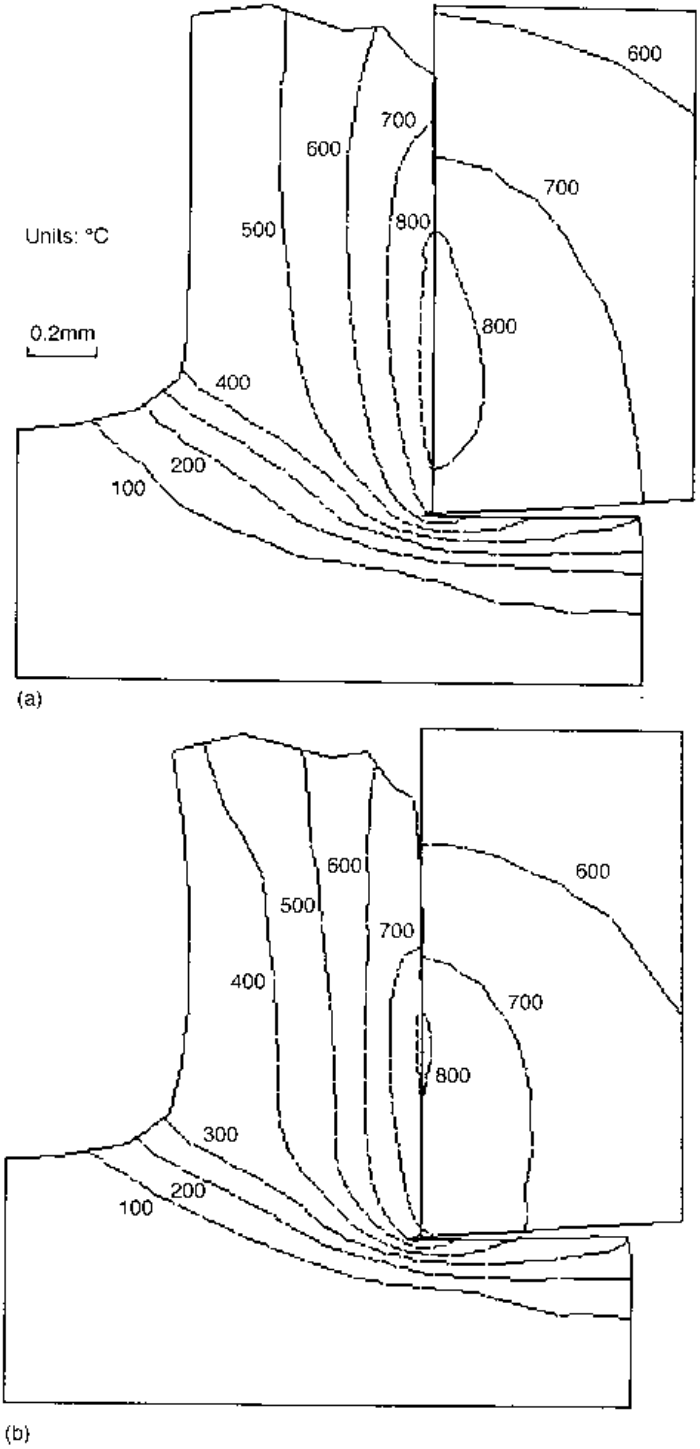
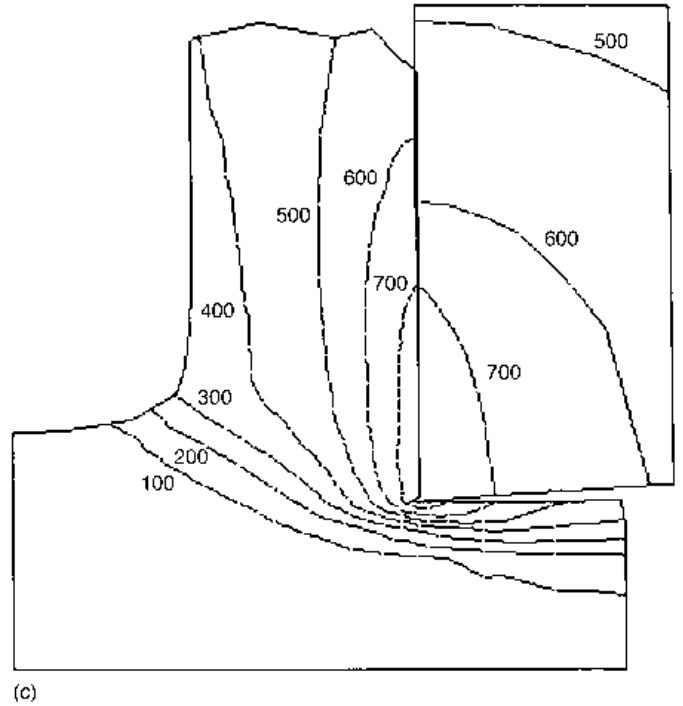
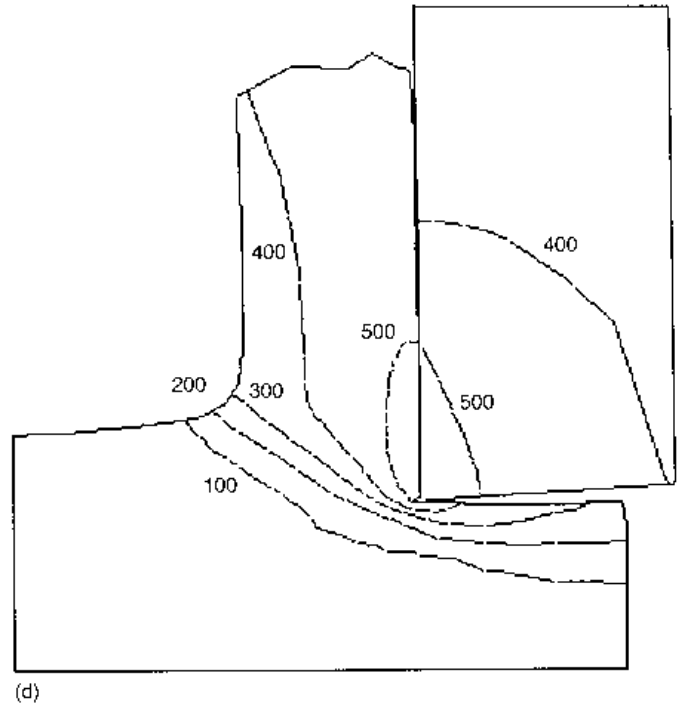


Fig. 8.17 Isotherms for the same cutting conditions as in Figure 8.16: (a) steel P; (b) steel X, (c) steel Y; and (d) steel L



(c)



(d)

Fig. 8.17 *continued*

mm and zero rake angle and without cutting fluid. Each consists of four panels representing (a) steel P, (b) steel X, (c) steel Y and (d) steel L. Since the steels show similar flow stresses as shown in Figure 8.13, it is certain that their friction differences differentiate their cutting mechanisms. As the friction becomes more severe in the order steel L, steel Y, steel X and steel P, the chip thickens, curls less and increases its contact length. The plastic deformation within the workpiece ahead of and below the cutting edge (getting larger in the order L to P, and resulting in a larger accumulated plastic strain in the chip), and also the larger secondary flow due to the friction, can also be seen. In the end this leads to a larger temperature rise on the tool rake face, although the maximum temperature occurs far from the cutting edge.

To investigate consistency between the real machining and the simulation, supplementary cutting experiments were carried out. Figure 8.18 shows the chip sections obtained from quick-stop tests on the four steels. The cutting conditions are the same as those used in the simulation. Changes in chip thickness and its curl are in agreement with those of

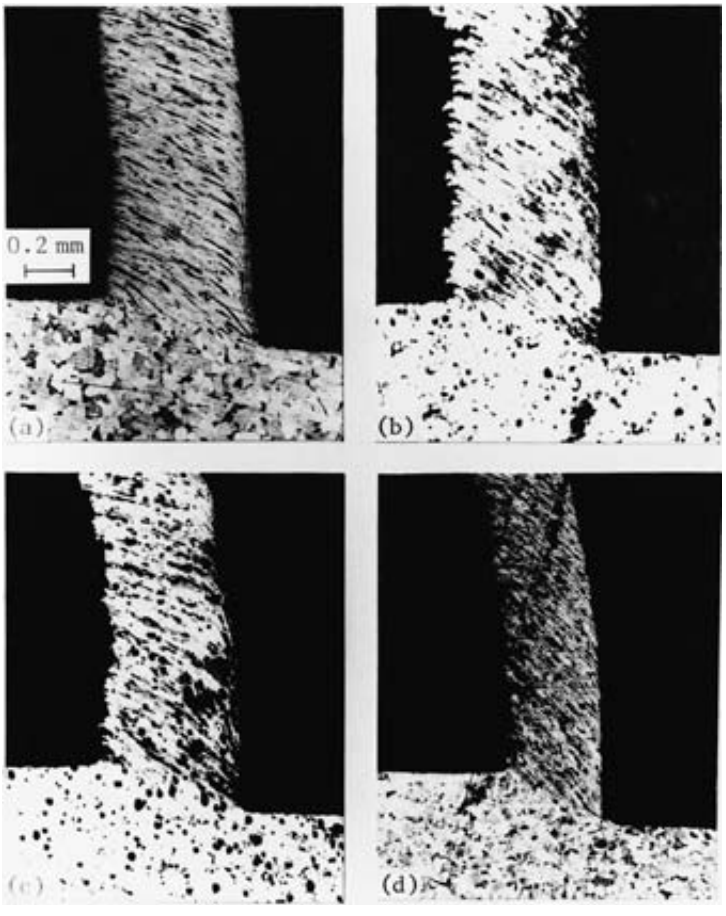


Fig. 8.18 Etched cross-sections obtained by quick stop tests in the same conditions as Figures 8.16 and 8.17: (a) steel P, (b) steel X, (c) steel Y; and (d) steel L

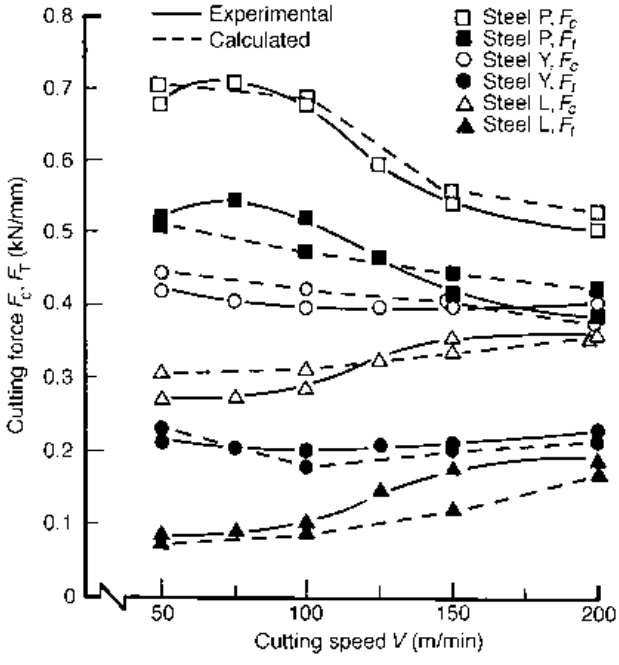


Fig. 8.19 Comparison of predicted and measured cutting force–cutting speed curves at f=0.2 mm, d=2 mm, α=10° and without coolant, for steels P, Y and L

Figure 8.16. A thinner and curlier chip is obtained when machining Steel L. Severe deformation in the secondary shear zone is particularly recognized when cutting Steel P, which leads to a high temperature rise along the rake face.

Figure 8.19 compares the predicted cutting force with the measured one when machining three of the steels at a rake angle of 10°. The friction constants shown in Table 8.3 for the cutting speed of 100 m/min were used in the simulations for cutting speeds from 50 m/min to 150 m/min, and those for 200 m/min at the cutting speed of 200 m/min. The solid lines denote experiment, whereas the dashed ones represent simulation. The different force characteristics of the three steels are entirely explained by their different friction behaviours. There is agreement with the general observation of Figure 3.16.

From the viewpoint of machinability assessment, the leaded resulphurized steel (steel L) is most effective in reducing cutting forces and tool temperature. Lower temperature will provide less tool wear. The second best is the MnS-based free cutting steel with finer MnS inclusions (steel Y). The primary reason for the better machinability lies in the lubrication effect of the inclusions. When steel L is machined at the cutting speed of 200 m/min, however, the lubrication effect is reduced to a similar level as Steel Y at 100 m/min. Probably, the lead is melted and partially vaporized with increasing cutting speed or cutting temperature. The rake temperature is predicted to reach 1000°C at 200 m/min.

In summary, on the basis of the experimental friction and flow stress characteristics, the finite element analysis has revealed that differences in friction characteristics mainly cause the chip flow, temperature and cutting force change in the free cutting steels. The MnS-based steel with smaller inclusions shows better machinability, including a thinner chip,

narrower deformation zone, lower rake temperature and smaller cutting force. The leaded resulphurized steel gives the best machinability at cutting speeds lower than 200 m/min, where lead is the most effective lubricant on the tool rake.

8.4 Cutting edge design

The importance of non-planar rake faces in controlling chip flow and reducing tool forces, wear and failure was briefly mentioned at the end of Chapter 3.

Chip controllability and disposability depends strongly on tool geometry as well as the cutting conditions. To design an optimum, high-performance cutting tool it is necessary to understand how chip flow is modified by machining with a cutting tool with a chip former in place. Many experimental observations have been carried out from this point of view (Nakayama, 1962; Jawahir, 1990; Jawahir and van Luttervelt, 1993). Section 8.4.1 describes a two-dimensional (orthogonal) finite element simulation of chip breaking when machining with grooved rake face tools (Shinozuka *et al.*, 1996a, 1996b; Shinozuka, 1998). Cutting force, temperature and tool wear reduction by rake face design are the subjects of Section 8.4.2, which describes a three-dimensional simulation of chip formation (Maekawa *et al.*, 1994).

8.4.1 Tool geometry design for chip controllability

A hybrid simulation is described here, in which a steady-state chip formation is first analysed by the ICM method and then modified approximately to a non-steady phase in order to study the development of chip breaking behaviour.

Steady-state (ICM) simulation

Figure 8.20 shows a tool rake face similar to that in Figure 3.30(c), but made more general by approximating the profile ABCD to a Bézier curve with w_G the width of a groove and h_B the height of a chip former. The effects on chip flow of varying w_G and h_B, while keeping the positions of A, B and C constant, as shown, have been studied for the machining of a 0.18%C plain carbon steel by a P20 carbide tool, the same materials as in Section 8.1.

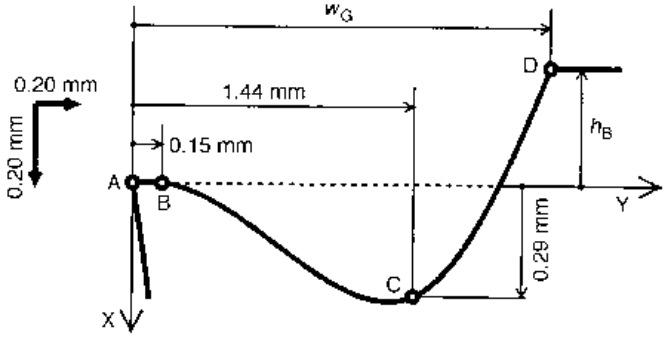
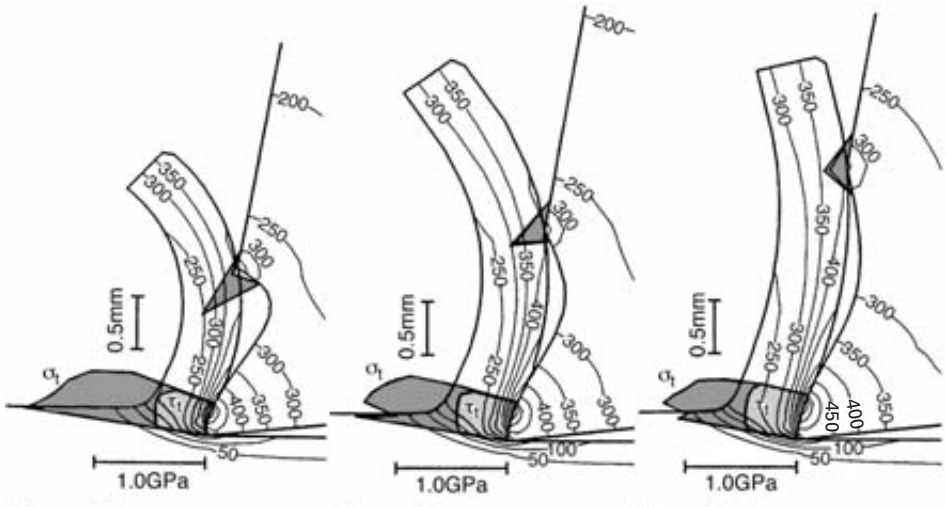


Fig. 8.20 Rake face geometry with chip former



(a) W_G = 1.6 mm (b) W_G = 2.0 mm (c) W_G = 2.6 mm

Fig. 8.21 Predicted chip shape and distributions of temperature (°C) and stresses with changing w_G (h_B=0) when machining 0.18%C steel with carbide P20 at V=100 m/min, undeformed chip thickness t_1=0.25 mm and without coolant

Unless otherwise specified, a cutting speed of 100 m/min, a feed (uncut chip thickness) of 0.25 mm, a primary rake angle of 10° and no coolant have been chosen for the simulation conditions.

Figure 8.21 shows the chip shape and the distributions of temperature and stresses acting on the rake face with changing w_G, when h_B = 0 mm. As w_G is reduced from 2.6 mm to 1.6 mm, chip curl radius, rake face temperature and chip/tool contact length are all reduced. Inversely, the magnitude of the normal stress σ_t is increased at the chip/tool contacts.

In the examples of Figure 8.21 the chips are so short that their ends are free, not curled round enough to contact the work ahead of the tool. The approximate analysis of what happens, once contact with the work does occur, is considered next.

Approximate unsteady flow and chip breaking

Figure 8.22 shows the instant at which the chip first touches the work ahead of the tool, at point C. In principle, the contact forces at C will change the flow in the primary shear zone; but that is neglected here. It is imagined that the chip continues to flow out of the region h–g–c–d, with a velocity prescribed to vary linearly, from V_h to V_g, along the surface h–g, where V_h and V_g are the chip surface velocities obtained from the ICM calculation; and that in the region h–g–c–d, the flow stress and temperature variations from the inside to the outside radius of the chip are also as obtained from the primary shear flow in the ICM analysis. As the chip grows in length, new elements are added at the boundary h–g. How the slender chip formed in this way deforms and breaks due to the contact forces at C and at the chip former (point B) is analysed next.

The contact force at B arises from the velocity boundary condition along h–g. The

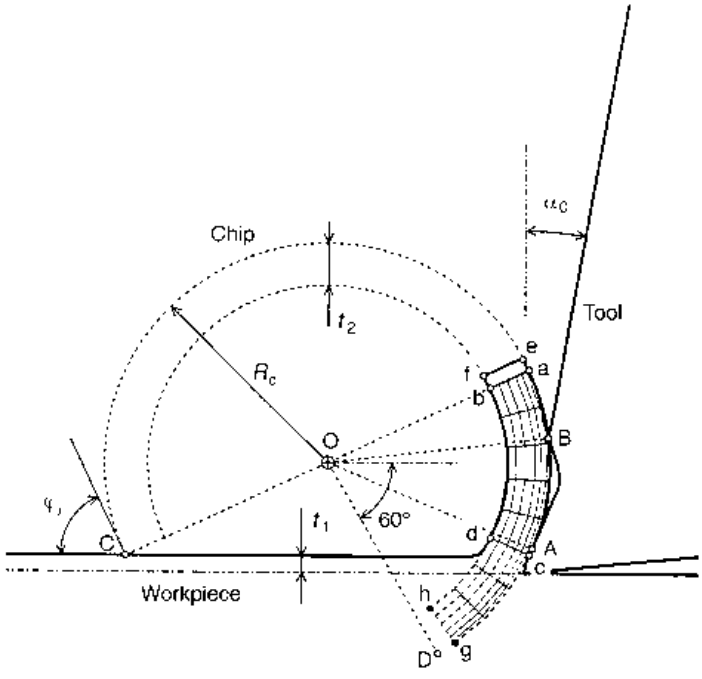


Fig. 8.22 Initial finite element model used for chip breakability analysis

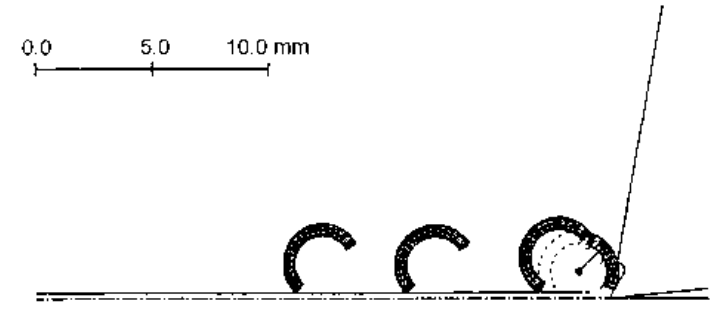
angular velocity $V_h/(R_c–t_2)$ at point h, from the ICM analysis, is always greater than that of V_g/R_c at point g: without a force at B, the chip path would penetrate the tool. The contact forces at C are assumed to obey Coulomb's laws of friction. While the chip slips on the work surface $\tau_t = \mu_d\sigma_t$ holds. If the relative velocity between the chip and workpiece becomes zero, then the chip is regarded as adhering to the surface. The adhered chip does not slide again until $\tau_t > \mu_s\sigma_t$. The static and dynamic friction coefficients μ_s and μ_d are assumed to be 0.3 and 0.2, respectively.

As for a fracture criterion in the chip beyond its formation region, it is assumed that a crack nucleates and develops from the chip's rough free surface when the maximum principal stress or the maximum shear stress exceeds a critical value σ_{1c} or τ_c. A crack that satisfies σ_{1c} propagates in the direction of minimum principal stress, whereas one that satisfies τ_c grows in the direction of maximum shear stress. In this work, σ_{1c} = 880 MPa and τ_c = 440 MPa have been found to give good representations of practice.

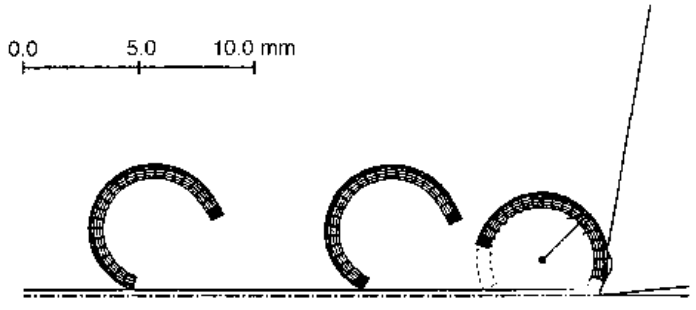
To follow the crack growth, it is necessary to subdivide the elements around the crack tip; and this requires reorganization of the node connectivity too. Remeshing around the point B is also required – and a small time step of $\leq 10^{-6}$ s (for the cutting speed of 100 m/min) is also needed.

Simulation results

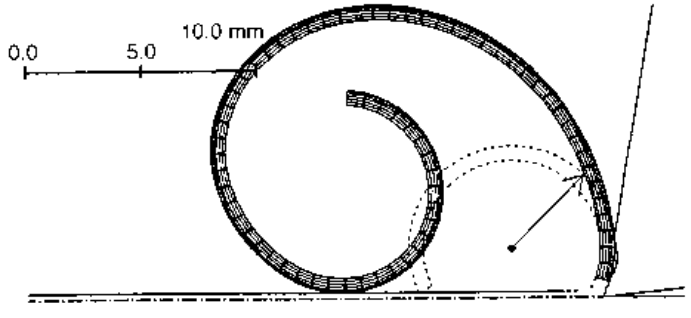
Figure 8.23 shows the chip shape simulated with changing w_G and $h_B = 0$. As w_G increases from 1.6 mm to 2.0 mm, the radius of the broken chips becomes larger; and at w_G = 2.6 mm chip breakage does not occur.



(a) w_G = 1.6 mm



(b) w_G = 2.0 mm



(c) w_G = 2.6 mm

Fig. 8.23 Predicted chip shape with changing w_G (h_B=0)

Increasing h_B aids chip breakage. Figure 8.24 shows the development of chip shape with time for w_G = 2.6 mm but h_B = 0.4 mm. Plastic deformation with $\dot{\bar{\varepsilon}} > 10\ s^{-1}$ takes place at the hatched regions in the figure and the chip breaks after 25 ms (the time t is measured from the instant at which the chip first collides with the workpiece surface). The figure also records the contact forces. F_{BH} and F_{BV} are the horizontal and vertical force components acting at point B, and F_{CH} and F_{CV} are those at point C. The small size of the forces at C and the almost constant forces at B throughout the chip breaking cycle support the approximation that contact of the chip with the work does not much alter the flow in the primary shear region.

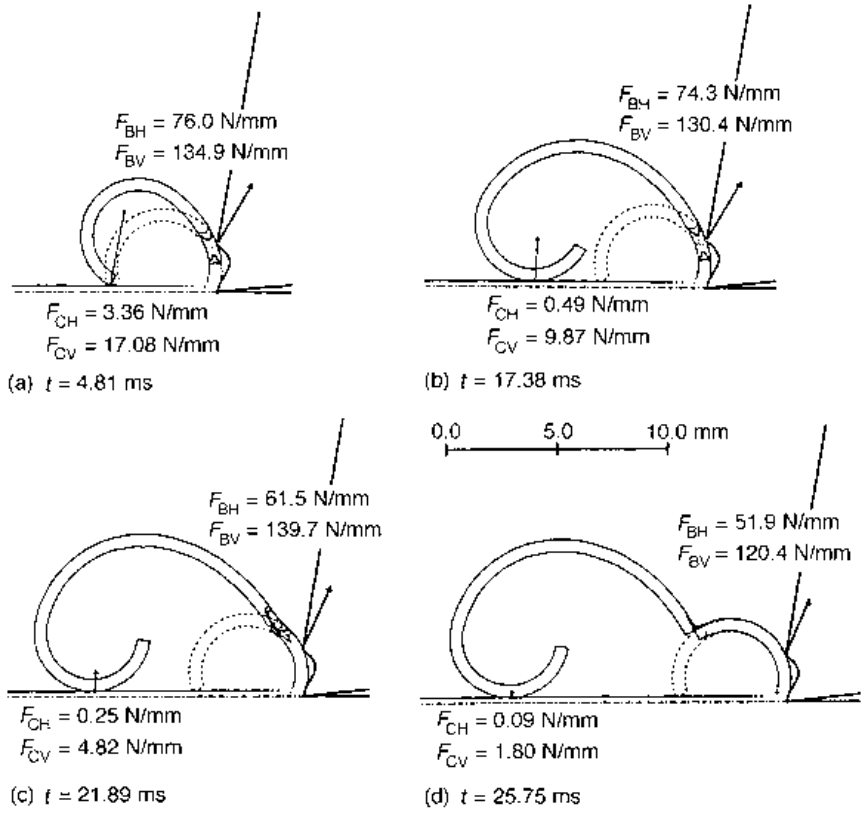


Fig. 8.24 Variation of chip shape and forces at w_G=2.6 mm and h_B=0.4 mm

The effect of undeformed chip thickness t_1 is considered in Figure 8.25, which compares the predicted chip shape with experiment at $w_G = 2.14$ mm and $h_B = 0$. When t_1 is increased from 0.10 mm to 0.36 mm, the chip shape is changed from continuous to segmented. In particular, an ear- (or ε-) type chip is generated at $t_1 = 0.25$ mm. The simulated chip morphology, including curl and thickness, is in good agreement with experiment (similar observations have been reported by Jawahir, 1990). When the rake angle is decreased, the segmentation is accelerated and chips with a smaller radius are produced (Shinozuka *et al.*, 1996b).

8.4.2 Three-dimensional cutting edge design

Tools with cut-away rake faces, to restrict the chip contact to be shorter than it would naturally be, have advantages beyond that of chip control considered in the previous section. Smaller cutting forces, lower cutting temperature, longer tool life, better surface finish and the prevention of tool breakage can be achieved in practice, provided the restriction is properly chosen (Chao and Trigger, 1959; Jawahir, 1988). Slip-line field plasticity theory has been applied to two-dimensional machining with a cut-away tool, to analyse the changes to chip flow caused by a restricted contact (Figure 6.6 – Usui *et al.*, 1964). Here, a closer-to-practice three-dimensional ICM finite element analysis is introduced of the

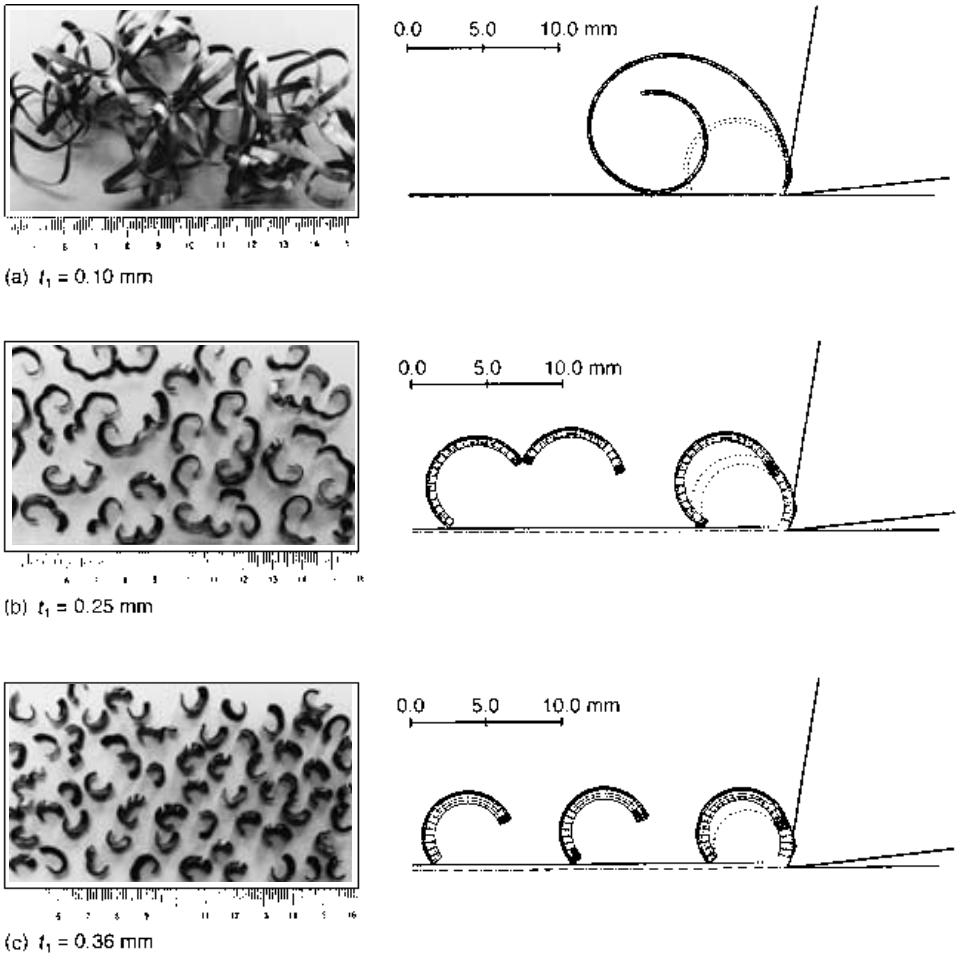


Fig. 8.25 Comparison of predicted chip shape with experiment (w_G=2.14 mm and h_B=0)

effect – on steady-state chip formation, tool temperature and wear – of varying a cut-away in the region of the nose radius of a single point P20-grade turning tool, used to turn an 18%Mn–18%Cr alloy steel, at a cutting speed of 60 m/min, a feed of 0.2 mm, and a depth of cut of 2 mm, without coolant. The mechanical and thermal properties and friction and wear behaviour, assumptions (from measurements) are listed in Table 8.4.

Simulation model

The three-dimensional analysis has been developed from the two-dimensional ICM scheme described in Section 7.3. Figure 8.26 shows the analytical model for machining with a single point tool at zero cutting edge inclination angle. The *x*- and *y*-axes are, respectively, parallel and perpendicular to the cutting direction, and the *z*-axis is set along the major cutting edge. The tool is assumed to be stationary and rigid, while the work-piece has boundaries moving towards the tool at the specified cutting speed. Apart from

Table 8.4 Mechanical and thermal properties used in simulation

Flow stress characteristics: $\bar{\sigma} = A\,(\dot{\bar{\varepsilon}}/1000)^M\,(\bar{\varepsilon}/0.3)^N$ GPa
where $A = 2.01\exp(-0.00177T)$, $M = 0.00468\exp(0.00355T)$,
$N = 0.346\exp(-0.000806T) + 0.111\exp\{-0.0000315(T-375)^2\}$

Friction characteristic: $\tau/k = 1 - \exp(-\mu\sigma_n/k)$

Wear characteristic: $\dfrac{dh}{ds} = C_1\sigma_n \exp\left(-\dfrac{C_2}{T}\right)$

Young's modulus $E = 206$ GPa, Poisson's ratio $\nu = 0.3$, Friction constant $\mu = 1.6$,
Wear constants $C_1 = 14.67$ MPa^{-1}, $C_2 = 21\,930$ K

	Thermal conductivity K [W m^{-1} K^{-1}]	Density, ρ [kg m^{-3}]	Specific heat, C [J kg^{-1} K^{-1}]
Workpiece (18%Mn–18% Cr steel)	12.6	7950	502
Insert (carbide P20)	66.9	11200	402
Shank (0.55% C steel)	36.0	7750	461

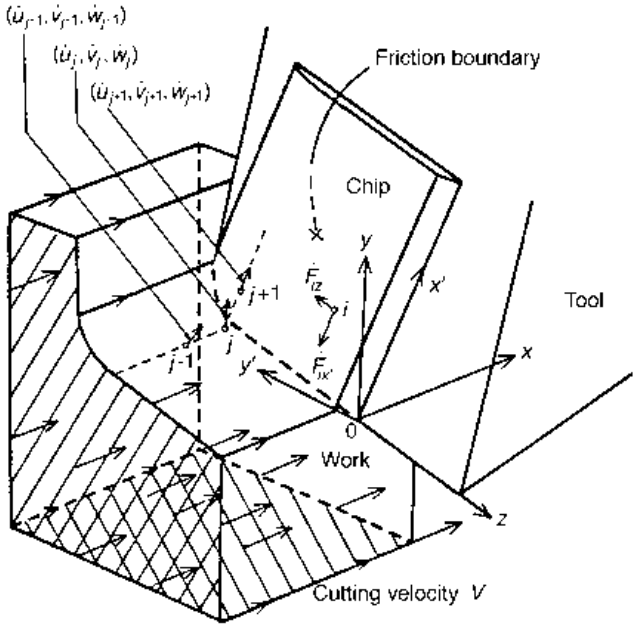


Fig. 8.26 Three-dimensional machining model and boundary conditions

the obvious differences stemming from converting two-dimensional finite element stiffness equations to three-dimensional ones, the main complication is allowing the chip to flow in the z-direction. The formulation of sliding friction behaviour at the tool–chip interface is modified to allow for this: for a node i contacting the rake face, the following conditions are imposed on the finite element stiffness equation:

$$\dot{F}_{ix'} = \left(\frac{\mathrm{d}\tau}{\mathrm{d}\sigma_\mathrm{n}}\right)\frac{\dot{u}_\mathrm{i}}{(\dot{u}_\mathrm{i}^2 + \dot{w}_\mathrm{i}^2)^{1/2}}\,\dot{F}_{iy'}, \qquad \dot{F}_{iz} = \left(\frac{\mathrm{d}\tau}{\mathrm{d}\sigma_\mathrm{n}}\right)\frac{\dot{w}_\mathrm{i}}{(\dot{u}_\mathrm{i}^2 + \dot{w}_\mathrm{i}^2)^{1/2}}\,\dot{F}_{iy'}, \qquad \dot{v}_i = 0 \tag{8.4}$$

where x' and y' are the local coordinate system as shown in the figure, $\dot{F}_{iy'}$ is the rate of normal force on node i, and $(\dot{u}'_i, \dot{v}'_i, \dot{w}'_i)$ are the velocities of node i in the (x', y', z') directions. $(\mathrm{d}\tau/\mathrm{d}\sigma_\mathrm{n})$ is the effective friction coefficient given by the differentiation of the friction characteristic, equation (2.24c). A further complication in description arises when a chip flows into a cut-away groove in a primary (plane) rake face of a tool. Although this has been dealt with in the example of Section 8.4.1, in the simulation in this section it is assumed that a chip makes contact only with the primary rake face.

Figure 8.27 shows the finite element structure of the model. It is an assembly of linear tetrahedral elements (7570 elements and 1887 nodes in all). The mesh shown is an ICM initial-guess for turning with a plane rake faced tool, with cutting occurring over the major and minor cutting edges and over the tool nose radius. The tool geometry is ($\alpha_\mathrm{p} = 0^\circ$, $\alpha_\mathrm{f} = 0^\circ$, $\gamma_\mathrm{p} = 6^\circ$, $\gamma_\mathrm{f} = 6^\circ$, $\kappa'_\mathrm{r} = 15^\circ$, $\psi = 15^\circ$, $R_\mathrm{n} = 1$ mm) where the terms are defined in Figure 6.16. The mesh is automatically generated from a specified shear plane angle and chip flow direction, the tool geometry, feed and depth of cut.

Cutting with this conventional, plane, tool is analysed as well as cutting with two cut-away tools derived from it. Views of the two cut-away tools, types I and II, are shown in Figure 8.28. Both of these tools have a secondary rake of angle 15° superimposed on the primary rake. The type I tool has a restricted primary land width r that is constant along the major cutting edge but reduces around the tool's nose radius, to zero at the minor cutting edge, in the same way that the uncut chip thickness varies. The type II tool has a restricted land width that is constant

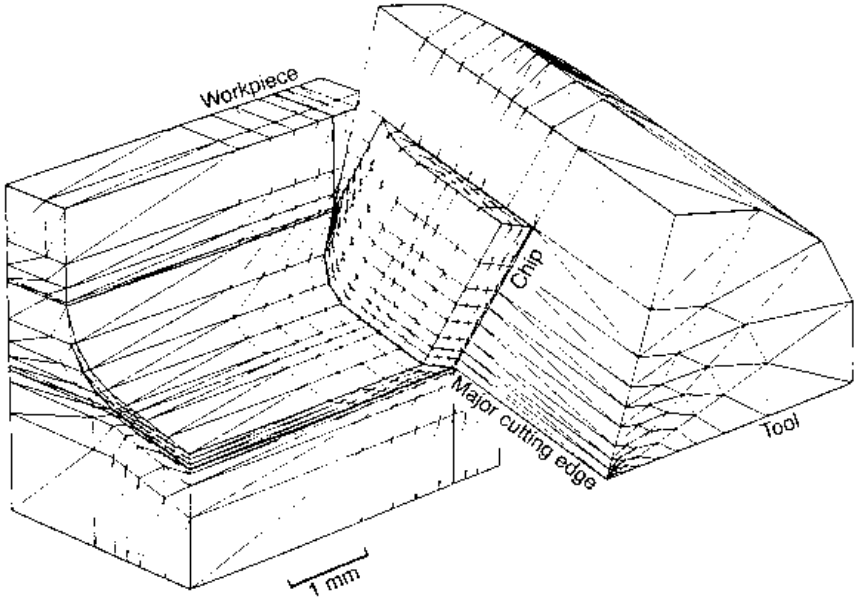


Fig. 8.27 Initial finite element mesh for a tool geometry (0°, 0°, 6°, 6°, 15°, 15°, 1.0mm) and f=0.2 mm, d=2 mm

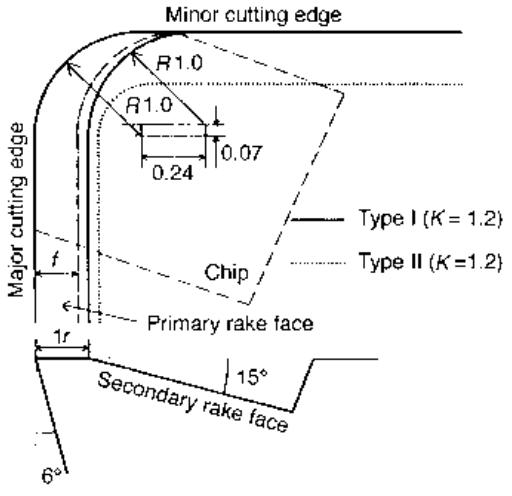


Fig. 8.28 Rake face geometries of types I and II cut-away tools

around both the major and minor cutting edges. The influences of these design differences, and also of varying the width r relative to the feed f are studied. The value of r over the major cutting edge, divided by f, is referred to as the restriction constant K.

Simulation and experimental results

The simulation predicts that type I tools should create lower deformation in the workpiece and lower tool temperature and wear than the plane faced or type II tools; and that $K = 1.2$ is a good value for the restriction constant. Experimental measurements, with tools of different rake face geometries created by electro-discharge machining, of tool forces, rake face temperatures – using a single-wire thermocouple (Figure 5.19(b), Usui *et al.*, 1978) – and tool wear, support this.

Figure 8.29 shows the final predicted chip shape and the distribution of equivalent plastic

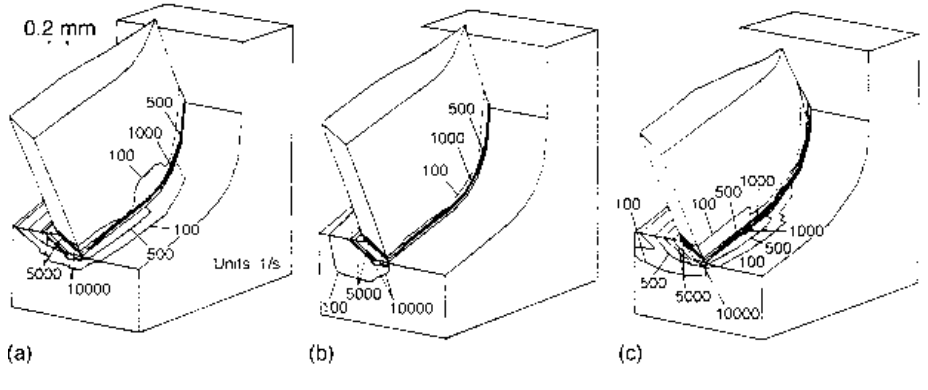


Fig. 8.29 Equivalent plastic strain rate distribution and chip configuration when machining 18%Mn–18%Cr steel at a cutting speed of 60 m/min, $f = 0.2$ mm and $d = 2$ mm with (a) conventional, (b) type I and (c) type II P20 carbide tools

strain rate for the plane-faced tool and type I and II tools with $K = 1.2$. The type I tool produces narrower plastic regions in the chip and workpiece, and less plastic deformation over the finished surface, than the type II and plane tools. As well as the plane and type II tools causing more deformation in the work surface beneath the major cutting edge, the type II tool generates a thicker chip at the minor cutting edge, and the chip flow angle is larger than for the other tools.

Figure 8.30 shows temperature distributions over the rake faces. The dark region represents the contact area with the chip, and the symbol * indicates the location of the highest temperature. The maximum temperature of the type I tool is 50 to 100°C lower than the others. The type II tool produces a higher temperature than the type I tool at the minor cutting edge and nose radius, where the chip contact area is wider. However, the distance of the highest temperature from the major cutting edge is almost the same for both.

Figure 8.31 shows the predicted contour lines of constant wear rate. The distribution and the isotherms in Figure 8.30 are closely correlated because temperature dominates the wear (Table 8.4 and equation (4.1c)). The wear of the type II tool is severe at the corner and near the major cutting edge, while the type I tool yields less wear along all its edges.

Comparisons with experiment are shown in Figures 8.32 to 8.34. Figure 8.32 shows experimental measurements of cutting force variation with restriction constant K for the type I tools. Experimentally, there is a minimum in all the force components at around $K = 1.2$. The predicted forces show a similar tendency: predictions for the conventional and type II tools are also included in the figure.

Figure 8.33 shows the measured and predicted rake face temperatures of the conventional and type I ($K = 1.2$) tools in the direction of chip flow at the midpoint of the depth of cut. A temperature difference of about 100°C can be seen in both the predictions and experiments, although prediction and experiment are not in absolute agreement with each other.

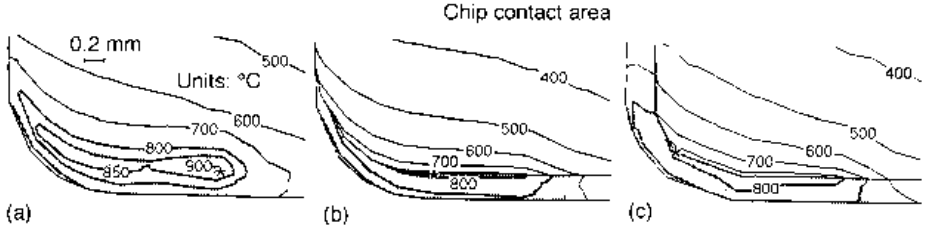


Fig. 8.30 Tool rake face isotherms in the conditions of Figure 8.29: (a) conventional, (b) type I; and (c) type II tools

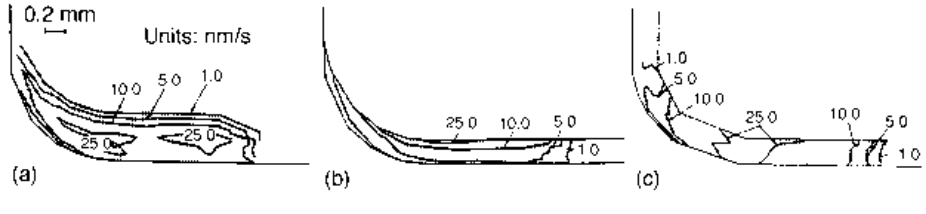


Fig. 8.31 Contours of constant crater wear rate, conditions of Figure 8.29: (a) conventional; (b) type I; and (c) type II tools

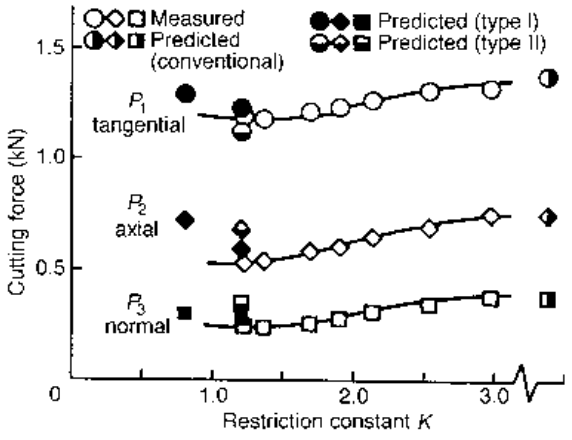


Fig. 8.32 Cutting force dependencies on restriction constant

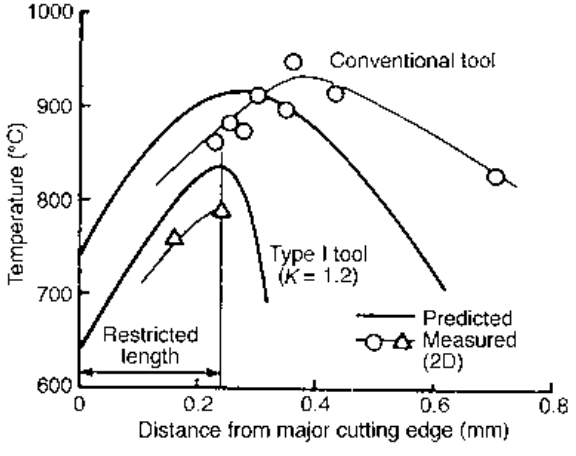


Fig. 8.33 Comparison of predicted and measured rake temperatures at the midpoint of the depth of cut, for plane and type I tools

Figure 8.34 compares the differences in wear profiles at a cut distance of 600 m, obtained both by profilometry and microphotography. The type I tool shows least tool wear, more than 10% less than with the conventional tool: the similarity in wear distribution with that predicted in Figure 8.31 is clear.

In summary, a finite element machining simulation has been employed to analyse the turning of a difficult-to-machine 18%Mn–18%Cr high manganese steel with a sintered carbide three-dimensional cut-away tool. A cut-away design in which the primary restricted contact length varies along the cutting edge in proportion to the uncut chip thickness has been found to give a better performance than one with a restricted contact which is constant along the major and minor cutting edges and around the tool nose radius; and it is also better than a plane rake faced tool. A restriction constant of around 1.2 has been found to give the least cutting forces, leading to reductions in cutting temperature and tool wear.

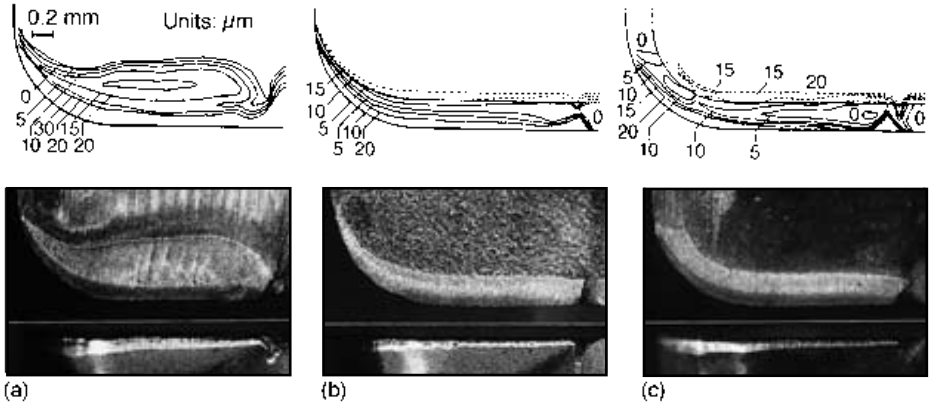


Fig. 8.34 Measured crater depth and optical micrographs of worn tools at a cut distance of 600 m: (a) conventional, (b) and (c) type I and II tools, $K = 1.2$

8.5 Summary

A new concept of computational or virtual machining simulation is starting to emerge, based on the theoretical background surveyed in Chapter 7, to support the increasing demands of high productivity, quality and accuracy of modern automated machining practice. There is no doubt that advances in computing capability and graphical visualization technologies will bring further developments in the field of machining simulation.

At present, finite element simulation is mainly of use to mechanical and materials engineers, as a tool to support process understanding, materials' machinability development and tool design. However, the computing time required by this method is too long for it to be of use in machine shops for online control and optimization, although it can help offline evaluation and rationalization of practical experience.

Online control requires other sorts of machining models. These and their relationships with finite element models are the subject of the next and final chapter of this book, which considers how to use modelling and monitoring in the production engineering context of process planning and improvement.

References

Chao, B. T. and Trigger, K. J. (1959) Controlled contact cutting tools. *Trans ASME J. Eng. Ind.* **81**, 139–151.

Hazra, J., Caffarrelli, D. and Ramalingam, S. (1974) Free machining steels – the behavior of type I MnS inclusions in machining. *Trans ASME J. Eng. Ind.* **97**, 1230–1238.

Jawahir, I. S. (1988) The tool restricted contact effect as a major influencing factor in chip breaking: an experimental analysis. *Annals CIRP* **37**(1), 121–126.

Jawahir, I. S. (1990) On the controllability of chip breaking cycles and modes of chip breaking in metal machining. *Annals CIRP* **39**(1), 47–51.

Jawahir, I. S. and van Luttervelt, C. A. (1993) Recent developments in chip control research and applications. *Annals CIRP* **42**(2), 659–693.

Johnson, W. and Kudo, H. (1962) *The Mechanics of Metal Extrusion.* Manchester: Manchester University Press.

Lemaire, J. C. and Backofen, W. A. (1972) Adiabatic instability in the orthogonal cutting of steel. *Metallurgical Trans.* **3**(2), 477–481.

Maekawa, K., Kitagawa, T. and Childs, T. H. C. (1991) Effects of flow stress and friction characteristics on the machinability of free cutting steels. *Proc. 2nd Int. Conf. on The Behaviour of Materials in Machining*, Institute of Materials, York, pp. 132–145.

Maekawa, K., Ohhata, H. and Kitagawa, T. (1994) Simulation analysis of cutting edge performance of a three-dimensional cut-away tool. In Usui, E. (ed.), *Advancement of Intelligent Production*. Tokyo: Elsevier, pp. 378–383.

Marusich, T. D. (1999) private communication.

Marusich, T. D. and Ortiz, M. (1995) Modelling and simulation of high speed machining. *Int. J. for Num. Methods in Engng*. **38**, 3675–3694.

Nakayama, K. (1962) A study on the chip breakers. *Bull. Japan Soc. Mech. Eng.* **5**(17), 142–150.

Naylor, D. J., Llewellyn, D. T. and Keane, D. M. (1976) Control of machinability in medium-carbon steels. *Metals Technol.* **3**(5/6), 254–271.

Obikawa, T. and Usui, E. (1996) Computational machining of titanium alloy – finite element modeling and a few results. *Trans ASME J. Manufacturing Sci. Eng.* **118**, 208–215.

Obikawa, T., Matumara, T. and Usui, E. (1990) Chip formation and exit failure of cutting edge (1st report). *J. Japan. Soc. Prec. Eng.* **56**(2), 336–342.

Obikawa, T., Sasahara, H., Shirakashi, T. and Usui, E. (1997) Application of computational machining method to discontinuous chip formation. *Trans ASME J. Manufacturing Sci. Eng.* **119**, 667–674.

Pekelharing, A. J. (1974) Built-up edge (BUE): is the mechanism understood? *Annals CIRP* **23**(2), 207–212.

Pekelharing, A. J. (1978) The exit failure in interrupted cutting. *Annals CIRP* **27**(1), 5–8.

Recht, R. F. (1964) Catastrophic thermoplastic shear. *J. Appl. Mechanics* **31**(2), 189–193.

Sandstrom, D. R. and Hodowany, J. N. (1998) Modeling the physics of metal cutting in high speed machining. *Int. J. Machining Sci. and Tech.* **2**, 343–353.

Shaw, M. C., Usui, E. and Smith, P. A. (1961) Free machining steel (part III). *Trans ASME J. Eng. Ind.* **82**, 181–192.

Shinozuka, J. (1998) Analytical prediction of cutting performance of grooved rake face tools. PhD Thesis. Tokyo: Tokyo Institute of Technology.

Shinozuka, J., Obikawa, T. and Shirakashi, T. (1996a) Chip breaking process simulation by thermo-elastic plastic finite element method. *J. Japan Soc. Prec. Eng.* **62**(8), 1161–1166.

Shinozuka, J., Obikawa, T. and Shirakashi, T. (1996b) Chip breaking analysis from the viewpoint of the optimum cutting tool geometry design. *J. Matls Processing Tech.* **62**, 345–351.

Trent, E. M. (1963) Cutting steel and iron with cemented carbide tools (part II). *J. Iron Steel Inst.* **201**, 923–932.

Usui, E., Kikuchi, K. and Hoshi, T. (1964) The theory of plasticity applied to machining with cut-away tools. *Trans ASME J. End. Ind.* **B86**, 95–104.

Usui, E., Shirakashi, T. and Kitagawa, T. (1978) Analytical prediction of three dimensional cutting process (part 3). *Trans ASME, J. Eng. Ind.* **100**, 236–243.

Usui, E., Ihara, T. and Shirakashi, T. (1979) Probabilistic stress-criterion of brittle fracture of carbide tool materials. *Bull. Japan Soc. Prec. Eng.* **13**(4), 189–194.

Usui, E., Maekawa, K. and Shirakashi, T. (1981) Simulation analysis of built-up edge formation in machining of low carbon steel. *Bull. Japan Soc. Prec. Eng.* **15**(4), 237–242.

Usui, E., Ihara, T., Kanazawa, K., Obikawa, T. and Shirakashi, T. (1982) An evaluation method of fracture strength of brittle materials with disk compression test. *J. Mat. Sci. Soc. Japan* **19**(4), 238–243.

Usui, E., Obikawa, T. and Matsumura, T. (1990) Chip formation and exit failure of cutting edge (2nd report). *J. Japan Soc. Prec. Eng.* **56**(5), 911–916.

Vyas, A. and Shaw, M. C. (1999) Mechanics of saw-tooth chip formation in metal cutting. *Trans ASME J. Manuf. Sci. and Engng.* **121**, 163–172.

Williams, J. E., Smart, E. F. and Milner, D. (1970) The metallurgy of machining, Part 2. *Metallurgia* **81**, 51–59.

Yamaguchi, K. and Kato, T. (1980) Friction reduction actions of inclusions in metal cutting. *Trans ASME J. Eng. Ind.* **103**, 221–228.

Yamane, Y., Usuki, H., Yan, B. and Narutaki, N. (1990) The formation of a protective oxide layer in machining resulphurised free-cutting steels and cast irons. *Wear* **139**, 195–208.

9

Process selection, improvement and control

9.1 Introduction

This final chapter deals with the planning and control of machining processes. Planning and control systems are composed of several modules, such as modules for process modelling, optimization and prediction; for selection of tools and cutting conditions; for tool path generation; for machine tool operation; for monitoring and recognition; for diagnosis and evaluation; for learning and tuning. Data and knowledge-base modules support a system's operation. There is overlap between the functions of some of these modules. In the interests of efficient construction and operation, some of the modules may be combined and some may be neglected in any particular system.

The quantitative modelling of machining processes, based on machining theory, with the prediction or simulation that this enables, greatly assists planning and control. Figure 9.1 shows examples of systems containing a simulation module at their heart. The subject of Section 9.2 is process models for prediction, simulation and control, but more widely defined than in previous chapters of this book.

Initial process optimization is the subject of Section 9.3. The tasks and tools of optimization depend on whether there is a single goal or whether there are conflicting goals (and in that case how clear are their priorities); and whether the process is completely or only partly modelled (how clear is the understanding). An example that approaches single goal optimization of a well understood system is optimization of speed, feed and depth of cut to minimize cost (or maximize productivity) once a cutting tool has been selected and part accuracy and finish have been specified. This is the subject of Section 9.3.1. Even in this case, all aspects of the process may not be completely modelled, or some of the coefficients of the model may be only vaguely known. Consequently, the skills of practical machinists are needed. Section 9.3.2 introduces how the optimization process may be recast to include such practical experience, by using fuzzy logic.

Optimization becomes more complicated if it includes selection of the tool (tool holder and cutting edge), as well as operation variables. The tool affects process constraints and, at the tool selection level, constraints and goals can overlap and be in conflict (a surface finish design requirement may be thought of both as a constraint and a goal, in conflict with cost reduction). As a result of this complexity, tool selection in machine shops currently depends more on experience than models. Section 9.3.3 deals with rule-based tool selection systems, a branch of knowledge-based engineering.

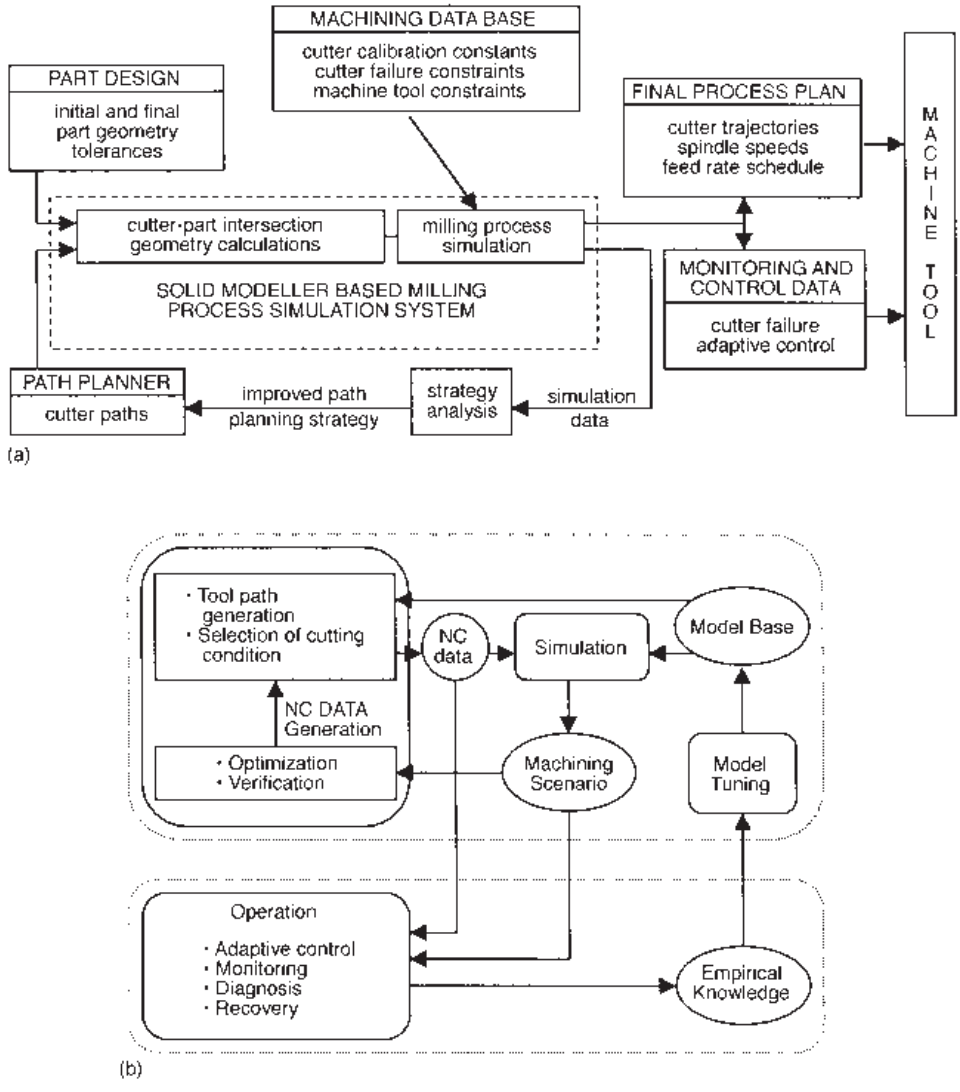


Fig. 9.1 Model-based systems for design and control of machining processes: (a) CAD assisted milling process simulator and planner (Spence and Altintas, 1994) and (b) machining-scenario assisted intelligent machining system (Takata, 1993)

Because what tool is selected depends in part on the speeds, feed and depth of cut that it will experience, tool selection systems commonly include rules on the expected ranges of these variables. However, combined optimization of these and the tool would be better. That is the topic of the last part of Section 9.3.

Section 9.4 is concerned with process monitoring. This is directly valuable for detecting process faults (either gradual, such as wear; or sudden, such as tool failure or wrong cutter path instructions). It may also be used, with recognition, diagnosis and evaluation of cutting states, to improve or tune an initial process model or set of rules. Finally, Section

9.5 is allocated to model (simulation) based control, which is one of the major destinations of machining theory.

9.2 Process models

Models of machining processes are essential for prediction, control and optimization. Especially important are models for cutting force, cutting temperature, tool wear, tool breakage and chatter. Physically based models of these are the main concern of previous chapters of this book. In this chapter, a broader view of modelling is taken, to include empirical and feature-based models constructed by regression or artificial intelligence methods. A model should be chosen appropriate for the purpose for which it is to be used; and modified if necessary. The more detailed (nearer-to-production) the purpose and the quicker the response required of the system, the more likely it is that an empirical model will be the appropriate one; but a physical model may guide the form of the empirical model and its limits of applicability. The different types of models are reviewed here.

Cutting force models are considered first, because of their general importance, both influencing tool breakage, tool wear and dimensional accuracy, as well as determining cutting power and torque. Tool paths in turning are more simple than in milling; and this leads to smaller force variations during a turning than during a milling process. For the purposes of control, force models applied to turning tend to be simpler than those applied to milling. However, accuracy control in milling processes, such as end milling, is very important technologically. Here, two sections are devoted to force models, the first generally to turning and the second specially to end milling.

9.2.1. Cutting force models (turning)

Cutting forces in turning $\mathbf{F}^T = \{F_d, F_f, F_c\}$ may be written in terms of a non-linear system $\mathbf{H}$ and operation variables $\mathbf{x}^T = \{V, f, d\}$:

$$\mathbf{F} = \mathbf{H}(\mathbf{x}) \tag{9.1}$$

The non-linear system $\mathbf{H}$ may be a finite element modelling (FEM) simulator $\mathbf{H}_{FEM}$, as described in Chapters 7 and 8, an analytical model $\mathbf{H}_A$ (for example the three-dimensional energy model described in Section 6.4), a regression model $\mathbf{H}_R$, or a neural network $\mathbf{H}_{NN}$ (Tansel, 1992). The coefficients and exponents of a regression model and the weights of a neural network are most often determined from experimental machining data, by linear regression or back propagation algorithms, respectively. However, they may alternatively be determined from calculated FEM or energy approach results. They then become the means of interpolating a limited amount of simulated data. In addition to the operation variables, a tool's geometric parameters, such as rake angles, tool nose radius and approach angle, may be included in the variables $\mathbf{x}$.

An extended set of variables $\bar{\mathbf{x}}$ can be developed, to include a tool's shape change due to wear $\mathbf{w}$, where $\mathbf{w}$ is a wear vector, the components of which are the types of wear considered: $\bar{\mathbf{x}}^T = \{\mathbf{x}^T, \mathbf{w}^T\}$. The cutting forces may be related to this extended variable set, similarly to equation (9.1):

$$\mathbf{F} = \bar{\mathbf{H}}(\bar{\mathbf{x}}) \tag{9.2a}$$

A regression model example of such a non-linear equation (to be used in Section 9.4), for machining a chromium molybdenum low alloy steel BS 709M40 (British Standard, 1991) with a triple-coated carbide tool insert of grade P30 and shape code SPUN 120312 (International Standard, 1991), held in a tool holder of code CSTPR T (International Standard, 1995), has been established as:

$$\left.\begin{aligned} F_d &= 500f^{0.46}d^{0.810} + 2377(VS^{1.93} - 0.007\ln V) \\ &\quad \times (VB^{0.26} - 0.007\ln V)\,(VN^{-0.33} - 0.007\ln V) \\ F_f &= 629f^{0.30}d^{0.720} + 1199(VS^{3.58} - 0.023\,V^{0.27}) \\ &\quad \times (VB^{-0.66} - 0.23\,V^{0.27})\,(VN^{0.03} - 0.23\,V^{0.27}) \\ F_c &= 1862f^{0.94}d^{1.11} + 2677(VS^{0.24} - 0.05\ln V) \\ &\quad \times (VB^{0.23} - 0.05\ln V)\,(VN^{0.16} - 0.05\ln V) \end{aligned}\right\} \quad (9.2b)$$

where F_d, F_f, and F_c are values in N; V, f and d are in m/min, mm/rev and mm, respectively; and the dimensions of flank wear VB (Chapter 4), notch wear VN and nose wear VS are in mm (Oraby and Hayhurst, 1991).

9.2.2 Cutting force models (end milling)

The end milling process is complex compared with turning, both because of its more complicated machine tool linear motions and its repeated intermittent engagement and disengagement of rotating cutting edges. However, as already written, it is very important from the viewpoint of process control in modern machining technologies. This section deals extensively with end milling because of this importance and also because some of the results will be used in Section 9.5, on model-based process control. A general model is first introduced, followed by particular developments in time varying, peak and average force models, and the use of force models to develop strategies for the control of cutter deflection and part accuracy.

A general model

The three basic operation variables, V, f, d, of turning are replaced by four variables V, f, d_R, d_A in end milling, where, from Chapter 2.2, the cutting speed $V = \pi D\Omega$, the feed f is the feed per tooth $U_{feed}/(N_f\Omega)$, and d_R and d_A are the radial and axial depths of cut. In terms of a non-linear system $\mathbf{H}'$ and operation variables $\mathbf{x}^T = \{V, f, d_R, d_A\}$, the cutting forces on an end mill may be written similarly to equation (9.1):

$$\mathbf{F} = \mathbf{H}'(\mathbf{x}) \quad (9.3)$$

where $\mathbf{F}$ is the combined effect of all the active cutting edges.

End milling's extra complexity relative to turning has led to regression force models $\mathbf{H}'_R$ being most developed and contributing most to its process control. FEM models as in Chapters 7 and 8, $\mathbf{H}'_{FEM}$ and analytical approaches $\mathbf{H}'_A$ (for example Shirakashi *et al*., 1998, 1999; Budak *et al*., 1996), are developing, but are not yet at a level of detail where they may usefully be applied to process control. Neural networks $\mathbf{H}'_{NN}$ have not been of interest.

Time-varying models

Implementations of equation (9.3), able to follow the variations of cutting force with time, may be constructed by considering the contributions of an end mill's individual cutting

edges to the total forces. Figure 9.2(a) – similar to Figure 2.3 but developed for the purposes of process control and which will be used further in Section 9.5 – shows a clockwise-rotating end mill with N_f flutes (four, in the figure). The end mill is considered to move over and cut a stationary workpiece, in the same way that the tool path is generated. A global coordinate system (x', y', z'), fixed in the workpiece, is necessary to define the relative positions of the end mill and workpiece so that instantaneous values of d_R and d_A may be determined. Cutting forces are expressed in a second coordinate system (x, y, z) with axes parallel to (x', y', z') but with the origin fixed in the end mill. The forces are obtained from the summation of force increments calculated in local coordinate systems (r, t_n, z_E) with axes in radial, tangential and axial directions and origins O_E on the helical cutting edges.

When the tool path is a straight line (as in Figure 9.2(a)), it is clear which dimension is

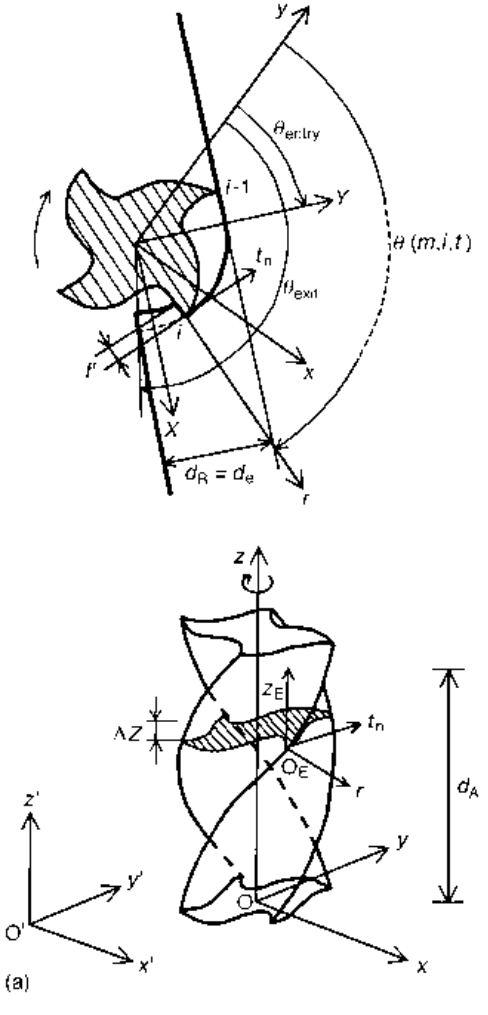
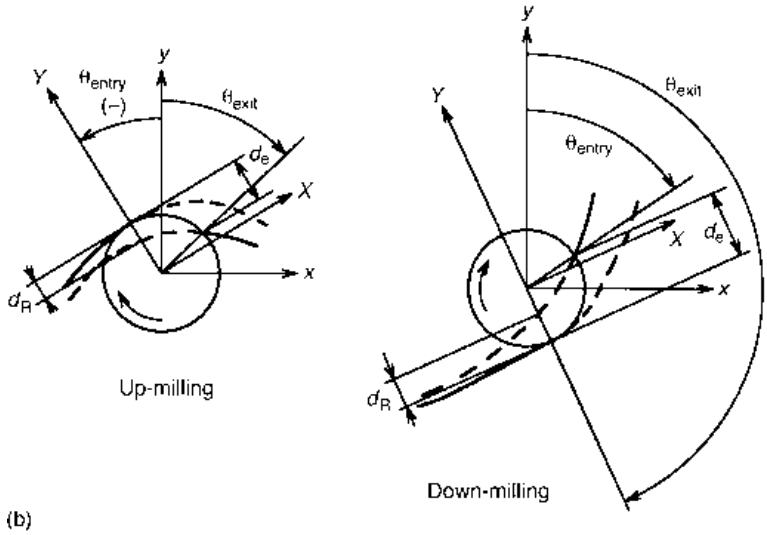


Fig. 9.2 Milling process: (a) coordinates and angles in a slice by slice model and (b) the effective radial depth of cut with curved cutter paths



(b)

Fig. 9.2 *continued*

the radial depth of cut, d_R; but when the tool path is curved (Figure 9.2(b)), there is a difference between the geometrical radial depth d_R and an effective radial depth d_e (described further in the next section): a fourth coordinate system (X, Y, Z) with the same origin as (x, y, z) but co-rotating with the instantaneous feed direction, so that the feed speed U_{feed} is always in the X direction, deals with this.

The starting point of the force calculation is to calculate the instantaneous values of uncut chip thickness f' in a r–t_n plane, along the end mill's cutting edges. For an end mill with non-zero helix angle λ_s, a cutting edge is discretized into M axial slices each with thickness $\Delta z = d_A/M$ (Kline *et al.*, 1982). The plan view in Figure 9.2(a) shows the cutting process in the mth slice from the end mill tip. An edge numbered i proceeds ones numbered less than i. An edge enters into and exits from the workpiece at angles θ_{entry} and θ_{exit} ($\theta_{entry} < \theta_{exit}$) measured clockwise from the y-axis, as shown. At a time t, the angular position of the point O_E on the ith edge of slice m is $\theta(m, i, t)$, also measured clockwise from the y-axis. Choosing the origin of time so that $\theta(1, 1, 0) = 0$,

$$\theta(m, i, t) = \Omega t + \frac{2\pi}{N_f}(i-1) - \frac{2(m-1)\Delta z}{D}\tan\lambda_s \tag{9.4}$$

For the cutting edge at O_E to be engaged in cutting,

$$\theta_{entry} + 2\pi n \le \theta(m, i, t) \le \theta_{exit} + 2\pi n \tag{9.5a}$$

where n is any integer. Then the cutting forces acting on the thin slice around O_E are

$$\Delta\mathbf{F}_{\mathbf{x}}(m, i, t) = \begin{Bmatrix} \Delta F_x \\ \Delta F_y \\ \Delta F_z \end{Bmatrix} = \begin{Bmatrix} -F_t^* \cos(\theta(m, i, t)) - F_r^* \sin(\theta(m, i, t)) \\ F_t^* \sin(\theta(m, i, t)) - F_r^* \cos(\theta(m, i, t)) \\ F_z^* \end{Bmatrix} f'(m, i, t)\Delta z \tag{9.5b}$$

where F_t^*, F_r^* and F_z^* are the specific cutting forces in the tangential, radial and axial directions, respectively.

On the other hand, when the cutting edge at O_E is not engaged in cutting,

$$\theta_{\text{exit}} + 2\pi(n-1) < \theta(m, i, t) < \theta_{\text{entry}} + 2\pi n \tag{9.5c}$$

and

$$\Delta\mathbf{F}_{\mathbf{x}}(m, \text{i}, t) = \mathbf{0} \tag{9.5d}$$

The total cutting forces are obtained from the sum of the forces on all the slices:

$$\mathbf{F}_{\mathbf{x}}(t) = \begin{Bmatrix} F_x(t) \\ F_y(t) \\ F_z(t) \end{Bmatrix} = \sum_{m=1}^{M} \sum_{i=1}^{N_{\text{f}}} \Delta\mathbf{F}_{\mathbf{x}}(m, i, t) \tag{9.6}$$

A physical force model would seek to express the specific forces in equation (9.5b) as functions of cutting speed, uncut chip thickness and depth of cut. The purpose of end milling process control force models is to determine force variations under conditions of varying d_R and d_A, commonly at constant cutting speed. The specific cutting forces are usually written as a regression model good for one speed only, in which the variables are chosen from d_R, d_A, f (feed per tooth) and f'; and the influence of cutting speed is subsumed in the regression coefficients. Equations (9.7) are three examples of regression equations, due respectively to Kline *et al.* (1982), Kline and De Vor (1983) and Moriwaki *et al.* (1995):

$$\begin{Bmatrix} F_t^* \\ F_r^*/F_t^* \end{Bmatrix} = \begin{Bmatrix} k_{t0} + k_{t1}d_R + k_{t2}d_A + k_{t3}f + k_{t4}d_Rd_A + k_{t5}d_Rf \\ \quad + k_{t6}d_Af + k_{t7}d_R^2 + k_{t8}d_A^2 + k_{t9}f^2 \\ k_{r0} + k_{r1}d_R + k_{r2}d_A + k_{r3}f + k_{r4}d_Rd_A + k_{r5}d_Rf \\ \quad + k_{r6}d_Af + k_{r7}d_R^2 + k_{r8}d_A^2 + k_{r9}f^2 \end{Bmatrix} \tag{9.7a}$$

$$\begin{Bmatrix} F_t^* \\ F_r^*/F_t^* \end{Bmatrix} = \begin{Bmatrix} k_{t1}(f'_{\text{av}})^{-k_{t2}} \\ k_{r1}(f'_{\text{av}})^{-k_{r2}} \end{Bmatrix} \tag{9.7b}$$

or

$$\begin{Bmatrix} F_t^* \\ F_r^*/F_t^* \end{Bmatrix} = \begin{Bmatrix} k_{t0} + k_{t1}(f')^{-k_{t2}} \\ k_{r0} + k_{r1}(f')^{-k_{r2}} \end{Bmatrix} \tag{9.7c}$$

where the k_{ij} ($i = t, r$; $j = 0$ to 9) are constants and f'_{av} is the average uncut chip thickness per cut. These formulations are used for the model (simulation) based process control to be described in Section 9.5.

Peak and average force models

If only the peak or mean cutting force is to be used for process control, the force equation (9.6) may be simplified, by working with the (X, Y, Z) coordinate system; and it becomes practical explicitly to re-introduce the influence of cutting speed. As the tool always feeds in the X direction, it is the depth of cut, d_e, in the Y direction, measured from the tool entry point, which enters into calculations of the uncut chip thickness and which acts as the effective radial depth of cut. It is this which should be used in force regression models. Consequently, the peak resultant cutting force $F_{\text{R, peak}}$ and its direction measured clockwise from the Y axis, $\theta_{\text{R, peak}}$, may be simply expressed as

$$F_{\mathrm{R,\,peak}} = F_{\mathrm{R}}^{*} f^{m_{\mathrm{R1}}} d_{\mathrm{e}}^{m_{\mathrm{R2}}} d_{\mathrm{A}}^{m_{\mathrm{R3}}} V^{m_{\mathrm{R4}}} + F_{\mathrm{R0}} \tag{9.8a}$$

$$\theta_{\mathrm{R,\,peak}} = \theta_{\mathrm{R}}^{*} f^{m_{\mathrm{R5}}} (D - d_{\mathrm{e}})^{m_{\mathrm{R6}}} d_{\mathrm{A}}^{m_{\mathrm{R7}}} V^{m_{\mathrm{R8}}} + \theta_{\mathrm{R0}} \tag{9.8b}$$

where F_{R}^{*}, F_{R0}, θ_{R}^{*}, θ_{R0} and $m_{\mathrm{R}j}$ ($j = 1$ to 8) are constants. (In a slotting process, when $d_{\mathrm{e}} = D$, the cutting conditions have the least influence on $\theta_{\mathrm{R,\,peak}}$.)

The X and Y force components obtained from equations (9.8a) and (9.8b) are

$$F_{X\mathrm{p}} = F_{\mathrm{R,\,peak}} \sin \theta_{\mathrm{R,\,peak}} \tag{9.8c}$$

$$F_{Y\mathrm{p}} = F_{\mathrm{R,\,peak}} \cos \theta_{\mathrm{R,\,peak}} \tag{9.8d}$$

The mean values may be expressed similarly to the peak values.

An example of a regression model in the form of equations (9.8) (to be used in the next section) can be derived from down-milling data for machining the nickel chromium molybdenum AISI 4340 steel (ASM, 1990), used by Kline in developing equation (9.7a) (Kline *et al.*, 1982). With $F_{\mathrm{R,\,mean}}$ in newtons and $\theta_{\mathrm{R,\,mean}}$ in degrees, the feed per tooth, the effective radial depth of cut and the axial depth of cut in mm, and no information on the influence of cutting speed,

$$F_{\mathrm{R,\,mean}} = 38 f^{0.7} d_{\mathrm{e}}^{1.2} d_{\mathrm{A}}^{1.1} + 222 \tag{9.8e}$$

$$\theta_{\mathrm{R,\,mean}} = 4.86 f^{0.15} (D - d_{\mathrm{e}})^{0.9} - 26 \tag{9.8f}$$

Dimensional accuracy and control

The force component F_{Y} causes relative deflection between the tool and workpiece normal to the feed direction. In principle, this gives rise to a dimensional error unless it is compensated. Figure 9.3 shows the direction of forces acting on an end mill: the force component

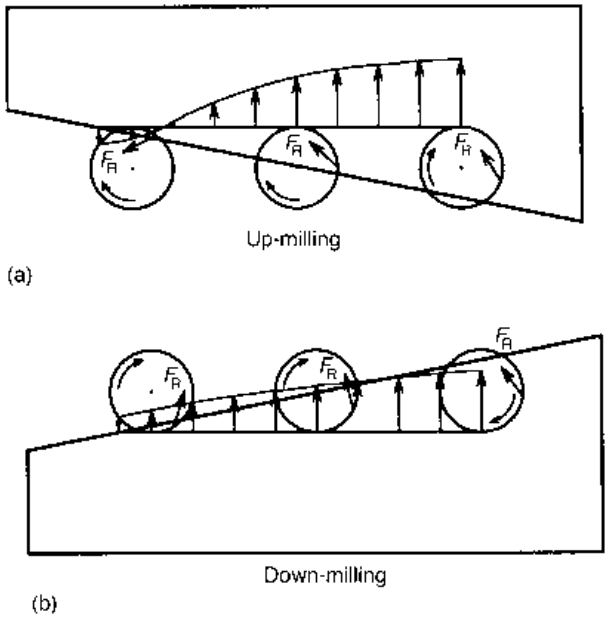


Fig. 9.3 Machining error and cutting force direction in up and down-milling

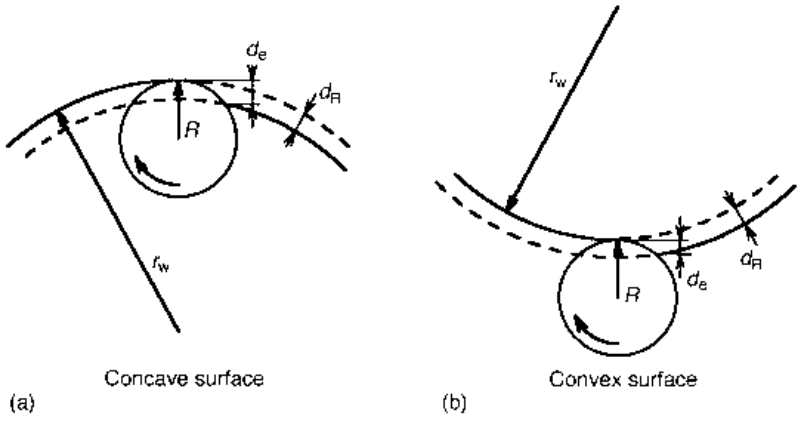


Fig. 9.4 The effective radial depth of cut in milling concave and convex surfaces

F_Y with a helical end mill is always positive, irrespective of up- or down-milling, except for up-milling with a small effective radial depth of cut. Hence, down-milling gives rise to undercut; and up-milling to overcut unless the radial depth is small – in which case, anyway, the deflection is small.

An additional factor, of practical importance, must be considered when end milling a curved surface. Other things being equal, the deflection in milling a concave surface is greater than in milling a convex one. Figure 9.4 shows two surfaces of constant curvature, one concave, one convex, both being end milled to a radius r_w by a cutter of radius R (or diameter D), by removing a radial depth d_R. The effective radial depth of cut, d_e, as defined previously, is greater than d_R for the concave surface and less than d_R for the convex one. According to equations (9.8), for the same values of f and d_A, the force (and hence the tool deflection) will be larger for milling the concave than for milling the convex surface.

The size of this effect is conveniently estimated after introducing a radial depth ratio, c_r, equal to d_e/d_R. From the geometry of Figure 9.4,

$$\left.\begin{array}{ll} \text{for a concave surface} & (r_w - d_R)^2 - (r_w - d_e)^2 = R^2 - (R - d_e)^2 \\ \text{for a convex survace} & (r_w + d_R)^2 - (r_w + d_e)^2 = R^2 - (R - d_e)^2 \end{array}\right\} \quad (9.9a)$$

Hence

$$\left.\begin{array}{ll} \text{for a concave surface} & c_r = \dfrac{d_e}{d_R} = \dfrac{2r_w - d_R}{2r_w - D} \\ \text{for a convex surface} & c_r = \dfrac{d_e}{d_R} = \dfrac{2r_w + d_R}{2r_w + D} \end{array}\right\} \quad (9.9b)$$

Since $d_R \le D$, $c_r \ge 1$ for a concave surface, $c_r \le 1$ for a convex surface and $c_r = 1$ for slotting ($d_R = D$) or for a flat surface ($r_w = \infty$).

It often happens in practical operations that the radius of curvature r_w decreases to the value of the end mill diameter D. Then the ratio c_r can increase up to a value of around

two. The consequent force change depends on the appropriate regression equation, such as equation (9.8e). Another way of explaining this effect is to note that the stock removal rate (which is the volume removed per unit time) increases as $(c_r - 1)$ at a constant feed speed and axial depth of cut.

The equations (9.9b) can be used, with equations (9.8), to control exactly the dimensional error of surfaces of constant curvature; and to control approximately the error when curvature changes only slowly along the end mill's path. Such a case occurs when cutting a scroll surface. As shown in Figure 9.5, the radius of curvature gradually reduces as a cutter moves from the outside to the centre. According to equations (9.9b), the decrease in the radius of curvature increases the effective radial depth of cut on a concave surface and decreases it on a convex one; and thus changes the cutting force and direction too. Since dimensional error is caused by the *Y* force component, a condition of constant error is

$$F_{Yp} = c_0 \tag{9.10a}$$

When the radial and axial depth of cut, d_R and d_A, and the cutting speed V are constant, the feed should be changed to satisfy the following (from equations (9.8)):

$$(c_1 f^{m_{R1}} d_e^{m_{R2}} + F_{R0}) \cos(c_2 f^{m_{R5}} (D - d_e)^{m_{R6}} + \theta_{R0}) = c_0 \tag{9.10b}$$

where c_1 and c_2 are constants. If the change in the direction of the peak resultant force due to a change in the effective radial depth of cut has only a small influence on the *Y* force component (as is often the case in down-milling), the feed should be changed by

$$f \approx c_3 (d_e)^{-m_{R2}/m_{R1}} \qquad \text{or} \qquad f \approx c_4 (c_r)^{-m_{R2}/m_{R1}} \tag{9.10c}$$

where c_3 and c_4 are constants. On a concave surface the feed must be decreased, but it should be increased on a convex surface provided an increase in feed does not violate other constraints, for example imposed by maximum surface roughness requirements.

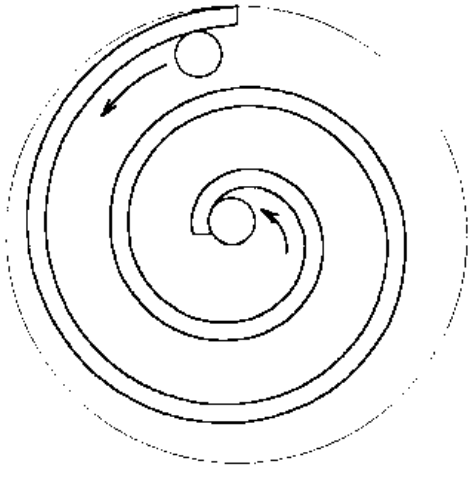


Fig. 9.5 Milling of scroll surfaces

Corner cutting

c_r values much larger than 2 occur when a surface's radius of curvature changes suddenly with position. An extreme and important case occurs in corner cutting. Figure 9.6(a) (an example from Kline *et al.*, 1982) shows corner cutting with an end mill of 25.4 mm diameter. The surface has been machined beforehand, leaving a radial stock allowance of 0.762 mm on both sides of the corner and a corner radius of 25.4 mm. The corner radius to be finished is 12.7 mm. Thus, there is no circular motion of the finish end mill's path, but just two linear motions. Figure 9.6(b) shows, for this case, the changes in the effective radial depth of cut d_e and the mean cutting forces F_X and F_Y with distance l_r from the corner. l_r is negative when the tool is moving towards the corner and positive when away from it. The mean cutting forces are calculated from equations (9.8e) and (9.8f). The effective radial depth of cut increases rapidly by a factor of more than 20 as the end mill approaches the corner; $c_r = 25.1$ at $l_r = 0$. The force component normal to the machined surface increases with the effective radial depth of cut to cause a large dimensional error.

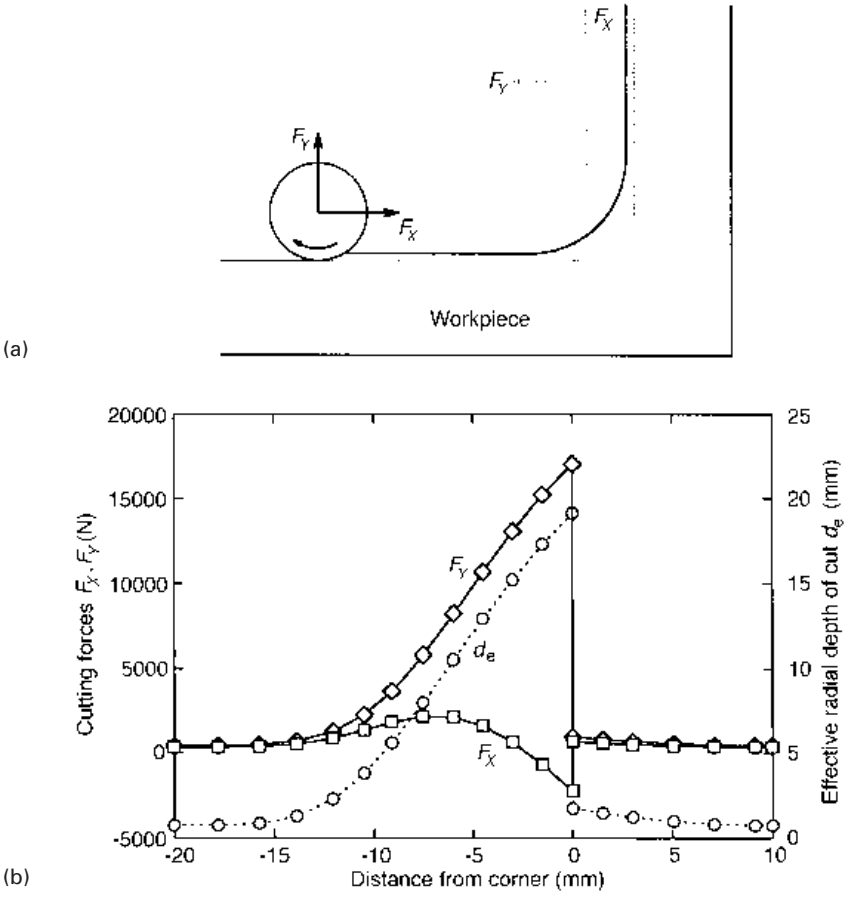
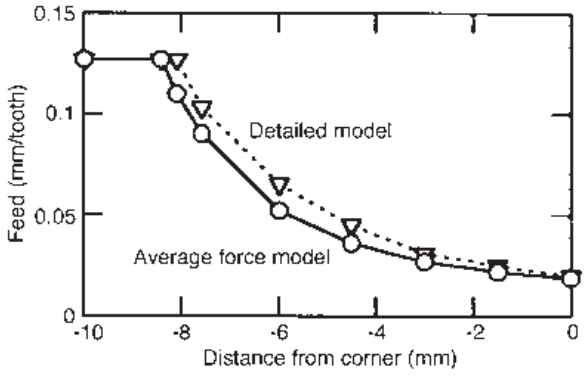


Fig. 9.6 Corner cutting: (a) tool path (Kline *et al.*, 1982); (b) calculated change in cutting forces (average force model with axial depth of cut $d_A = 38.1$ mm) and (c) feed control under constant cutting force $F_Y = 4448$ N



(c)

Fig. 9.6 *continued*

Even if the pre-machined corner has the same radius (12.7 mm) as the end mill and the nominal stock allowance is small, the maximum value of c_r during corner cutting, which is then given by

$$c_r = \frac{D}{2d_R} + \left(\frac{D}{d_R} - 1 \right)^{½} \tag{9.11}$$

is very large: $c_r = 22.4$ at $l_r = 0$, when $D = 25.4$ mm and $d_R = 0.762$ mm. It follows from equation (9.11) that a decrease in radial depth of cut does not lead to decreases in cutting force and dimensional error if corner cutting is included in finish end milling. The dimensional accuracy (error) should be controlled by changing the feed, as in the case of machining a scroll surface. In order for the mean force component to be constant during the corner cut in Figure 9.6(a), the feed is recommended (from equations (9.8)) to decrease as shown in Figure 9.6(c). Kline's results, from detailed modelling based on equations (9.6) and (9.7a), are plotted for comparison. The more simple model may be preferred for control, because of its ease and speed of use.

9.2.3 Cutting temperature models

Cutting temperature is a controlling factor of tool wear at high cutting speeds. Thermal shock and thermal cracking due to high temperatures and high temperature gradients cause tool breakage. Thermal stresses and deformation also influence the dimensional accuracy and surface integrity of machined surfaces. For all these reasons, cutting temperature θ has been modelled, in various ways, using the operation variables $\mathbf{x}$ and a non-linear system Θ:

$$\theta = \Theta(x) \tag{9.12}$$

The non-linear system may be an FEM simulator $\Theta_{FEM,}$ as described in Chapters 7 and 8, a finite difference method (FDM) simulator Θ_{FDM} (for example Usui *et al.*, 1978, 1984), an analysis model Θ_A as described in Section 2.3, a regression model Θ_R, or a neural network Θ_{NN}. An extended temperature model, in terms of extended variables $\bar{\mathbf{x}}$ and a non-linear system $\bar{\Theta}$ may be developed to include the effects of wear – similar to the extended cutting force model of equation (9.2a).

If only the average tool–chip interface temperature is needed, analysis models are often sufficient, as has been assessed by comparisons with experimental measurements (Stephenson, 1991). However, tool wear is governed by local temperature and stress: to obtain the details of a temperature distribution, a numerical simulator is preferable – and regression or neural net simulators are not useful at all.

Advances in personal computers make computing times shorter. The capabilities of FEM simulators have already been reported in Chapters 7 and 8. An FDM simulator $\bar{\Theta}_{FDM}$, using a personal computer with a 200 MHz CPU clock, typically requires only about ten seconds to calculate the temperature distribution on both the rake face and flank wear land in quasi-steady state orthogonal cutting; while with a 33 MHz clock, the time is around two minutes (Obikawa *et al.*, 1995). An FDM simulator can, in a short time, report the influences of cutting conditions and thermal properties on cutting temperature (Obikawa and Matsumura, 1994).

9.2.4 Tool wear models

A wear model for estimating tool life and when to replace a tool is essential for economic assessment of a cutting operation. Taylor's equation (equation (4.3)) is an indirect form of tool wear model often used for economic optimization as described in Chapter 1.4 and again in Section 9.3. However, it is time-consuming to obtain its coefficients because it requires much wear testing under a wide range of cutting conditions. This may be why Taylor's equation has been little written about since the 1980s. Instead, the non-linear systems $\mathbf{W}$ and $\dot{\mathbf{W}}$ directly relating wear and wear rate to the operation variables of cutting speed, feed and depth of cut

$$\mathbf{w} = \mathbf{W}(\mathbf{x}) \tag{9.13a}$$

$$\dot{\mathbf{w}} = \dot{\mathbf{W}}(\mathbf{x}) \tag{9.13b}$$

have been intensively studied, not only for wear prediction but for control and monitoring of cutting processes as well.

Although wear mechanisms are well understood qualitatively (Chapter 4), a comprehensive and quantitative model of tool wear and wear rate with multi-purpose applicability has not yet been presented. However, wear rate equations relating to a single wear mechanism, based on quantitative and physical models, and used for a single purpose such as process understanding or to support process development, have been presented since the 1950s (e.g. Trigger and Chao, 1956). In addition to the operation parameters, the variable $\mathbf{x}$ typically includes stress and temperature on the tool rake and/or clearance faces, and tool-geometric parameters. The thermal wear model of equation (4.1c) (Usui *et al.*, 1978, 1984) has, in particular, been applied successfully to several cutting processes. For example, Figure 9.7 is concerned with the prediction, at two different cutting speeds, of flank wear rate of a carbide P20 tool at the instant when the flank wear land VB is already 0.5 mm (Obikawa *et al.*, 1995). Because the wear land is known experimentally to develop as a flat surface, the contact stresses and temperatures over it must be related to give a local wear rate independent of position in the land. In addition, the heat conduction across the wear land, between the tool and finished surface, depends on how the contact stress influences the real asperity contact area (as considered in Appendix 3). The temperature distributions in Figure 9.7(a) and the flank contact temperatures and stresses in Figure 9.7(b)

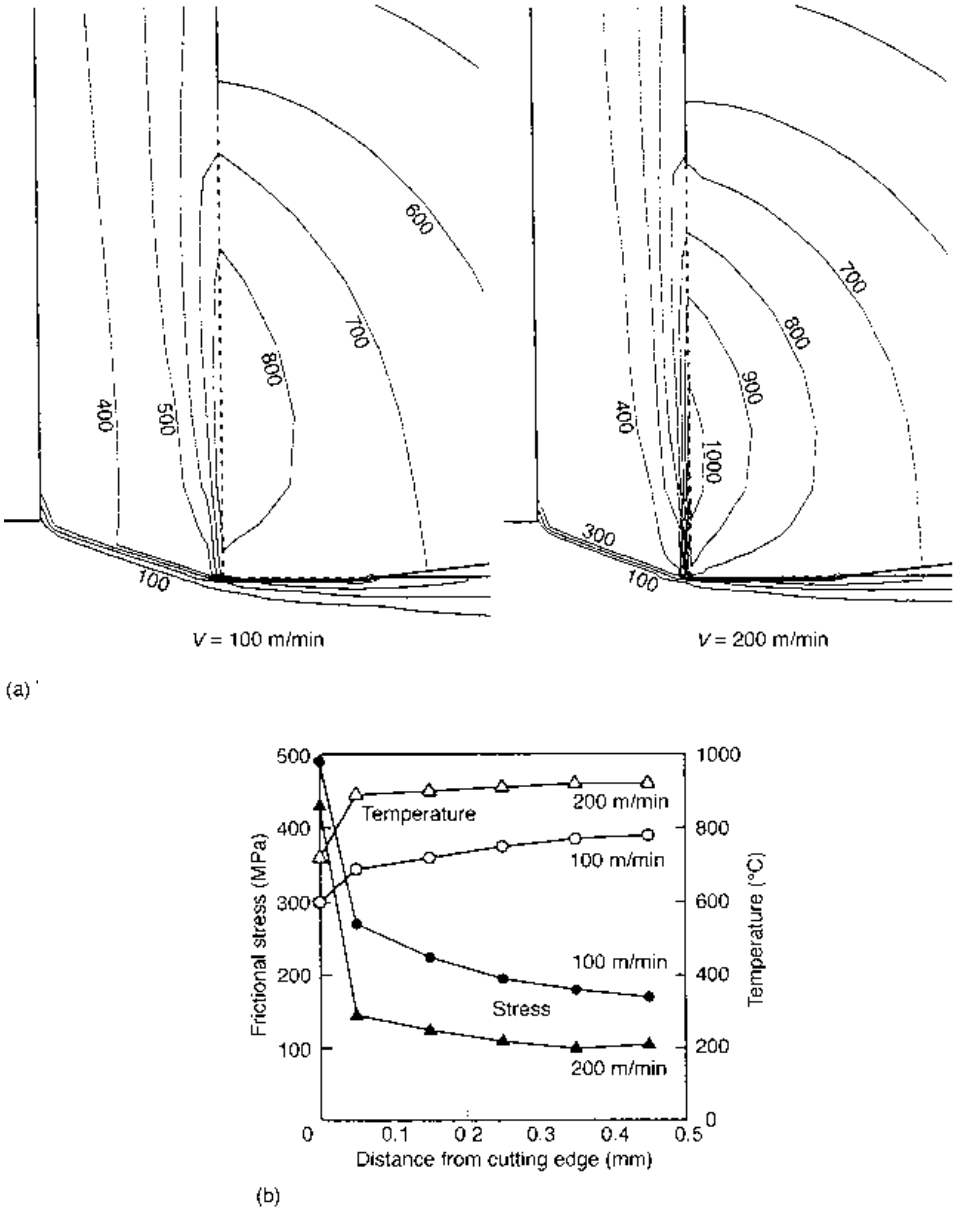


Fig. 9.7 An example of calculated results by a simulator $\bar{\Theta}_{FDM}$ (a) temperature distribution in chip and tool and (b) temperature and frictional stress on the worn flank (Obikawa *et al.*, 1995)

have been obtained from an FDM simulator, $\bar{\Theta}_{FDM}$, of the cutting process in which these conditions were considered simultaneously. The flank wear rate d(*VB*)/d*t* was estimated (from the stresses and temperatures; and for *VB* = 0.5 mm) to be 0.0065 mm/min at a cutting speed of 100 m/min and 0.024 mm/min at 200 m/min, and its change as *VB* increased could be followed.

When control and monitoring of wear are the main purposes of modelling, other variables are added to **x**, such as tool forces and displacements and acoustic emission signals – sometimes in the form of their Fourier or wavelet transform spectra (or expansion coefficients in the case of digital wavelet transforms) – as will be considered in more detail in Section 9.4. In the absence of a quantitative model between **w** or $\dot{\mathbf{w}}$ and **x**, the non-linear system is usually represented by a neural network $\mathbf{W}_{NN}$ or $\dot{\mathbf{W}}_{NN}$. Even when a quantitative relation is known, neural networks are often used because of their rapid response. For example, an empirical model relating cutting forces and wear, such as that of equation (9.2b), may be transformed inversely by neural network means to

$$\mathbf{w} = \mathbf{W}_{NN}(\bar{\mathbf{F}}) \tag{9.13c}$$

where $\bar{\mathbf{F}}^{T} = \{\mathbf{x}^{T}, \mathbf{F}^{T}\}$. In the conditions to which it applies, equation (9.13c) may be used with force measurements to monitor wear (Section 9.4.3).

9.2.5 Tool fracture models

Tool breakage is fatal to machining and difficult to plan against in production (other than extremely conservatively) because of the strong statistically random nature of its occurrence. Once a tool is broken, machining must stop for tool changing and possibly the workpiece may also be damaged and must be changed. Models of fracture during cutting, based on fundamental principles of linear fracture mechanics, attempting to relate failure directly to the interaction of process stresses and tool flaws, have met with only marginal success. It is, in practice, most simply assumed that tool breakage occurs when the cutting force F exceeds a critical value $F_{critical}$, which may decrease with the number of impacts N_i between an edge and workpiece, as expected of fatigue (as considered earlier, in Figure 3.25). A first criterion of tool breakage is then

$$F = F_{critical}(N_i) \tag{9.14a}$$

However, there is a significant scatter in the critical force level at any value of N_i. It is well known that the probability statistics of fracture and fatigue of brittle materials, such as cemented carbides, ceramics or cermets, may be described by the Weibull distribution function. The Weibull cumulative probability, p_f, of tool fracture by a force F, at any value of N_i, is

$$p_f = 1 - \exp\left[-\left(\frac{F - F_l}{F_0}\right)^b\right] \equiv 1 - \exp\left[-a\left(\frac{F - F_l}{F_h - F_l}\right)^b\right] \tag{9.14b}$$

where F_l and F_h are forces with a low and high expectation of fracture after N_i impacts and F_0, a and b are constants. Alternatively, and as considered further in Section 9.3, p_f may be identified with the membership function μ of a fuzzy set (fuzzy logic is introduced in Appendix 7)

$$\mu(F) = S(F, F_l, F_h) \tag{9.14c}$$

where the form of S is chosen from equations like (A7.4a) or (A7.4b) to approximate p_f.

Statistical fracture models in terms of cutting force are useful for the economic planning of cutting operations, supporting tool selection and change strategies once a tool's dependencies of F_l and F_h on N_i have been established. They are not so useful for tool

design, where one purpose is to develop tool shape to reduce and resist forces. Then, more physically-based modelling is needed, to assess how tool shape affects tool stresses; and then how stresses affect failure. An approximate approach of this type has already been considered in Chapter 3, supported by Appendix 5, to relate a tool's required cutting edge included angle to its material's transverse rupture stress.

A more detailed approach is to estimate, from surface contact stresses obtained by the machining FEM simulators of Chapters 7 and 8, the internal tool stress distribution – also by finite element calculation – and then to assess from a fracture criterion whether the stresses will cause failure. This is the approach used in Chapter 8.2.2 to study failure probabilities in tool–work exit conditions. The question is: what is an appropriate tri-axial fracture stress criterion? A deterministic criterion introduced by Shaw (1984) is shown in Figure 9.8(a), whilst a probabilistic criterion developed from work by Paul and Mirandy (1976) and validated for the fatigue fracture of carbide tools by Usui *et al.* (1979) is shown in Figure 9.8(b). Both show fracture loci in (σ_1,σ_3) principal stress space when the third principal stress $\sigma_2 = 0$. Whereas Figure 9.8(a) shows a single locus for fracture, Figure 9.8(b) shows a family of surfaces T to U. σ_c is a critical stress above which fracture

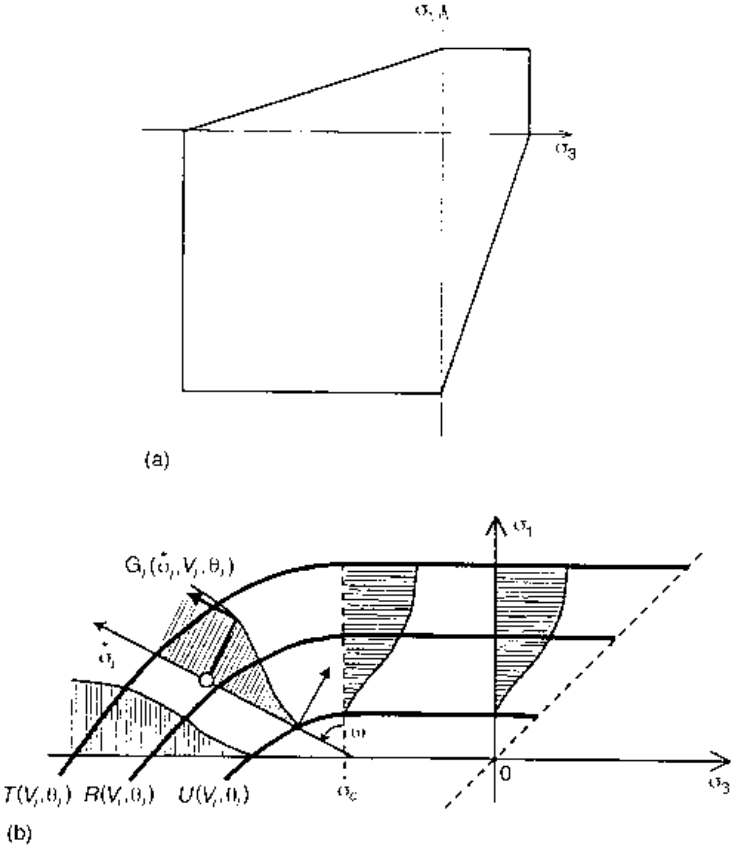


Fig. 9.8 Fracture criteria of cutting tools: (a) Shaw's (1984) deterministic criterion and (b) Usui *et al.*'s (1979) probabilistic one

depends only on the maximum principal stress. T represents 90%, R 50% and U 0% probability of failure of a volume V_i of material after N_i impacts at temperature θ_i. The loci contract with increasing V_i and N_i and θ_i (Shirakashi *et al.*, 1987). The use of these criteria for the design of tool geometry has been demonstrated by Shinozuka *et al.* (1994) and Shinozuka (1998). The approach will become appropriate for tool selection once FEM cutting simulation can be conducted more rapidly than it currently can.

9.2.6 Chatter vibration models

It is possible for periodic force variations in the cutting process to interact with the dynamic stiffness characteristics of the machine tool (including the tool holder and workpiece) to create vibrations during processing that are known as chatter. Chatter leads to poor surface finish, dimensional errors in the machined part and also accelerates tool failure. Although chatter can occur in all machining processes (because no machine tool is infinitely stiff), it is a particular problem in operations requiring large length-to-diameter ratio tool holders (for example in boring deep holes or end milling deep slots and small radius corners in deep pockets) or when machining thin-walled components. It can then be hard to continue the operation because of chatter vibration. The purpose of chatter vibration modelling is to support chatter avoidance strategies. One aspect is to design chatter-resistant machine tool elements. After that has been done, the purpose is then to advise on what feeds, speeds and depths of cut to avoid. This section only briefly considers chatter, to introduce some constraints that chatter imposes on the selection of cutting conditions. More detailed accounts may be found elsewhere (Shaw, 1984; Tlusty, 1985; Boothroyd and Knight, 1989).

The most commonly studied form of chatter is known as regenerative chatter. It can occur when compliance of the machine tool structure allows cutting force to displace the cutting edge normal to the cut surface and when, as is common, the path of a cutting edge over a workpiece overlaps a previous path. It depends on the fact that cutting force is proportional to uncut chip thickness (with the constant of proportionality equal to the product of cutting edge engagement length ($d/\cos\psi$) and specific cutting force k_s). If both the previous and the current path are wavy, say with amplitude a_0, it is possible (depending on the phase difference between the two paths) for the uncut chip thickness to have a periodic component of amplitude up to $2a_0$. The cutting force will then also have a periodic component, up to $[2a_0(d/\cos\psi)k_s]$, at least when the two paths completely overlap. The component normal to the cut surface may be written $[2a_0(d/\cos\psi)k_d]$ where k_d is called the cutting stiffness. This periodic force will in turn cause periodic structural deflection of the machine tool normal to the cut surface. If the amplitude of the deflection is greater than a_0, the surface waviness will grow – and that is regenerative chatter. If the machine tool stiffness normal to the cut surface is written k_m (but see the next paragraph for a more careful definition), chatter is avoided if

$$\frac{2dk_d}{\cos\psi k_m} < 1 \qquad \text{or} \qquad d < \frac{k_m \cos\psi}{2k_d} \tag{9.15a}$$

The maximum safe depth of cut increases with machine stiffness and reduces the larger is the cutting stiffness (i.e. it is smaller for cutting steels than aluminium alloys).

Real machine tools contain damping elements. It is their dynamic stiffness, not their static stiffness, that determines their chatter characteristics. k_m above is frequency and

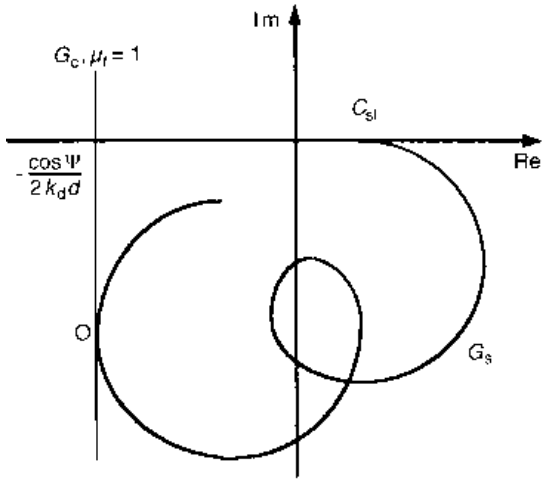


Fig. 9.9 Unconditional chatter limit

damping dependent. A structure's dynamic stiffness is often described in terms of its compliance transfer function G_s – how the magnitude of its amplitude-to-force ratio, and the phase between the amplitude and force, vary with forcing frequency. Figure 9.9 represents a possible G_s in a polar diagram. It also shows the compliance transfer function G_c of the cutting process when there is total overlap ($\mu_f = 1$) between consecutive cutting paths (the real part of G_c is $-\cos\psi/(2k_d d)$, as considered above, and the minus sign has been introduced as chatter occurs when positive tool displacements give decreases of uncut chip thickness). The physical description leading to equation (9.15a) may be recast in the language of dynamics modelling, to take properly into account the frequency dependence of both the amplitude and phase of the structural response, via the statement that cutting is unconditionally stable if G_c and G_s do not intersect (Tlusty, 1985). At the unconditional stability limit, the two transfer functions touch (as shown in the figure).The maximum depth of cut d_{uc} which is unconditionally stable is then

$$d_{uc} = -\frac{\cos\psi}{2k_d[\mathrm{Re}(G_s)]_{min}} \tag{9.15b}$$

where $[\mathrm{Re}(G_s)]_{min}$ is the minimum real part of the transfer function of the structure: it more exactly defines the inverse of k_m in equation (9.15a).

If the structure is linear with a single degree of freedom, the minimum real part $[\mathrm{Re}(G_s)]_{min}$ is proportional to the static compliance C_{st}. In that case, d_{uc} may be written, with c_d a constant, as

$$d_{uc} = \frac{c_d}{C_{st}} \tag{9.15c}$$

Equations (9.15b) or (9.15c) provide a constraint on the maximum allowable depth of cut in a machining process. Another type of constraint may occur in the absence of regenerative

chatter, if periodic variation of the cutting force occurs due to discontinuous, serrated or shear localized chip formation. This may cause forced chatter vibration. For chips of similar shape, the chip formation frequency f_{chip} is linearly proportional to the cutting speed and inversely proportional to the feed:

$$f_{\text{chip}} = c_{\text{f}} \frac{V}{f} \tag{9.15d}$$

where c_{f} is a constant. Chatter vibration can occur when the chip formation frequency is close to one of the natural frequencies of the structure. Hence, the ratio of cutting speed to feed that should be avoided in that case is given by

$$\frac{f_{\text{n}i}}{c_{\text{f}}} (1 - \Delta) \leq \frac{V}{f} \leq \frac{f_{\text{n}i}}{c_{\text{f}}} (1 + \Delta) \tag{9.15e}$$

where $f_{\text{n}i}$ is the ith natural frequency and $f_{\text{n}i}\Delta$ is the half width of the unavailable frequency band.

9.3 Optimization of machining conditions and expert system applications

Previous chapters and sections have described aspects of machining that are amenable to theoretical modelling. Some cutting phenomena have been modelled quantitatively, others described qualitatively. As is well known, however, not all details of machining technology have yet been captured in machining theories. Heuristic (practical experience) knowledge and the skills of machinists are still needed to optimize conditions in production. Although a computer aided manufacturing (CAM) system, by reference to a database, can automatically provide tool paths, and recommend tools and cutting conditions, sorted according to workpiece materials, cutting operations (turning, milling, boring, drilling, etc) and cutting types (finishing, light roughing, roughing, heavy roughing), the heuristic knowledge and skills of machinists are indispensable for trouble shooting and final optimization of cutting conditions, beyond the capabilities of data-based recommendations.

Almost by definition, the heuristic knowledge and skills of machinists for selecting tools and cutting conditions are hard to describe explicitly or quantitatively. Moreover, skilled machinists have not much interest in self-analysis, nor in such descriptions; nor, typically, in the economics of machining. Machinists' goals are somehow to find a better solution that satisfies all the constraints to a particular problem: cutting time, dimensional accuracy, power, tool life, stability, etc. Satisfactory cutting performance is their subjective measure of evaluation, especially such aspects as good surface finish, avoidance of chatter and excellent chip control.

The dependence of optimization on heuristic knowledge implies that the objectives and rules of machining may not all be explicitly stated. In that sense machining is a typical ill-defined problem. Reducing the lack of definition by representing machinists' knowledge and skills in some form of model description must be a step forward. Fortunately, for the

last two decades, knowledge-based engineering (e.g. Barr and Feigenbaum, 1981, 1982) and fuzzy logic (e.g. Zimmermann, 1991) have been developed and applied to machining problems. Three application areas are considered here, first the optimization of cutting conditions for given tool and work materials, from an economic point of view; then the selection of cutting tools; and finally the simultaneous selection of tools and cutting conditions. Rational (theoretical) knowledge economic optimization methods, under the assumption that they are sufficient, are reviewed and developed in Section 9.3.1 before their supplementation by subjective, fuzzy, optimization, in Section 9.3.2. Tool selection methods (by heuristic means) and simultaneous selection of tools and cutting conditions (with the integration of rational and heuristic knowledge) are the topics of Sections 9.3.3 and 9.3.4.

9.3.1 Model-based optimization of cutting conditions

When everything is known about a process, a criterion by which to judge its optimization can be objectively defined and constraints on the optimization can be established. A feasible region from which optimal operating conditions may be selected can be established and finally an optimal set of conditions can be chosen. These activities are illustrated by the selection of cutting conditions from an economic point of view, as introduced in Section 1.4.

Objective function

As described in Section 1.4, minimum cost, maximum productivity or maximum efficiency are typical criteria of economic optimization. In considering the minimum cost criterion for optimizing a turning operation (the maximum productivity criterion will also be briefly treated), the same analysis that leads to equation (1.8) for the operation cost per part C_p, but before constraining it by substituting the form of Taylor's tool life (equation (1.3)), gives

$$C_p = C_c t_{load} + C_c \frac{t_{mach}}{f_{mach}} + C_c t_{ct} \frac{t_{mach}}{T} + C_t \frac{t_{mach}}{T} \tag{9.16a}$$

where $C_c = M_t + M_w$. When cutting a cylindrical workpiece of diameter D and length L, the cutting time t_{mach} is

$$t_{mach} = \frac{\pi D L d_a}{Vfd} \tag{9.16b}$$

where d_a is the radial stock allowance. The substitution of Taylor's equation (equation (4.3)) (dependent on f and d as well as V, whereas equation (1.3) only included V dependence) and equation (9.16b) into equation (9.16a), yields the objective function for minimum cost:

$$C_p = C_c t_{load} + \frac{\pi D L d_a}{Vfd} \left\{ \frac{C_c}{f_{mach}} + (C_c t_{ct} + C_t) \frac{V^{1/n_1} f^{1/n_2} d^{1/n_3}}{C'} \right\} \tag{9.16c}$$

The objective function to be minimized for maximum productivity is the total time

$$\left.\begin{aligned} t_{\text{total}} &= t_{\text{load}} + \frac{t_{\text{mach}}}{f_{\text{mach}}} + t_{\text{ct}} \frac{t_{\text{mach}}}{T} \\ &= t_{\text{load}} + \frac{\pi D L d_{\text{a}}}{V f d} \left(\frac{1}{f_{\text{mach}}} + t_{\text{ct}} \frac{V^{1/n_1} f^{1/n_2} d^{1/n_3}}{C'} \right) \end{aligned}\right\} \tag{9.16d}$$

Constraints

For a given combination of tool, workpiece and machine tool, the cutting conditions become optimal when the operation cost is minimized, subject to constraints g on the operation variables $\mathbf{x}$ given by

$$g_i(\mathbf{x}) \le g_{ic} \qquad (i = 1, 2, \ldots, N_{\text{c}}) \tag{9.17}$$

where N_{c} is the number of constraints. In modern machining systems there may be many constraints, for example the following.

Chip breakability. This, the first constraint (C1), must be taken into consideration in modern machining systems, leading to:

$$\text{(C1)} \quad g_1(f, d) \le g_{1c} \tag{9.18a}$$

If depth of cut affects chip breakability independently of the feed, (C1) becomes

$$\text{(C1}') \quad d_{\text{l}} \le d \le d_{\text{u}} \tag{9.18b}$$

$$\text{(C1}'') \quad f_{\text{l}} \le f \le f_{\text{u}} \tag{9.18c}$$

where d_{l} and d_{u} are the lower and upper limits of depth of cut, and f_{l} and f_{u} are the lower and upper limits of feed, respectively. These limits depend on the type of chip breaker described in Section 3.2.8.

Tool geometry and stock allowance. The depth of cut and feed are limited by tool geometry and the stock allowance as well:

$$\text{(C2)} \quad d \le a_1 l_{\text{c}} \cos \psi \tag{9.19}$$

$$\text{(C3)} \quad f \le a_2 r_{\text{n}} \tag{9.20}$$

$$\text{(C4)} \quad d \le d_{\text{a}} \tag{9.21}$$

where a_1 and a_2 are constants, and l_{c} is the effective edge length of an insert.

Surface finish. In finishing, the surface finish should be an important constraint: when the finish is geometrically determined by the tool nose radius (Diniz *et al.*, 1992)

$$\text{(C5)} \quad R_{\text{a}} = \frac{f^2}{31.3 r_{\text{n}}} \le R_{\text{a, max}} \tag{9.22}$$

where $R_{\text{a,max}}$ is the required surface finish.

Chatter. Chatter limits (C6) have been given by equation (9.15b), (9.15c) or (9.15e), and are often critical when the workpiece or tool is not rigid.

Maximum operation time per part. t_{max} may be a constraint:

$$\text{(C7)} \quad t_{\text{total}} = t_{\text{load}} + \frac{\pi D L d_{\text{a}}}{VFd}\left(\frac{1}{f_{\text{mach}}} + \frac{t_{\text{ct}}}{T}\right) \le t_{\max} \tag{9.23a}$$

If tool life T is much longer than the time for tool change t_{ct},

$$\text{(C7}') \quad t_{\text{total}} = t_{\text{load}} + \frac{\pi D L d_{\text{a}}}{VFd}\left(\frac{1}{f_{\text{mach}}}\right) \le t_{\max} \tag{9.23b}$$

Maximum rotational speed N_{max}. This limits the cutting speed. Writing spindle speed as N_{s}:

$$\text{(C8)} \quad V = 2\pi D N_{\text{s}} \le 2\pi D N_{\max} \tag{9.24}$$

Maximum spindle motor power P_{lim}. This also provides constraints

$$\text{(C9)} \quad F_{\text{c}} V \le P_{\text{lim}} \tag{9.25a}$$

When a regression model of cutting force with a non-linear system $\mathbf{H}_{\text{R}}$ is given, this may take a form such as

$$\text{(C9}') \quad F_{\text{c}} V = k_{\text{s}} f^{m_1} d^{m_2} V^{m_3+1} \le P_{\text{lim}} \tag{9.25b}$$

where k_{s} is the specific cutting force, and m_1, m_2 and m_3 are constants (here the regression model differs from that in equation (9.2b)).

Force limits. The cutting forces are limited by factors such as, among others, tool breakage, slip between the chuck and workpiece, and dimensional accuracy due to tool and workpiece deflection

$$\text{(C10)} \quad F_{\text{j}} = k_j f^{m_{\text{j}1}} d^{m_{\text{j}2}} V^{m_{\text{j}3}} \le F_{j,\max} = \min\{F_{j1,\max}, \ldots, F_{ji,\max}, \ldots\} \tag{9.26a}$$

$$\text{(C11)} \quad R = \sqrt{F_1^2 + F_2^2 + F_3^2} \le R_{\max} = \min\{R_{1,\max}, \ldots, R_{i,\max}, \ldots\} \tag{9.26b}$$

where $j = 1, 2, 3$ represents the three orthogonal directions of force components; $F_{ji,\max}$ and $R_{i,\max}$ are the maximum force component and maximum resultant force permissible for factor i, respectively, and $\min\{\ldots\}$ is the minimum operator. For tool breakage, equation (9.14a) may be used for a set of deterministic constraints.

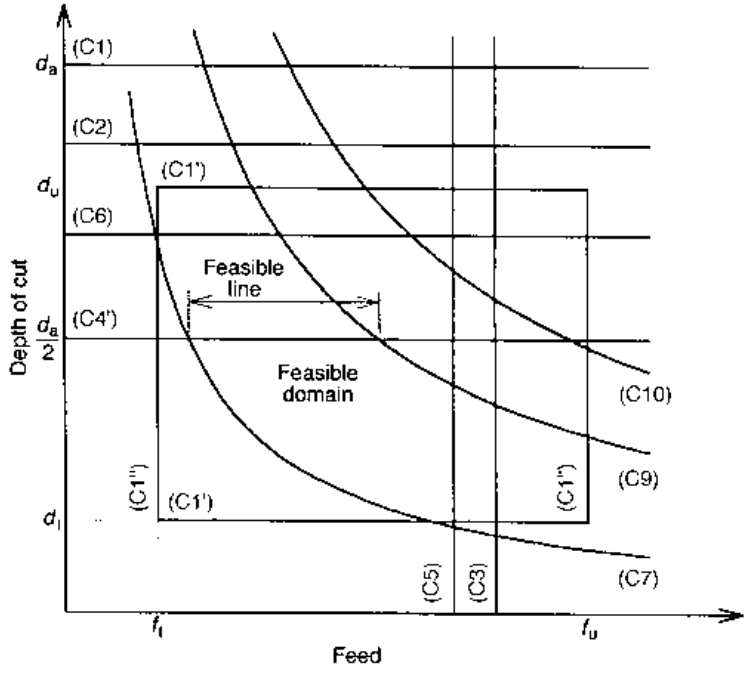
Other limits. There may be other constraints, depending on the cutting operation.

Feasible space

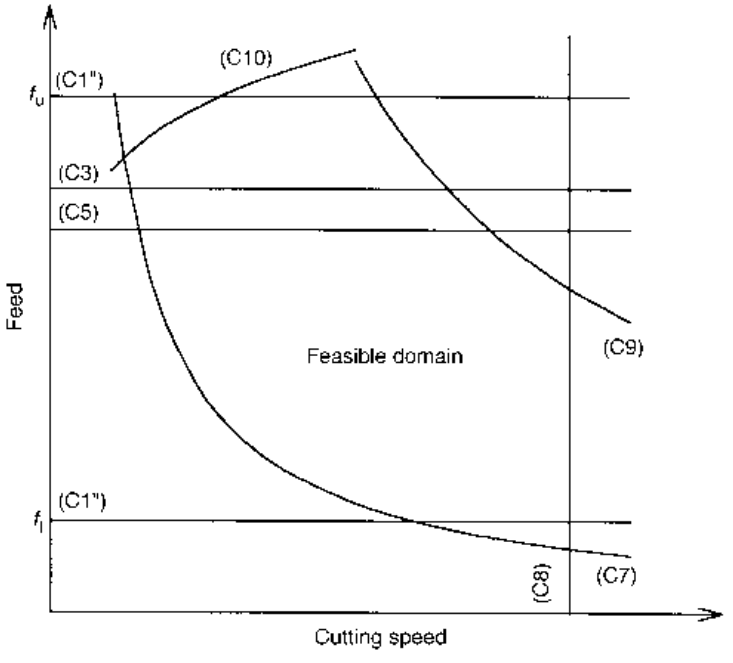
The feasible feed, depth of cut and cutting speed space for a particular cutting operation is the space that satisfies all the constraints. It is inside and on a closed surface:

$$h(V, f, d) \le h_{\text{c}} \tag{9.27a}$$

When the cutting speed or the depth of cut are specified, the feasible domains in the (f, d) or (V, f) planes respectively are given inside and on closed lines as shown in Figures 9.10(a) or (b):



(a)



(b)

Fig. 9.10 Constraints and feasible regions of machining conditions in (a) (f, d) and (b) (V, f) planes

$$h_V(f, d) \le h_{Vc} \tag{9.27b}$$

$$h_d(V, f) \le h_{dc} \tag{9.27c}$$

Each segment of the closed lines represents a limit due to one constraint. Lines numbered C1 to C10 represent the corresponding constraints described by equations (9.18) to (9.26).

In the case of roughing operations, if n passes are chosen for removing the stock allowance d_a, the depth of cut in each pass is usually taken as $d_n = d_a/n$. In this case

$$A_d = \{d_n\} = \{d_a, d_a/2, d_a/3, \ldots\} \tag{9.27d}$$

is a set of depths of cut available for machining. Then the feasible space and domain shrink to a finite number of planes and lines, respectively:

$$h(V, f, d_i) \le h_c \quad (d_i \in A_d) \tag{9.27e}$$

$$h_V(f, d_i) \le h_{Vc} \quad (d_i \in A_d) \tag{9.27f}$$

The lines of equation (9.27f) are schematically shown in Figure 9.10(a).

Optimum conditions

Equation (9.16c), with C_p constant, represents a surface of constant operation cost in (V, f, d) space. The surface may be superimposed on the surfaces of feasible space, as shown in Figure 9.11. The form of equation (9.16c) ensures that the operation cost is minimum where the surface of constant cost just touches the boundary of feasible space $h(V, f, d) = h_c$ or the set of planes $h(V, f, d_i) = h_c$.

Since the constants n_1, n_2 and n_3 of Taylor's equation (4.3) have the relation, $n_1 < n_2 < n_3$ – e.g. $n_1/n_2 \approx 0.77$ and $n_1/n_3 \approx 0.37$ for HSS tools (Stephenson and Agapiou, 1997) – tool life is most sensitive to cutting speed and second most sensitive to feed, among the operation variables. Therefore, the point of tangency M_{opt} between the surface of constant cost and the boundary of feasible space will locate at a coordinate of large depth of cut d_{opt}, large feed f_{opt} and intermediate cutting speed V_{opt}. In Figures 9.11(a) and (b), this point is, as is usual, placed at the upper right corner (vertex) M_V (f_{opt}, d_{opt}) and at a point M_d (V_{opt}, f_{opt}) on the upper boundary of the respective feasible domains.

When the upper boundary of the feasible domain $h_d(V, f)$ is represented by a straight line, $f = f_{opt}$, the minimization of the operation cost with respect to the cutting speed (with constant feed f_{opt} and constant depth of cut d_{opt}),

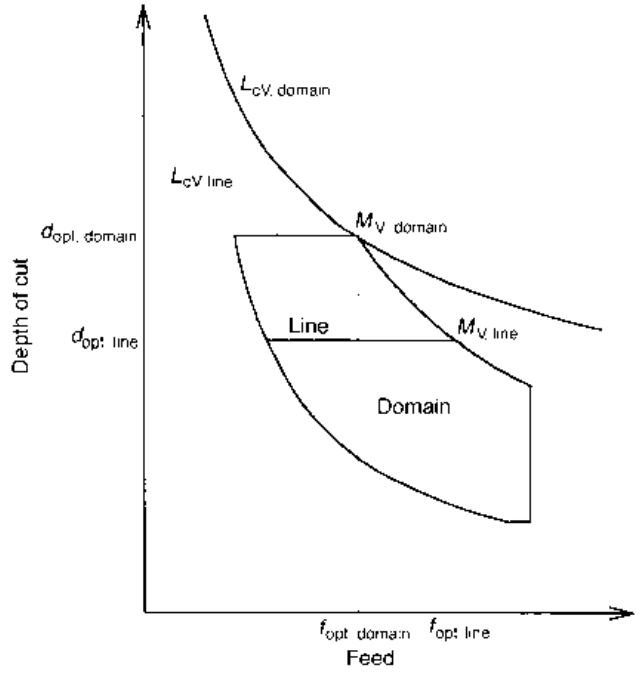
$$(\partial C_p/\partial V)_{f=f_{opt}, d=d_{opt}} = 0 \tag{9.28a}$$

yields the optimum cutting speed

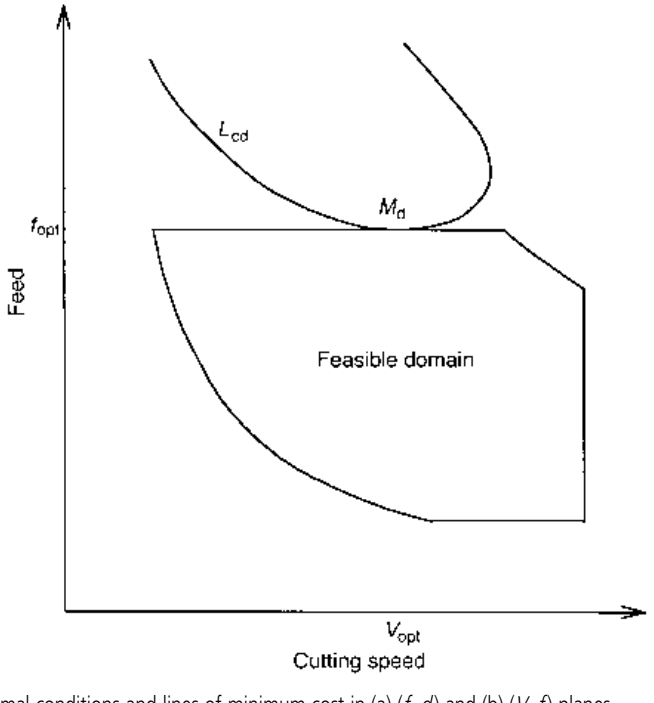
$$V_{opt} = \left(\frac{n_1}{1 - n_1} \cdot \frac{C_c C'}{(C_c t_{ct} + C_t) f_{mach}} \frac{1}{f_{opt}^{1/n_2} d_{opt}^{1/n_3}} \right)^{n_1} \tag{9.28b}$$

It is assumed that the optimum point M_d (V_{opt}, f_{opt}) is not outside the feasible domain. The insertion of equation (9.28b) in Taylor's equation (4.3) and equation (9.16c) leads to the optimum tool life, T_{opt}, and the minimum cost, C_{opt}, respectively:

$$T_{opt} = \frac{1 - n_1}{n_1} \cdot \frac{(C_c t_{ct} + C_t) f_{mach}}{C_c} \tag{9.28c}$$



(a)



(b)

Fig. 9.11 Optimal conditions and lines of minimum cost in (a) (f, d) and (b) (V, f) planes

$$C_{opt} = C_c t_{load} + \pi DLd_a \left(\frac{f_{mach}(1-n_1)}{C_c} \right)^{n_1-1} \left(\frac{C_c t_{ct} + C_t}{C' n_1} \right)^{n_1} f_{opt}^{(n_1-n_2)/n_2} d_{opt}^{(n_1-n_3)/n_3} \tag{9.28d}$$

By replacing f_{opt} and d_{opt} by f and d, respectively, equation (9.28d) expresses the line of the minimum cost L_{cV} in an (f, d) plane:

$$d = \left(\frac{\pi DLd_a}{C_{opt} - C_c t_{load}} \right)^{\frac{n_3}{n_3-n_1}} \left(\frac{f_{mach}(1-n_1)}{C_c} \right)^{\frac{n_3(n_1-1)}{n_3-n_1}} \left(\frac{C_c t_{ct} + C_t}{C' n_1} \right)^{\frac{n_1 n_3}{n_3-n_1}} f^{\frac{n_3(n_1-n_2)}{n_2(n_3-n_1)}} \tag{9.29a}$$

Since the exponents of Taylor's equation have relations $n_1/n_2 \approx 0.77$ and $n_1/n_3 \approx 0.37$ for HSS tools, and exponents of the force model (equations (9.2b), (9.25b)) have a relation $m_1/m_2 \approx 0.85$ for an alloy steel, the exponent of f in equation (9.29a), which is negative, may satisfy the relation

$$\left| \frac{n_3(n_1-n_2)}{n_2(n_3-n_1)} \right| = \frac{1-n_1/n_2}{1-n_1/n_3} < \frac{m_1}{m_2} < 1 \tag{9.29b}$$

Thus, even if the constraint C9′ or C10 in equations (9.25b) or (9.26a) is the boundary segment of the feasible domain $h_V(f, d) \le h_{Vc}$, the line of minimum cost L_{cV} passes through the vertex M_V as described above.

On the other hand, the substitution $C_p = C_{opt}$ and $d = d_{opt}$ into equation (9.16c) yields the line of minimum cost L_{cd} in the (V, f) plane:

$$\frac{(C_{opt} - C_c t_{load}) d_{opt}}{\pi DLd_a} Vf - \frac{(C_c t_{ct} + C_t)}{C'} d_{opt}^{1/n_3} V^{1/n_1} f^{1/n_2} - \frac{C_c}{f_{mach}} = 0 \tag{9.29c}$$

Advances in tool materials, tool geometrical design and tool making technologies decrease the cost of consuming cutting edges C_t. This results in increases in the optimal cutting speed. The lines of minimum cost L_{cV} and L_{cd} are respectively shown schematically in Figures 9.11(a) and (b).

If maximum productivity rather than minimum cost is specified as the criterion for optimization, a faster cutting speed is always the result. The optimum depth of cut d_{opt} and the optimum feed f_{opt}, being fixed at M_V in Figure 9.11, are not affected by changing the criterion. Minimization of the operation time t_{total} in equation (9.16d) with respect to the cutting speed yields the optimum cutting speed V'_{opt},

$$V'_{opt} = \left(\frac{n_1}{1-n_1} \cdot \frac{C'}{t_{ct} f_{mach}} \frac{1}{f_{opt}^{1/n_2} d_{opt}^{1/n_3}} \right)^{n_1} \tag{9.30}$$

When $C_t = 0$, $V'_{opt} = V_{opt}$.

Generally, the minimization of an objective function under the action of constraints may be solved by non-linear programming methods.

Critical constraints

A constraint, the limit line of which contains the optimum point M_V or M_d, is called a critical constraint and the limit line a critical line. Two critical constraints are possible for each feasible domain in the (f, d) and (V, f) planes, whilst three are possible in (V, f, d) feasible space. Since different tools have different constraint coefficients, the feasible space may change when a specified tool is changed. If a critical line moves outward, there is always a possibility to find better cutting conditions that further decrease the operation cost C_{opt}. This is why the tool and cutting conditions must be optimized simultaneously. The simultaneous optimization of tool and cutting conditions is described later, in Section 9.3.4.

9.3.2 Fuzzy logic based optimization of cutting conditions

The best tool cutting conditions may be obtained if all the coefficients in the objective function and constraints are known *a priori*. However, the cost data associated with machining and the constants of Taylor's equation for a particular combination of tool, workpiece and machine tool, for example, are not always accurate. Additionally, not all the constraints due to chip breakability, chatter limit, surface finish, etc are precisely modelled.

Vagueness in the coefficients and constraints may be naturally modelled by fuzzy logic, as in the case of the Weibull distribution for tool breakage, already considered in equation (9.14c). Modelling by fuzzy logic brings about a new way to optimize cutting conditions, and also tool selection (Zimmermann, 1976).

Fuzzy optimization

The constraints of equation (9.17) may be considered to be a crisp or conventional (the terminology is described in Appendix 7) set $\mathbf{R}_i$ of functions of the operation variables $\mathbf{x}$ (V, f, d):

$$\mathbf{R}_i = \{\mathbf{x} \mid g_i(\mathbf{x}) \le g_{ic}\} \qquad (i = 1, 2, \ldots, N_c) \tag{9.31a}$$

Then, the feasible space of machining, $h(f, d, V) \le h_c$, is given by the intersection $\mathbf{H}_c$ of sets $\mathbf{R}_i$:

$$\mathbf{H}_c = \{\mathbf{x} \mid h(\mathbf{x}) \le h_c\} \equiv \bigcap_{i=1}^{N_c} \mathbf{R}_i \tag{9.31b}$$

In these (crisp) terms, the feasibility of machining f_m may be defined as

$$f_m(\mathbf{x}) = \begin{cases} 1 & \mathbf{x} \in \mathbf{H}_c \\ 0 & \mathbf{x} \notin \mathbf{H}_c \end{cases} \tag{9.31c}$$

On the other hand, cutting operation constraints may be represented by a fuzzy set $\tilde{\mathbf{R}}_i$ with membership functions

$$\mu_i(g_i(\mathbf{x})) = \begin{cases} 1 & 0 \le g_i(\mathbf{x}) \le g_{ic-} \\ \dfrac{g_{ic+} - g_i(\mathbf{x})}{g_{ic+} - g_{ic-}} & g_{ic-} < g_i(\mathbf{x}) \le g_{ic+} \\ 0 & g_{ic+} < g_i(\mathbf{x}) \end{cases} \tag{9.32a}$$

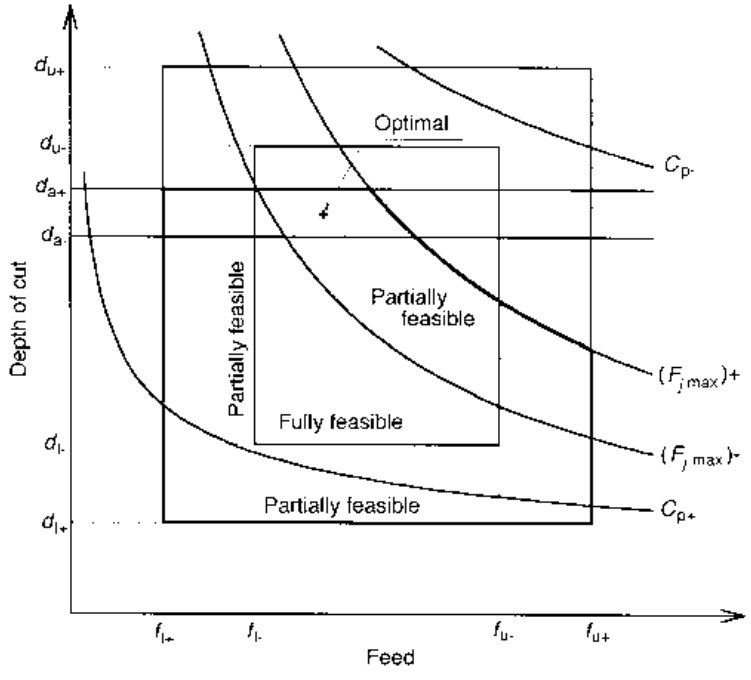


Fig. 9.12 Fuzzy optimization of cutting conditions; only three constraints 1, 4 and 10 are considered

where g_{ic-} and g_{ic+} are constants. The maximum tolerance of the fuzziness is $g_{ic+} - g_{ic-}$. If $g_{ic-} = g_{ic+}$, the fuzzy set $\tilde{\mathbf{R}}_i$ is identical to the crisp set $\mathbf{R}_i$. When a constraint has a probabilistic nature, such as the tool breakage criterion in equation (9.14b) it is naturally modelled by a membership function as in equation (9.14c). Similar to equation (9.31b), the feasible space of machining is given by the intersection $\tilde{\mathbf{H}}_c$ of the fuzzy set $\tilde{\mathbf{R}}_i$:

$$\mu_{\tilde{H}_c}(\mathbf{x}) = \bigwedge_{i=1}^{N_c} \mu_i(g_i(\mathbf{x})) \tag{9.32b}$$

where Λ is the fuzzy operator representing the minimum operation.

The membership $\mu_{\tilde{H}_c}(\mathbf{x})$ represents the feasibility of machining as well: $f_m(\mathbf{x}) = \mu_{\tilde{H}_c}(\mathbf{x})$. Figure 9.12 shows schematically the feasibility space of equation (9.32b). It is seen that there is an intermediate space with feasibility $0 < f_m < 1$ between the fully feasible ($f_m(\mathbf{x}) = 1$) and unfeasible ($f_m(\mathbf{x}) = 0$) spaces.

Like constraints, the objective function (9.16c) is represented by a fuzzy set $\tilde{\mathbf{R}}_0$ with membership functions:

$$\mu_0(g_0(\mathbf{x})) \equiv \begin{cases} 1 & 0 \le C_p(\mathbf{x}) \le C_{p-} \\ \dfrac{C_{p+} - C_p(\mathbf{x})}{C_{p+} - C_{p-}} & C_{p-} < C_p(\mathbf{x}) \le C_{p+} \\ 0 & C_{p+} < C_p(\mathbf{x}) \end{cases} \tag{9.32c}$$

Regarding the objective function in equation (9.32c) as one of the constraints for fuzzy optimization, optimal conditions are found from the value of the variable $\mathbf{x}(V, f, d)$ that maximizes the membership

$$\mu_s(\mathbf{x}) = \bigwedge_{i=0}^{N_c} \mu_i(g_i(\mathbf{x})) \tag{9.32d}$$

An example of fuzzy optimization of tool and cutting conditions will be presented in Section 9.3.4.

9.3.3 Knowledge-based expert systems for tool selection

The previous two sections assume that there is a feasible space in which optimization can be implemented. It is in the interests of cutting tool manufacturers to make sure that that is so, by designing tool holders and inserts – which give chip control, stability, low wear at high speeds, and so on – that are not too constraining on process operation. As there are many constraints on the boundaries of feasible space, and usually it is not initially clear which are critical, tool selection currently relies more on the skills of machinists than does the choice of subsequent operation conditions. Tool selection systems mirror this, in relying strongly on knowledge-based engineering. (In addition, if no tool can be selected, that is a matter for process research and development rather than for process optimization.)

A number of different reasoning systems have developed in the field of knowledge-based engineering – names such as production, blackboard, semantic network, frame, object and predicate calculus are used to describe them (Barr and Feigenbaum, 1981, 1982). Tool selection systems to be described in this section are if (a condition is met) – then (take an action) rule-based (or 'production') expert systems. They all have three essential elements: a workpiece description file (or working memory), to hold a description of a required shape change to be machined; a set of rules relating machining operations and conditions to tool selection (a rule base or file, or production memory); and a way of selecting, interpreting and acting upon the rules (an inference engine or interpreter).

They model the human thinking process in that a rule can be added to or deleted from the rule base, or be modified by experience, without necessarily affecting other rules. This makes them easy to develop. They differ in complexity, depending on whether the rules are complete and well-established, each leading to single actions not in conflict with each other; or whether they are vague and overlap, with possibilities of conflict between them. In the first case, application of the rules will lead to a single (monotonic) route of reasoning, ending up with a right answer. In the second case, methods of compromise are necessary and different experts might reach different answers.

They also, like experts, have a range of points of view. Some (most simple) systems are workpiece oriented, making a recommendation of ideal tool characteristics, leaving it to the user to determine if such a tool is available. These systems only need a working memory, a production memory and an interpreter. Other systems are tool oriented, recommending a specific tool that is available. These require a tool database in addition to workpiece information, selection rules and an interpreter. An issue then arises about how the system interrogates the tool database: exhaustively or selectively (intelligently).

Finally, some rules may require modelling and calculation (rational knowledge) for their interpretation, in addition to or instead of heuristic (qualitative) expertise. Then the

expert system also needs a process modelling capability and, in that sense, may be described as a hybrid (rational/heuristic) system.

In the following, three examples are described that span these ranges of functionality and viewpoint: a monotonic, workpiece oriented system; a non-monotonic (weighted rule), exhaustive tool search system; and a hybrid, selective tool search system. The last, by simplifying its rules, makes it practical, simultaneously, to find acceptable (not necessarily optimal) combinations of tools and their operation variables.

A monotonic rule, workpiece oriented system

The basic, three element, architecture of such a system is shown in Figure 9.13, in this case with feedback that changes the shape information in the working memory, according to the actions of the selected tools. If–then tool selection rules are stored in the production memory. When data about a shape change to be machined are presented to the working memory, the interpreter picks up every rule that is even partly relevant to them. This is the first step of inference, named matching. Next, according to some strategy, one rule is selected from the matched rules. This is the second step, deciding which is the most relevant rule. Meta-knowledge, or knowledge about knowledge, is used for determining the strategy of rule selection. In the third, action step, the process selected by the rule is carried out. As a result, the shape data are altered. If the alteration has not achieved the complete change required, the new data are returned to the working memory and the cycle is repeated.

One expert system of this sort selects tools for drilling (SITC, 1987). It not only generates a sequence of boring operations and tools, but also records its reasoning processes. In fact, it infers boring operations inversely to their practical sequence. Figure 9.14 shows its recommended steps for how to create a 20 mm diameter hole of good finish ($\nabla\nabla$) in a blank plate, from finishing with a reamer to initial centring. The actual order of shape change is shown at the left-hand side and the inversely inferred boring operations at the right-hand side. How it reached its recommendations is shown in Figure 9.15. The left column shows the production (P) rules that it used. The condition (if) and action (then) parts of each rule are separated by an arrow. Each is quite simple and natural: P rule 1 is that if a reamed hole exists, of diameter D, it should be made by letting a reamer of diameter D pass through a hole of diameter D-0.5 (mm); P rule 2 is that if a hole has diameter D between 13 mm and 32 mm, then select a drill of diameter D for enlarging a hole of diameter 0.6D to

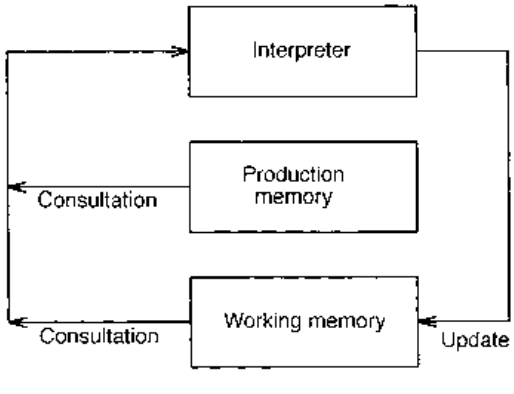


Fig. 9.13 Basic architecture of 'production system'

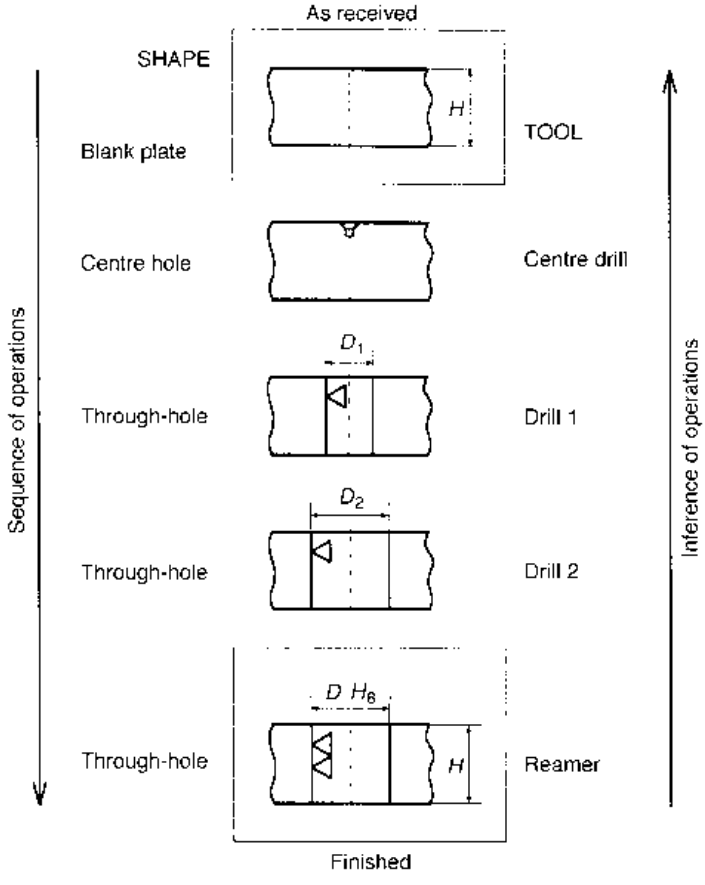


Fig. 9.14 Inference of drilling operations in an expert system (SITC, 1987)

D; P rule 3 is that if $D < 13$ mm, select a drill to make a through hole of diameter D following centre drilling; finally P rule 4 is that if there is a centre hole of 2 mm diameter, make it in a solid plate, using a centre drill. The right column of the figure shows, for each rule, the tool selected and, as a result of its action, the start and end features of the machined plate, i.e. hole shape, hole diameter and surface finish. The tools selected are, in operation order, a centre drill 2 mm∅, two drills 11.7 mm∅ and 19.5 mm∅, and a reamer 20 mm∅. The system is not concerned about whether such tools are available.

A weighted rule, exhaustive tool search system

In the previous example, only two aspects of a tool were being selected: type (centre drill, drill or reamer) and diameter. In many cases, tool geometry needs to be selected in much more detail, and also the tool material or grade. In turning, for example, a range of angles (approach, rake, inclination, etc), tool nose radius and chip breaker form should be chosen. What is chosen may be a compromise between conflicting requirements. For example, a decrease in approach angle in turning leads to a lower radial force but a weakening of the insert (because of a lower included angle). What is then a best approach angle depends at least on how those two effects influence a process. Additionally, what is a best approach

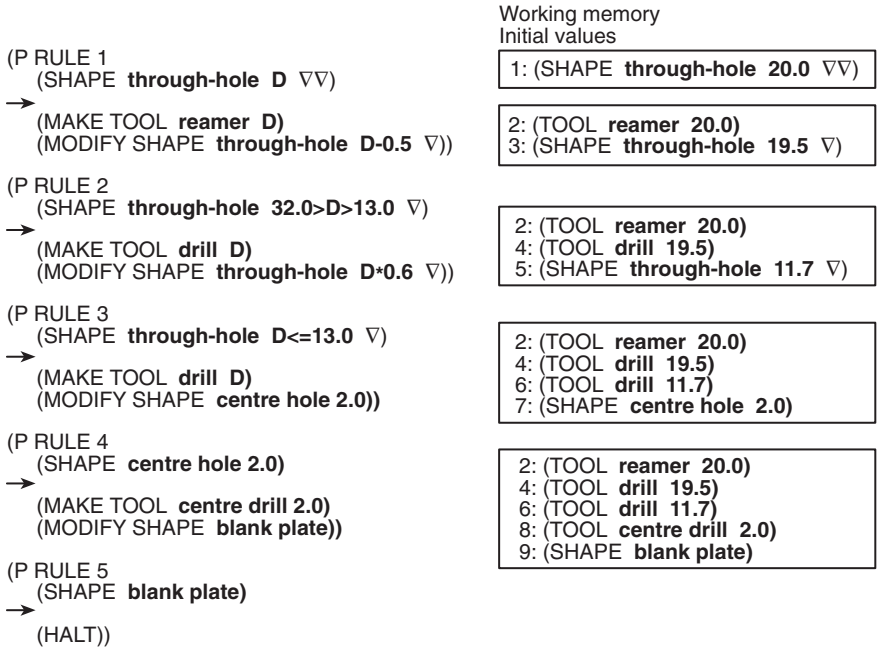


Fig. 9.15 Applied rules and reasoning processes (SITC, 1987)

angle may depend also on what is the rake angle (also for overall force and insert strength reasons) – and so on for other tool material and geometry features. In the absence of a rational model, judgement is needed. One of the simplest methods for introducing judgement is to weight rules according to their perceived importance. The recommendations of all the rules that match a given application can then be assembled as a weighted profile of desirable features. Finally, a tool that best matches the profile can be selected from a database.

This is the approach taken by COATS, an expert module for COmputer Aided Tool Selection, within a larger computer aided process planning (CAPP) system (Giusti *et al.*, 1986). This module recommends tools based on a total evaluation of some particular aspects of a given cutting situation. Figure 9.16 shows the machining of a slender workpiece, an example for which COATS has been asked to recommend tool holders and cutting inserts. In this case, the reduction of radial force is required to decrease workpiece deflection as much as possible. As a negative approach angle ψ very effectively achieves this, rules that deduce a negative approach angle in their action part have high weight. In the following example, the rule weight is 5:

APPROACH ANGLE (ψ)		RULE No. 13
IF	workpiece slenderness is ≥ 12	
AND	workpiece clamping is between centres	
AND	operation is finishing	
THEN	approach angle is ≤ 0°	
RULE WEIGHT: 5.		

(Giusti *et al.*, 1986)

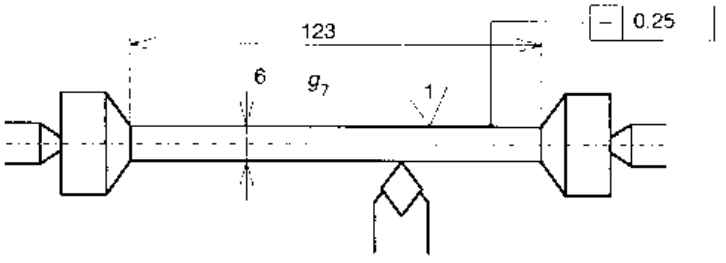


Fig. 9.16 Finishing of a slender workpiece: depth of cut 0.5 mm (Giusti *et al.*, 1986)

When several rules part match a situation, for example rules on approach angle in the situation of Figure 9.16, COATS gives a score s_i equal to the weight w_i of the matched rule i to the range of the variable (for example approach angle $(\psi)_{i-} \le \psi \le (\psi)_{i+}$) which rule i specifies:

$$s_i(\psi) = \begin{cases} 0 & \psi < (\psi)_{i-} \\ w_i & (\psi)_{i-} \le \psi \le (\psi)_{i+} \\ 0 & (\psi)_{i+} < \psi \end{cases} \tag{9.33a}$$

It then sums the scores s_i in a design range $\psi_{\min} \le \psi \le \psi_{\max}$ to give a sub-total score $S(\psi)$:

$$\mathrm{S}(\psi) = \sum_i s_i(\psi) \tag{9.33b}$$

To continue with the same example, COATS also has rules for the normal relief angle γ_n, normal rake angle α_n, cutting edge inclination angle λ_s, tool included angle ε_r ($\varepsilon_r = \pi/2 + \psi - \kappa'_r$), nose radius r_n, grade and type of insert, and feed range, among others. Sub-total scores $S(\gamma_n)$, $S(\alpha_n)$, $S(\lambda_s)$, $S(\varepsilon_r)$ and $S(r_n)$ are estimated as well as $\mathrm{S}(\psi)$. All are shown in Figure 9.17. Their distributions can be understood in terms of force and cutting edge strength effects.

As a final operation, COATS searches its library of tools and their holders to determine which have the largest total scores, estimated as the sum of the sub-scores:

$$S_{\text{Total}} = \sum_{j=1}^{N} S(N) \tag{9.33c}$$

where $j = 1$ to N are all the tool features such as ψ, γ_n, α_n and so on. Table 9.1 lists, in order of decreasing total score, COATS's recommendations for finish turning the slender workpiece in Figure 9.16. The maximum and minimum feeds in the table were determined by the chip breakability properties of the selected inserts at the given depth of cut. All the recommended tools have high normal rake. Negative approach angles are not recommended as they reduce cutting edge strength too much.

A hybrid rule, selective tool search system

A system differently structured to COATS, and applied to rough turning operations, has been described by Chen *et al.* (1989). Expertise about the usability of tools is introduced at an early stage to eliminate many unlikely-to-be-chosen tool holder and insert combinations from the eventual detailed search of the tool database. In addition, the eventual search

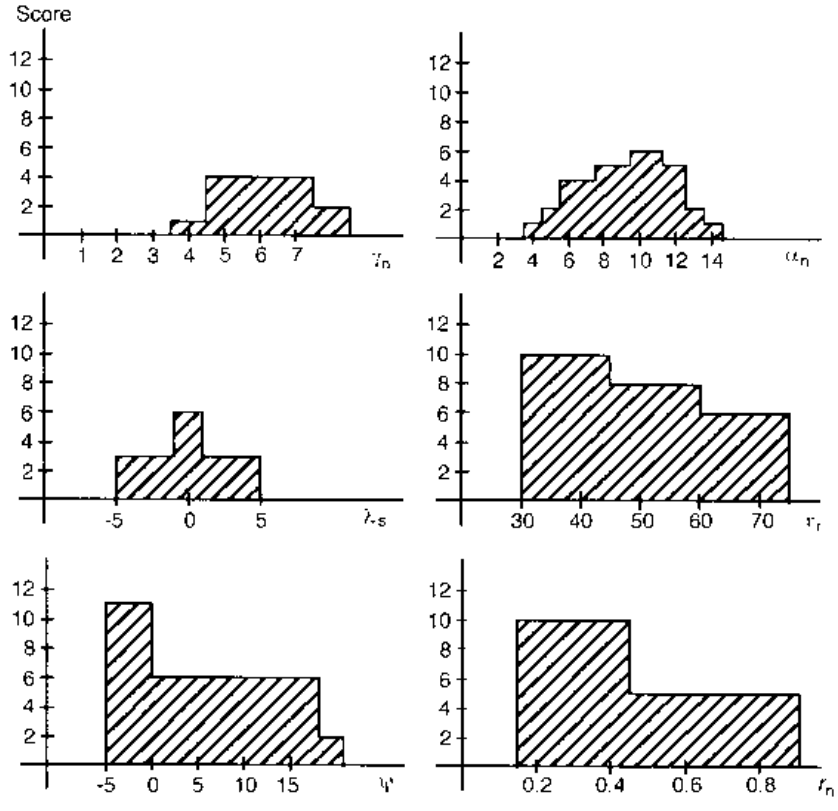


Fig. 9.17 Distributions of subtotal scores of tool's geometric parameters (Giusti *et al.*,1986)

Table 9.1 Recommended tools by COATS

Tool holder (ISO code)	Insert (ISO code)	Insert Grade	Score	Min. feed [mm]	Max. feed [mm]	γ_n [deg]	α_n [deg]	ψ [deg]	ε_r [deg]	r_n [mm]
SVVBN2525M16	VBMM160404 53	P10	49	0.10	0.33	5	12	17	35	0.4
MVVNN2020M16	VNMG160404 53	P10	45	0.20	0.48	4	11	17	35	0.4
MVVNN2020M16	VNMG160408 53	P10	38	0.40	0.70	4	11	17	35	0.8

is model-based, with constrained cost minimization as the criterion for selection (in principle, as in Section 9.3.1, but with differences in detail). It is not claimed that the system's eventual recommendation is optimal, but that it is unlikely that a substantially better recommendation exists.

The elimination and eventual search strategy is split up into six stages or levels, as listed in Table 9.2. Levels 1 to 3 and 6 use heuristic knowledge and levels 4 and 5 are model-based. Starting with level 1, only tool holders that are compatible with the specified operation are considered further: for example, if an insert's approach angle is limited by steps on a turned part, only holders that present a less than critically oriented insert to the work are considered. At level 2, if there are holders identical but for their insert clamping

Table 9.2 Search tree levels (Chen *et al.*, 1989)

Level	Parameters
1	Tool function
2	Insert clamping method
3	Holder dimension, i.e. shank height and width, and tool length
4	Holder type, i.e. approach angle, insert shape, size and thickness
5	Insert type, i.e. chip breaker type and carbide grade
6	Nose radius and insert tolerance

system, only that holder with the stiffest clamping system is considered further (unless the clamp interferes with the work, when the next stiffest is chosen). At level 3, only those holders whose shank height is suitable to the machine tool are considered further. If there are holders otherwise identical but for their length and shank width, only the shortest and broadest is considered further, because of its greatest stiffness.

The cost model is entered at level 4. At this stage, all that is known about an insert is that it must fit one of the holders still being considered. This determines, for each holder, the insert shape, size and orientation but not the insert grade or chip breaking features. Chen *et al.* suggested, reasonably, that a good choice of shape, size and orientation could be made without knowing the grade and chip breaking detail, by supposing some average-costing grade and chip breaker geometry to have been chosen already.

Insert shape, size and orientation most strongly affect cost through C_t (the tool cost per edge, equation (9.16a)), after that by being associated with different approach angles and hence tool life, and finally by influencing the cutting forces and insert strength, and hence the operational critical constraints and feasible space. The constraints that are affected at this level are C2, C6, C9, C10 and C11 (Section 9.3.1). In their selection procedure, Chen *et al.* first ranked holder and insert combinations in increasing order of C_t:

$$C_t = \frac{C_i}{0.75 n_e} + \frac{C_h}{400} \tag{9.34}$$

where C_i, C_h and n_e are the insert cost, the holder cost and the number of cutting edges; and the coefficients 0.75 and 400 are from experience. If two holder/insert combinations had the same C_t, they regarded the one with the larger approach angle as effectively cheaper because it would have a longer tool life. They argued that a more expensive combination could only reduce machining cost if it enlarged the feasible machining space. Starting with the cheapest C_t combination, they therefore checked whether any of the constraints C2 . . . C11 (above) were critical for the next cheapest. If they were not, the selection procedure was moved on to level 5, with the current holder/insert combination, on the grounds that more expensive combinations were unlikely to reduce cost.

At level 5, the carbide grade and type of insert chip breaker are selected, for the predetermined holder/insert size combination. A grade and chip breaker type not likely to lower the cost relative to a previously considered combination is quickly eliminated from the search, by establishing whether, with it, the previous cost could be bettered at feasible feeds and depths of cut. This is achieved by drawing, in the (f,d) plane, for the grade/breaker combination being considered, its line of constant cost equal to the previously established lowest cost, C_o. (This line is obtained from equation (9.29a), with coefficients valid for the

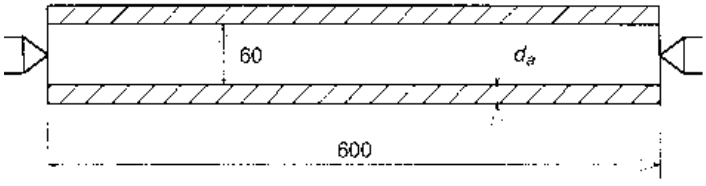


Fig. 9.18 Rough turning of a cylindrical bar (Chen *et al.*, 1989)

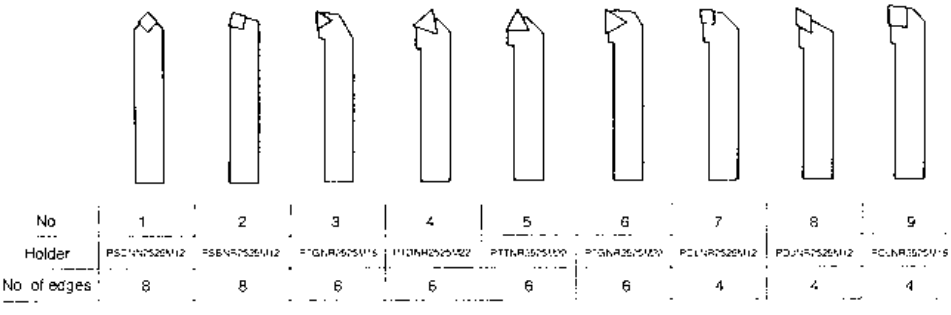


Fig. 9.19 Nine tool holders arranged in increasing order of cost (Chen *et al.*, 1989)

considered combination, by replacing C_{opt} by C_o.) If this line falls outside the feasible domain $h_V(f, d) \leq h_{V0}$ or the reduced domain $h_V(f, d_i) \leq h_{V0}$ for the combination, the combination is ignored as it is not able to reduce the cost and the next combination is considered. If it falls inside the feasible domain, a lower cost will be achievable by altering the operation variables: then the new minimum cost (and optimal cutting conditions) are evaluated and the search continued.

Finally, at level 6, if chatter provides one of the critical constraints, an insert with a smaller nose radius is selected to reduce the thrust force; otherwise a large nose radius is selected to increase strength and wear resistance; and an insert of lowest acceptable tolerance is always chosen because of low cost.

Figure 9.18 shows an example of rough turning, for which the optimum tool and machining conditions have been determined by the system. The workpiece was specified as a 0.4% plain carbon steel, the stock to be machined (d_a) as 10 mm or 3 mm from the radius and the maximum permissible operation time to be infinite. Figure 9.19 shows the nine tool holders considered by the system. All the holders have a stiff, P type (International Standard, 1995) clamping system and a shank height and width of 25 mm. They are arranged in increasing order of tool cost C_t: it can be seen that the number of edges n_e has a great influence on this.

293 inserts in the library could fit in these holders, with 11 types of chip breaker, 3 grades of carbide and 4 nose radii. By applying the search strategy just described, detailed cost calculations at level 5 needed to be carried out only for 8 inserts when d_a = 10 mm: the optimal selection was a combination of holder no.7 and a coated insert of grade P10–P20 and nose radius 0.8 mm. When d_a = 3 mm, the grade was unchanged but the tool holder and nose radius were altered to no. 3 and 1.2 mm; and the chip breaker style changed too. The search time was only 5% of that required in a parallel study in which detailed costings were carried out, unintelligently, on all 293 possibilities.

Summary

These expert systems examples illustrate the diversity of practical considerations that influence production machining; and the range of viewpoints taken and range of skills applied by an expert in recommending tools and operating conditions. The range of views span work-centred to tool-centred (from what does the work need? – to what can the tool do?): the first and last examples just considered are at the extremes of the span; while COATS offers a balanced view. The range of skills covers monotonic and non-monotonic heuristic and rational reasoning. It is a real problem to replace real experts by a single expert system, both for these reasons of diversity and the huge number of rules that are involved. A limited expert is not so useful. That is perhaps the reason why expert systems are not currently more widely used in industry and why human experts are still heavily relied upon. Nevertheless, expert system development continues to be worthwhile, both because human experts are scarce and expensive; and because it helps to increase the organization of knowledge about machining. Any tool that might help to unify expert system structures must be useful: fuzzy logic, because of its ability to handle vagueness and rational constraints in the same form (as introduced in Section 9.3.2) is a possible one.

9.3.4 Fuzzy expert systems

A fuzzy expert system for the design of turning operations, with three modules – for tool selection, cutting condition design and learning – and given the name SAM (Smart Assistant to Machinists) is shown in Figure 9.20 (Chen *et al.*, 1995). The system's inputs

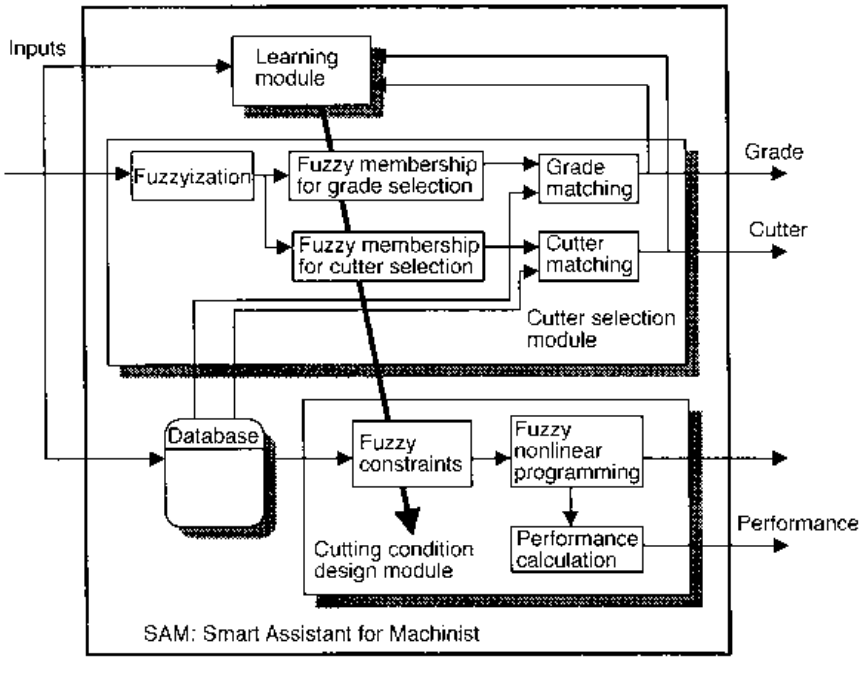


Fig. 9.20 A fuzzy expert system for the design of cutting operations (Chen *et al.*, 1995)

Table 9.3 Breadth of input data for a fuzzy expert system (Chen *et al.*, 1995)

(1)	Work material	
	(1.1)	material code: (*ISO code = P, CMC code = 02.1*, ANSI standard)
	(1.2)	material type: {steel alloy, stainless steel, . . .}
	(1.3)	hardness: (Rockwell C scale, Rockwell B scale, *Brinell scale 180*)
	(1.4)	machinability: *0.98*
(2)	Machine tool	
	(2.1)	power and maximum power: (*25 kW*, HP)]
	(2.2)	torque and maximum torque: (N m, lb. ft)
	(2.4)	maximum cutting speed: (m/min, ft/min, *1450 rpm*)
	(2.6)	power efficiency: (*95%*)
(3)	Machining plan	
	(3.1)	machining
		(3.1.1) turning: {*general turning*, contouring, tapering, grooving, . . .}
	(3.2)	machining type: {heavy roughing, *roughing*, light roughing, finishing, . . .}
	(3.3)	material removal rate:{large, medium, small} or (mm^3/min, $inch^3$/min)
	(3.4)	surface finish: {rough, good, fine, extreme fine} or (μm, μinch)
	(3.5)	cutting speed: {fast, medium, slow} or (m/min, inch/min)
	(3.6)	feed: {fast, medium, slow} or (mm, inch)
	(3.7)	depth of cut: {large, medium, small} or (*2.5 mm*, inch)
	(3.8)	length of cut: (*100 mm*, inch)
	(3.9)	diameter of the workpiece: (*25 mm*, inch)
	(3.10)	cost
		(3.10.1) machining cost with overhead: (*1–2 $/min*)
	(3.11)	time factor
		(3.11.1) tool change time: (*1.5–2.5 min*)
(4)	Cutter and cutter holder	
	(4.1)	cost: (*$ 12*)
	(4.2)	supplier: {. . .}
	(4.3)	cutter geometry: tool nose radius, thickness, . . .
	(4.4)	tool life: {long, *average*, short}
	(4.5)	cutter holder
		(4.5.1) geometry: lead angle, rake angle, side rake angle, relief angle, . . .
		(4.5.2) size:
	(4.6)	availability

are listed in Table 9.3. They can be defined by either numerical values or qualitative terms or not defined at all. (The italicized values in the table define an example for which the system has recommended a cutting tool, cutting speed and feed, as described later).

Tool selection is performed in three stages. First, all the system's inputs are made fuzzy by assigning fuzzy membership functions to them. A numerical input $x = \bar{x}$, is transformed to a fuzzy membership function

$$\mu(x, a_1, a_2, a_3, a_4) = \begin{cases} SF(x, a_1, a_2), & x < a_2 \\ 1 & a_2 \le x < a_3 \\ 1 - SF(x, a_3, a_4) & a_3 \le x \end{cases} \tag{9.35a}$$

as shown in Figure 9.21, where the parameters a_1, a_2, a_3 and a_4 are constants spanning the value $\bar{x}$ and, in this example, the function *SF* is defined by equation (A7.4b).

When a qualitative term is input, such as 'finishing' for machining type (under machining plan in Table 9.3), a fuzzy membership function is assigned after the manner:

$$\mu(MT_2) = 0.8/MT_1 + 1.0/MT_2 + 0.8/MT_3 + 0.4/MT_4 + 0.0/MT_5 \tag{9.35b}$$

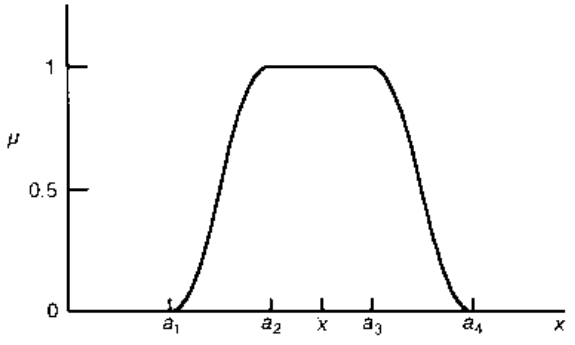


Fig. 9.21 Fuzzification of a numerical value $\bar{x}$

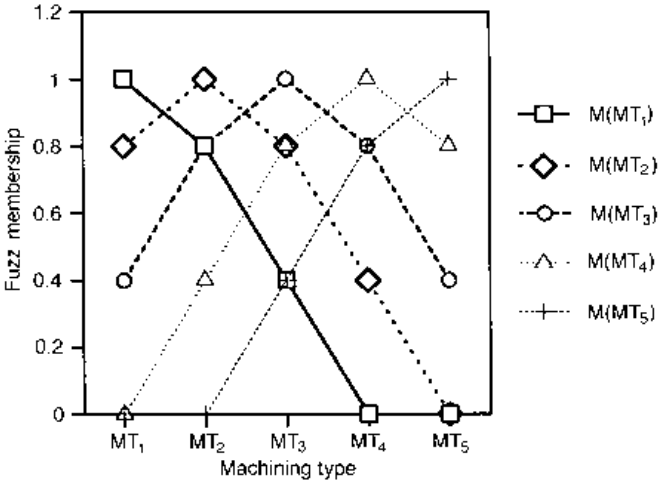


Fig. 9.22 Fuzzification of a qualitative term, e.g. machining type (Chen *et al.*, 1995)

where MT_1 is extreme finishing, MT_2 finishing, MT_3 light roughing, MT_4 roughing and MT_5 heavy roughing and the membership functions assigned to the five machining types MT_i (i = 1 to 5) are shown in Figure 9.22.

In the second stage, the applicability of inserts to the specified inputs is determined, also in fuzzy logic terms. Inserts are described by a series of fields, such as Y^i in Table 9.4 (i = 1 to 8 in this case), and by their grade G. Each field i has k elements $y^i{}_k$ and a grade has m elements g_m. The applicability of an element $y^i{}_k$ or g_m to an input variable x_j is defined by a membership function. For example, field Y^6 (insert thickness) has elements $T_1 \equiv y^6{}_1$ = 6.3 mm., $T_2 \equiv y^6{}_2$ = 9.5 mm, and so on. The applicability of insert thickness 6.3 mm, or element $y^6{}_1 = T_1$ to the depth of cut d (mm) may then be written after the manner:

$$\mu(T_1|d) = \begin{cases} SF(d,\ 0.76,\ 1.27), & d < 1.27 \\ 1 & 1.27 \le d < 1.78 \\ 1 - SF(d,\ 1.78,\ 2.29) & 1.78 \le d \end{cases} \qquad (9.36a)$$

where the coefficients' values reflect a strength constraint.

Table 9.4 Eight fields describing an insert (Chen *et al.*, 1995)

Field	Descriptions (Elements)
1: shape	R: round, S: square, T: triangle, . . .
2: relief angle	N: 0°, A: 3°, B: 5°, . . .
3: tolerances	A: ± 0.0002, B: ± 0.0005, . . .
4: type	A: with hole, B: with hole and one countersink, . . .
5: size	4: 1/2 in. I.C., 5: 5/8 in. I.C., . . .
6: thickness	number of 1/32nds on inserts less than 1/4 in. I.C., . . .
7: cutter nose radius	1: 1/64 in., 2: 1/32 in., . . ., A: square 45 ° chamfer, . . .
8: special tool only	T: negative land, . . .

In SAM's system, over 100 functions of element applicability to input variables are defined, based on metal cutting principles and various tool manuals, handbooks and technical reports. Using these functions, the applicability of an element y^i_k to a given machining operation with n inputs is given by

$$\mu(y^i_k) = \frac{1}{n} \sum_{j=1}^{n} \mu(y^i_k \mid x_j) \wedge \mu(x_j) \tag{9.36b}$$

where $\wedge$ is the minimum operator. As an example, the insert thickness is closely related to workpiece material *WM*, machining type *MT* and depth of cut. Thus, the applicability of elements $T_k \equiv y^6_k$ is given (with $n = 3$) as follows:

$$\left.\begin{aligned} \mu(T_1) &= \{\mu(T_1 \mid WM) \wedge \mu(WM) + \mu(T_1 \mid MT) \wedge \mu(MT) + \mu(T_1 \mid d) \wedge \mu(d)\}/3 \\ \mu(T_2) &= \{\mu(T_2 \mid WM) \wedge \mu(WM) + \mu(T_2 \mid MT) \wedge \mu(MT) + \mu(T_2 \mid d) \wedge \mu(d)\}/3 \\ &\vdots \end{aligned}\right\} \tag{9.36c}$$

As a second example, the applicability of nose radius elements $C_k \equiv y^7_k$ to the machining operation is defined as follows: in heavy roughing, for which the nose radius is selected according to the feed and depth of cut ($n = 2$)

$$\left.\begin{aligned} \mu(C_1) &= \{\mu(C_1 \mid f) \wedge \mu(f) + \mu(C_1 \mid d) \wedge \mu(d)\}/2 \\ \mu(C_2) &= \{\mu(C_2 \mid f) \wedge \mu(f) + \mu(C_2 \mid d) \wedge \mu(d)\}/2 \\ &\vdots \end{aligned}\right\} \tag{9.36d}$$

but in finishing, with the nose radius selected according to required surface finish ($n = 1$)

$$\left.\begin{aligned} \mu(C_1) &= \mu(C_1 \mid \text{surface\_finish}) \wedge \mu(\text{surface\_finish}) \\ \mu(\mathrm{C}_2) &= \mu(\mathrm{C}_2 \mid \text{surface\_finish}) \wedge \mu(\text{surface\_finish}) \\ &\vdots \end{aligned}\right\} \tag{9.36e}$$

After determining the applicability to a planned operation, $\mu(y^i_k)$, of each element k in all the fields i, SAM simplifies (de-fuzzifies) final tool selection by retaining only the highest valued $\mu(y^i_k)$ and assigning it to a new membership $M(y^i_k)$:

Table 9.5 Four candidate inserts for rough turning as in Table 9.3 (Chen *et al.* 1995)

No.	Shape	Size	Thickness	Nose radius	Applicability
1	C**	5	4	3	0.7856
2*	K	16	04	08	0.7534
3	C**	5	4	3	0.7027
4	C**	5	4	3	0.7027

*: this is coded in ISO standard; **: types of chip are different.

$$M(y^i_k) = \max\mu(y^i_k) \tag{9.37a}$$

If the new membership $M(y^i_k)$ has its maximum at $k = k^*$, $y^i_{k^*}$ is the best choice. The applicability M of a chosen tool m, CT_m, with specified tool parameters y^i_m is then given by

$$M(CT_m) = \frac{1}{8}\sum_{i=1}^{8} M(y^i_m) \tag{9.37b}$$

For a most applicable tool $M(CT_m) = 1$; for a least applicable tool, $M(CT_m) = 0$.

The applicability of the tool material grade is established in a similar manner; and in a final stage, a tool database is searched to select tools that maximize their grade applicability separately from their shape and size. For the rough turning example specified by the italicized elements in Table 9.3, the system recommended coated tools from its database of grades P20 and P30, both with an applicability of unity. No insert shape and size was found with unit applicability. Table 9.5 shows four types of insert recommended with applicability greater than 0.7. The parameters in this table are defined in Table 9.4, except for insert no. 2 which is coded according to ISO1832 (International Standard, 1991).

Among the operation variables, the depth of cut is specified in Table 9.3 as 2.5 mm, but the cutting speed and feed are not specified. They are determined in the cutting condition design module, by the fuzzy optimization described in Section 9.3.2. An optimum cutting speed and feed are recommended as 119 m/min and 0.13 mm/rev.

9.4 Monitoring and improvement of cutting states

In modern machining systems, the monitoring of cutting states, including tool condition monitoring, is regarded as a key technology for achieving reliable and improved machining processes, free from fatal damage and trouble (Micheletti *et al.*, 1976; Tlusty and Andrews, 1983; Tonshoff *et al.*, 1988; Dan and Mathew, 1990; Byrne *et al.*, 1995). Tool wear, tool breakage and chatter vibration are the tool conditions of major concern, as already introduced from the point of view of process modelling in Section 9.2. Sources of signals used for monitoring are the cutting forces, cutting torque, acoustic emission from the tool, workpiece and the interface between them, tool and workpiece displacements, cutting temperature, cutting sound, tool face images, etc. Methods for measuring process signals have been described in Chapter 5.

The monitoring of cutting states may be classified into direct and indirect methods. In direct monitoring, the width of flank wear, crater depth, chipped edge shape, displacements

of tool or workpiece, etc, are measured in-process or out-of-process. In-process monitoring that does not require the machining process to be stopped is preferable to out-of-process monitoring, other things being equal. However, chips being produced and cutting fluid are obstacles to measurement; the space available for measurement is limited; and direct measurement sensors may disturb the process. The continuing development of ingenious measurement methods is indispensable for reliable monitoring, for example the in-process and direct monitoring of worn or chipped end mill edges by laser-based tool image reconstruction, in the presence of cutting fluid (Ryabov *et al.*, 1996).

Indirect monitoring, which interprets signals related to a particular cutting state, can be free from the obstacles and space limitations of direct monitoring. Instead of ingenious measurement methods, process modelling (Section 9.2) plays a significant role. In this section, indirect monitoring – which is closely related to process models – and its application to the improvement of cutting states are described although the treatment is not comprehensive.

9.4.1 Monitoring procedures

There are three activities in monitoring cutting states, as shown in Figure 9.23: sensing, processing and recognition. Guidance on what signals to sense is obtained, if possible, from process models. For example, for monitoring tool wear, equations (9.13a) and (9.13b) specify non-linear systems $\mathbf{W}$ and $\dot{\mathbf{W}}$ relating tool wear or its rate, $\mathbf{w}$ or $\dot{\mathbf{w}}$, to the variable $\mathbf{x}$. The components of $\mathbf{x}$ – the operation variables, tool and workpiece geometry, etc – are what need to be monitored for the indirect assessment of wear. If a physical model is incomplete or weak, so that there is uncertainty as to what should be measured, more reliable monitoring is achieved by selecting redundant signals. The monitoring of cutting

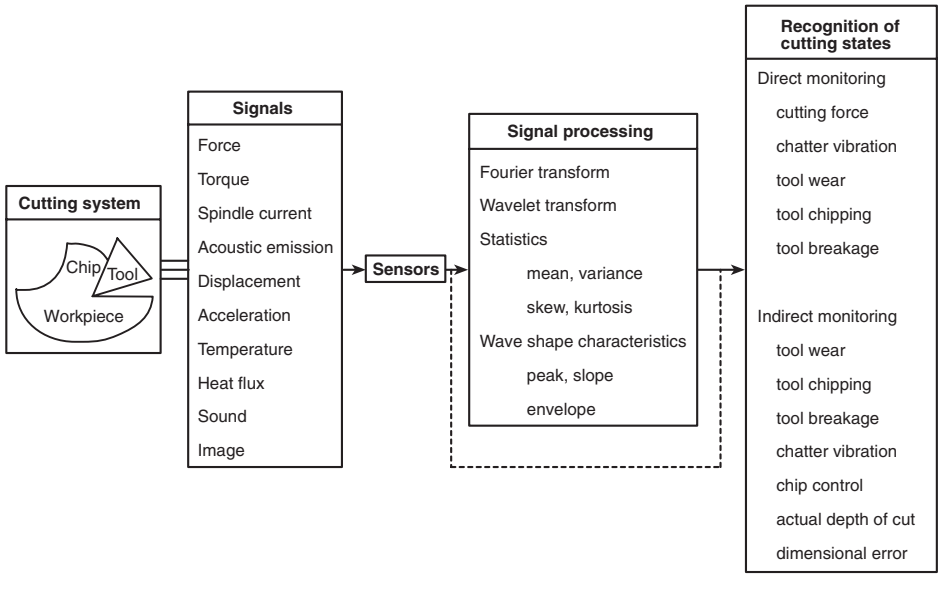


Fig. 9.23 Monitoring of cutting states

states based on multiple signals with more than one sensor is called sensor fusion or sensor integration (Dornfeld, 1990; Rangwala and Dornfeld, 1990).

Measured signals are usually processed to clarify their features: Fourier analysis (Cheng, 1972), wavelet analysis (Daubechies, 1988; Koornwinder, 1993), statistical analysis and filtering (for noise reduction) are typical signal processing methods. After signal processing, the cutting states can be characterized by two kinds of representation. One is a quantitative value, obtained from the cutting state process model: for example, the output of a wear monitoring system may be the width of flank wear. The other is a status, for example normal or abnormal, classified by pattern recognition using such tools as threshold or linear discriminant functions, artificial neural networks, or fuzzy logic.

For an operator, pattern output with one bit of information is easy to deal with. What should be done, in response to normal or abnormal, is to continue or stop, respectively. However, to control a machining process by changing operation variables, the quantitative output of a numerical value is preferable. The next section deals with methods of recognizing cutting states in ever-increasing detail, and the section after takes up the topic of model-based quantitative monitoring.

9.4.2 Recognition of cutting states

Pattern recognition by the threshold method

When the value of a particular cutting state increases or decreases monotonously with a feature of the processed signal, the normal and abnormal statuses can easily be classified by a threshold set at a particular signal level. The value of the threshold may be determined either from experimental results or by prediction based on a process model.

Tool life due to wear is often monitored by this classification method, using cutting force as the only input signal **x**, either directly or as a ratio of the force components F_d/F_c, F_f/F_c or F_d/F_f. The latter are more effective because small changes in cutting conditions (not associated with wear) have less influence on the ratios than on the individual components (Konig *et al.*, 1972). Figure 9.24 shows schematically the more direct situation of

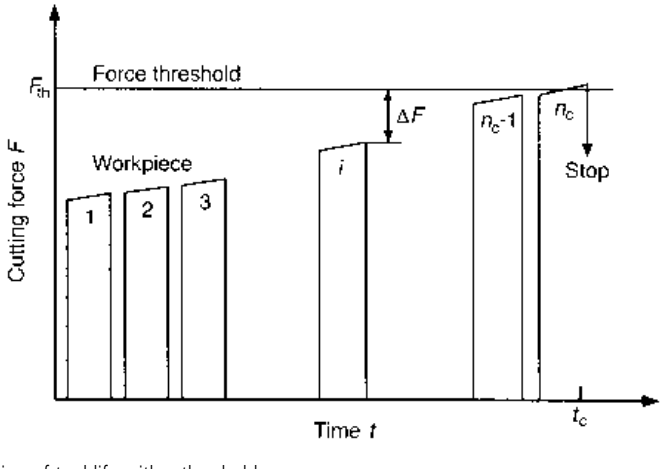


Fig. 9.24 Detection of tool life with a threshold

cutting force change due to turning many workpieces. The cutting force increases to a threshold F_{th} with cutting time and the number of workpieces machined. A simple production strategy may specify the cutting time t_c, or the number of machined parts n_c, before expecting to change a tool edge. In the first case, if production is completed before tool life is exhausted, the difference ΔF between the force threshold F_{th} and the current value of the cutting force F,

$$\Delta F = F_{th} - F \tag{9.38}$$

may be used as an index of remaining tool life. In the second case, if F_{th} is reached before n_c parts are made, the cutting conditions must be modified.

Tool breakage and chatter vibration are also detected by threshold classification. Tool breakage monitoring uses cutting force as a signal, as does wear monitoring. Chatter is detected by a threshold amplitude of vibration (displacement) or by a peak value of power in the vibration spectrum, appearing near the chatter frequency.

In many practical operations, machined parts have steps, tapers and other irregular shapes. The cutting conditions, particularly depth of cut and sometimes feed, can change during machining one part. When the resulting change in cutting force is known by experiment or model-based simulation, thresholds for breakage as well as wear can be set to be time-dependent. Figure 9.25 shows cutting force estimates in turning the *i*th workpiece of a batch. $F_i(t)$ is the expected force variation and F_{th} is the allowed threshold due to wear. $F_i(t)^u_{th}$ and $F_i(t)^l_{th}$ are more widely separated upper and lower thresholds, the measurement of force outside which indicates tool breakage.

Tool wear is usually gradual over a time scale of machining one workpiece. It is then good enough for life detection by threshold force monitoring to monitor only the peak force in the machining cycle. F_{th} may be set relative to the force F^p_1 expected with a fresh edge:

$$F_{th} = (1 + \beta_1)F^p_1 \qquad \text{or} \qquad F_{th} = F^p_1 + F_0 \tag{9.39a}$$

where β_1 and F_0 are constants. The introduction of two constants β_1 and F_0 allows a choice to be made about the way in which wear changes the cutting force, either offsetting it or scaling it.

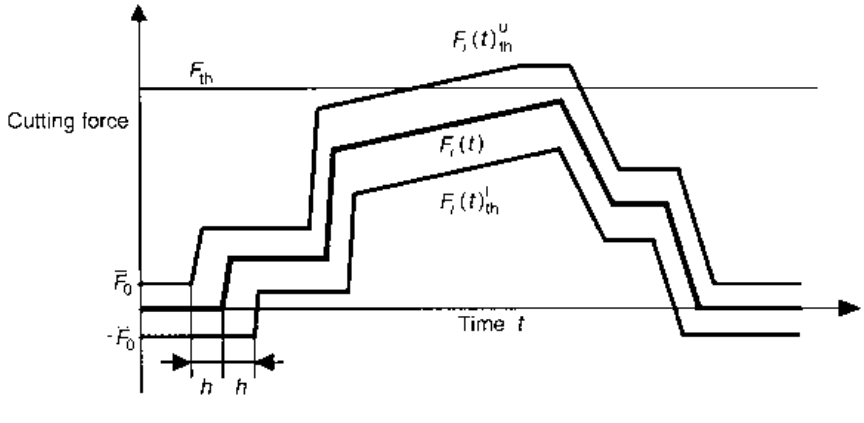


Fig. 9.25 Detection of tool breakage and wear with time dependent thresholds

On the other hand, tool breakage occurs suddenly. The loss of the tool tip, which causes the cutting force to change widely, makes it of the greatest importance to stop machining immediately. To achieve this, the upper and lower thresholds may be set respectively:

$$F_i(t)^{\mathrm{u}}_{\mathrm{th}} = \max\{(1 + \beta_2)F_{i,\mathrm{est}}(\tau)_{\max}, F_{i,\mathrm{est}}(\tau)_{\max} + \bar{F}_0\} \tag{9.39b}$$

and

$$F_i(t)^{\mathrm{l}}_{\mathrm{th}} = \min\{(1 - \beta_2)F_{i,\mathrm{est}}(\tau)_{\min}, F_{i,\mathrm{est}}(\tau)_{\min} - \bar{F}_0\} \tag{9.39c}$$

where $F_{i,\mathrm{est}}(\tau)_{\max}$ and $F_{i,\mathrm{est}}(\tau)_{\min}$ are the maximum and minimum values of estimated cutting force $F_{i,\mathrm{est}}(\tau)$ on the current workpiece during the time width $t - h \le \tau \le t + h$, and β_2 and $\bar{F}_0$ are constants. The selection of the half time width h allows updated feed-forward monitoring. By setting h to be a small fraction of the cycle time (but greater than the sampling time), the monitor, if it is fast enough, may follow force changes within a cycle and respond to abnormality within the time h.

These methods may be applied to the monitoring of tool wear and failure in end milling with varying radial depths of cut, as well as in turning, and also to drilling (where the expected force cycle is more simple). The key is to select values of the constants β_2, $\bar{F}_0$ and h appropriate to the purpose.

Pattern recognition with linear discriminant functions

A little better than recognizing a cutting state only as normal and abnormal, for purposes of control, is to classify it into more statuses, for example four. Linear discriminant functions have been used for this. A linear discriminant function has the form (Rosenblatt, 1961)

$$G_i(\mathbf{x}) = \sum_{k=1}^{N_{\mathrm{input}}} w_{ik}x_k + w_{i0} \tag{9.40}$$

where i is the status number (1 to 4 in the present case), N_{input} is the number of monitored inputs, $\mathbf{x} = [x_1, x_2, \ldots]^{\mathrm{T}}$ is the input vector, and w_{ik} ($k = 0$ to N_{input}) are weights, which are tuned by training patterns. If $G_i(\mathbf{x}) > G_j(\mathbf{x})$ for all $j \neq i$, a cutting state is assigned to the status i.

Monitoring and linear discriminant function analysis have been considered by Sata, *et al.* (1973) for assessing the status of a cutting process as one of the four of chatter generation, built-up edge formation, or either long continuous or properly broken chips. After investigating the relation between these four statuses and inputs $\mathbf{x}$, they selected six inputs for the linear classifier: (1) the total power of the cutting force spectrum; (2) the power of the spectrum in a very low frequency range; (3) the power and (4) the frequency of the highest peak in the spectrum; (5) the cutting speed; and (6) the uncut chip cross-section. They applied their recognition scheme to online chip control (Matsushima and Sata, 1974). The objective was to find a feed at which properly broken chips would be formed when machining a 0.45%C carbon steel (type S45C) with a P20 carbide tool with a chip breaker. The feed, initially set at 0.12 mm/rev, was increased in 20% steps, sampling the six signals at each step until the cutting state was classified as the formation of properly broken chips. This occurred when the feed reached 0.207 mm/rev. Figure 9.26 shows how the chip shape changed from long continuous to properly broken with increasing feed.

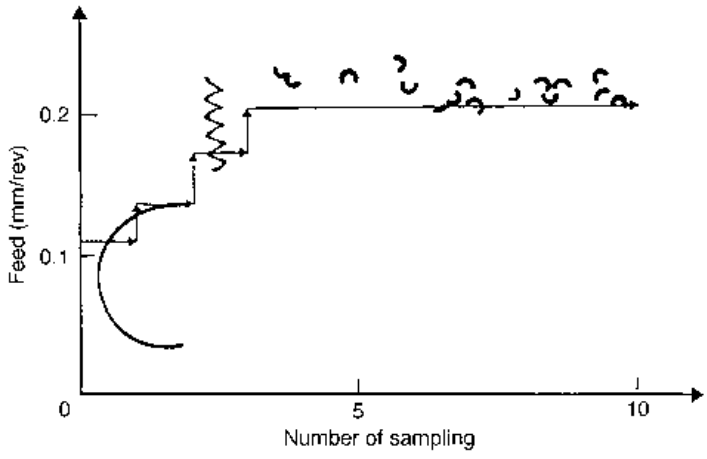


Fig. 9.26 Control of chip formation based on pattern recognition (Matsushima and Sata, 1974)

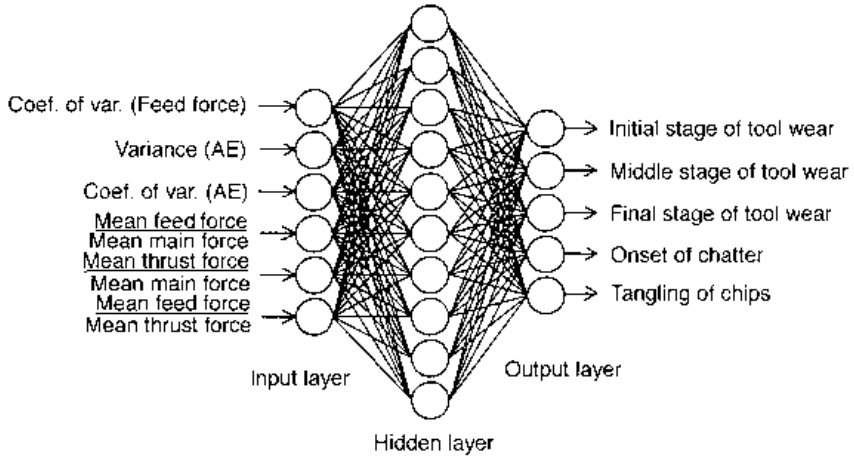


Fig. 9.27 Neural network classification of cutting states (Moriwaki and Mori, 1993)

Pattern recognition with artificial neural networks

It is now known that linear classification, with linear discriminant functions, has only limited use in pattern recognition. In particular, linear discriminant functions cannot deal with simple 'exclusive or' relations (an 'exclusive or' relation between two input statements A and B has a 'true' output if A or B, but not both, are true; and a 'false' output if A and B are both true or both false). Instead, a growth in applications of artificial neural networks, highly non-linear classifiers, has taken place.

An example of classification of cutting states by artificial neural networks is the monitoring of turning an S45C carbon steel with a coated tool (Moriwaki and Mori, 1993). Figure 9.27 shows the non-linear neural network classifier. The input variables **x** to the neural network were the monitored variance of the AE signal, the coefficient of variance (the ratio of the standard deviation to the average) of the AE signal and also of the feed

force, and the average cutting force ratios F_f/F_c, F_d/F_c, and F_f/F_d. The cutting statuses that were classified, or the outputs of the neural network, were the initial, middle and final stages of tool wear, the onset of chatter, and the tangling of chips. The initial, middle and final stages of tool wear were defined by the ranges of flank wear, $0 \leq VB \leq 0.2$, $0.2 < VB < 0.3$ and $VB \geq 0.3$ mm, respectively.

Figure 9.28 shows the signals from the three tool wear outputs, over a 40 minute cutting period. The change in tool wear status from the initial to the middle stage at around 23 minutes is clear: the heavy output activity changes from part (a) to part (b) of Figure 9.28. The change from the middle to the final stage occurs at around 32 minutes, although an early warning classification into the final stage was made at around 28 minutes.

9.4.3 Model-based quantitative monitoring

If an output of process monitoring is a quantitative value of a current cutting state, and if a process model exists that gives an expected value of that state, a comparison of the two may be used to predict future process behaviour and to improve it. Two examples are given in this section, the first about prediction, the second about improvement, to illustrate the direction of modern monitoring strategies. They span the topics of monitoring and control, the latter being developed further in Section 9.5.

The first example concerns the possibility of predicting tool wear rate in conditions of changing cutting speed and feed, from the values of monitored cutting force signals, when a model relating wear and forces, such as equation (9.2b), exists. The problem to be overcome is that the model in this case relates current forces only to current wear dimensions and operation variables, and has no element of time variation in it. However, change of wear changes the forces: monitoring the changes of force with time provides a way of including time in a modified model.

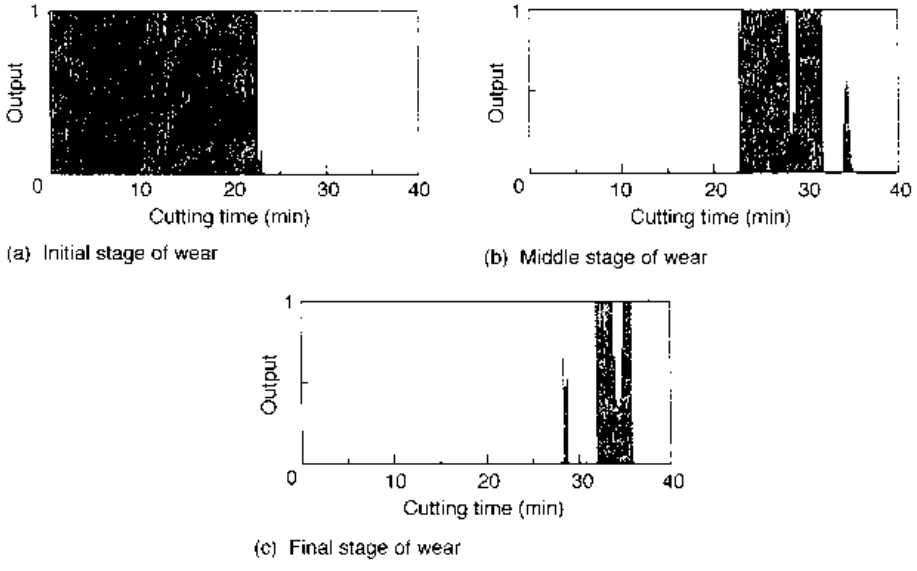


Fig. 9.28 Recognized tool wear status states (Moriwaki and Mori, 1993)

A strategy for combining the wear model and force monitoring, to create a wear rate model, using two separate neural networks, has been described, and tested in a simulation, by Ghasempoor *et al.* (1998). In a first stage, equation (9.2b) was cast in neural network form (network 1), to relate the current levels of flank, notch and nose wear (*VB*, *VN* and *VS*) and operation variables to current forces. The levels of *VB*, *VN* and *VS*, *V*, *f* and *d* were the inputs and F_d, F_f and F_c were outputs of the net; and equation (9.2b) was used to train it.

Time, measured in increments of Δt, was introduced in a second stage, by supposing that the wear vector **w** at time $k\Delta t$ depended on the wear at time $(k–1)\Delta t$ and *V*, *f* and *d*:

$$\mathbf{w}(k) = \mathbf{W}(V, f, d, \mathbf{w}(k–1)) \tag{9.41}$$

A second neural net (network 2) was created, with *VB*, *VN* and *VS* at time interval $(k–1)$, *V*, *f* and *d* as inputs; and *VB*, *VN* and *VS* at time interval *k* as outputs.

The two networks were hierarchically related: the outputs of network 2 were input to network 1 – the final outputs were the three cutting force components. During a cutting operation, only the second net was trained online, continuously, using the cutting force error signal from network 1. It was proposed that, after online training under varying conditions of the operation variables, network 2 (separated from network 1) would have the ability to predict the development of wear, step by step at time intervals Δt, from its initial level at $t = 0$.

The capabilities of this approach and its robustness were tested by simulation of a turning process in which it was supposed that the cutting forces were monitored and the cutting speed and feed were changed continuously with time. The wear expected from the forces (equation (9.13c)) and estimated from the wear rate formulation (network 2) were compared. Figure 9.29 shows the close agreement between the expected and estimated values, after about 10 min of cutting.

In this case, wear values at zero time were input to the second network that were intentionally much higher than expected from the forces. The 10 min is the time that the coefficients of the second net took to adapt themselves, by learning, to the actual state. The input signals to the second net were also degraded by white noise, as might be expected in a real monitoring situation. The level of noise is seen in the expected signals. The system can obviously cope with this.

This first example demonstrates only that the combined monitoring and modelling method can assess wear, under pre-set variations of speed and feed. The possibility of adaptively altering the rates of change of speed and feed to meet some goal (for example optimization) is an obvious extension, requiring only production planning to be added to monitoring and modelling (prediction), as in Figure 9.30. The second example is concerned with this, although there are also differences between it and the first example with respect to its monitoring (calibrated by measurement, not by modelling) and modelling (physical rather than empirical) parts. It is concerned with the situation in which batches of workpieces are to be machined with a maximimum allowable tool wear per batch, but there may be differences in machinability between batches that require a different speed or feed for each in order that the wear constraint be met. The example is made specific by considering turning a batch of 30 bars under the initial conditions listed in Table 9.6, with the constraints that cutting speed and feed may be altered by up to ±50% from their nominal values, that all of a batch must be turned with one edge (corner) of an insert, without *VB* exceeding 0.2 mm, and that conditions should be set to minimize the cutting time (Obikawa *et al.*, 1996).

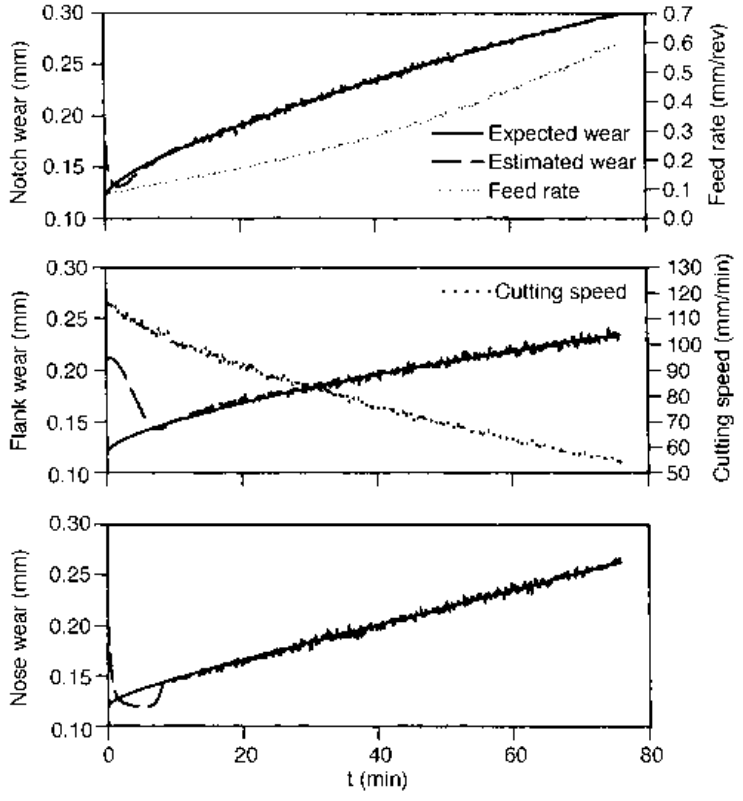


Fig. 9.29 Wear development estimated online for continuous change of both cutting speed and feed (Ghasempoor *et al.*, 1998)

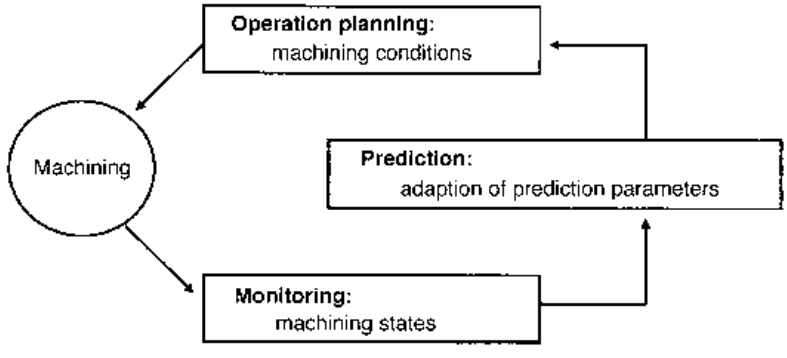


Fig. 9.30 Integration of monitoring, prediction and operation planning of cutting processes (Obikawa *et al.*, 1996)

As in the first example, wear was monitored indirectly by force measurements, using neural nets, but it was found that the axial and radial cutting position of the tool on the workpiece influenced the nets' predictions: dynamic force signals were influenced by the workpieces' compliance. One net was used to monitor wear while a second wear rate; both were trained by direct measurement (rather than, as in the first example, by model predictions).

Table 9.6 Operation planning conditions and initial turning conditions

Tool life	VB = 0.2 mm
Number of workpieces	30
Longitudinal cutting length of workpiece	150 mm
Diameter of workpiece	100 mm
Work material	0.45%C plain carbon steel
Tool material	carbide P20
Tool geometry	(–5, –6, 5, 6, 15, 15, 0.8)
Cutting speed	150 m/min
Feed rate	0.15 mm/rev
Depth of cut	1.0 mm
Lubrication	dry

Because of this, a large amount of redundancy (robustness) was built into the nets, with 34 inputs to each net, as shown in Figure 9.31. Thirty of these were auto-regression (AR) coefficients of the feed force power spectrum (as much information as could be extracted from it), two were the total power of the spectrum of feed force and cutting force, and two were the axial and radial positions of the cutting tool, as already mentioned.

Under the assumptions of the AR model, the power spectrum was defined as

$$P_s(j\omega) = \frac{pP_n(j\omega)}{\pi} \left| \frac{1}{1 + \sum_{k=1}^{p} a_k z^{-k}} \right|^2 \tag{9.42}$$

where p is its order (and also the number of peaks in the spectrum), $P_n(j\omega)$ is the white noise power spectrum, a_k is the kth AR coefficient and $z = e^{j\omega}$. In this case p was chosen to be 30 by the Akaike Information Criterion (AIC – Akaike, 1974).

The outputs from the two nets, the flank wear $(VB)_t$ and its rate$(\dot{V}B)_t$, were combined as follows, with Δt being the time interval between estimates, to give an even more robust estimate:

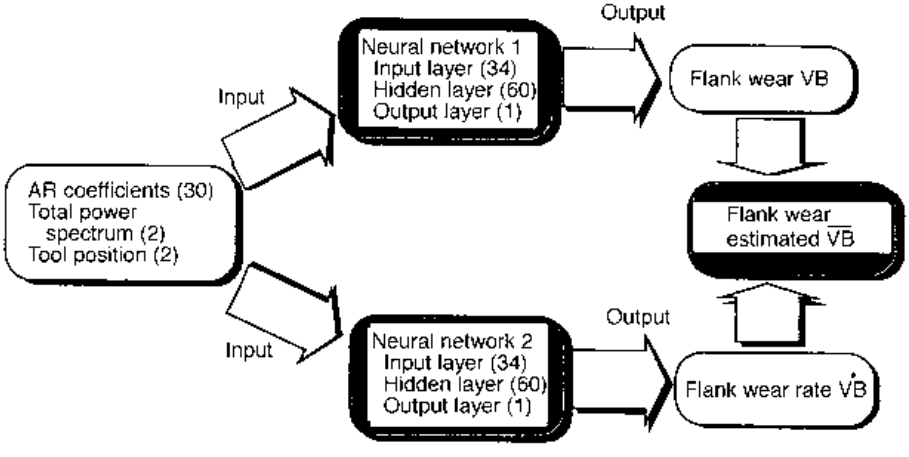


Fig. 9.31 Neural networks for predicting flank wear (Obikawa *et al.*, 1996)

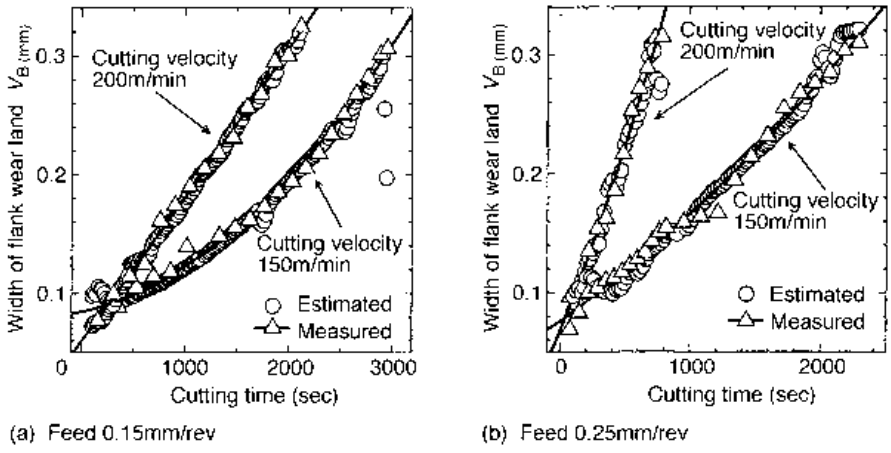


Fig. 9.32 Flank wear development predicted by neural network (Obikawa *et al.*, 1996)

$$(\overline{VB})_t = \frac{1}{2}\left[(VB)_t + \{(\dot{VB})_t \Delta t + (VB)_{t-\Delta t}\}\right] \tag{9.43}$$

Figure 9.32 shows a comparison between estimated and measured flank wear in four different speed and feed cutting conditions. Training the nets was carried out on one batch of material and the estimates and measurements on another.

The predominant wear in the conditions of this example could be modelled by equation (4.1c). The prediction element of Figure 9.30 was the physical model already described in Section 9.2.4, with an example of its outputs given in Figure 9.7. Precise prediction of flank wear rate requires accurate values of the constants C_1 and C_2 in equation (4.1c): they can vary from batch to batch of the tool and workpiece. Optimization needs them to be continually tuned and identified. In this example, wear rate was calculated by the FDM simulator $\bar{\Theta}_{FDM}$ (Section 9.2.4) beforehand, for many combinations of C_1 and C_2, cutting speed, feed rate and width of flank wear, to create a look-up table. When the wear rate in an actual turning operation was estimated by the monitoring system, the values of C_1 and C_2, which gave agreement with the estimate, were identified quickly by referring to the table.

After tuning the constants, the cutting speed and feed could be optimized. Figure 9.33 shows, for one batch, the width of flank wear V_{Bend} at the end of turning all the workpieces, predicted for different speeds and feeds. Under the constraint of maximum wear land length of 0.2 mm and shortest cutting time, a cutting speed of 130 m/min and a feed of 0.225 mm would be chosen in this case. These conditions could be set, adaptively, after tuning the constants while turning the first bar of the batch.

9.4.4 The development of monitoring methods

The direction of development of monitoring methods during the 1990s can be understood from the list of reported studies in Table 9.7. Force continues to be the dominant signal to be monitored. In the area of signal processing, there is a slow growth in the application of

wavelet transforms (wt), which translate a signal in the time domain into a representation localized not only in frequency but in time as well. Neural networks are becoming a standard method for the recognition of cutting states. For pattern recognition, unsupervised ART 2 type neural networks (Carpenter and Grossberg, 1987) have been effectively used (Tansel *et al.*, 1995; Niu *et al.*, 1998).The integration of wavelet transform coefficients as

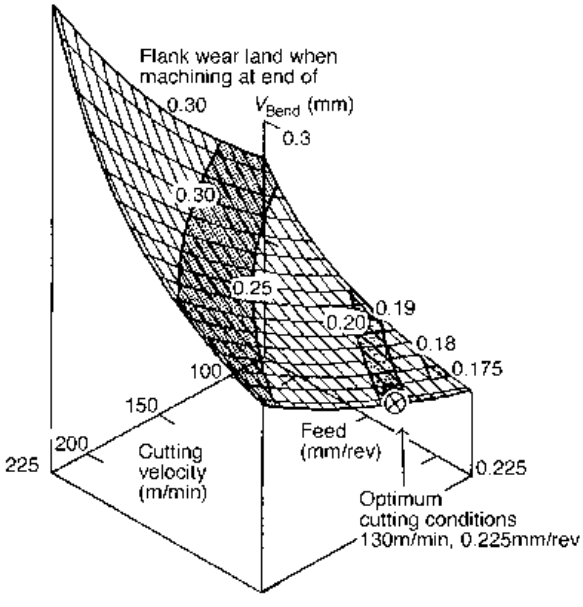


Fig. 9.33 Optimized cutting conditions using a tuned wear equation (Obikawa *et al.*, 1996)

Table 9.7(a) Recent approaches to cutting state monitoring – abbreviations given in Table 9.7(b)

Processes and monitored states	Sensor signals	Signal processing features and/or models	Recognition methods	References
Turn: w	A	am	Pa, TH	Blum and Inasaki (1990)
Tapp: a, s, w	F, Q	cr, cv, me, pe, rm, va	Pa, PV	Chen *et al.* (1990)
Turn: w	A, C, F	ar, rm, pd (FFT)	Pa, NN	Dornfeld (1990)
Turn: w	A	me, rm, sk, vc	Pa, CL	Moriwaki and Tobita (1990)
Turn: w	A, F	cs, fr, sf	Pa, NN	Rangwala and Dornfeld (1990)
Turn: w	F	fw	Qv, AN	Koren *et al.* (1991)
Turn: w	A, F, T	aw, fw	Qv, NN, ST	Chryssolouris *et al.* (1992)
Turn: t, v, w	A, F	rf, va, vc	Pa, NN	Moriwaki and Mori (1993)
Drill: w	F	wt	Pa, NN	Tansel *et al.* (1993)
Face: b	F	af, vf	Pa, NN	Tarng *et al.* (1994)
Face: b	F	wt	Pa, TH	Kasashima *et al.* (1994)
Turn: w	F	df, tp (AR model)	Pa, FL	Ko and Cho (1994)
Drill: w	F, Q	me, pe, rm, ft, tt	Pa, Qv, NN	Liu and Anantharaman (1994)
Turn: w	A, F	cs, fr, ku, me, sd, sf, sk	Pa, NN	Leem *et al.* (1995)
EndM: b	F	wt	Pa, NN	Tansel *et al.* (1995)
Turn: w	F	ar, cp, tp	Qv, NN	Obikawa *et al.* (1996)
Turn: w	F	wt	Pa, TH	Gong *et al.* (1997)
Turn: b, c, r, w	A	ku, sk , fb, me, sd, wt	Pa, NN	Niu *et al.* (1998)
Turn: w	F	fw	Qv, NN	Ghasempoor *et al.* (1998)

Table 9.7(b) Abbreviations used in Table 9.7(a)

Processes and monitored states		
Drill: drilling	a: misalignment	t: chip tangling
EndM: end milling	b: tool breakage	v: chatter vibration
Face: face milling	c: tool chipping	w: tool wear
Tapp: tapping	r: chip breakage	
Turn: turning	s: hole size error	
Sensor signals		
A: accoustic emission	F: cutting forces	T: temperature
C: spindle motor current	Q: cutting torque	
Signal processing features and/or models		
af: cutting force moving average per revolution	df: dispersion in frequency ranges	sd: standard deviation
am: AE mode (amplitude with maximum probability density)	fb: frequency band power	sf: power spectrum feature
ar: AR coefficients	fr: feed rate	sk: skew
aw: acoustic emission-wear model	ft: force-time area	tp: total power
cp: cutting positions	fw: force-wear model	tt: torque-time area
cr: correlation	ku: kurtosis	va: variance
cs: cutting speed	me: mean	vc: coefficient of variation
cv: covariance	pe: peak	vf: variable cutting force averaged per tooth period
	pd: power spectral density	wt: coefficients of wavelet transform
	rf: ratio of force components	FFT: fast Fourier transform
	rm: root mean square	
Recognition methods		
Pa: pattern recognition	CL: Cluster analysis based on mean square distance	NN: neural network
Qv: quantitative value	FL: fuzzy logic	PV: probability voting
AN: analytical		ST: statistical
		TH: threshold

inputs with neural networks as classifiers can be expected to lead to more detailed and reliable recognition of cutting states in the future.

9.5 Model-based systems for simulation and control of machining processes

In this final section, the application of machining theory to complicated machining tasks is described. As larger and larger applications, taking more time, or more and more complex components, requiring more operations, are considered, the need for more rational planning and operation becomes greater. A total or global optimization is needed, in contrast to optimizing the production of a single feature. Optimization in such conditions needs machining times, machining accuracy, tool life, etc, to be known over a wide range of cutting conditions. If the machining process is monitored, for example based on cutting force, the expected change in force with cutter path (in the manner of Figure 9.25) must also be known over a long machining time. Once the time scale reaches hours, force measurement and its total storage in a memory become unrealistic. For these reasons, cutting process simulation based on rational models, namely model-based simulation, is expected to have a significant role in the design and control of machining processes and to give solutions to rather complicated processes.

9.5.1 Advantages of model-based systems

Consider some of the optimization issues associated with the roughing of the aerospace component shown in Figure 9.34 (Tarng *et al.*, 1995). Figures 9.34(b) and (c) show end mill tool paths that convert the block (a) to the rough shape (d). First, machining is conducted smoothly along *Y–Z* plane tool paths, then along *X–Z* planes. In the *X–Z* plane paths, the end mill must remove steps left by machining along the *Y–Z* plane paths, as shown schematically in Figure 9.35: step changes in axial depth of cut are unavoidable. The major constraints to the roughing operation may be: (1) the peak cutting force, F_{peak}, must be less than a critical value, $F_{critical}$, which causes the tool to fail and (2) the finishing allowance left on the machined surface must be less than a given amount (depending

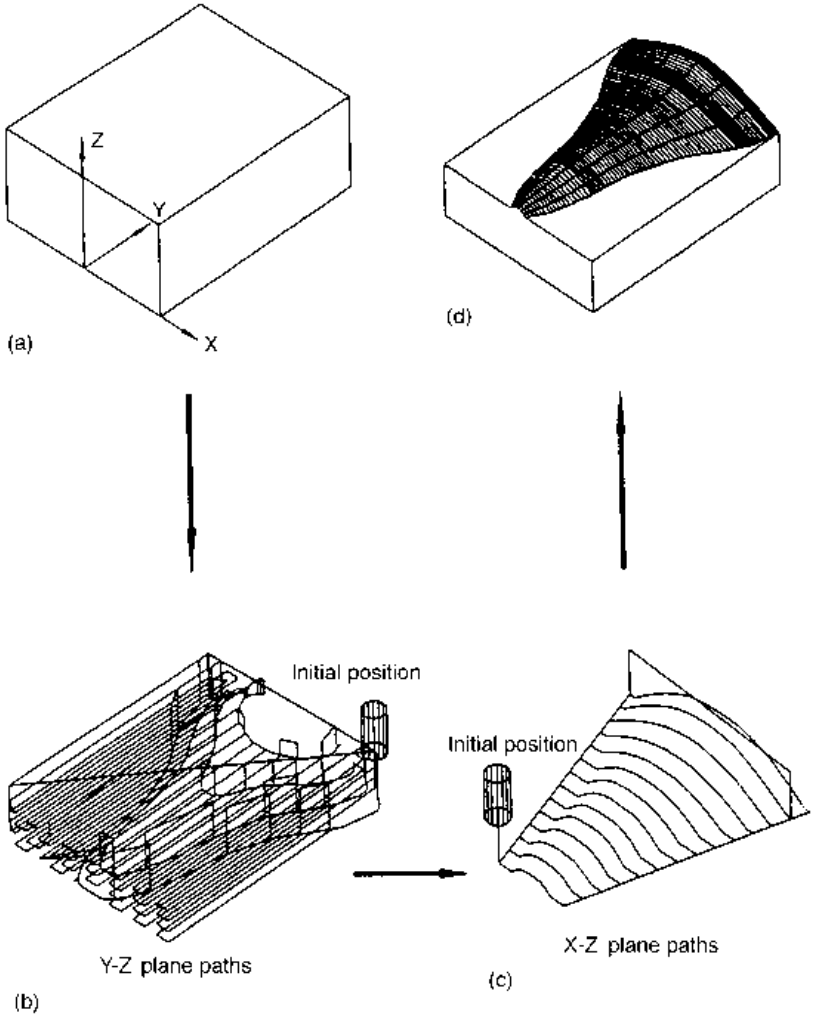


Fig. 9.34 Tool path for machining an aerospace component (Tarng *et al.*, 1995): (a) original workpiece, (b) tool paths in the *Y–Z* planes; (c) tool paths in the *X–Z* planes; and (d) machined workpiece

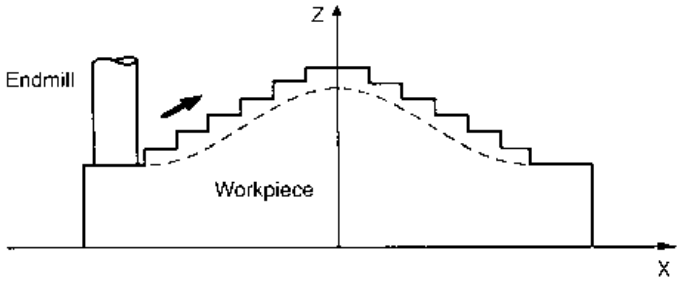


Fig. 9.35 A schematic of a tool path and pre-machined steps in an *X–Z* plane

on the required finished accuracy): this constraint eventually determines the *Y* cross feed for the *X–Z* plane machining strokes. The objective in selecting the cutting conditions may be to find the minimum machining time under these constraints.

To simplify the problem of cutting condition optimization, the axial depth of cut in each *Y–Z* plane path and the cross feeds in the *X* and *Y* directions may be set constant. If the cutting speed is also held constant, the feed speed (U_{feed}, Chapter 2) becomes the single variable that controls the cutting states. The feed per tooth may change in a specified range with an upper limit $f_{\max}$; that too is one of the constraints.

There are two methods to find optimal feed changes in the above milling operation. One is online adaptive control; the other is model-based simulation and control. Adaptive control (Centner and Idelsohn, 1964; Bedini and Pinotti, 1982) is a method that adjusts cutting conditions until they approach optimal, based on monitored cutting states. However, it has some response time, reliability and stability difficulties. Although tool wear rate, chatter vibration, chip form, surface finish and dimensional accuracy are all candidate states for control, they are seldom used in adaptive control because of insufficient reliability. Cutting forces and torque are usually the only states that are selected.

As in the cornering cut described in Section 9.2.2, the cutting force is effectively controlled by feed. Therefore, to minimize machining time, it might be decided, in an adaptive control strategy, to maximize the peak cutting force by adjusting the feed from an initial value f, with a measured force F_{peak}, to a new value $f_{\text{a,force}}$:

$$f_{\text{a, force}} = F_{\text{critical}} \frac{f}{F_{\text{peak}}} \tag{9.44a}$$

where F_{critical} is the largest safe value.

If a model-based system is used to control f, force change with cutting time is simulated based on one of the force models: generally equation (9.6) is recommended. Then feed is adjusted to raise the simulated peak force to the critical level. It may be necessary in practice to allow for feed servo control delays that are inevitable in numerical control.

If no trouble arises in a machining process, adaptive and model-based control should yield the same results. However, if a sudden increase in the axial depth of cut or the effective radial depth of cut occurs, as at steps in Figure 9.35 or at corners in Figure 9.6, adaptive control may not function well, because of the response time limitation mentioned above. Under adaptive control, with time minimization as its goal, an end mill is probably moving at its highest feed rate before it meets a step or a corner. The sudden increase in

the axial depth of cut or effective radial depth of cut is likely to yield a very large cutting force, causing tool damage, before the adaptive controller can command the reduction of feed rate and the feed is actually reduced. Tool damage due to sudden overloading is more likely to be avoidable if the force change is predicted by model-based simulation. The cutting conditions may be optimally designed beforehand to decrease the feed to a value low enough to anticipate the changes at steps and corners.

In short, the principal difference between the two control methods is that model-based simulation is feed-forward in its characteristics, whilst adaptive control is a feedback method. Its feed-forward nature is one great advantage of model based simulation.

A second advantage of model-based simulation is that prediction of change in cutting states can support monitoring and diagnosis of cutting state problems in complicated machining processes. In the absence of an expected response, a monitoring system cannot distinguish a normal from an abnormal change. A third advantage is that the machining time under optimized conditions is always estimated beforehand. This helps the scheduling of machining operations.

From all this, a model-based system is a tool for global optimization. In this sense, adaptive control is a tool for local optimization.

9.5.2 Optimization and diagnosis by model-based simulation

Model-based simulation has been applied to the end milling example of Figure 9.34 (Tarng *et al.*, 1995). Figure 9.36(a) shows the simulated resultant cutting force in fixed feed conditions. The detail force model of equation (9.6) and the specific cutting force model of equation (9.7b) (Kline and DeVor, 1983) were used. The spindle speed selected was 1200 rev/min, the maximum axial depth of cut (the depth of cut in *Y–Z* plane paths) was 6 mm, the maximum radial depth of cut was the full immersion of 12 mm, and the feed rate was fixed at 105 mm/min.

Figure 9.36(b) shows a simulation under variable feed. Compared with Figure 9.36(a), peak forces are more uniform; and the machining time has been reduced by about a third. Furthermore, the simulated result was confirmed experimentally, when the operation was actually carried out with the planned strategy (Figure 9.36(c)).

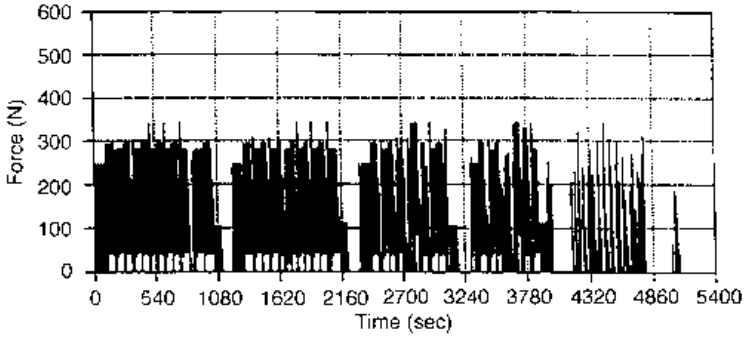
The strategy was to adjust the feed to

$$f_{\text{a, force}} = \left(-2 \frac{F_{\text{peak}}}{F_{\text{critical}}} + 3 \right) f \tag{9.44b}$$

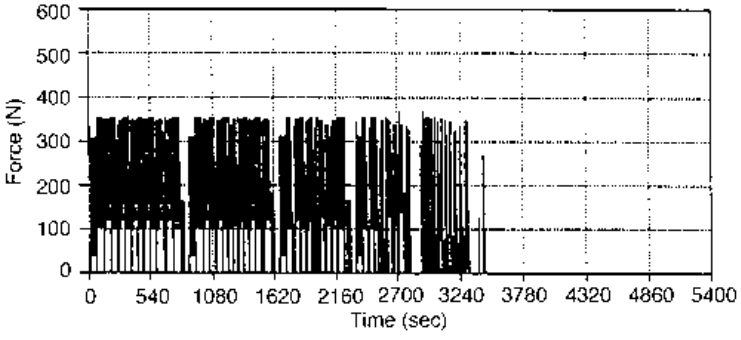
where f is the original constant feed. By this means, the feed rate $f_{\text{a,force}} = f$ when $F_{\text{peak}} = F_{\text{critical}}$ and rises linearly to $3f$ as F_{peak} reduces to zero.

Similar pre-machining feed rate adjustment in end milling and face milling has been applied to the control of the average torque, average cutting force, and maximum dimensional surface error caused by tool deflection, as well as the maximum resultant cutting force (Spence and Altintas, 1994). It is the Spence and Altintas (1994) model-based system that is illustrated in Figure 9.1(a).

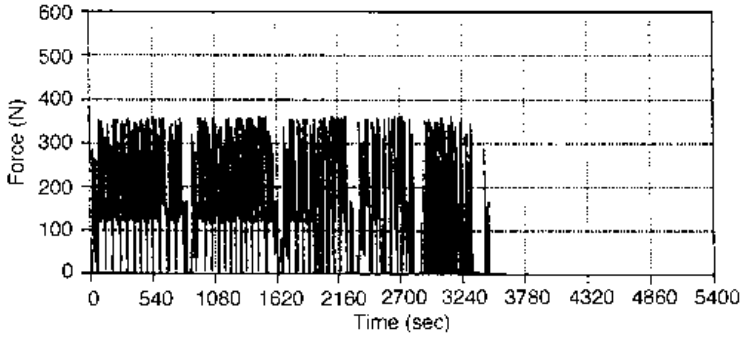
Figure 9.1(b) shows a machining operation system that can generate a machining scenario for a given operation (Takata, 1993). The machining scenario describes changes in cutting situations predicted by geometric and physical simulation. Cutting situations include both machining operations and cutting states. For end milling, five types of



(a) Simulated force at fixed feed rate



(b) Simulated force at adjusted feed rate

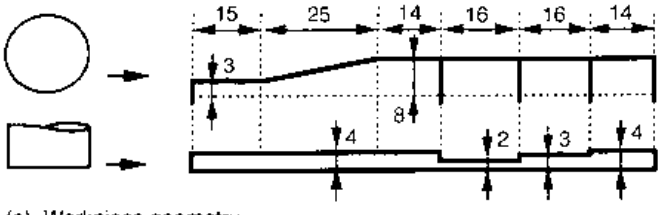


(c) Measured force with adjusted feed rate

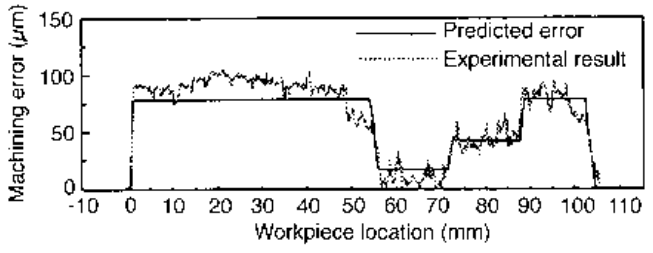
Fig. 9.36 Variation of resultant cutting force (Tarng *et al.*, 1995)

operations are recognized: slotting, down-milling, up-milling, centring and splitting. The machining scenario is used to control cutting force and machining error by pre-machining feed adjustment, and to diagnose machining states.

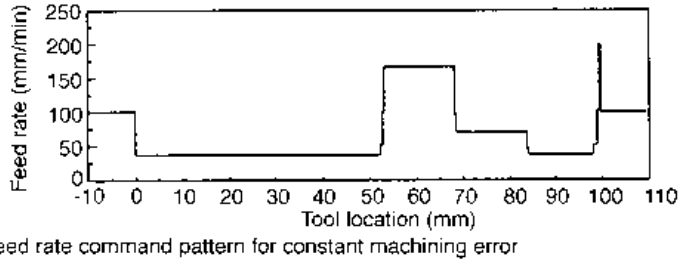
Figure 9.37 (from Takata, 1993) shows an example of the effectiveness of pre-machining feed adjustment in controlling dimensional errors in end milling. Figure 9.37(a) shows plan



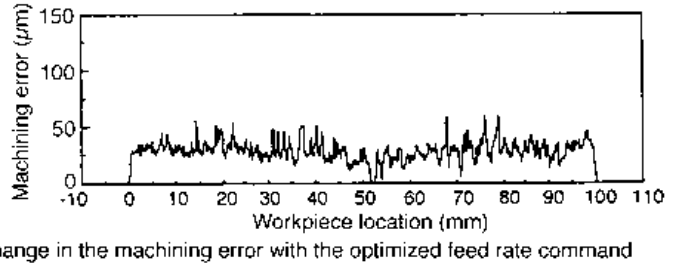
(a) Workpiece geometry



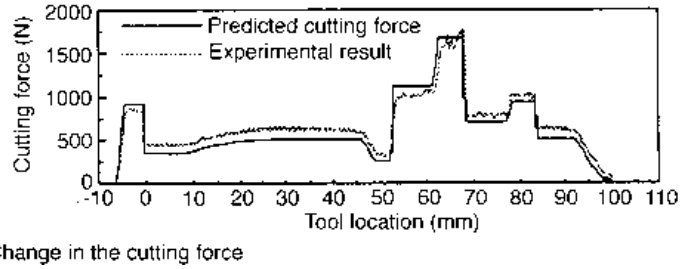
(b) Change in the machining error with the constant feed rate command



(c) Feed rate command pattern for constant machining error



(d) Change in the machining error with the optimized feed rate command



(e) Change in the cutting force

Fig. 9.37 Effectiveness of pre-machining feed adjustment in controlling dimensional error (Takata, 1993)

and side views of the stock to be removed by a two-flute square end mill 16 mm in diameter, rotating with a spindle speed of 350 rev/min. When the feed rate was set at 100 mm/min in a trial cut, the dimensional error varied with large amplitude, as shown in Figure 9.37(b). Then, using an equation similar to equation (9.44a), the feed was adjusted as follows:

$$f_{\text{a,error}} = \frac{E_{\text{critical}}}{E_{\text{siml}}} f \tag{9.44c}$$

where $f_{\text{a,error}}$ is the feed adjusted for the limit of dimensional error E_{critical}, and E_{siml} is the error simulated under the trial conditions. Figures 9.37(c), (d) and (e) show the adjusted feed rate, measured error, and simulated and measured cutting forces. The dimensional error is almost constant over the workpiece as expected. The simulated and measured cutting forces show good agreement

Figure 9.38 shows the principle of a second use of the machining scenario, to diagnose faults in an operation. A fault may be excessive tool wear, tool breakage, chatter vibration, tangling of chips, incorrect workpiece positioning, incorrect tool geometry, workpiece geometry incorrectly pre-machined, incorrect tool preset, among others. In any case, it will cause the measured force variation with time to differ from the expected one. If a measured wave form differs from the expected one by more than a set amount, a fault hypothesis library is activated. It holds information on how different types of fault may be expected to change an expected pattern. A fault simulation routine modifies the expected pattern accordingly. This is compared with the measured pattern and a fault diagnosis is produced from the best match between measured and simulated alternatives.

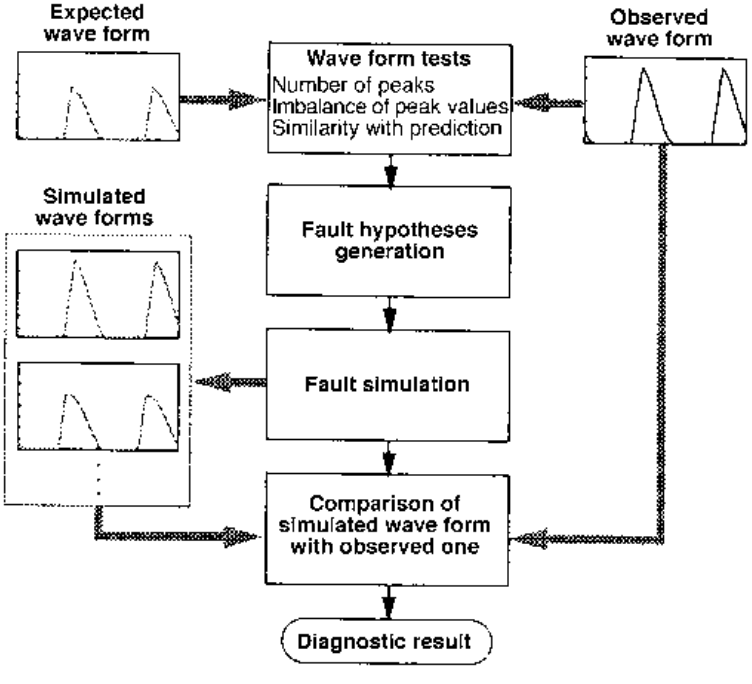


Fig. 9.38 Diagnosis procedure for faulty machining states (Takata, 1993)

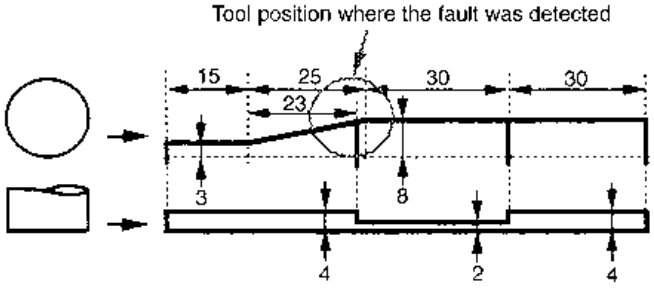


Fig. 9.39 Diagnosis of machining a different workpiece (Takata, 1993)

To demonstrate the system's abilities, the workpiece shown in Figure 9.39 was prepared and machined instead of the intended workpiece shown in Figure 9.37(a). The diagnosis system detected the difference between the two workpieces when the centre of the end mill had travelled 37 mm from the left end. It diagnosed the force error as arising from too small an axial depth of cut and that this was due to an error in the workpiece shape. Details of the comparator algorithm are given in Takata (1993).

9.5.3 Conclusions

A huge number of experiments have been carried out and many theoretical approaches have been developed to support machining technologies. Nevertheless, it is often felt that the available experimental and theoretical data are insufficient for determining the machining conditions for a particular workpiece and operation.

These days, partly because of a decrease in the number of experts and partly because of the demands of unmanned and highly flexible machining systems, machine tool systems are expected to have at least a little intelligence to assist decision making. For this purpose, expert systems for determining initial cutting conditions and cutting state monitoring technologies are increasingly being implemented. Up to now, monitoring technologies in particular have been intensively studied for maintaining trouble-free machining. Nowadays, they are regarded as indispensable in the development of intelligent machining systems.

However, machining systems have not yet been equipped with effective functions for diagnosing and settling machining troubles and revising cutting conditions by themselves. To develop such a system, prediction, control, design and monitoring of cutting processes should be integrated by sharing the same information on cutting states. A model-based system, with advanced process models, provides a way of enabling that integration. This integration will help the further development of autonomous and distributed machining systems with increased intelligence and flexibility. The theory of machining can contribute greatly to this.

References

Akaike, H. (1974) A new look at the statistical model identification. *IEEE Trans. on Auto. Cont.* **19**, 716–722.

ASM (1990) Classification and designation of carbon and low-alloy steels. *In ASM Handbook* 10th edn. Vol. 1, 140–194. Ohio: American Society of Metals.

Barr, A. and Feigenbaum, E. A. (eds) (1981) *The Handbook of Artificial Intelligence* Vol. I. Los Altos: William Kaufmann.

Barr, A. and Feigenbaum, E. A. (eds) (1982) *The Handbook of Artificial Intelligence* Vol. II. Los Altos: William Kaufmann.

Bedini, R. and Pinotti, P. C. (1982) Experiments on adaptive constrained control of a CNC lathe. *Trans. ASME. J. Eng. Ind.* **104**, 139–149.

Blum, T. and Inasaki, I. (1990) A study on acoustic emission from the orthogonal cutting process. *Trans. ASME J. Eng. Ind.* **112**, 203–211.

Boothroyd, G. and Knight, W. A. (1989) *Fundamentals of Machining and Machine Tools*. New York: Marcel Dekker.

British Standard (1991) BS 970: Part 1: *Wrought Steels for Mechanical and Allied Engineering Purposes*. London: British Standards Institution.

Budak, E., Altintas, Y. and Armarego, E. J. A. (1996) Prediction of milling force coefficients from orthogonal cutting data. *Trans. ASME. J. Manufact. Sci. Eng.* **118**, 216–224.

Byrne, G., Dornfeld, D., Inasaki, I., Ketteler, G., Konig, W. and Teti, R. (1995) Tool condition monitoring (TCM) – the status of research and industrial application. *Annals CIRP* **44**(2), 541–567.

Carpenter, G. and Gossberg, S. (1987) ART 2: self-organisation of stable category recognition codes for analog input patterns. *Appl. Opt.* **26**, 4919–4930.

Centner, R. M. and Idelsohn, J. M. (1964) Adaptive controller for a metal cutting process. *IEEE Trans. Appl. Ind.* **83**, 154–161.

Chen, S. J., Hinduja, S. and Barrow, G. (1989) Automatic tool selection for rough turning operation. *Int. J. Machine Tools Manufact.* **29**, 535–553.

Chen, Y. B., Sha, J. L. and Wu, S. M. (1990) Diagnosis of the tapping process by information measure and probability voting approach. *Trans. ASME J. Eng. Ind.* **112**, 319–325.

Chen, Y., Hui, A. and Du, R. (1995) A fuzzy expert system for the design of machining operations. *Int. J. Machine Tools Manufact.* **35**, 1605–1621.

Cheng, D. K. (1972) *Analysis of Linear Systems*. Reading: Addison-Wesley.

Chryssolouris, G., Domroese, M. and Beaulieu, P. (1992) Sensor synthesis for control of manufacturing processes. *Trans ASME J. Eng. Ind.* **114**, 158–174.

Dan, L. and Mathew, J. (1990) Tool wear and failure monitoring techniques for turning – a review. *Int. J. Machine Tools Manufact.* **30**, 579–598.

Daubechies, I. (1988) Orthonormal bases of compactly supported wavelets. *Comm. Pure Applied Math.* **41**, 909–996.

Diniz, A. E., Liu, J. J. and Dornfeld, D. A. (1992) Correlating tool life, tool wear and surface roughness by monitoring acoustic emission in finish turning. *Wear* **152**, 395–407.

Dornfeld, D. A. (1990) Neural network sensor fusion for tool condition monitoring. *Annals of the CIRP* **39**, 101–104.

Ghasempoor, A., Moore, T. N. and Jeswiet, J. (1998) On-line wear estimation using neural networks. *Proc. I. Mech. E. Lond.* **212** Pt.B, 105–112.

Giusti, F., Santochi, M. and Dini, G. (1986) COATS: An expert module for optimal tool selection. *Annals of the CIRP* **35**, 337–340.

Gong, W., Obikawa, T. and Shirakashi, T. (1997) Monitoring of tool wear states in turning based on wavelet analysis. *JSME Int. J.* (Series C) **40**, 447–453.

International Standard (1991) *ISO 1832 3rd edn.: Indexable Inserts for Cutting Tools – Designation*. Geneva: International Organization for Standardization.

International Standard (1995) *ISO 5608: Turning and Copying Tool Holders and Cartridges for Indexable Inserts – Designation*. Geneva: International Organization for Standardization.

Kasashima, N., Mori, K. and Ruiz, G. H. (1994) Diagnosing cutting tool conditions in milling using wavelet transform. In Usui, E. (ed.), *Advancement of Intelligent Production*. Amsterdam: Elsevier Science BV, 339–344.

Kline, W. A. and DeVor, R. E. (1983) The effect of runout on cutting geometry and forces in end milling. *Int. J. Mach. Tool Des. Res.* **23**, 123–140.

Kline, W A., DeVor, R. E. and Lindberg, J. R. (1982) The prediction of cutting forces in end milling with application to cornering cuts. *Int. J. Mach. Tool Des. Res.* **22**, 7–22.

Ko, T. J. and Cho, D. W. (1994) Tool wear monitoring in diamond turning by fuzzy pattern recognition. *Trans ASME J. Eng. Ind.* **116**, 225–232.

Koornwinder, T. H. (ed) (1993) *Wavelets: An Elementary Treatment of Theory and Application*. Singapore: World Scientific.

Konig, W., Langhammer, K. and Schemmel, H. U. (1972) Correlation between cutting force components and tool wear. *Annals of the CIRP* **21**, 19–20.

Koren, Y., Ko, T. R., Ulsoy, A. G. and Danai, K. (1991) Flank wear estimation under varying cutting conditions. *Trans ASME J. Dyn. Sys. Meas. Control* **113**, 300–307.

Leem, C. S., Dornfeld, D. A. and Dreyfus, S. E. (1995) A customised neural network for sensor fusion in on-line monitoring of cutting tool wear. *Trans ASME J. Eng. Ind.* **117**, 152–159.

Liu, T. I. and Anantharaman, K. S. (1994) Intelligent classification and measurement of drill wear. *Trans ASME J. Eng. Ind.* **116**, 392–397.

Matsushima, K. and Sata, T. (1974) On-line control of the cutting state by the pattern recognition technique. *Annals of the CIRP* **23**, 151–152.

Micheletti, G. F., Koenig, W. and Victor, H. R. (1976) In process tool wear sensors for cutting operations. *Annals of the CIRP* **25**, 483–496.

Moriwaki, T. and Tobito, M. (1990) A new approach to automatic detection of life of coated tool based on acoustic emission measurement. *Trans. ASME J. Eng. Ind.* **112**, 212–218.

Moriwaki, T. and Mori, Y. (1993) Recognition of cutting state based on neural network sensor fusion. *J. Japan Soc. Prec. Eng.* **59**, 779–784.

Moriwaki, T, Shamoto, E., Kou, D. and Sugihara, K. (1995) Milling state recognition based on dynamic cutting force model (1st report). *Trans. Japan Soc. Mech. Eng.* **61-C**, 2592–2598.

Niu, Y. M., Wong, Y. S., Hong, G. S. and Liu, T. I. (1998) Multi-category classification of tool conditions using wavelet packets and ART2 network. *Trans. ASME J. Manufact. Sci. Eng.* **120**, 807–816.

Obikawa, T. and Matsumura, T. (1994) Visualization, adaptation and optimization of machining technology. *J. Japan Soc. Prec. Eng.* **60**, 641–646.

Obikawa, T., Liu, W. N. and Shirakashi, T. (1995) Prediction of cutting temperature and flank wear rate for practical use using personal computer. *Proc. 7th Int. Manufac. Conf. China,* Harbin, **2**, 55–60.

Obikawa, T., Kaseda, C., Matsumura, T., Gong, W. G. and Shirakashi, T. (1996) Tool wear monitoring for optimizing cutting conditions. *J. Mat. Proc. Tech.* **62**, 374–379.

Oraby, S. E. and Hayhurst, D. R. (1991) Development of models for tool wear force relationships in metal cutting. *Int. J. Mech. Sci.* **33**, 125–138.

Paul, B. and Mirandy, L. (1976) An improved fracture criterion for three-dimensional stress states. *Trans. ASME, J. Eng. Mat. Tech.* **98**, 159–163.

Rangwala, S. and Dornfeld, D. (1990) Sensor integration using neural networks for intelligent tool condition monitoring. *Trans. ASME J. Eng. Ind.* **112**, 219–228.

Rosenblatt, F. (1961) *Principles of Neurodynamics: Perceptron and the Theory of Brain Mechanics*. Washington DC: Spartan Books.

Ryabov, O., Mori, K., Kasashima, N and Uehara, K. (1996) An in-process direct monitoring method for milling tool failures using a laser sensor. *Annals of the CIRP* **45**(1), 97–100.

Sata, T., Matsushima, K., Nagakura, T. and Kono, E. (1973) Learning and recognition of the cutting states by the spectrum analysis. *Annals of the CIRP* **22**, 41–42.

Shaw, M. C. (1984) *Metal Cutting Principles*. New York: Oxford University Press.

Shinozuka, J. (1998) Analytical prediction of cutting performance of grooved rake face tools. PhD Thesis. Tokyo: Tokyo Institute of Technology.

Shinozuka, J., Obikawa, T. and Shirakashi, T. (1994) Cutting performance of tools with curved rake face. In Usui, E. (ed), *Advancement of Intelligent Production*. Amsterdam: Elsevier Science BV, 379–384.

Shirakashi, T., Ihara, T., Kanazawa, K. and Usui, E. (1987) Analytical prediction of chipping occurrence of carbide tool in interrupted turning operation with temperature rise (1st report). *J. Japan Soc. Prec. Eng.* **53**, 1589–1595.

Shirakashi, T., Obikawa, T. and Shinozuka, J. (1998) Development of intelligent milling head. In Chen, D. *et al.* (eds), *Progress of Cutting and Grinding.* Beijing: International Academic Publishers, pp. 344–349.

Shirakashi, T., Obikawa, T., Shinozuka, J. and Yoshino, M. (1999) Development of reaction-controllable milling head and its cutting performance (2nd Report). *Trans. Japan Soc. Mech. Eng.* **65-C**, 1223–1228.

SITC (1987) Application of expert system to preparation of MC processing. In *Technological Development of Collaborative Research*. Kawaguchi: Saitama Industrial Technology Center, p. 27.

Spence, A. D. and Altintas, Y. (1994) A solid modeller based milling process simulation and planning system. *Trans. ASME J. Eng. Ind.* **116**, 61–69.

Stephenson, D. A. (1991) Assessment of steady-state metal cutting temperature models based on simultaneous infrared and thermocouple data. *Trans. ASME, J. Eng. Ind.* **113**, 121–128.

Stephenson, D. A. and Agapiou, J. S. (1997) *Metal Cutting Theory and Practice*. New York: Marcel Dekker, p. 601.

Takata, S. (1993) Generation of a machining scenario and its applications to intelligent machining operation. *Annals of the CIRP* **42**(1), 531–534.

Tansel, I. N. (1992) Modelling 3-D dynamics with neural networks. *Int. J. Mach. Tools Manufact.* **32**, 829–853.

Tansel, I. N., Mekdeci, C., Rodriguez, O. and Uragun, B. (1993) Monitoring drill conditions with wavelet based encoding and neutral networks. *Int. J. Mach. Tools Manufact.* **33**, 559–575.

Tansel, I. N., Mekdeci, C. and McLaughlin, C. (1995) Detection of tool failure in end milling with wavelet transformations and neural networks (WT-NN). *Int. J. Mach. Tools and Manufact.* **35**, 1137–1147.

Tarng, Y. S., Hseih, Y. W. and Hwang, S. T. (1994) Sensing tool breakage in face milling with a neural network. *Int. J. Mach. Tools and Manufact.* **34**, 341–350.

Tarng, Y. S., Cheng, C. I. and Kao, J. Y. (1995) Modeling of three-dimensional numerically controlled end milling operation. *Int. J. Mach. Tools and Manufact.* **35**, 939–950.

Tlusty, J. and Andrews, G. C. (1983) A critical review of sensors for unmanned machining. *Annals of the CIRP* **32**, 563–572.

Tlusty, J. (1985) Machine dynamics. In King, R. I. (ed), *Handbook of High Speed Machining Technology*. New York: Chapman and Hall, pp. 48–153.

Tonshoff, H. K., Wulfsberg, J. P., Kals, H. J. J., Konig, W. and van Luttervelt, C. A., (1988) Developments and trends in monitoring and control of machining processes. *Annals of the CIRP* **37**(2), 611–622.

Trigger, K. J. and Chao, B. T. (1956) The mechanism of crater wear of cemented carbide tools. *Trans. ASME.* **78**, 1119–1126.

Usui, E., Shirakashi, T. and Kitagawa, T. (1978) Analytical prediction of three dimensional cutting process. *Trans. ASME., J. Eng. Ind.* **100**, 236–243.

Usui, E., Ihara, T. and Shirakashi, T. (1979) Probabilistic stress-criterion of brittle fracture of carbide tool materials. *Bull. Japan Soc. Prec. Eng.* **13**, 189–194.

Usui, E. Shirakashi, T. and Kitagawa, T. (1984) Analytical prediction of cutting tool wear. *Wear* **100**, 129–151.

Zimmermann, H. J. (1976) Description and optimization of fuzzy systems, *Int. J. General Systems* **2**, 209–215.

Zimmermann, H. J. (1991) *Fuzzy Set Theory and Applications*. Boston: Kluwer Academic Publishers.

Appendix 1

Metals' plasticity, and its finite element formulation

This appendix supports Chapters 2 and 6 and subsequent chapters. More complete descriptions of plasticity mechanics can be found in any of the excellent texts from the early works of Hill (1950) and Prager and Hodge (1951), through books such as by Thomsen *et al*. (1965) and Johnson and Mellor (1973), to more recent finite element oriented work (Kobayashi *et al*. 1989).

Section A1.1 answers the questions, initially in terms of principal stresses and strains (Figure A1.1) concerning (i) what combinations of principal stresses σ_1, σ_2, and σ_3 will cause yielding of a metal; (ii) if a metal has yielded, and the stress state is changed to cause further plastic strain increments $d\varepsilon_1$, $d\varepsilon_2$, and $d\varepsilon_3$, what are the relations between the strain increments and the stresses; and (iii) what is the work rate in a plastic field? Extension of the answers to non-principal stress state descriptions is briefly introduced. In Section A1.1, elastic components of deformation are ignored. Any anisotropy of flow, such as is important for example in sheet metal forming analysis, is also ignored.

To analyse flow in any particular application, the yielding and flow laws (constitutive laws) are combined with equilibrium and compatibility equations and boundary conditions. If the flow is in plane strain conditions and when a metal's elastic responses and work hardening can be ignored, the equilibrium and compatibility equations take a particularly simple form if they are referred to maximum shear stress directions. The analysis of flow in this case is known as slip-line field theory and is introduced in Section A1.2.

Apart from the circumstances of slip-line field theory, the simultaneous solution of

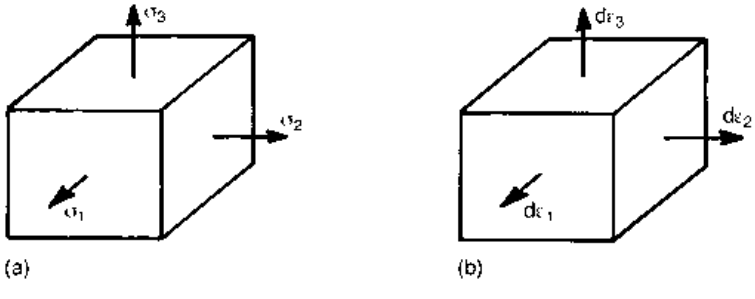


Fig. A1.1 (a) Principal stresses and (b) principal strain increments

constitutive, equilibrium and compatibility equations is difficult. Finite element approximations are needed to solve metal machining problems. Further analysis of stress, needed to support finite element methods, is found in Section A1.3. Section A1.4 extends the constitutive laws to include elastic deformation, and manipulates both rigid–plastic and elastic–plastic laws to forms suitable for numerical analysis. Section A1.5 considers finite element methods in particular.

A1.1 Yielding and flow under triaxial stresses: initial concepts

A1.1.1 Yielding and the description of stress

The principal stresses acting on a metal may be written as the sum of a hydrostatic (or mean) part σ_m and a deviation from the mean, or deviatoric part, which will be written as σ':

$$\left.\begin{aligned} \sigma_m &= (\sigma_1 + \sigma_2 + \sigma_3)/3 \\ \sigma_1' &= \sigma_1 - \sigma_m \equiv 2\sigma_1/3 - (\sigma_2 + \sigma_3)/3 \\ \sigma_2' &= \sigma_2 - \sigma_m \equiv 2\sigma_2/3 - (\sigma_3 + \sigma_1)/3 \\ \sigma_3' &= \sigma_3 - \sigma_m \equiv 2\sigma_3/3 - (\sigma_1 + \sigma_2)/3 \end{aligned}\right\} \quad \text{(A1.1)}$$

Hydrostatic stress plays no part in the yielding of cast or wrought metals, if they have no porosity. (They are incompressible; any hydrostatic volume change is elastic and is recovered on unloading.) An acceptable yield criterion must be a function only of the deviatoric stresses. Inspection of equation (A1.1) shows that the sum $(\sigma_1' + \sigma_2' + \sigma_3')$ is always zero: yielding cannot be a function of this. However, the resultant deviatoric stress σ_r':

$$\sigma_r' = (\sigma_1'^2 + \sigma_2'^2 + \sigma_3'^2)^{1/2} \quad \text{(A1.2)}$$

has been found by experiment to form a suitable yield function. That yielding occurs when σ_r' reaches a critical value is now known as the von Mises yield criterion.

The magnitude of the critical value can be related to the yield stress Y in a simple tension test. In simple tension, two of the principal stresses, say σ_2 and σ_3, are zero. Substituting these and $\sigma_1 = Y$ into equations (A1.1) for the deviatoric stresses and then these into equation (A1.2) gives for the yield criterion

$$\sigma_r' = Y\sqrt{2/3} \quad \text{(A1.3a)}$$

Alternatively, the critical value may be related to the yield stress k in a simple shear test, in which for example $\sigma_1 = -\sigma_2 = k$ and $\sigma_3 = 0$. By substituting these values in equations (A1.1) and (A1.2),

$$\sigma_r' = k\sqrt{2} \quad \text{(A1.3b)}$$

That the yield stress in tension is $\sqrt{3}$ times that in shear is just one consequence of the von Mises yield criterion.

It is customary to introduce a quantity known as the equivalent stress, $\bar{\sigma}$, equal to $\sqrt{(3/2)}$ times the resultant deviatoric stress. The critical value of the equivalent stress for yielding to occur is then identical to the yield stress in simple tension. The von Mises yield criterion becomes

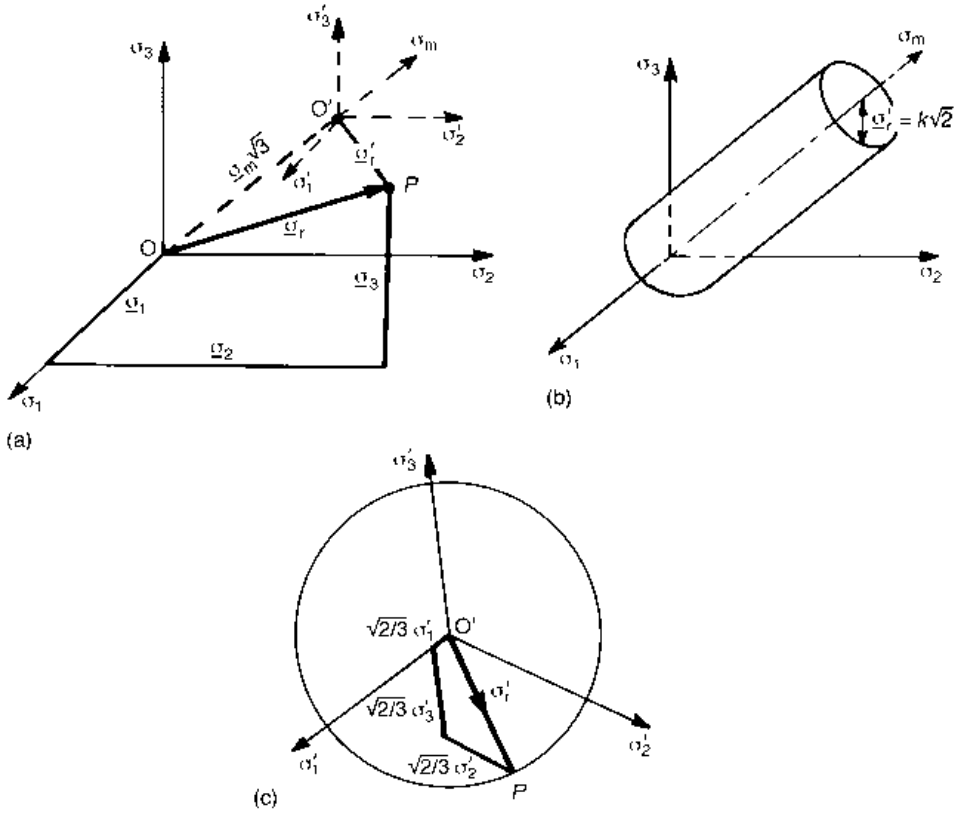


Fig. A1.2 Geometrical representations of principal stresses and yielding

$$\left.\begin{aligned} \bar{\sigma} &\equiv \sqrt{3/2}\,\sigma'_{\mathrm{r}} = Y \\ \bar{\sigma} &\equiv \sqrt{3/2}\,\sigma'_{\mathrm{r}} = k\sqrt{3} \end{aligned}\right\} \tag{A1.4}$$

The equivalent stress and the yield criterion may be represented in a number of different ways. Figure A1.2(a) is a geometrical view of a state of stress P in principal stress space, origin O. The vector OP is the resultant stress σ_{r}. It has principal components (σ_1, σ_2, σ_3). Alternatively, it has components OO′ and O′P along and perpendicular to the hydrostatic line $\sigma_1 = \sigma_2 = \sigma_3$. This line has direction cosines $1/\sqrt{3}$ with the principal axes, so $\mathrm{OO}' = \sigma_{\mathrm{m}}\sqrt{3}$. OP is σ'_{r}. By vector addition

$$\sigma'^{2}_{\mathrm{r}} = \sigma^2_{\mathrm{r}} - 3\sigma^2_{\mathrm{m}} = (\sigma_1^2 + \sigma_2^2 + \sigma_3^2) - 3\sigma^2_{\mathrm{m}} \tag{A1.5}$$

After substituting for σ_{m} from equation (A1.1),

$$3\sigma'^{2}_{\mathrm{r}} = (\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2 \tag{A1.6}$$

The yield criterion may be restated in terms of the principal stresses:

$$\bar{\sigma}^2 = \frac{1}{2}\left[(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2\right] = Y^2 \text{ or } 3k^2 \tag{A1.7}$$

The yield criterion, equation (A1.3) or (A1.7), can be represented (Figure A1.2(b)) by the cylinder, σ'_r = constant. For a material to yield, its stress state must be raised to lie on the surface of the cylinder. A simpler diagram (Figure A1.2(c)) is produced by projecting the stress state on to the deviatoric plane: that is the plane perpendicular to σ_m through the point O′. The principal deviatoric stress directions have direction cosines $\sqrt{(2/3)}$ with their projections in the deviatoric plane. Figure A1.2(c) shows the projected deviatoric stress components as well as the resultant deviatoric stress. Yield occurs when the resultant deviatoric stress lies on the yield locus of radius $k\sqrt{2}$.

A1.1.2. Plastic flow rules and equivalent strain

Suppose that material has been loaded to a plastic state *P* (Figure A1.3(a)) and is further loaded to *P** to cause more yielding, so that the yield locus expands by work hardening to a new radius σ'^*_r: what further plastic principal strain increments ($d\varepsilon_1$, $d\varepsilon_2$, $d\varepsilon_3$) then occur?

It is found (Figure A1.3(b)) that the strain increments are in proportion to the deviatoric stresses. A resultant strain increment $d\varepsilon_r$, is defined analogously to σ'_r as

$$d\varepsilon_r = (d\varepsilon_1^2 + d\varepsilon_2^2 + d\varepsilon_3^2)^{1/2} \tag{A1.8}$$

$d\varepsilon_r$ is parallel to σ'_r. It is as if the change of deviatoric stress, $d\sigma'_r$ in Figure A1.3(a) has a component tangential to the yield locus that causes no strain and one normal to the locus which is responsible for the plastic strain. In fact, the tangential component causes elastic strain, but this is neglected until Section A1.4.

The proportionalities between $d\varepsilon_r$ and σ'_r may be written

$$d\varepsilon_1 = c\sigma'_1; \qquad d\varepsilon_2 = c\sigma'_2; \qquad d\varepsilon_3 = c\sigma'_3 \tag{A1.9}$$

where the constant *c* depends on the material's work hardening rate. By substituting equations (A1.9) into (A1.8), $c = d\varepsilon_r/\sigma'_r$.

To simplify the description of work hardening, an equivalent strain increment $d\bar{\varepsilon}$ is

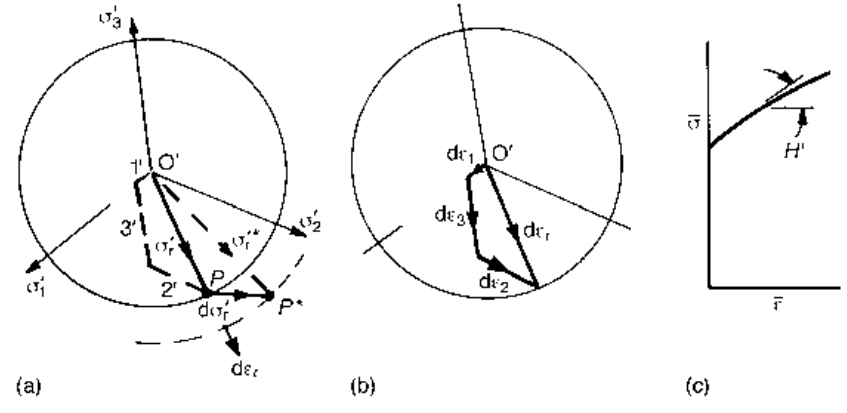


Fig. A1.3 (a) A plastic stress increment, P to P*; (b) the resulting strain increment; and (c) the linking work-hardening relationship

introduced, proportional to $d\varepsilon_r$, just as $\bar{\sigma}$ has been introduced proportional to σ'_r. $d\bar{\varepsilon}$ is defined as

$$d\bar{\varepsilon} = \sqrt{2/3}\, d\varepsilon_r \tag{A1.10}$$

Then, in a simple tension test (in which $d\varepsilon_2 = d\varepsilon_3 = -0.5 d\varepsilon_1$), $d\bar{\varepsilon} = d\varepsilon_1$. A plot of equivalent stress against equivalent strain (Figure A1.3(c)), gives the work hardening of the material along any loading path. H' is the work hardening rate $d\bar{\sigma}/d\bar{\varepsilon}$. $d\varepsilon_r$ and σ'_r in the expression for c may be replaced by $\sqrt{3/2}\, d\bar{\varepsilon}$ and $\sqrt{2/3}\bar{\sigma}$, and $d\bar{\varepsilon}$ by $d\bar{\sigma}/H'$, to give

$$c = \frac{3\, d\bar{\sigma}}{2\, H'\bar{\sigma}} \tag{A1.11}$$

Equations (A1.9) and (A1.11) are known as the Levy–Mises flow laws.

A1.1.3 Extended yield and flow rules, and the plastic work rate

The yield criterion must be able to be formulated in any set of non-principal axes, with equation (A1.7) as a special case. Consider the expression

$$(\sigma_x - \sigma_y)^2 + (\sigma_y - \sigma_z)^2 + (\sigma_z - \sigma_x)^2 + 6(\tau_{xy}^2 + \tau_{yz}^2 + \tau_{zx}^2) = 6k^2 \text{ or } 2Y^2 \tag{A1.12}$$

When the shear stresses τ are zero, it is identical to equation (A1.7). When the direct stresses are zero, the factor 6 cancels out and the equation states that yielding occurs when the resultant shear stress reaches k. Equation (A1.12) thus is possible as an expression for the yield criterion generalized to non-principal stress axes. It is established more rigorously in Section (A1.3).

Similarly, the Levy–Mises flow rules may be written more generally as

$$\frac{d\varepsilon_x}{\sigma'_x} = \frac{d\varepsilon_y}{\sigma'_y} = \frac{d\varepsilon_z}{\sigma'_z} = \frac{d\varepsilon_{xy}}{\tau_{xy}} = \frac{d\varepsilon_{yz}}{\tau_{yz}} = \frac{d\varepsilon_{zx}}{\tau_{zx}} = \frac{3\, d\bar{\varepsilon}}{2\, \bar{\sigma}_x} \quad \text{or} \quad \frac{3\, d\bar{\sigma}}{2\, H'\bar{\sigma}} \tag{A1.13}$$

Care must be taken to interpret the shear strains. $d\varepsilon_{xy} = d\varepsilon_{yx} = 1/2(\partial u/\partial y + \partial v/\partial x)$, for example, where u and v have the usual meanings as displacement increments in the x and y directions respectively. This differs from the definition $\gamma = (\partial u/\partial y + \partial v/\partial x)$ by a factor of 2.

Finally, the work increment dU per unit volume in a plastic flow field is

$$\begin{aligned} dU &= \sigma_{xx} d\varepsilon_{xx} + \sigma_{yy} d\varepsilon_{yy} + \sigma_{zz} d\varepsilon_{zz} + 2(\sigma_{xy} d\varepsilon_{xy} + \sigma_{yz} d\varepsilon_{yz} + \sigma_{zx} d\varepsilon_{zx}) \\ &\equiv \bar{\sigma} d\bar{\varepsilon} + \sigma_m (d\varepsilon_{xx} + d\varepsilon_{yy} + d\varepsilon_{zz}) \end{aligned} \tag{A1.14}$$

but because the material is incompressible, the last term is zero: the work increment per unit volume is simply $\bar{\sigma} d\bar{\varepsilon}$.

A1.2 The special case of perfectly plastic material in plane strain

Section A1.1 is concerned with a plastic material's constitutive laws. Material within a plastically flowing region is also subjected to equilibrium and compatibility (volume conservation) conditions, for example in Cartesian coordinates

$$\left.\begin{aligned}\frac{\partial\sigma_{xx}}{\partial x}+\frac{\partial\sigma_{xy}}{\partial y}+\frac{\partial\sigma_{xz}}{\partial z}=\frac{\partial\sigma_{yx}}{\partial x}+\frac{\partial\sigma_{yy}}{\partial y}+\frac{\partial\sigma_{yz}}{\partial z}=\frac{\partial\sigma_{zx}}{\partial x}+\frac{\partial\sigma_{zy}}{\partial y}+\frac{\partial\sigma_{zz}}{\partial z}=0\\ \frac{\partial\dot{u}}{\partial x}+\frac{\partial\dot{v}}{\partial y}+\frac{\partial\dot{w}}{\partial z}=0\end{aligned}\right\} \tag{A1.15}$$

where $\dot{u}$, $\dot{v}$ and $\dot{w}$ are the x, y and z components of the material's velocity. The general three-dimensional situation is complicated. However, in plane strain conditions, and if the work hardening of the material is negligible, the integration of the equilibrium and compatibility equations, under the constraint of the constitutive equations, is simplified by describing the stresses and velocities not in a Cartesian coordinate system but in a curvilinear system that is everywhere tangential to the maximum shear stress directions. The net of curvilinear maximum shear stress lines is known as the slip-line field. Determining the shape of the net for any application and then the stresses and velocities in the field is achieved through slip-line field theory. This theory is now outlined.

A1.2.1 Constitutive laws for a non-hardening material in plane strain

When the strain in one direction, say the z-direction, is zero, from the flow rules (equation (A1.13)) the deviatoric stresses in that direction are also zero. Then $\sigma_{zz} = \sigma_m = (1/2)(\sigma_{xx} + \sigma_{yy})$. The yield criterion, equation (A1.12), and flow rules, equation (A1.13), become

$$\left.\begin{aligned}(\sigma_{xx}-\sigma_{yy})^2+4\sigma_{xy}^2=4k^2\\ \frac{d\varepsilon_{xx}}{1/2(\sigma_{xx}-\sigma_{yy})}=\frac{-d\varepsilon_{yy}}{1/2(\sigma_{xx}-\sigma_{yy})}=\frac{d\varepsilon_{xy}}{\sigma_{xy}}\end{aligned}\right\} \tag{A1.16}$$

When the material is non-hardening, the shear yield stress k is independent of strain. If, in a plastic region, the x, y directions are chosen locally to coincide with the maximum shear stress directions, σ_{xx} becomes equal to σ_{yy} (and equal to σ_m), so $(\sigma_{xx} - \sigma_{yy}) = 0$. Equation (A1.16) becomes a statement that (i) the maximum shear stress is constant throughout the plastic region and (ii) there is no extension along maximum shear stress directions. The consequences of these statements for stress and velocity variations throughout a plastic region are developed in the next two subsections.

A1.2.2 Stress relations in a slip-line field

Figure A1.4(a) shows a network of slip-lines in a plastic field. The pressure p $(= -\sigma_m)$ and the shear stress k is shown at a general point O in the field. The variation of pressure throughout the field may be found by integrating the equilibrium equations along the slip-lines. How this is done, and some consequences for the shape of the field, are now described.

First, the two families of lines, orthogonal to each other, are labelled α and β. Which is labelled α and which is β is chosen, by convention, so that, if α and β are regarded as a right-handed coordinate system, the direction of the largest principal stress lies in the first quadrant. (This means that the shear stresses k are positive as shown in the figure.) The

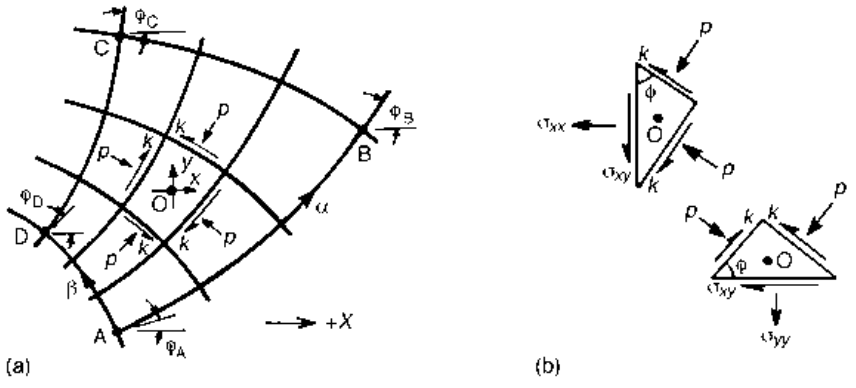


Fig. A1.4 (a) A slip-line net and (b) free body equilibrium diagrams around O

direction of an α line at any point is described by its anticlockwise rotation from a datum direction, for example ϕ from the $+X$ direction.

By stress analysis (Figure A1.4(b)), the stresses p and k at O have (x,y) components (σ_{xx}, σ_{yy}, σ_{xy})

$$\sigma_{xx} = -p - k\sin 2\phi; \qquad \sigma_{yy} = -p + k\sin 2\phi; \qquad \sigma_{xy} = k\cos 2\phi \tag{A1.17}$$

Substituting these into the equilibrium equations

$$\frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xy}}{\partial y} = 0; \qquad \frac{\partial \sigma_{yy}}{\partial y} + \frac{\partial \sigma_{xy}}{\partial x} = 0 \tag{A1.18a}$$

gives, after noting that k is a constant,

$$\left.\begin{aligned} -\frac{\partial p}{\partial x} - 2k\cos 2\phi\,\frac{\partial \phi}{\partial x} - 2k\sin 2\phi\,\frac{\partial \phi}{\partial y} = 0 \\ -\frac{\partial p}{\partial y} + 2k\cos 2\phi\,\frac{\partial \phi}{\partial y} - 2k\sin 2\phi\,\frac{\partial \phi}{\partial x} = 0 \end{aligned}\right\} \tag{A1.18b}$$

If the direction of X is chosen so that $\phi = 0$, that is so that the α slip-line is tangential to X, $\sin 2\phi = 0$ and $\cos 2\phi = 1$ and

$$\frac{\partial}{\partial x}(p + 2k\phi) = 0; \qquad \frac{\partial}{\partial y}(p - 2k\phi) = 0 \tag{A1.18c}$$

or

$$p + 2k\phi = \text{constant along an } \alpha\text{-line}; \qquad p - 2k\phi = \text{constant along a } \beta\text{-line}. \tag{A1.18d}$$

If the geometry of the slip-line field and the pressure at any one point is known, the pressure at any other point can be calculated. Equation (A1.19) relates, for the example of Figure A1.4, pressures along the α-lines AB and DC, and along the β-lines AD and BC

$$\left.\begin{array}{ll} p_A + 2k\phi_A = p_B + 2k\phi_B; & p_C + 2k\phi_C = p_D + 2k\phi_D \\ p_A + 2k\phi_A = p_D - 2k\phi_D; & p_B - 2k\phi_B = p_C + 2k\phi_C \end{array}\right\} \quad \text{(A1.19)}$$

Geometry of the field

The inclinations ϕ_A, ϕ_B, ϕ_C and ϕ_D are not independent. The pressure p_C at C may be calculated from that at A in two ways from equations (A1.19), either along the path ABC or ADC. For p_C to be single valued

$$\phi_B - \phi_A = \phi_C - \phi_D; \qquad \phi_D - \phi_A = \phi_C - \phi_B \quad \text{(A1.20)}$$

Figure A1.5(a) gives some common examples of curvilinear nets that satisfy this condition: a grid of straight lines in which the pressure is constant, a centred fan and a net constructed on two circular arcs. Systematic methods for constructing more complicated fields are described by Johnson *et al*. (1982).

Force boundary conditions

Figure A1.5(b) shows α and β slip-lines meeting a tool surface on which there is a friction stress τ_f. Equilibrium of forces on the triangle ABC, in the direction of τ_f, gives

$$\tau_f = k \cos 2\zeta \quad \text{(A1.21)}$$

Thus, the magnitude of the friction stress relative to k determines the angle ζ at which the α-line intersects the tool face. Similarly, α and β slip-lines meet a free surface at 45° ($\tau_f/k = 0$). Because there is no normal stress on a free surface, $p = \pm k$ there, depending on the direction of k.

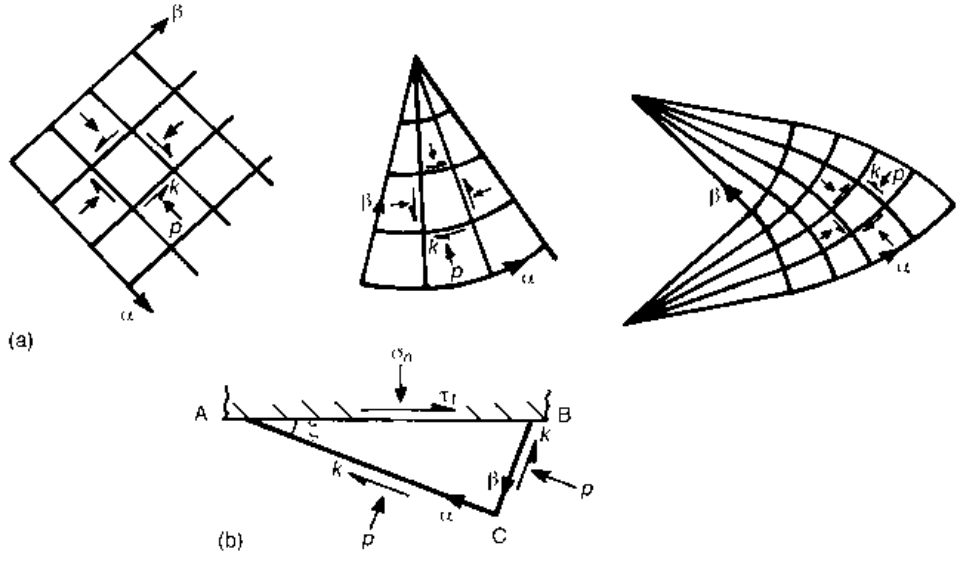


Fig. A1.5 (a) Nets satisfying internal force equilibrium and (b) slip-lines meeting a friction boundary

A1.2.3 Velocity relations in a slip-line field

Analogous equations to equations (A1.18d) exist for the variation of velocity along the slip lines. However, the statement that there is no extension along a slip line (Section A1.2.1) may directly be used to develop velocity relations and further rules for the geometry of a slip-line field. Figure A1.6(a) repeats the net of Figure A1.4(a). Figure A1.6(b) represents, in a velocity diagram, possible variations of velocity in the field. Because there is no extension along a slip-line, every element of the velocity net is perpendicular to its corresponding element in the physical plane of Figure A1.6(a). Thus, equations (A1.20) also apply in the velocity diagram.

Velocity boundary conditions

Other constraints on slip-line fields may be derived from velocity diagrams (in addition to the obvious boundary condition that the velocity of work material at an interface with a tool must be parallel to the tool surface). Figure A1.7(a) shows proposed boundaries AB and CDE between a plastic region and a rigid region in a metal forming process. Because this is a book on metal machining, the example is of continuous chip formation, but any example could have been chosen in which part of the work is plastically deformed and part is not.

First, the boundary between a plastic and a rigid region must be a slip-line. Secondly, the boundary between a plastic region and a rotating rigid region (for example CDE in Figure A1.7(a) must have the same shape in the physical plane as in the velocity diagram. Both these can be shown by considering the second case.

Suppose that any boundary such as CD is not a slip-line. Then any point such as H inside the plastic region can be joined to the boundary in two places by two slip-lines, for example to F and G by HF and HG. Figure A1.7(b) is the velocity diagram. The velocities v_F and v_G of points F and G are determined from the rigid body rotation of the chip to be ωOF and ωOG, where ω is the angular velocity of the chip. The velocity v_H relative to v_F is perpendicular to HF and that of v_H relative to v_G is perpendicular to HG. By comparing the positions of v_F, v_G and v_H relative to v_O, the origin of the velocity diagram, with the positions of F, G and H relative to the centre of rotation O in the physical diagram, it is

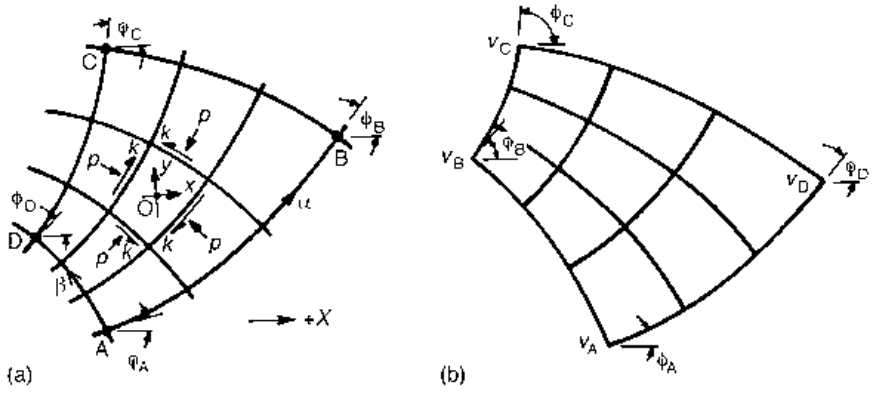


Fig. A1.6 (a) The physical net of Figure A1.4(a) and (b) a possible associated velocity diagram

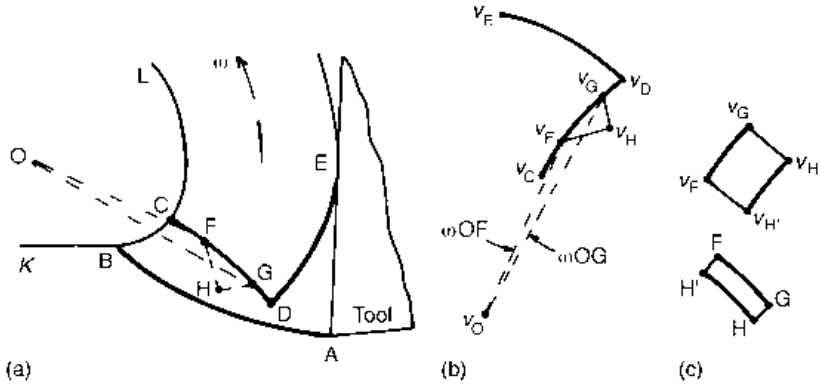


Fig. A1.7 (a) A possible machining process with (b) a partial velocity diagram and (c) an illustration of a velocity discontinuity across a slip-line

seen that the velocity of H is part of the rigid-body rotation: if the boundary CD is not a slip-line, it cannot accommodate velocity changes that must occur in a plastic field.

If the boundary is a slip-line, a point H can only be joined to the boundary in two places by three slip-lines: thus, the argument above can no longer be made. For continuity of flow between a plastic and a rigid region, the boundary between the two must be a slip-line. Figure A1.7(b) also shows the whole boundary $v_C\, v_D\, v_E$. It is visually obvious that only if it has the same shape relative to the origin of velocity that CDE has relative to O, can it be consistent with a rigid body rotation.

Velocity discontinuities

The usual procedure in slip-line field analysis is to construct fields that satisfy the geometry and force requirements of a problem and then to check that the velocity requirements are met. In this last part, one more feature of the theory must be introduced: the possibility of velocity jumps (discontinuities) occurring. Figure A1.7(c) returns to the consideration of the velocity of a point H in the plastic field. H is connected to the boundary by slip-lines, both directly to G and indirectly to F through H′. It is possible for there to be a finite velocity difference between H and G, however short is the length HG, i.e. a discontinuity. If there is a discontinuity, then the rules for constructing the velocity net require that there be a discontinuity of equal size between H′ and F. A velocity discontinuity can exist across a slip line, but only if it is of constant size along the line. It is not implied that there is a discontinuity in the condition of the example described here: examples of actual machining slip-line fields are given in Section A1.2.5.

A1.2.4 Further considerations

Slip-line fields must satisfy more than the force and velocity conditions considered in Sections A1.2.2 and A1.2.3. First, they must (as must every plastic flow) satisfy a work criterion, that everywhere the work rate on the flow is positive. This means that the direction of the shear stresses in the physical diagram must be the same as the direction of the shear strain rates deduced from the velocity diagram.

It must also be checked that it is possible that regions in the work material that are assumed to be rigid can in fact be rigid. For example, in Figure A1.7(a), in the rigid regions KBA and LCD, the loads change from values on BA and CD determined by the plastic flow to zero on the free surfaces KB and LC. It must be checked that such load changes can be accommodated without the yield stress being exceeded in the rigid regions in the neighbourhood of the vertices B and C. Checking for overstressing is introduced in another context in Appendix A5. The overstressing limits developed in Appendix 5 (Hill, 1954) apply here too.

A1.2.5. Machining slip-line fields

Figure A1.8 collects a range of slip-line fields, and their velocity diagrams (due to Lee and Shaffer, 1951, Kudo, 1965, and Dewhurst, 1978), which describe steady state chip formation by a flat-faced cutting tool.

The first is Lee and Shaffer's field. It describes formation of a straight chip. The work velocity U_{work} is transformed to the chip velocity U_{chip} by a discontinuous change U_{OA} tangential to the slip-line OA. The angle at which OA meets the free surface is not set by a free surface boundary condition. A is a singularity where the surface direction is not defined. Instead, the direction of OA is determined by its being perpendicular to BD. The inclination of BD to the rake face is given by equation (A1.21). Because all the slip-lines are straight, the hydrostatic pressure is constant along them (equation (A1.19)). The chip region above ADB is free, i.e. there are no forces acting on it. This determines that $p = k$ and AD = DB.

The second is due to Kudo. It may be thought of as a modification of Lee and Shaffer's field in which the primary shear plane OA is replaced by a fan-shaped zone of angular extent δ, still with a singularity at the free surface A. It still describes a straight chip. The slip-lines intersecting the rake face do so at a constant angle ζ: the field therefore continues to describe a condition of constant friction stress along the rake face. The free-chip boundary condition still requires $p = k$ on AD and DB and AD = DB. However δ can take a range of values, from zero up to a maximum at which the region below AE becomes overstressed. For the same friction condition, tool rake angle and feed, f, as in the Lee and Shaffer field, the Kudo field describes chip formation with a larger shear plane angle and a shorter contact length.

Two further fields suggested by Kudo are the third and fourth examples in Figure A1.8. These describe rotating chips. The boundaries ADB in the physical plane between the fields and the chips can be seen to transform into their own shapes in their velocity diagrams. The third field may be thought of as a distortion of the Lee and Shaffer field and the fourth as a distortion of Kudo's first field. The slip-lines in the secondary shear zone intersect the rake face at angles ζ which vary from O to B: these fields describe conditions of friction stress reducing from O to B. Because the slip-lines are curved, the hydrostatic stress now varies throughout the field. Again the allowable fields are limited by the requirement that material assumed rigid outside the flow zone around A must be able to be rigid. However, the possibility arises that it is the chip material downstream of A that becomes overstressed.

The last example shows another way in which a rotating chip may be formed. A fan region OED is centred on the cutting edge O and the remainder DA of the primary shear region is a single plane. With this field, the slip-lines intersect the rake face at a constant angle, so that it describes constant friction stress conditions. The fan angle ψ can take a

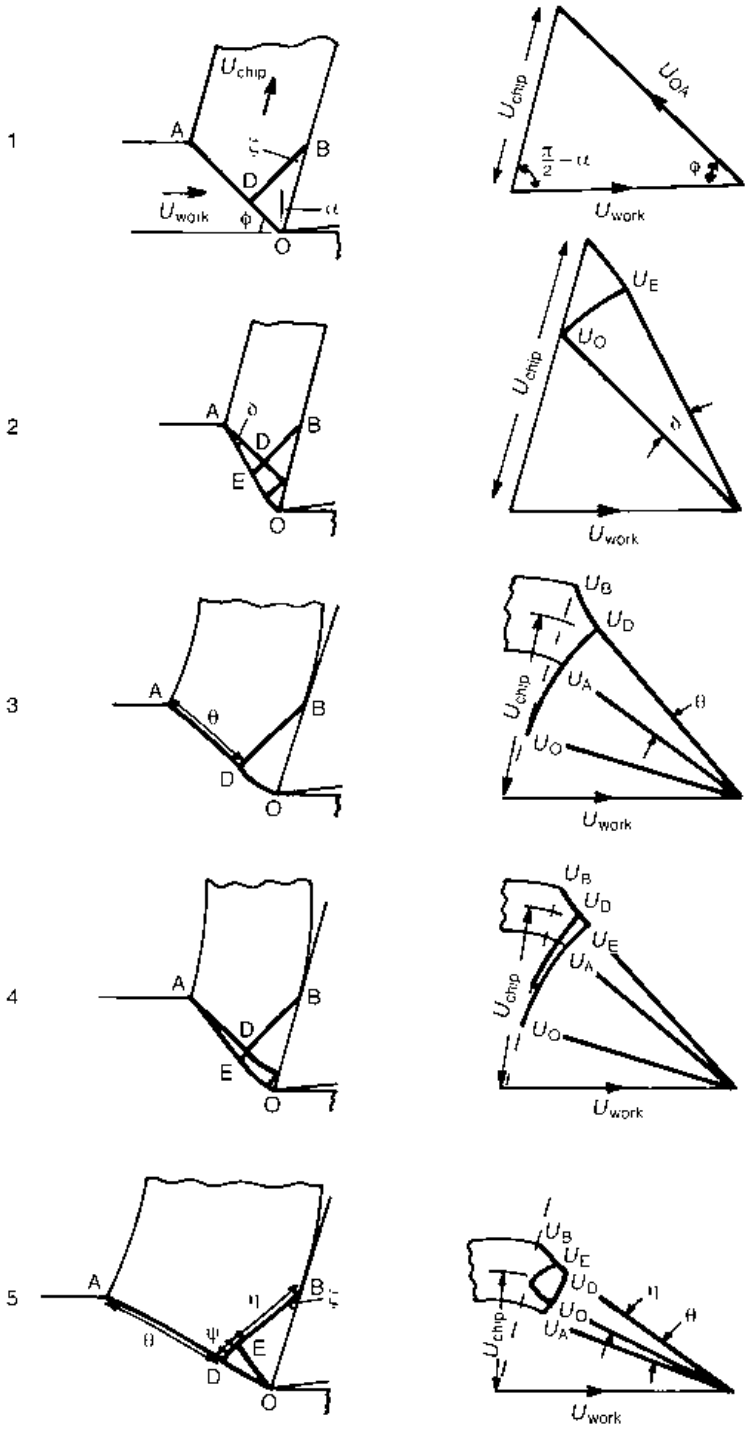


Fig. A1.8 Metal machining slip-line fields (left) and their velocity diagrams (right), due to (1) Lee and Shaffer (1951), (2–4) Kudo (1965) and (5) Dewhurst (1978)

range of values, limited only by its effect on overstressing material around A. For the same friction condition, tool rake angle and feed, *f*, as in the Lee and Shaffer field, this last field describes chip formation with a lower shear plane angle and a longer contact length.

A1.3 Yielding and flow in a triaxial stress state: advanced analysis

A1.3.1 Yielding and flow rules referred to non-principal axes – analysis of stress

The yield criterion is stated in equation (A1.7) in principal stress terms. It is extended to non-principal stresses in equation (A1.12): this has been justified in the two special cases when it represents principal stress and maximum shear stress descriptions of stress. It is now justified more generally, by showing that the function of stress which is the left-hand side of equation (A1.12) has a magnitude that is independent of the orientation of the (*x*,*y*,*z*) coordinate system. If it is valid in one case (as it is when the axes are the principal axes), it is valid for all cases. Tensor analysis is chosen as the tool for proving this, in part to introduce it for later use.

Tensor notation and the summation convention

Figure A1.9 shows two Cartesian coordinate systems (*x*,*y*,*z*) and (*x**,*y**,*z**) rotated arbitrarily with respect to each other. In the (*x*,*y*,*z*) system the stresses are σ_{ij} with *i* and *j* denoting

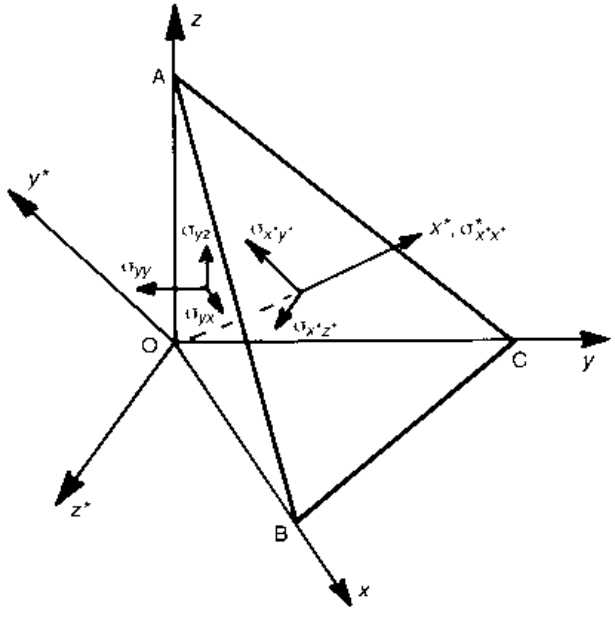


Fig. A1.9 (*x*,*y*,*z*) and (*x**,*y**,*z**) co-ordinate systems

x, y or z as appropriate. In the (x^*,y^*,z^*) system the stresses are σ^*_{kl}, with k,l denoting x^*, y^* or z^*. The figure also shows a tetrahedron OABC, the faces of which are normal to the x, y, z and x^* directions. Writing the direction cosines of x^* with x, y and z as a_{x^*x}, a_{x^*y} and a_{x^*z}, with similar quantities a_{y^*x}, a_{y^*y}, a_{y^*z} and a_{z^*x}, a_{z^*y}, a_{z^*z} for the direction cosines of y^* and z^* with x, y and z, first, by geometry, the ratios of the areas OAC, OAB and OBC to ABC are respectively a_{x^*x}, a_{x^*y} and a_{x^*z}. Then, from force equilibrium on the tetrahedron, for example

$$\begin{aligned}\sigma^*_{x^*x^*} = {} & a_{x^*x}a_{x^*x}\sigma_{xx} + a_{x^*x}a_{x^*y}\sigma_{yx} + a_{x^*x}a_{x^*z}\sigma_{zx} \\ & + a_{x^*y}a_{x^*y}\sigma_{yy} + a_{x^*y}a_{x^*x}\sigma_{xy} + a_{x^*y}a_{x^*z}\sigma_{zy} \\ & + a_{x^*z}a_{x^*z}\sigma_{zz} + a_{x^*z}a_{x^*y}\sigma_{yz} + a_{x^*z}a_{x^*x}\sigma_{xz}\end{aligned} \tag{A1.22a}$$

In general and more compactly, any of the stresses σ^*_{kl} may be written

$$\sigma^*_{kl} = \sum_{j=1}^{3}\sum_{i=1}^{3} a_{ki}a_{lj}\sigma_{ij} \tag{A1.22b}$$

Quantities which transform like this are called tensors, and the study of the properties of the transformation is tensor analysis.

By the summation convention, the summation signs are omitted, but are implied by the repetition of the suffixes i and j among the coefficients a_{ki} and a_{lj}. Thus equation (A1.22b) becomes

$$\sigma^*_{kl} = a_{ki}a_{lj}\sigma_{ij} \tag{A1.22c}$$

Furthermore, the repetition of k and l, between the left and right-hand sides of the equation, implies that it represents all nine equations for the components of σ^*. The meaning of the equation is unchanged by substituting another pair of letter suffixes, say m and n, for i and j: suffixes such as i and j, repeated on the same side of an equation, are called dummy suffixes and are said to be interchangeable. Suffixes such as k and l are called free suffixes. In the special case when $k = l$, the summation convention extends to include

$$\sigma^*_{kk} = \sigma^*_{x^*x^*} + \sigma^*_{y^*y^*} + \sigma^*_{z^*z^*} \tag{A1.22d}$$

Properties of the direction cosines

Because the angle between a direction i and another direction k is the same as the angle between the direction k and the direction i, $a_{ik} = a_{ki}$.

Because the scalar product of two unit vectors is unity if they are parallel and zero if they are perpendicular to each other, the same is true of the sum of the scalar products of their components in any other coordinate system. In repeated suffix notation, $a_{ik}\,a_{jk} = 1$ if $i = j$ and 0 if $i \neq j$. This can be written

$$a_{ik}a_{jk} = \delta_{ij} \tag{A1.23}$$

where δ_{ij} is defined as 1 or 0 depending respectively on whether or not $i = j$.

Transformations of stress

Now consider the summation of the direct stresses

$$\sigma^*_{kk} = a_{ki}a_{kj}\sigma_{ij} = \delta_{ij}\sigma_{ij} = \sigma_{ii} \tag{A1.24}$$

This demonstrates that the sum of direct stresses σ^*_{kk} in the (x^*,y^*,z^*) system equals the sum σ_{ii} in the (x,y,z) system. One of the systems could be the principal stress

system: thus, the hydrostatic stress σ_m is a stress invariant (it is known as the first stress invariant).

Consider now the product of stresses $\sigma^*_{kl}\,\sigma^*_{lk}$, with the transformations of equation (A1.22c), the interchangeability of dummy suffixes and equation (A1.23):

$$\begin{aligned}\sigma^*_{kl}\sigma^*_{lk} &= (a_{ki}a_{lj}\sigma_{ij})(a_{lm}a_{kn}\sigma_{mn}) = (a_{ki}a_{kn}\sigma_{ij})(a_{lm}a_{lj}\sigma_{mn}) \\ &= (\delta_{in}\sigma_{ij})(\delta_{mj}\sigma_{mn}) = \sigma_{nj}\sigma_{jn} = \sigma_{ij}\sigma_{ji}\end{aligned} \tag{A1.25}$$

In principal stress space, $\sigma_{ij}\sigma_{ji} = \sigma_r^2$. So σ_r is also a stress invariant (it is known as the second stress invariant). From equation (A1.25)

$$\sigma_r^2 = \sigma_{xx}^2 + \sigma_{yy}^2 + \sigma_{zz}^2 + 2(\sigma_{xy}^2 + \sigma_{yz}^2 + \sigma_{zx}^2) \tag{A1.26a}$$

As σ_m and σ_r are stress invariants, so is σ'_r. From Figure A1.2(a) $\sigma'^2_r = \sigma_r^2 - 3\sigma_m^2$. From this, equation (A1.4) and similar manipulations as in equations (A1.6) to (A1.7), the yield criterion becomes

$$\begin{aligned}&2\bar{\sigma}^2 \equiv 3(\sigma_r^2 - 3\sigma_m^2) \equiv 3(\sigma'^2_{xx} + \sigma'^2_{yy} + \sigma'^2_{zz}) + 6(\sigma^2_{xy} + \sigma^2_{yz} + \sigma^2_{zy}) \equiv \\ &(\sigma_{xx} - \sigma_{yy})^2 + (\sigma_{yy} - \sigma_{zz})^2 + (\sigma_{zz} - \sigma_{xx})^2 + 6(\sigma^2_{xy} + \sigma^2_{yz} + \sigma^2_{zy}) = 6k^2 \text{ or } 2Y^2\end{aligned} \tag{A1.26b}$$

which is the same as equation (A1.12) of Section A1.1.

Strain transformations

The strain increments also transform as a tensor:

$$\mathrm{d}\varepsilon^*_{ij} = a_{ik}a_{jl}\mathrm{d}\varepsilon_{kl} \tag{A1.27}$$

It follows, as for stress, that the resultant strain increment and the equivalent strain increment are invariants of the strain. The extension of the definition of resultant strain to a general strain state is

$$\mathrm{d}\varepsilon_r^2 = \mathrm{d}\varepsilon_{xx}^2 + \mathrm{d}\varepsilon_{yy}^2 + \mathrm{d}\varepsilon_{zz}^2 + 2(\mathrm{d}\varepsilon_{xy}^2 + \mathrm{d}\varepsilon_{yz}^2 + \mathrm{d}\varepsilon_{zx}^2) \tag{A1.28}$$

where, as in equation (A1.13), $\mathrm{d}\varepsilon_{xy} = \mathrm{d}\varepsilon_{yx} = (1/2)(\partial u/\partial \mathrm{y} + \partial v/\partial \mathrm{x})$ and similarly for $\mathrm{d}\varepsilon_{yz}$ and $\mathrm{d}\varepsilon_{zx}$. Equivalent strain increments are $\sqrt{(2/3)}$ times resultant strain increments.

1.3.2 Further developments

The repeated suffix notation may be used to write the plastic flow rules (equation (A1.13)) more compactly and to express various relations between changes in equivalent stress and the deviatoric stress components that will be of use in Section A1.4. First, from equation (A1.13),

$$\mathrm{d}\varepsilon_{ij} = \frac{3}{2}\,\sigma'_{ij}\,\frac{\mathrm{d}\bar{\varepsilon}}{\bar{\sigma}} \qquad \text{or} \qquad \frac{3}{2}\,\frac{\sigma'_{ij}}{H'\bar{\sigma}}\,\mathrm{d}\bar{\sigma} \tag{A1.29}$$

The dependence of $\mathrm{d}\bar{\sigma}$ on its components σ'_{kl} is

$$\mathrm{d}\bar{\sigma} = \frac{\partial\bar{\sigma}}{\partial\sigma'_{kl}}\,\mathrm{d}\sigma'_{kl} \tag{A1.30}$$

From inspection of equation (A1.26b),

$$\frac{\partial \bar{\sigma}}{\partial \sigma'_{kl}} = \frac{3}{2} \frac{\sigma'_{kl}}{\bar{\sigma}} \tag{A1.31}$$

Equation (A1.31) substituted into (A1.30) gives the first equation of, and multiplying it by σ'_{kl} gives the second equation of

$$\mathrm{d}\bar{\sigma} = \frac{3}{2} \frac{\sigma'_{kl}}{\bar{\sigma}} \mathrm{d}\sigma'_{kl}; \qquad \sigma'_{kl} \frac{\partial \bar{\sigma}}{\partial \sigma'_{kl}} = \frac{3}{2} \frac{\sigma'_{kl}\sigma'_{kl}}{\bar{\sigma}} = \bar{\sigma} \tag{A1.32}$$

A1.4 Constitutive equations for numerical modelling

In Sections A1.1 and A1.2, the flow rules are developed for a rigid plastic material, relating strain increments to deviatoric stresses. Most plasticity calculations, however, are formulated to obtain stresses from strain increments (or strain rates). The flow rules must be inverted. This section reviews how this may be done, first for a rigid plastic material, then for combined elastic and plastic deformation.

A1.4.1. Rigid plastic flow rule inversion

Rearranging the first of equations (A1.29) gives

$$\sigma'_{ij} = \frac{2}{3} \frac{\bar{\sigma}}{\mathrm{d}\bar{\varepsilon}} \mathrm{d}\varepsilon_{ij} \tag{A1.33}$$

In principle, this can be used directly to determine the deviatoric stress field from a given strain increment field. However, the right-hand side is non-linear in the strain increments as both $\bar{\sigma}$ and $\mathrm{d}\bar{\varepsilon}$ depend on them. Practical rigid plastic finite element methods (Section A1.5) seek actual strain increment fields from initial guesses, by iteration. For efficiency of operation, they use a linearized form of equation (A1.33), to calculate variations of σ'_{ij} caused by variations in strain increment. They must also find some way to estimate hydrostatic stresses, not involved in yielding. Both these practical matters are introduced in this section.

Linearization

The sensitivity of σ'_{ij} to variations in strain increments about a base value $(\mathrm{d}\varepsilon_{ij})_0$ may be expressed, to a first approximation, in a linear way as

$$\mathrm{d}\sigma'_{ij} = \frac{\partial \sigma'_{ij}}{\partial(\mathrm{d}\varepsilon_{kl})} \mathrm{d}(\mathrm{d}\varepsilon_{kl}) = \frac{2}{3} \frac{\bar{\sigma}_0}{\mathrm{d}\bar{\varepsilon}_0} \mathrm{d}(\mathrm{d}\varepsilon_{ij}) + \frac{2}{3} \frac{\partial \bar{\sigma}}{\partial(\mathrm{d}\bar{\varepsilon})} \frac{\partial(\mathrm{d}\bar{\varepsilon})}{\partial(\mathrm{d}\varepsilon_{kl})} \frac{\mathrm{d}(\mathrm{d}\varepsilon_{kl})(\mathrm{d}\varepsilon_{ij})_0}{\mathrm{d}\bar{\varepsilon}_0} - \frac{2}{3} \frac{\bar{\sigma}_0}{(\mathrm{d}\bar{\varepsilon}_0)^2} \frac{\partial(\mathrm{d}\bar{\varepsilon})}{\partial(\mathrm{d}\varepsilon_{kl})} \mathrm{d}(\mathrm{d}\varepsilon_{kl})(\mathrm{d}\varepsilon_{ij})_0 \tag{A1.34}$$

For a material whose flow stress depends only on $\bar{\varepsilon}$, for example whose flow stress does not depend on strain rate,

$$\frac{\partial \bar{\sigma}}{\partial(\mathrm{d}\bar{\varepsilon})} = 0 \tag{A1.35}$$

Furthermore, from equation (A1.28) and by analogy with equation (A1.31),

$$\frac{\partial(\mathrm{d}\bar{\varepsilon})}{\partial(\mathrm{d}\varepsilon_{kl})} = \frac{2}{3}\,\frac{(\mathrm{d}\varepsilon_{kl})_0}{(\mathrm{d}\bar{\varepsilon})_0} \tag{A1.36}$$

Substituting equations (A1.35) and (A1.36) into equation (A1.34), and slightly rearranging, gives

$$\mathrm{d}\sigma'_{ij} = \frac{2}{3}\,\frac{\bar{\sigma}_0}{\mathrm{d}\bar{\varepsilon}_0}\,\mathrm{d}(\mathrm{d}\varepsilon_{ij}) - \frac{4}{9}\,\frac{\bar{\sigma}_0}{(\mathrm{d}\bar{\varepsilon}_0)^3}\,(\mathrm{d}\varepsilon_{ij})_0(\mathrm{d}\varepsilon_{kl})_0\mathrm{d}(\mathrm{d}\varepsilon_{kl}) \tag{A1.37a}$$

Hydrostatic stresses

Hydrostatic stress terms do not arise in rigid plasticity theory due to the assumption of incompressibility. If a small amount of compressibility is introduced, hydrostatic stress values may be calculated without seriously altering the magnitudes of the deviatoric stresses. Following Osakada *et al*. (1982), suppose a material is slightly plastically compressible, so that equation (A1.9) is added to by

$$\mathrm{d}\varepsilon_{\mathrm{v}} = g(\mathrm{c}\sigma_{\mathrm{m}}) \tag{A1.38a}$$

where $\mathrm{d}\varepsilon_{\mathrm{v}} = (\mathrm{d}\varepsilon_1 + \mathrm{d}\varepsilon_2 + \mathrm{d}\varepsilon_3)$ and g is a small positive constant; and introduce modified definitions of equivalent stress and equivalent strain (where now, because of compressibility, one must distinguish between total strain increment and deviatoric strain increment components)

$$\bar{\sigma}^2 = \frac{3}{2}\,(\sigma'_{ij}\sigma'_{ji} + g\sigma_{\mathrm{m}}^2); \qquad \mathrm{d}\bar{\varepsilon}^2 = \frac{2}{3}\left(\mathrm{d}\varepsilon'_{ij}\mathrm{d}\varepsilon'_{ji} + \frac{\mathrm{d}\varepsilon_{\mathrm{v}}^2}{g}\right) \tag{A1.38b}$$

then by the same procedure that led to equation (A1.11), c remains equal to (3/2) $(\mathrm{d}\bar{\varepsilon}/\bar{\sigma})$. Then, noting that $\sigma_{ij} = \sigma'_{ij} + \delta_{ij}\sigma_{\mathrm{m}}$ and $\mathrm{d}\varepsilon_{ij} \equiv \mathrm{d}\varepsilon'_{ij} + \delta_{ij}\mathrm{d}\varepsilon_{\mathrm{v}}/3$,

$$\sigma_{ij} = \sigma'_{ij} + \delta_{ij}\sigma_{\mathrm{m}} = \frac{2\bar{\sigma}}{3\mathrm{d}\bar{\varepsilon}}\left(\mathrm{d}\varepsilon'_{ij} + \delta_{ij}\,\frac{\mathrm{d}\varepsilon_{\mathrm{v}}}{g}\right) \equiv \frac{2\bar{\sigma}}{3\mathrm{d}\bar{\varepsilon}}\left(\mathrm{d}\varepsilon_{ij} + \delta_{ij}\left(\frac{1}{g} - \frac{1}{3}\right)\mathrm{d}\varepsilon_{\mathrm{v}}\right) \tag{A1.39}$$

Equation (A1.39) is identical in form to equation (A1.33) but for a modified strain increment term. The linearization leading to equation (A1.37a) for deviatoric stress components may be repeated to produce the equivalent equation for total stress components. In addition to equation (A1.37a),

$$\mathrm{d}\sigma_{\mathrm{m}} = \frac{2}{3}\left(\frac{1}{g} - \frac{1}{3}\right)(\mathrm{d}\varepsilon_{\mathrm{v}})_0\left\{\frac{\bar{\sigma}_0}{\mathrm{d}\bar{\varepsilon}_0}\,\frac{\mathrm{d}(\mathrm{d}\varepsilon_{\mathrm{v}})}{(\mathrm{d}\varepsilon_{\mathrm{v}})_0} - \frac{2}{3}\,\frac{\bar{\sigma}_0}{(\mathrm{d}\bar{\varepsilon}_0)^3}\,(\mathrm{d}\varepsilon_{kl})_0\mathrm{d}(\mathrm{d}\varepsilon_{kl})\right\} \tag{A1.37b}$$

Osakada *et al*. (1982) have found that choosing g in the range 0.01 to 0.0001 gives a practically usable equation.

A1.4.2 Elastic–plastic flow rules

Introducing elastic deformation creates advantages and disadvantages relative to rigid plastic modelling. There is no problem with respect to including hydrostatic stress effects (the material is elastically compressible). However, immediate linearization of the flow rules, after the manner of equation (A1.33), is no longer possible. First, one must separate the elastic from the total strain increment, to isolate the plastic strain increment.

Elastic strains

Consider the geometrical representation of the stress state in Figure A1.2. If the point P lies within the yield cylinder, only elastic strains occur. The relation between strain and stress is commonly given as (Hooke's law)

$$\left.\begin{aligned}
\varepsilon_{xx} &= \frac{1}{E}\left[\sigma_{xx} - \nu(\sigma_{yy} + \sigma_{zz})\right] \\
\varepsilon_{yy} &= \frac{1}{E}\left[\sigma_{yy} - \nu(\sigma_{zz} + \sigma_{xx})\right] \\
\varepsilon_{zz} &= \frac{1}{E}\left[\sigma_{zz} - \nu(\sigma_{xx} + \sigma_{yy})\right] \\
\varepsilon_{xy} = \varepsilon_{yx} &= \frac{1+\nu}{E}\sigma_{xy}; \qquad \varepsilon_{yz} = \varepsilon_{zy} = \frac{1+\nu}{E}\sigma_{yz}; \qquad \varepsilon_{zx} = \varepsilon_{xz} = \frac{1+\nu}{E}\sigma_{xz}
\end{aligned}\right\} \tag{A1.40}$$

where E and ν are Young's modulus and Poisson's ratio. The strains have deviatoric and hydrostatic parts ε' and ε_{m} ($\varepsilon_{\mathrm{m}} = \varepsilon_{\mathrm{v}}/3$). More compactly, in tensor notation

$$\varepsilon_{ij} \equiv \varepsilon'_{ij} + \delta_{ij}\varepsilon_{\mathrm{m}} = \frac{1+\nu}{E}\sigma'_{ij} + \delta_{ij}\frac{1-2\nu}{E}\sigma_{\mathrm{m}} \tag{A1.41}$$

The inversion of this relation is

$$\sigma_{ij} \equiv \sigma'_{ij} + \delta_{ij}\sigma_{\mathrm{m}} = \frac{E}{1+\nu}\varepsilon'_{ij} + \delta_{ij}\frac{E}{1-2\nu}\varepsilon_{\mathrm{m}} \tag{A1.42a}$$

or in incremental terms

$$\mathrm{d}\sigma_{ij} \equiv \mathrm{d}\sigma'_{ij} + \delta_{ij}\mathrm{d}\sigma_{\mathrm{m}} = \frac{E}{1+\nu}\mathrm{d}\varepsilon'_{ij} + \delta_{ij}\frac{E}{1-2\nu}\mathrm{d}\varepsilon_{\mathrm{m}} \tag{A1.42b}$$

Plastic and elastic strains

When the point P lies on the yield cylinder (Figure A1.2), and a stress increment causes further yielding, the total strain change has an elastic part proportional to it, and a plastic part normal to the yield locus (Figure A1.3 (a)):

$$(\mathrm{d}\varepsilon_{ij})_{\text{total}} \equiv (\mathrm{d}\varepsilon'_{ij})_{\text{total}} + (\delta_{ij}\mathrm{d}\varepsilon_{\mathrm{m}})_{\text{elastic}} \equiv (\mathrm{d}\varepsilon'_{ij})_{\text{plastic}} + (\mathrm{d}\varepsilon'_{ij})_{\text{elastic}} + (\delta_{ij}\mathrm{d}\varepsilon_{\mathrm{m}})_{\text{elastic}}$$
$$= \frac{3}{2}\frac{\mathrm{d}\bar{\sigma}}{H'\bar{\sigma}}\sigma'_{ij} + \frac{1+\nu}{E}\mathrm{d}\sigma'_{ij} + \delta_{ij}\frac{1-2\nu}{E}\mathrm{d}\sigma_{\mathrm{m}} \tag{A1.43}$$

The complexity in inverting this flow rule is caused by the presence in the right-hand side of both total stress and stress increment terms.

Elastic–plastic flow rule inversion

The elastic deviatoric strain increment of equation (A1.43) is the difference between the total deviatoric and the plastic strain increment. It causes a deviatoric stress increment given by the deviatoric part of equation (A1.42b):

$$\mathrm{d}\sigma'_{ij} = \frac{\mathrm{E}}{1+\nu}\left[(\mathrm{d}\varepsilon'_{ij})_{\text{total}} - \frac{3}{2}\frac{\mathrm{d}\bar{\sigma}}{H'\bar{\sigma}}\sigma'_{ij}\right] \tag{A1.44}$$

This may be substituted into the first of equations (A1.32) to give

$$\mathrm{d}\bar{\sigma} = \frac{3}{2}\frac{E}{1+\nu}\left[\frac{\sigma'_{kl}}{\bar{\sigma}}(\mathrm{d}\varepsilon'_{kl})_{\text{total}} - \frac{3}{2}\frac{\mathrm{d}\bar{\sigma}}{H'\bar{\sigma}}\frac{\sigma'_{kl}\sigma'_{kl}}{\bar{\sigma}}\right] \tag{A1.45a}$$

After simplifying, by means of the second of equations (A1.32), and rearranging, an expression for $\mathrm{d}\bar{\sigma}$ is found:

$$\mathrm{d}\bar{\sigma} = \frac{\frac{3}{2}\frac{E}{1+\nu}\sigma'_{kl}(\mathrm{d}\varepsilon'_{kl})_{\text{total}}}{\bar{\sigma}\left(1+\frac{3}{2}\frac{E}{1+\nu}\frac{1}{H'}\right)} \tag{A1.45b}$$

Finally, substituting this back into the elastic deviatoric stress equation (A1.44) and adding the hydrostatic stress term (equation (A1.42b)), the total stress increment becomes (after dropping the subscript $_{\text{total}}$, so $\mathrm{d}\varepsilon'_{kl}$ is the total deviatoric strain)

$$\mathrm{d}\sigma_{ij} = \frac{E}{1+\nu}\mathrm{d}\varepsilon'_{ij} + \delta_{ij}\frac{E}{1-2\nu}\mathrm{d}\varepsilon_{\mathrm{m}} - \frac{\frac{9}{4}\left(\frac{E}{1+\nu}\right)^2\sigma'_{ij}\sigma'_{kl}\mathrm{d}\varepsilon'_{kl}}{\bar{\sigma}^2\left(H' + \frac{3}{2}\frac{E}{1+\nu}\right)} \tag{A1.46}$$

Because the sum of deviatoric stress terms $\delta_{kl}\sigma'_{kl}$ is zero, the total deviatoric strain increment $\mathrm{d}\varepsilon'_{kl}$ in the last term may be replaced by the total strain increment $\mathrm{d}\varepsilon_{kl}$.

A1.4.3 Matrix notation

Tensor notation with the summation convention enables the most compact writing and analysis of relations between stress and strain. When it comes to applying the results, a

different representation is more useful. The stress tensor σ_{ij} has nine components but, because $\sigma_{ij} = \sigma_{ji}$, only six are independent. The same applies to the strain tensor. The six independent stress components, regarded as a vector, may be obtained from the six independent strain components, also regarded as a vector, by matrix multiplication:

$$\begin{Bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sigma_{xy} \\ \sigma_{yz} \\ \sigma_{zx} \end{Bmatrix} = \begin{bmatrix} \quad \mathbf{D} \quad \end{bmatrix} \begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \\ \varepsilon_{xy} \\ \varepsilon_{yz} \\ \varepsilon_{zx} \end{Bmatrix} \quad \text{or} \quad \{\sigma\} = [\mathbf{D}]\{\varepsilon\} \tag{A1.47}$$

Here, [**D**] is a 6×6 matrix. The values of its elements (as well as the detail of whether the vectors should be stress or stress increment, strain or strain increment) depend on whether the relation between stress and strain is elastic, elastic–plastic, or rigid–plastic: they can be found from equations (A1.42), (A1.46) or (A1.37) as appropriate.

Elastic conditions

Equation (A1.47) can be written either in total or increment form:

$$\{\sigma\} = [\mathbf{D}^{\mathrm{e}}]\{\varepsilon\} \qquad \text{or} \qquad \{\mathrm{d}\sigma\} = [\mathbf{D}^{\mathrm{e}}]\{\mathrm{d}\varepsilon\} \tag{A1.48a}$$

where, from equation (A1.42)

$$[\mathbf{D}^{\mathrm{e}}] = \frac{E}{1+\nu} \begin{bmatrix} \dfrac{1-\nu}{1-2\nu} & \dfrac{\nu}{1-2\nu} & \dfrac{\nu}{1-2\nu} & 0 & 0 & 0 \\ \dfrac{\nu}{1-2\nu} & \dfrac{1-\nu}{1-2\nu} & \dfrac{\nu}{1-2\nu} & 0 & 0 & 0 \\ \dfrac{\nu}{1-2\nu} & \dfrac{\nu}{1-2\nu} & \dfrac{1-\nu}{1-2\nu} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{A1.48b}$$

(γ_{ij} can be used instead of ε_{ij} for shear strains; then the diagonal coefficients equal to 1 above are replaced by 1/2.)

Elastic–plastic conditions

Equation (A1.47) must be used in incremental (or rate) form:

$$\{\mathrm{d}\sigma\} = [\mathbf{D}^{\mathrm{e-p}}]\{\mathrm{d}\varepsilon\} \qquad \text{or} \qquad \{\dot{\sigma}\} = [\mathbf{D}^{\mathrm{e-p}}]\{\dot{\varepsilon}\} \tag{A1.49a}$$

where, from equation (A1.46), after noting that the shear modulus $G = 0.5E/(1 + \nu)$

$$[\mathbf{D}^{e-p}] = [\mathbf{D}^{e}] - \frac{9G^2}{\bar{\sigma}^2(H' + 3G)} \begin{bmatrix} \sigma'_{xx}\sigma'_{xx} & \sigma'_{xx}\sigma'_{yy} & \sigma'_{xx}\sigma'_{zz} & 2\sigma'_{xx}\sigma'_{xy} & 2\sigma'_{xx}\sigma'_{yz} & 2\sigma'_{xx}\sigma'_{zx} \\ \sigma'_{yy}\sigma'_{xx} & \sigma'_{yy}\sigma'_{yy} & \sigma'_{yy}\sigma'_{zz} & 2\sigma'_{yy}\sigma'_{xy} & 2\sigma'_{yy}\sigma'_{yz} & 2\sigma'_{yy}\sigma'_{zx} \\ \sigma'_{zz}\sigma'_{xx} & \sigma'_{zz}\sigma'_{yy} & \sigma'_{zz}\sigma'_{zz} & 2\sigma'_{zz}\sigma'_{xy} & 2\sigma'_{zz}\sigma'_{yz} & 2\sigma'_{zz}\sigma'_{zx} \\ \sigma'_{xy}\sigma'_{xx} & \sigma'_{xy}\sigma'_{yy} & \sigma'_{xy}\sigma'_{zz} & 2\sigma'_{xy}\sigma'_{xy} & 2\sigma'_{xy}\sigma'_{yz} & 2\sigma'_{xy}\sigma'_{yz} \\ \sigma'_{yz}\sigma'_{xx} & \sigma'_{yz}\sigma'_{yy} & \sigma'_{yz}\sigma'_{zz} & 2\sigma'_{yz}\sigma'_{xy} & 2\sigma'_{yz}\sigma'_{yz} & 2\sigma'_{yz}\sigma'_{zx} \\ \sigma'_{zx}\sigma'_{xx} & \sigma'_{zx}\sigma'_{yy} & \sigma'_{zx}\sigma'_{zz} & 2\sigma'_{zx}\sigma'_{xy} & 2\sigma'_{zx}\sigma'_{yz} & 2\sigma'_{zx}\sigma'_{zx} \end{bmatrix} \tag{A1.49b}$$

If γ_{ij} instead of ε_{ij} is used for the shear strains the factors 2 are replaced by 1.

Rigid–plastic conditions

The basic relation between stress and strain increment, equation (A1.39), for a base set of stresses σ_o and strain increments $d\varepsilon_o$ leads to

$$\{\sigma_o\} = [\mathbf{D}_o^{r-p}]\{d\varepsilon_o\} \tag{A1.50a}$$

with

$$[\mathbf{D}_o^{r-p}] = \frac{2\bar{\sigma}_o}{3d\bar{\varepsilon}_o} \begin{bmatrix} 1+a & a & a & 0 & 0 & 0 \\ a & 1+a & a & 0 & 0 & 0 \\ a & a & 1+a & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \text{ where } a = \frac{1}{3}\left(\frac{1}{g} - \frac{1}{3}\right) \tag{A1.50b}$$

Linearization about the base σ_o and $d\varepsilon_o$ leads to

$$\{d\sigma\} = [\mathbf{D}^{r-p}]\{d(d\varepsilon)\} \tag{A1.51}$$

where $[D^{r-p}]$ can be extracted from equations (A1.37a) and (A1.37b).

A1.5 Finite element formulations

The basic ideas of the finite element method are given in many general texts, for example Zienkiewicz (1989), as well as in Kobayashi *et al*. (1989). A continuum throughout which the solution to some problem is required is divided into an assembly of finite-sized elements, filling the continuum without leaving any gaps. Each element is identified by the positions of its nodes. The nodes are the vertices of the elements, and for some elements additional points too.

Instead of seeking an exact solution to a problem, over the whole continuum, an approximate solution is sought at the positions of the nodes, with some form of interpolation of the solution between the nodes. First, the governing equations of a problem are applied to each element alone, to obtain relations between the problem variables at the element nodes. These element nodal equations are then assembled to describe the whole continuum. The global assembly of all equations is finally solved numerically.

In mechanics problems, the nodal displacements, changes in displacements, or velocities, $\{u\}$,$\{du\}$ or $\{\dot{u}\}$, are usually the unknowns. They cause element strains, strain increments or

strain rates $\{\varepsilon\}$, $\{d\varepsilon\}$ or $\{\dot{\varepsilon}\}$, which depend on the element geometry. Transformation of the nodal displacement quantities to the element strain quantities may be carried out by a matrix multiplication operation: the transformation matrix is known as the [**B**] matrix. Once element strain expressions are created, they may further be transformed to stress quantities by operating with the [**D**] matrix described in the previous section. Finally, once element stress expressions are created, external force quantities on the element's nodes may be obtained either by direct consideration of force equilibrium or by virtual work methods. For the special cases of 3-node triangular or 4-node tetrahedral elements (for which stress and strain quantities are constant throughout an element), the transformation from stress to force is achieved by multiplication by the product of the element's volume, V_e, and the transpose of the [**B**] matrix; the chain of activities that relates nodal displacement and force quantities may be summarized as

$$\begin{matrix} & [\mathbf{B}] & & [\mathbf{D}] & & V_e[\mathbf{B}]^T & \\ \{u, du, \text{ or } \dot{u}\} & \Rightarrow & \{\varepsilon, d\varepsilon, \text{ or } \dot{\varepsilon}\} & \Rightarrow & \{\sigma, d\sigma, \text{ or } \dot{\sigma}\} & \Rightarrow & \{F, dF, \text{ or } \dot{F}\} \end{matrix}$$

and the resulting finite element assembly of equations to be solved becomes

$$\{F, dF \text{ or } \dot{F}\} = [\mathbf{K}]\{u, du \text{ or } \dot{u}\} \text{ where } [\mathbf{K}] = V_e[\mathbf{B}]^T[\mathbf{D}][\mathbf{B}] \tag{A1.52}$$

In general, for any shape of finite element, $[\mathbf{K}] = \int_{V_e}[\mathbf{B}]^T[\mathbf{D}][\mathbf{B}]dV$. [**K**] is known as the stiffness matrix and equation (A1.52) as the stiffness equation.

It is not the purpose of this appendix to develop all aspects of the finite element method applied to metal machining problems, but only to indicate differences that arise from the differences in the [**D**] matrix between elastic, elastic–plastic and rigid–plastic formulations.

Elastic conditions

The coefficients of $[\mathbf{D}^e]$, equation (A1.48b), are material constants. There is therefore a linear relation between the nodal forces and displacements, and force and displacement change. From equation (A1.48a),

$$\{F\} = [\mathbf{K}]\{u\} \quad \text{or} \quad \{dF\} = [\mathbf{K}]\{du\} \quad \text{or} \quad \{\dot{F}\} = [\mathbf{K}]\{\dot{u}\} \tag{A1.53}$$

Elastic–plastic conditions

The coefficients of $[\mathbf{D}^{e-p}]$, (equation (A1.49b)) include deviatoric stress terms. Integration of equation (A1.49a) to obtain current total stresses and strains depends on the path by which the current state has been reached. Thus, after creating, from equation (A1.49a),

$$\{dF\} = [\mathbf{K}]\{du\} \quad \text{or} \quad \{\dot{F}\} = [\mathbf{K}]\{\dot{u}\} \tag{A1.54}$$

displacement increments are calculated in steps along the problem's loading path. After each increment has been obtained, the accompanying strain and stress increments are calculated. The stress increments are added to the stresses that existed at the start of the step, to update $[\mathbf{D}^{e-p}]$ and hence [**K**]. The new value is used for the next step. The non-linearity of the calculation requires each step to be very small. Elastic–plastic calculations for large strain problems, as in metal machining, are inherently lengthy and time consuming.

Rigid–plastic conditions

Larger steps than in the elastic–plastic case can be taken with the rigid–plastic formulation; and hence the computing effort is less. However, it is not possible to simulate

some aspects of chip formation, for example the elastic contact region of the chip/tool contact.

Suppose that over some time interval dt, the velocity of a plastic flow is guessed to be $\{\dot{u}_o\}$, and a better guess is $\{\dot{u}_o + \delta\dot{u}\}$. The stiffness equation for the better guess is

$$\{F\} = [\mathbf{K}_o]\{\dot{u}_o\}\mathrm{d}t + [\mathbf{K}]\{\delta\dot{u}\}\mathrm{d}t \tag{A1.55}$$

where $[\mathbf{K}_o] = V_e[\mathbf{B}]^T[\mathbf{D}_o^{r-p}][\mathbf{B}]$ and $[\mathbf{K}] = V_e[\mathbf{B}]^T[\mathbf{D}^{r-p}][\mathbf{B}]$. $[\mathbf{D}_o^{r-p}]$ and $[\mathbf{D}^{r-p}]$ are given by equations (A1.50b) and (A1.51) (from equations (A1.37a and (b)). Their coefficients are known in terms of the initial guess.

Equation (A1.55) may be rearranged to

$$\{F\} - [\mathbf{K}_o]\{\dot{u}_o\}\mathrm{d}t = [\mathbf{K}]\{\delta\dot{u}\}\mathrm{d}t \tag{A1.56}$$

The left-hand side is known in terms of the applied forces and the initial guess of velocities. Equation (A1.56) may be solved for the unknown velocities $\{\delta\dot{u}\}$. If they are significant relative to $\{\dot{u}_o\}$, they may be added to $\{\dot{u}_o\}$ to create a better initial guess; and the cycle of calculation is repeated until $\{\delta\dot{u}\}$ becomes negligibly small.

This approach to calculating flows, and hence stresses, in plastic problems, ignoring elastic deformation, with the modification of the yield criterion and flow rules to include a small amount of compressibility, to enable hydrostatic stresses to be calculated, follows Osakada *et al*. (1982). It has been chosen because of the easy physical interpretation that can be given to the method of introducing hydrostatic stresses. Other methods, based on Lagrange multipliers and penalty functions (for example Kobayashi *et al*., 1989 and Zienciewicz and Godbole, 1975) give the same results.

References

Dewhurst, P. (1978) On the non-uniqueness of the machining process. *Proc. Roy. Soc. Lond.* **A360**, 587–610.

Hill, R. (1950) *Plasticity*. Oxford: Clarendon Press.

Hill, R. (1954) On the limits set by plastic yielding to the intensities of singularities of stress. *J. Mech. Phys. Solids,* **2**, 278–285.

Kobayashi, S., Oh, S-I. and Altan, T. (1989) *Metal Forming and the Finite Element Method*. New York: Oxford University Press.

Kudo, H. (1965) Some new slip-line solutions for two-dimentional steady state machining. *Int. J. Mech. Sci.* **7**, 43–55.

Johnson, W. and Mellor, P. B. (1973) *Engineering Plasticity*. London: van Nostrand.

Johnson, W., Sowerby, R. and Venter, R. D. (1982) *Plane-Strain Slip-Line Fields for Metal-Deformation Processes*. Oxford: Pergamon Press.

Lee, E. H. and Shaffer, B. W. (1951) The theory of plasticity applied to a problem of machining. *Trans. ASME J. Appl. Mech.* **18**, 405–413.

Osakada, K., Nakano, N. and Mori, K. (1982) Finite element method for rigid plastic analysis of metal forming. *Int. J. Mech. Sci.* **24**, 459–468.

Prager, W. and Hodge, P. G. (1951) *The Theory of Perfectly Plastic Solids*. New York: Wiley.

Thompsen, E. G., Yang, C. T. and Kobayashi, S. (1965) *Mechanics of Plastic Deformation in Metal Processing*. New York: MacMillan.

Zienkiewicz, O. C. (1989) *The Finite Element Method*, 4th edn. London: McGraw-Hill.

Zienkiewicz, O. C. and Godbole, P. N. (1975) A penalty function approach to problems of plastic flow of metals with large surface deformation. *J. Strain Analysis* **10**, 180–183.

Appendix 2

Conduction and convection of heat in solids

This appendix supports the thermal aspects of Chapters 2, 6 and subsequent chapters. A more complete description of heat transport in solids is given in Carslaw and Jaeger (1959). The basic law of heat conduction in an isotropic material is assumed; namely that the rate q of heat transfer per unit area normal to an isothermal surface is proportional to the temperature gradient in that direction and with K the thermal conductivity and T the temperature:

$$q = -K\frac{\partial T}{\partial n} \tag{A2.1}$$

A2.1 The differential equation for heat flow in a solid

Figure A2.1(a) shows a control volume dx.dy.dz fixed relative to axes $Oxyz$. A solid (density ρ and specific heat C – heat capacity ρC – and with thermal conductivity that varies with temperature) passes through it with velocity $\dot{u}_z$. The differential equation relating temperature in the solid to position and time is first derived for the special case when a temperature gradient exists only in the z-direction.

Consider the heat flow into and out of the control volume through the two surfaces of area dxdy, at $z = z$ and $z + \mathrm{d}z$. The heat accumulating in time dt due to conduction, H_{cond}, is (equation (A2.1) with n as z, allowing K and $\partial T/\partial z$ to vary with z):

$$\begin{aligned} H_{\mathrm{cond}} &= \left\{-K\frac{\partial T}{\partial z} + \left(K + \frac{\mathrm{d}K}{\mathrm{d}T}\frac{\partial T}{\partial z}\mathrm{d}z\right)\left(\frac{\partial T}{\partial z} + \left(\frac{\partial^2 T}{\partial z^2}\right)\mathrm{d}z\right)\right\}\mathrm{d}x\mathrm{d}y\,\mathrm{d}t \\ &\cong \left(K\frac{\partial^2 T}{\partial z^2} + \frac{\mathrm{d}K}{\mathrm{d}T}\left(\frac{\partial T}{\partial z}\right)^2\right)\mathrm{d}x\mathrm{d}y\,\mathrm{d}z\,\mathrm{d}t \end{aligned} \tag{A2.2a}$$

The heat accumulating due to convection, H_{conv}, is

$$H_{\mathrm{conv}} = \left\{\dot{u}_z\rho CT - \dot{u}_z\rho C\left(T + \frac{\partial T}{\partial z}\mathrm{d}z\right)\right\}\mathrm{d}x\mathrm{d}y\mathrm{d}t \cong -\dot{u}_z\rho C\frac{\partial T}{\partial z}\mathrm{d}x\mathrm{d}y\mathrm{d}z\,t \tag{A2.2b}$$

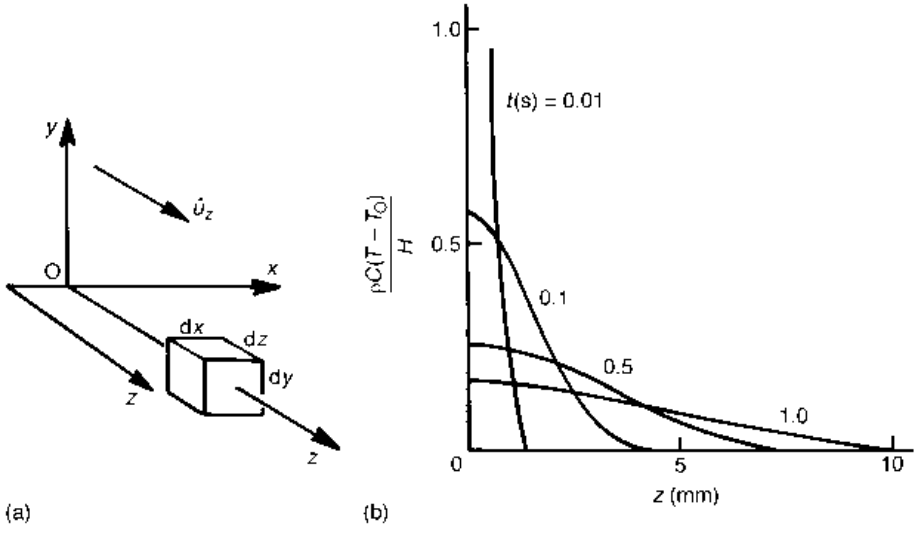


Fig. A2.1 (a) A control volume for temperature analysis and (b) dependence of temperature on position and time for the example of Section A2.2.1 ($\kappa = 10$ mm^2/s)

Internal heat generation at a rate q^* per unit volume causes an accumulation, H_{int}:

$$H_{\text{int}} = q^* \mathrm{d}x\mathrm{d}y\mathrm{d}z\mathrm{d}t \tag{A2.2c}$$

Equating the sum of the terms (equations (A2.2a) to (A2.2c)) to the product of temperature rise and heat capacity of the volume:

$$\rho C \frac{\partial \mathrm{T}}{\partial t} = K \frac{\partial^2 T}{\partial z^2} + \frac{\mathrm{d}K}{\mathrm{dT}} \left(\frac{\partial T}{\partial z} \right)^2 - \dot{u}_z \rho C \frac{\partial T}{\partial z} + q^* \tag{A2.3a}$$

The extension to three dimensions is straightforward:

$$\rho C \frac{\partial T}{\partial t} = K \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} \right) + \frac{\mathrm{d}K}{\mathrm{d}T} \left(\left(\frac{\partial T}{\partial x} \right)^2 + \left(\frac{\partial T}{\partial y} \right)^2 + \left(\frac{\partial T}{\partial z} \right)^2 \right) -$$
$$-\rho C \left(\dot{u}_x \frac{\partial T}{\partial x} + \dot{u}_y \frac{\partial T}{\partial y} + \dot{u}_z \frac{\partial T}{\partial z} \right) + q^* \tag{A2.3b}$$

When thermal conductivity does not vary with temperature, equation (A2.3b) reduces to

$$\rho C \frac{\partial T}{\partial t} = K \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} \right) - \rho C \left(\dot{u}_x \frac{\partial T}{\partial x} + \dot{u}_y \frac{\partial T}{\partial y} + \dot{u}_z \frac{\partial T}{\partial z} \right) + q^* \tag{A2.4}$$

A2.2 Selected problems, with no convection

When $\dot{u}_x = \dot{u}_y = \dot{u}_z = 0$, and $q^* = 0$ too, equation (A2.4) simplifies further, to

$$\frac{1}{\kappa}\frac{\partial T}{\partial t} = \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}\right) \tag{A2.5}$$

where the diffusivity κ equals $K/\rho C$. In this section, some solutions of equation (A2.5) are presented that give physical insight into conditions relevant to machining.

A2.2.1 The semi-infinite solid $z > 0$: temperature due to an instantaneous quantity of heat H per unit area into it over the plane $z = 0$, at $t = 0$; ambient temperature T_o

It may be checked by substitution that

$$T - T_0 = \frac{H}{\rho C}\frac{1}{\sqrt{\pi\kappa t}}\,\mathrm{e}^{-\frac{z^2}{4\kappa t}} \tag{A2.6}$$

is a solution of equation (A2.5). It has the property that, at $t = 0$, it is zero for all $z > 0$ and is infinite at $z = 0$. For $t > 0$, $\partial T/\partial z = 0$ at $z = 0$ and

$$\int_0^\infty \rho C(T - T_0)\mathrm{d}z = H \tag{A2.7}$$

Equation (A2.6) thus describes the temperature rise caused by releasing a quantity of heat H per unit area, at $z = 0$, instantaneously at $t = 0$; and thereafter preventing flow of heat across (insulating) the surface $z = 0$. Figure A2.1(b) shows for different times the dimensionless temperature $\rho C(T - T_0)/H$ for a material with $\kappa = 10$ mm^2/s, typical of metals. The increasing extent of the heated region with time is clearly seen.

At every time, the temperature distribution has the property that 84.3% of the associated heat is contained within the region $z/\sqrt{4\kappa t} < 1$. This result is obtained by integrating equation (A2.6) from $z = 0$ to $\sqrt{4\kappa t}$. Values of the error function erf p,

$$\mathrm{erf}\, p = \frac{2}{\sqrt{\pi}}\int_0^p \mathrm{e}^{-u^2}\mathrm{d}u \tag{A2.8}$$

that results are tabulated in Carslaw and Jaeger (1959). Physically, one can visualize the temperature front as travelling a distance $\approx \sqrt{4\kappa t}$ in time t. This is used in considering temperature distributions due to moving heat sources (Section A2.3.2).

A2.2.2 The semi-infinite solid $z > 0$: temperature due to supply of heat at a constant rate q per unit area over the plane $z = 0$, for $t > 0$; ambient temperature T_o

Heat $\mathrm{d}H = q\mathrm{d}t'$ is released at $z = 0$ in the time interval t' to $t' + \mathrm{d}t'$. The temperature rise that this causes at z at a later time t is, from equation (A2.6)

$$d(T - T_0) = \frac{q dt'}{\rho C} \frac{1}{(\pi\kappa(t - t'))^{1/2}} e^{-\frac{z^2}{4\kappa(t-t')}} \tag{A2.9}$$

The total temperature is obtained by integrating with respect to t' from 0 to t. The temperature at $z = 0$ will be found to be of interest. When q is independent of time

$$(T - T_0) = \frac{2}{\sqrt{\pi}} \frac{q}{K} \sqrt{\kappa t} \tag{A2.10}$$

The average temperature at $z = 0$, over the time interval 0 to t, is 2/3rds of this.

A2.2.3 The semi-infinite solid $z > 0$: temperature due to an instantaneous quantity of heat H released into it at the point $x = y = z = 0$, at $t = 0$; ambient temperature T_0

In this case of three-dimensional heat flow, the equivalent to equation (A2.6) is

$$T - T_0 = \frac{H}{4\rho C} \frac{1}{(\pi\kappa t)^{3/2}} e^{-\frac{x^2+y^2+z^2}{4\kappa t}} \tag{A2.11}$$

Equation (A2.11) is a building block for determining the temperature caused by heating over a finite area of an otherwise insulated surface, which is considered next.

A2.2.4 The semi-infinite solid $z > 0$: uniform heating rate q per unit area for $t > 0$, over the rectangle $-a < x < a$, $-b < y < b$ at $z = 0$; ambient temperature T_0

Heat flows into the solid over the surface area shown in Figure (A2.2a). In the time interval t' to $t' + dt'$, the quantity of heat dH that enters through the area $dA = dx'dy'$ at (x', y') is $q dA dt'$. From equation (A2.11) the contribution of this to the temperature at any point (x, y, z) in the solid at time t is

$$d(T - T_0) = \frac{q dx' dy' dt'}{4\rho C(\pi\kappa)^{3/2}(t - t')^{3/2}} e^{-\frac{(x-x')^2+(y-y')^2+z^2}{4\kappa(t-t')}} \tag{A2.12}$$

Integrating over time first, in the limit as t and t' approach infinity (the steady state),

$$d(T - T_0) = \frac{q}{2\pi K} \int_{-a}^{+a} \int_{-b}^{+b} \frac{dy'}{((x - x')^2 + (y - y')^2 + z^2)^{1/2}} dx' \tag{A2.13}$$

Details of the integration over area are given by Loewen and Shaw (1954). At the surface $z = 0$, the maximum temperature (at $x = y = 0$) and average temperature over the heat source are respectively

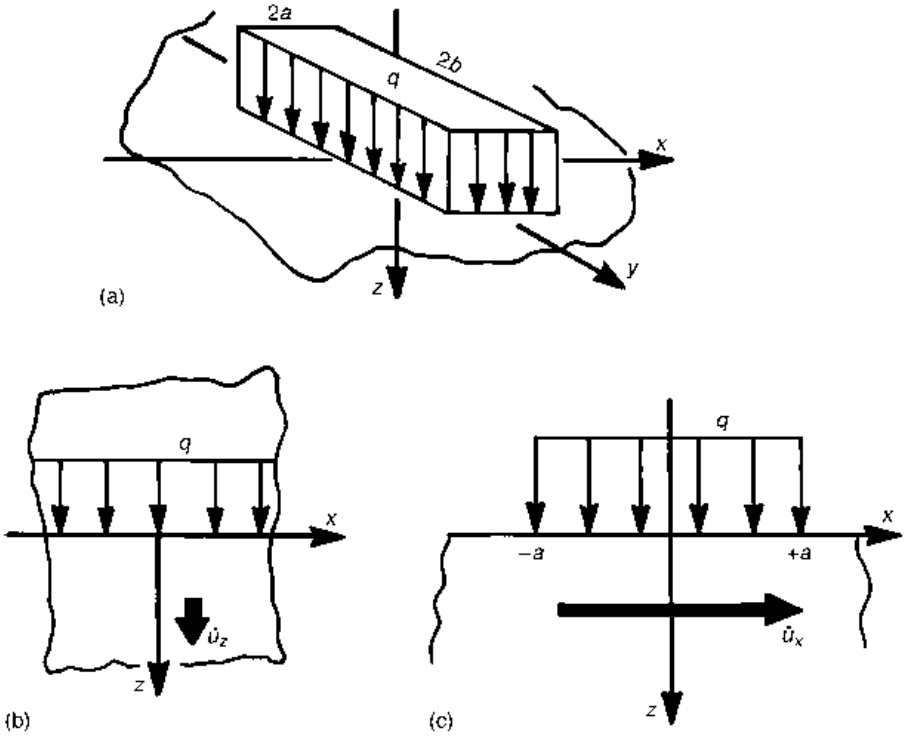


Fig. A2.2 Some problems relevant to machining: (a) surface heating of a stationary semi-infinite solid; (b) an infinite solid moving perpendicular to a plane heat source; (c) a semi-infinite solid moving tangentially to the plane of a surface heat source

$$\left.\begin{aligned}(T-T_0)_{\max} &= \frac{2qa}{\pi K}\left(\sinh^{-1}\frac{b}{a}+\frac{b}{a}\sinh^{-1}\frac{a}{b}\right)\\ (T-T_0)_{\mathrm{av}} &= (T-T_0)_{\max}-\frac{2qa}{3\pi K}\left[\left(\frac{a}{b}+\frac{b}{a}\right)\left(1+\frac{b^2}{a^2}\right)^{1/2}-\frac{b^2}{a^2}-\frac{a}{b}\right]\end{aligned}\right\} \tag{A2.14}$$

A2.3 Selected problems, with convection

Figures A2.2(b) and (c) show two classes of moving heat source problem. In Figure A2.2(b) heating occurs over the plane $z = 0$, and the solid moves with velocity $\dot{u}_z$ through the source. In Figure A2.2(c), heating also occurs over the plane $z = 0$, but the solid moves tangentially past the source, in this case with a velocity $\dot{u}_x$ in the x-direction.

A2.3.1 The infinite solid with velocity $\dot{u}_z$: steady heating at rate q per unit area over the plane $z = 0$ (Figure A2.2b); ambient temperature T_o

In the steady state, the form of equation (A2.4) (with $q^* = 0$) to be satisfied is

$$\kappa \frac{\partial^2 T}{\partial z^2} = \dot{u}_z \frac{\partial T}{\partial z} \tag{A2.15}$$

The temperature distribution

$$(T - T_0) = \frac{q}{\rho C \dot{u}_z}, \quad z \geq 0; \qquad (T - T_0) = \frac{q}{\rho C \dot{u}_z} \, e^{\frac{\dot{u}_z z}{\kappa}}, \quad z \leq 0 \tag{A2.16}$$

satisfies this. For $z > 0$, the temperature gradient is zero: all heat transfer is by convection. For $z = -0$, $\partial T/\partial z = q/K$: from equation (A2.1), all the heating rate q is conducted towards $-z$. It is eventually swept back by convection towards $+z$.

A2.3.2 Semi-infinite solid $z > 0$, velocity: $\dot{u}_x$ steady heating rate q per unit area over the rectangle $-a < x < a$, $-b < y < b$, $z = 0$ (Figure A2.2(c)); ambient temperature T_o

Two extremes exist, depending on the ratio of the time $2a/\dot{u}_x$, for an element of the solid to pass the heat source of width $2a$ to the time a^2/κ for heat to conduct the distance $2a$ (Section A2.2.1). This ratio, equal to $2\kappa/(\dot{u}_x a)$, is the inverse of the more widely known Peclet number P_e.

When the ratio is large ($P_e << 1$), the temperature field in the solid is dominated by conduction and is no different from that in a stationary solid, see Section A2.2.4. Equations (A2.14) give maximum and average temperatures at the surface within the area of the heat source. When $b/a = 1$ and 5, for example,

$$\left.\begin{array}{l} u_x^{\cdot} a/(2\kappa) << 1: \\ \end{array}\right\} \quad \begin{array}{l} \dfrac{b}{a} = 1{:}(T - T_o)_{max} = 1.12 \dfrac{qa}{K}; (T - T_o)_{av} = 0.94 \dfrac{qa}{K} \\ \dfrac{b}{a} = 5{:}(T - T_o)_{max} = 2.10 \dfrac{qa}{K}; (T - T_o)_{av} = 1.82 \dfrac{qa}{K} \end{array} \tag{A2.17a}$$

At the other extreme ($P_e >> 1$), convection dominates the temperature field. Beneath the heat source, $\partial T/\partial z >> \partial T/\partial x$ or $\partial T/\partial y$; heat conduction occurs mainly in the z-direction and temperatures may be found from Section A2.2.2. At $z = 0$, the temperature variation from $x = -a$ to $x = +a$ is given by equation (A2.10), with the heating time t from 0 to $2a/\dot{u}_x$. Maximum and average temperatures are, after rearrangement to introduce the dimensionless group (qa/K),

$$\dot{u}_x a/(2\kappa) >> 1: \; (T - T_0)_{max} = 1.13 \frac{qa}{K} \left(\frac{2\kappa}{u_x a} \right)^{1/2}; \quad (T - T_0)_{av} = 0.75 \frac{qa}{K} \left(\frac{2\kappa}{u_x a} \right)^{1/2} \tag{A2.17b}$$

Because these results are derived from a linear heat flow approximation, they depend only on the dimension a and not on the ratio b/a, in contrast to $P_e << 1$ conditions.

A more detailed analysis (Carslaw and Jaeger, 1959) shows equations (A2.16) and (A2.17) to be reasonable approximations as long as $\dot{u}_x a/(2\kappa) < 0.3$ or > 3 respectively. Applying them at $\dot{u}_x a/(2\kappa) = 1$ leads to an error of ≈20%.

A2.4 Numerical (finite element) methods

Steady state ($\partial T/\partial t = 0$) solutions of equation (A2.4), with boundary conditions

$T = T_s$ on surfaces S_T of specified temperature,
$K\partial T/\partial n = 0$ on thermally insulated surfaces S_{qo},
$K\partial T/\partial n = -h(T-T_o)$ on surfaces S_h with heat transfer (heat transfer coefficient h),
$K\partial T/\partial n = -q$ on surfaces S_q with heat generation q per unit area.

may be found throughout a volume V by a variational method (Hiraoka and Tanaka, 1968). A temperature distribution satisfying these conditions minimizes the functional

$$I(T) = \int_V \left[\frac{K}{2} \left\{ \left(\frac{\partial T}{\partial x} \right)^2 + \left(\frac{\partial T}{\partial y} \right)^2 + \left(\frac{\partial T}{\partial z} \right)^2 \right\} - \left\{ q^* - \rho C \left(\dot{u}_x \frac{\partial \bar{T}}{\partial x} + \dot{u}_y \frac{\partial \bar{T}}{\partial y} + \dot{u}_z \frac{\partial \bar{T}}{\partial z} \right) \right\} T \right] \mathrm{d}V + \int_{S_q} qT \mathrm{d}S + \int_{S_h} \frac{h}{2} (T^2 - 2T_0 T) \mathrm{d}S \tag{A2.18}$$

where the temperature gradients $\partial\bar{T}/\partial x$, $\partial\bar{T}/\partial y$, $\partial\bar{T}/\partial z$, are not varied in the minimization process. The functional does not take into account possible variations of thermal properties with temperature, nor radiative heat loss conditions.

Equation (A2.18) is the basis of a finite element temperature calculation method if its volume and surface integrations, which extend over the whole analytical region, are regarded as the sum of integrations over finite elements:

$$I(T) = \sum_{e=1}^{m} I^e(T) \tag{A2.19}$$

where $I^e(T)$ means equation (A2.18) applied to an element and m is the total number of elements. If an element's internal and surface temperature variations with position can be written in terms of its nodal temperatures and coordinates, $I^e(T)$ can be evaluated. Its variation δI^e with respect to changes in nodal temperatures can also be evaluated and set to zero, to produce an element thermal stiffness equation of the form

$$[\mathbf{H}]_e \{T\} = \{F\}_e \tag{A2.20a}$$

where the elements of the nodal F-vector depend on the heat generation and loss quantities q^*, q and h, and the elements of $[\mathbf{H}]_e$ depend mainly on the conduction and convection terms of $I^e(T)$. Assembly of all the element equations to create a global equation

$$[\mathbf{H}]\{T\} = \{F\} \tag{A2.20b}$$

and its solution, completes the finite element calculation. The procedure is particularly simple if four-node tetrahedra are chosen for the elements, as then temperature variations are linear within an element and temperature gradients are constant. Thermal properties varying with temperature can also be considered, by allowing each tetrahedron to have different thermal properties. In two-dimensional problems, an equally simple procedure may be developed for three-node triangular elements (Tay *et al*., 1974; Childs *et al*., 1988).

A2.4.1 Temperature variations within four-node tetrahedra

Figure A2.3 shows a tetrahedron with its four nodes *i*, *j*, *k*, *l*, ordered according to a right-hand rule whereby the first three nodes are listed in an anticlockwise manner when viewed from the fourth one. Node *i* is at (x_i, y_i, z_i) and so on for the other nodes. Temperature T^{e} anywhere in the element is related to the nodal temperatures $\{T\} = \{T_i\ T_j\ T_k\ T_l\}^{\mathrm{T}}$ by

$$T^{\mathrm{e}} = [N_i\ N_j\ N_k\ N_l]\{\mathrm{T}\} = [\mathbf{N}]\{T\} \tag{A2.21}$$

where [**N**] is known as the element's shape function.

$$N_i = \frac{1}{6V_{\mathrm{e}}}(a_i + b_i x + c_i y + d_i z)$$

where

$$a_i = \begin{vmatrix} x_j & y_j & z_j \\ x_k & y_k & z_k \\ x_l & y_l & z_l \end{vmatrix}, \qquad b_i = -\begin{vmatrix} 1 & y_j & z_j \\ 1 & y_k & z_k \\ 1 & y_l & z_l \end{vmatrix}$$

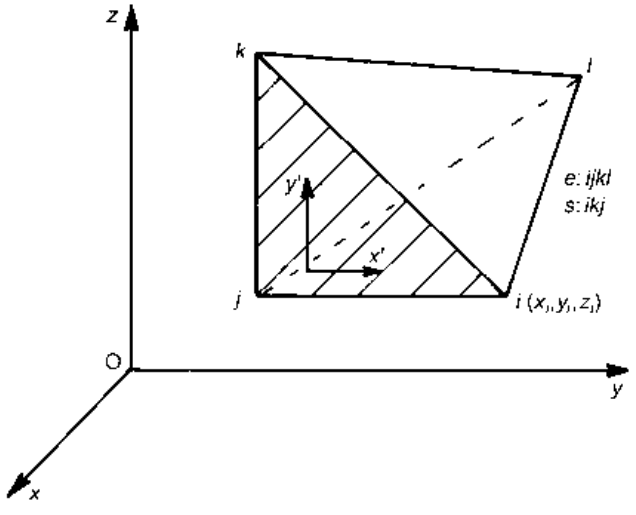


Fig. A2.3 A tetrahedral finite element

$$c_i = -\begin{vmatrix} x_k & 1 & z_j \\ x_k & 1 & z_k \\ x_l & 1 & z_l \end{vmatrix}, \quad d_i = -\begin{vmatrix} x_j & y_j & 1 \\ x_k & y_k & 1 \\ x_l & y_l & 1 \end{vmatrix}$$

and

$$V_c = \frac{1}{6}\begin{vmatrix} 1 & x_i & y_i & z_i \\ 1 & x_j & y_j & z_j \\ 1 & x_k & y_k & z_k \\ 1 & x_l & y_l & z_l \end{vmatrix} \tag{A2.22}$$

This may be checked by showing that, at the nodes, T^{e} takes the nodal values. N_j, N_k and N_l are similarly obtained by cyclic permutation of the subscripts in the order i, j, k, l. V_{e} is the volume of the tetrahedron.

In the same way, temperature T^{s} over the surface ikj may be expressed as a linear function of the surface's nodal temperatures:

$$T^{\mathrm{s}} = [N_i' N_j' N_k'] \{T\} = [\mathbf{N}'] \{T\} \tag{A2.23}$$

where

$$N_i' = \frac{1}{2\Delta_{ikj}} (a_i' + b_i' x' + c_i' y')$$

and

$$a_i' = x_k' y_j' - x_j' y_k'; \qquad b_i' = y_k' - y_j'; \qquad c_i' = x_j' - x_k' \tag{A2.24}$$

$$\Delta_{ikj} = \frac{1}{2}\begin{vmatrix} 1 & x_i' & y_i' \\ 1 & x_k' & y_k' \\ 1 & x_j' & y_j' \end{vmatrix}$$

The other coefficients are obtained by cyclic interchange of the subscripts in the order i, k, j. x', y' are local coordinates defined on the plane ikj. Δ_{ikj} is the area of the element's triangular face: it may also be written in global coordinates as

$$\Delta_{ikj} = \frac{1}{2}\left(\begin{vmatrix} y_k - y_i & y_j - y_j \\ z_k - z_i & z_j - z_i \end{vmatrix}^2 + \begin{vmatrix} z_k - z_i & z_j - z_i \\ x_k - x_i & x_j - x_i \end{vmatrix}^2 + \begin{vmatrix} x_k - x_i & x_j - x_i \\ y_k - y_i & y_j - y_i \end{vmatrix}^2 \right)^{1/2} \tag{A2.25}$$

A2.4.2 Tetrahedral element thermal stiffness equation

Equation (A2.21), after differentiation with respect to x, y and z, and equation (A2.23) are substituted into $I^{\mathrm{e}}(T)$ of equation A2.19. The variation of $I^{\mathrm{e}}(T)$ with respect to T_i, T_j, T_k and T_l is established by differentiation and set equal to zero. $[\mathbf{H}]_{\mathrm{e}}$ and $\{F\}_{\mathrm{e}}$ (equation (A2.20a)) are

$$[\mathbf{H}]_{\mathrm{e}} = \frac{K}{36V_{\mathrm{e}}}\begin{bmatrix} b_ib_i + c_ic_i + d_id_i & b_jb_i + c_jc_i + d_jd_i & b_kb_i + c_kc_i + d_kd_i & b_lb_i + c_lc_i + d_ld_i \\ b_ib_j + c_ic_j + d_id_j & b_jb_j + c_jc_j + d_jd_j & b_kb_j + c_kc_j + d_kd_j & b_lb_j + c_lc_j + d_ld_j \\ b_ib_k + c_ic_k + d_id_k & b_jb_k + c_jc_k + d_jd_k & b_kb_k + c_kc_k + d_kd_k & b_lb_k + c_lc_k + d_ld_k \\ b_ib_l + c_ic_l + d_id_l & b_jb_l + c_jc_l + d_jd_l & b_kb_l + c_kc_l + d_kd_l & b_lb_l + c_lc_l + d_ld_l \end{bmatrix}$$

$$+\frac{\rho C}{24}\begin{bmatrix} \dot{u}_x b_i + \dot{u}_y c_i + \dot{u}_z d_i & \dot{u}_x b_j + \dot{u}_y c_j + \dot{u}_z d_j & \dot{u}_x b_k + \dot{u}_y c_k + \dot{u}_z d_k & \dot{u}_x b_l + \dot{u}_y c_l + \dot{u}_z d_l \\ \dot{u}_x b_i + \dot{u}_y c_i + \dot{u}_z d_i & \dot{u}_x b_j + \dot{u}_y c_j + \dot{u}_z d_j & \dot{u}_x b_k + \dot{u}_y c_k + \dot{u}_z d_k & \dot{u}_x b_l + \dot{u}_y c_l + \dot{u}_z d_l \\ \dot{u}_x b_i + \dot{u}_y c_i + \dot{u}_z d_i & \dot{u}_x b_j + \dot{u}_y c_j + \dot{u}_z d_j & \dot{u}_x b_k + \dot{u}_y c_k + \dot{u}_z d_k & \dot{u}_x b_l + \dot{u}_y c_l + \dot{u}_z d_l \\ \dot{u}_x b_i + \dot{u}_y c_i + \dot{u}_z d_i & \dot{u}_x b_j + \dot{u}_y c_j + \dot{u}_z d_j & \dot{u}_x b_k + \dot{u}_y c_k + \dot{u}_z d_k & \dot{u}_x b_l + \dot{u}_y c_l + \dot{u}_z d_l \end{bmatrix}$$

$$+\frac{h\Delta_{ikj}}{12}\begin{bmatrix} 2 & 1 & 1 & 0 \\ 1 & 2 & 1 & 0 \\ 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{A2.26}$$

and

$$\{F\}_e = \frac{q^* V_e}{4}\begin{Bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{Bmatrix} - \frac{q\Delta_{ikj}}{3}\begin{Bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{Bmatrix} - \frac{hT_0\Delta_{ikj}}{3}\begin{Bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{Bmatrix} \tag{A2.27}$$

Global assembly of equations (A2.20a), with coefficients equations (A2.26) and (A2.27), to form equation (A2.20b), or similarly in two-dimensions, forms the thermal part of closely coupled steady state thermal–plastic finite element calculations.

A2.4.3 Approximate finite element analysis

Finite element calculations can be applied to the shear-plane cutting model shown in Figure A2.4. There are no internal volume heat sources, q^*, in this approximation, but internal surface sources q_s and q_f on the primary shear plane and at the chip/tool interface. If experimental measurements of cutting forces, shear plane angle and chip/tool contact length have been carried out, q_s and the average value of q_f can be determined as follows:

$$q_s = \tau_s V_s \tag{A2.28a}$$

$$q_f = \tau_f V_c \tag{A2.28b}$$

where

$$\left.\begin{aligned} \tau_s &= \frac{F_C \cos\phi - F_T \sin\phi}{fd}\sin\phi; & \tau_f &= \frac{F_C \sin\alpha + F_T \cos\alpha}{l_c d} \\ V_s &= \frac{\cos\alpha}{\cos(\phi-\alpha)}U_{work}; & V_c &= \frac{\sin\phi}{\cos(\phi-\alpha)}U_{work} \end{aligned}\right\} \tag{A2.29}$$

In general, q_s is assumed to be uniform over the primary shear plane, but q_f may take on a range of distributions, for example triangular as shown in Figure A2.4.

A2.4.4 Extension to transient conditions

The functional, equation (A2.18), supports transient temperature calculation if the q^* term is replaced by $(q^* - \rho C\partial\bar{T}/\partial t)$. Then the finite element equation (A2.20a) becomes

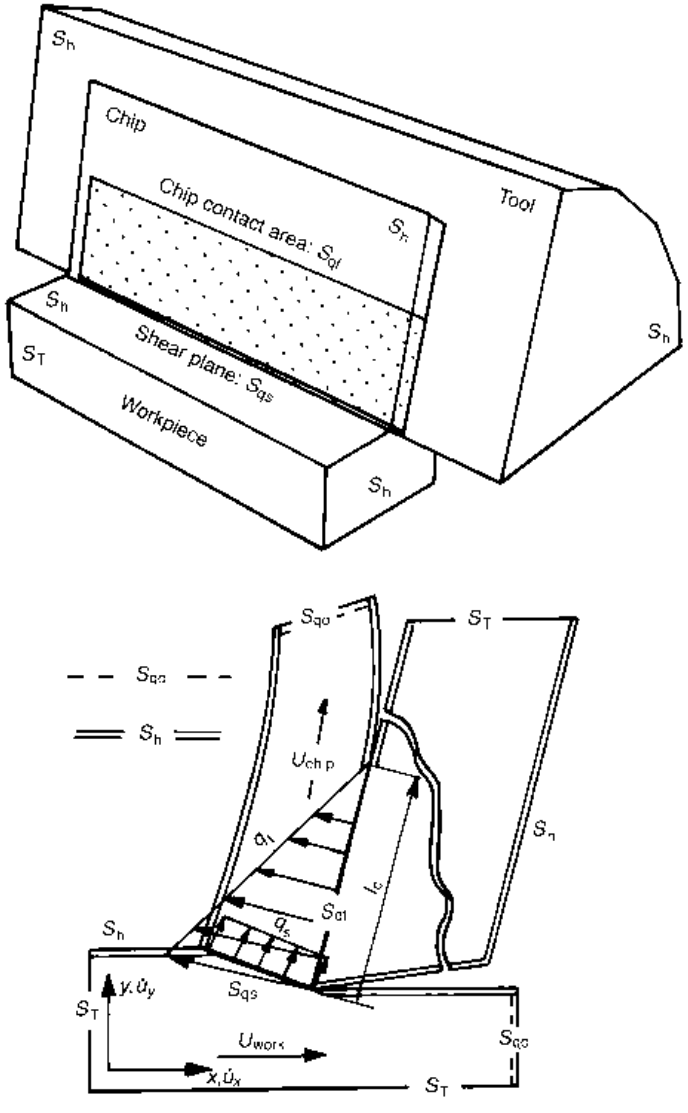


Fig. A2.4 Thermal boundary conditions for a shear plane model of machining

$$[\mathbf{C}]_e \left\{ \frac{\partial T}{\partial t} \right\} + [\mathbf{H}]_e \{T\} = \{F_e\} \tag{A2.30}$$

with

$$[\mathbf{C}]_e = \frac{\rho C V_e}{20} \begin{vmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{vmatrix}$$

([**C**] is given here for a four-node tetrahedron).

Over a time interval Δt, separating two instants t_n and t_{n+1}, the average values of nodal rates of change of temperature can be written in two ways

$$\left\{\frac{\partial T}{\partial t}\right\}_{\text{av}} = (1-\theta)\left\{\frac{\partial T}{\partial t}\right\}_{n} + \theta\left\{\frac{\partial T}{\partial t}\right\}_{n+1} \tag{A2.31a}$$

or

$$\left\{\frac{\partial T}{\partial t}\right\}_{\text{av}} = \left\{\frac{T_{n+1} - T_{\text{n}}}{\Delta t}\right\} \tag{A2.31b}$$

where θ is a fraction varying between 0 and 1 which allows the weight given to the initial and final values of the rates of change of temperature to be varied. After multiplying equations (A2.31) by [**C**], substituting $[\mathbf{C}]\{\partial T/\partial t\}$ terms in equation (A2.31a) for $(\{F\}–[\mathbf{H}]\{T\})$ terms from equation (A2.30), equating equations (A2.31a) and (A2.31b), and rearranging, an equation is created for temperatures at time t_{n+1} in terms of temperatures at time t_n: in global assembled form

$$\left(\frac{[\mathbf{C}]}{\Delta t} + \theta[\mathbf{K}]\right)\{T\}_{n+1} = \left(\frac{[\mathbf{C}]}{\Delta t} - (1-\theta)[\mathbf{K}]\right)\{T\}_{n} + \{F\} \tag{A2.32}$$

This is a standard result in finite element texts (for example Huebner and Thornton, 1982). Time stepping calculations are stable for $\theta \geq 0.5$. Giving equal weight to the start and end rates of change of temperature ($\theta = 0.5$) is known as the Crank–Nicolson method (after its originators) and gives good results in metal cutting transient heating calculations.

References

Carslaw, H. S. and Jaeger, J. C. (1959) *Conduction of Heat in Solids*, 2nd edn. Oxford: Clarendon Press.

Childs, T. H. C., Maekawa, K. and Maulik, P. (1988) Effects of coolant on temperature distribution in metal machining. *Mat. Sci. and Technol*. **4**, 1006–1019.

Hiraoka, M. and Tanaka, K. (1968) A variational principle for transport phenomena. *Memoirs of the Faculty of Engineering*, Kyoto University **30**, 235–263.

Huebner, K. H. and Thornton, E. A. (1982) *The Finite Element Method for Engineers*, 2nd edn. New York: Wiley.

Loewen, E. G. and Shaw, M. C. (1954) On the analysis of cutting tool temperatures. *Trans. ASME* **76**, 217–231.

Tay, A. O., Stevenson, M. G. and de Vahl Davis, G. (1974) Using the finite element method to determine temperature distributions in orthogonal machining. *Proc. Inst. Mech. Eng. Lond.* **188**, 627–638.

Appendix 3

Contact mechanics and friction

A3.1 Introduction

This appendix summarizes, in the context of metal machining, understanding of the stresses that occur at the contacts between sliding bodies. These stresses, with materials' responses to them, are responsible for materials' friction (and wear).

All engineering components – for example slideways, gears, bearings, and cutting tools – have rough surfaces, characteristic of how they are made. When such surfaces are loaded together, they touch first at their high spots. Figure A3.1 is a schematic view of two rough surfaces placed in contact under a load W, the top one sliding to the right under the action of a friction force F.

Figure A3.1(a) shows a contact, the material properties and roughness of which are such that the surfaces have deformed to bring the direction of sliding into the planes of the real areas of contact A_r. Resistance to sliding then comes from the surface shear stresses s. Friction that arises from shear stresses is called adhesive friction. If the real areas of contact on average support a normal contact stress p_r, the adhesive coefficient of friction μ_a is given by

$$\mu_a = \frac{s}{p_r} \tag{A3.1}$$

Figure A3.1(b) shows surfaces for which the real areas of contact are inclined to the sliding direction. Each contact is divided into two parts, ahead of (leading) and behind

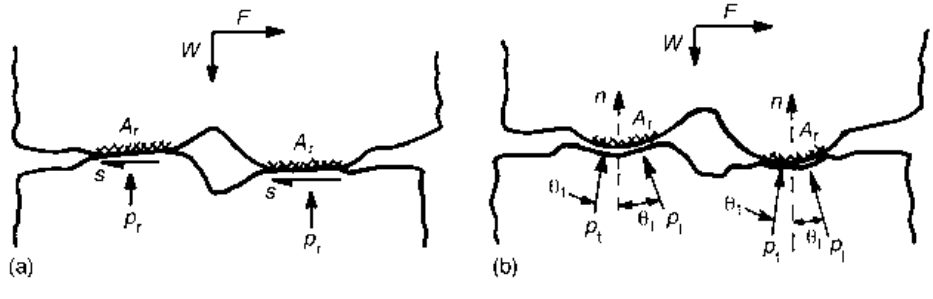


Fig. A3.1 Friction caused (a) by shear stresses s and (b) by direct stresses p

(trailing) the real contact mean normal n. Even in the absence of surface shear stresses, a resistance to sliding occurs if the normal forces on the leading and trailing portions of the contacts differ from one another. Friction arising from contact normal stresses is called deformation friction. If, on average, the normal stress p_l on the leading part of a contact of sub-area A_l is inclined at θ_l to the direction of the load W, and on the trailing part of the contact the equivalent variables are p_t, A_t and θ_t, force resolution in the directions of W and F give the deformation friction coefficient μ_d as

$$\mu_d = \frac{p_l A_l \sin\theta_l - p_t A_t \sin\theta_t}{p_l A_l \cos\theta_l + p_t A_t \cos\theta_t} \tag{A3.2a}$$

Special cases occur. If the contact is symmetrical ($p_l = p_t$; $A_l = A_t$; $\theta_l = \theta_t$), equation (A3.2a) simplifies to $\mu_d = 0$: this is the case of perfectly elastic deformation. At the other extreme, when the indenting surface plastically scratches (abrades) the other, there may be no trailing portion contact: $A_t = 0$. Then, equation (A3.2a) becomes

$$\mu_d = \tan\theta_l \tag{A3.2b}$$

This type of deformation friction (abrasion of metals) is of most relevance to this book. (There is a third situation, of visco-elastic contact, intermediate between perfectly elastic and totally plastic contact, when μ_d may be shown to depend on both $\tan\theta_l$ and $\tan\delta$, the loss factor for the contact deformation cycle.)

Equation (A3.1) shows that adhesive friction depends mainly on material properties s and p_r although, as will become clear, p_r also depends on surface contact geometry. By contrast, equation (A3.2b) shows that abrasive deformation friction depends mainly on surface geometry, insofar as the angle θ_l is the same as the slope of the leading part of the contact, but this could be modified by material properties if, for example, the real pressure distribution over A_l is not uniform.

The main focus of this appendix is to review how the friction coefficient varies with material properties and contact geometry, in adhesive and deformation friction conditions, and when both act together.

Two further points can usefully be introduced before proceeding with this review. The real contact stress p_r in equation (A3.1) is the natural quantity to be part of a friction law, but in practice it is the nominal stress, the load divided by the apparent, or nominal, contact area A_n, which is set in any given application. In Chapter 2, this stress has been written σ_n. The first point is that, from load force equilibrium, the ratio of σ_n to p_r is the same as the ratio of the real to apparent contact area (A_r/A_n):

$$\left(\frac{A_r}{A_n}\right) = \frac{\sigma_n}{p_r} \tag{A3.3a}$$

The second point is that, in Chapter 2, σ_n is normalized with respect to some shear flow stress k of the work or chip material. The dimensionless ratios p_r/k and s/k can be introduced into equation (A3.1) and further p_r/k eliminated in favour of σ_n/k by means of equation (A3.3a):

$$\mu_a = \frac{(s/k)}{(p_r/k)} = \frac{(s/k)}{(\sigma_n/k)}\left(\frac{A_r}{A_n}\right) \tag{A3.3b}$$

In the following sections, a view of how sliding friction depends on material properties, contact geometry and intensity of loading is developed, by concentrating on how p_r/k and A_r/A_n vary in adhesive and deformation friction conditions. A more detailed account of much of the contact mechanics is in the standard text by Johnson (1985). Reference will be made to this work in the abbreviated form (KLJ Ch.*x*).

A3.2 The normal contact of a single asperity on an elastic foundation

As a first step in building up a view of asperity contact, consider the normal loading of a single asperity against a flat counterface. At the lightest loading, the deformation may be elastic. At some heavier load, plastic deformation may set in. The purpose of this section is to establish how transition from an elastic to a plastic state varies with material properties and asperity shape; and what real contact pressures p_r are set up.

A3.2.1 Elastic contact

Figure A3.2 shows asperities idealized as a sphere or cylinder of radius R, or as a blunt cone or wedge of slope β, pressed on to a flat. The dashed lines show the asperity and flat penetrating each other to a depth δ, as if the other was not there. The solid lines show the deformation required to eliminate the penetration. How p_r varies with the contact width $2a$, or with δ; and with R or β; and with Young's modulus E_1 and E_2 and Poisson's ratio ν_1 and ν_2 of the asperity and counterface respectively, is developed here.

The contact of an elastic sphere or cylinder on a flat in the absence of interface shear is the well-known Hertzian contact problem. A dimensional approach gives insight into the contact conditions more simply than does a full Hertzian analysis.

In the left-hand part of Figure A3.2, the asperity is shown flattened by a depth δ_1, and the flat by a depth δ_2, in accommodating the total overlap δ and creating a contact width $2a$. From the geometry of overlap, supposing $2a$ to be a fixed fraction of the chordal length $2a_c$, and when $a_c << R$,

$$\delta = \delta_1 + \delta_2 = \frac{a_c^2}{2R} \propto \frac{a^2}{2R} \tag{A3.4}$$

The surface deformations in the asperity and flat cause sub-surface strains. In the asperity, these are in proportion to the dimensionless ratio δ_1/a and in the flat to δ_2/a. When the

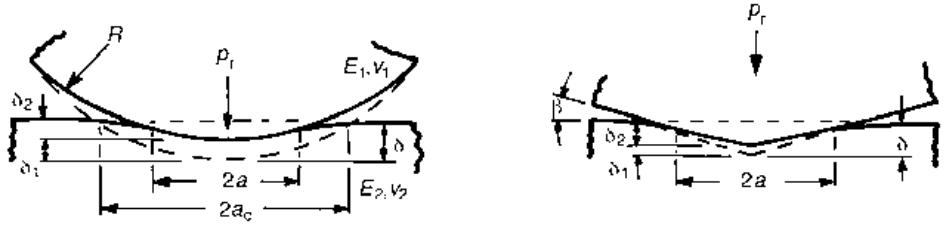


Fig. A3.2 Models of elastic asperity deformation

Table A3.1 Elastic contact parameters, from Johnson (1985, Chs 4 and 5)

Asperity peak shape	c, eqns (A3.8) and (3.9)	(p_r/τ_{max})	$(p_r/\tau_{max})/c$
Spherical	0.42	2.6	6.2
Cylindrical	0.39	2.2	5.6
Conical	0.50	1.6	3.2
Wedge-like	0.50	1.0	2.0

asperity and flat obey Hooke's law, the mean contact stress p_r will increase in proportion to the product of Young's modulus and strain in each:

$$\begin{aligned} &\text{from the asperity's point of view,} \quad && p_r \propto E_1(\delta_1/a) \\ &\text{from the flat's point of view,} && p_r \propto E_2(\delta_2/a) \end{aligned} \tag{A3.5}$$

Combining equations (A3.4) and (A3.5) gives

$$p_r = cE^*(a/R) \tag{A3.6}$$

where $1/E^* = (1/E_1+1/E_2)$ and the constant of proportionality c requires the full Hertz analysis for its derivation. The full analysis in fact shows that the proper definition of E^* involves Poisson's ratio:

$$\frac{1}{E^*} = \frac{1 - \nu_1^2}{E_1} + \frac{1 - \nu_2^2}{E_2} \tag{A3.7}$$

and c depends on whether the circular profile of radius R represents a spherically or a cylindrically capped asperity (Table A3.1).

Similarly, the pressing together of two spherical or two cylindrical asperities with parallel axes, of radii R_1 and R_2, creates a normal contact stress p_r:

$$p_r = cE^*(a/R^*) \quad \text{where} \quad 1/R^* = 1/R_1 + 1/R_2 \tag{A3.8}$$

The elastic contact of a wedge or cone on a flat (right-hand part of Figure A3.2(a)) generates a contact pressure p_r (KLJ Ch. 5):

$$p_r = cE^* \tan\beta \tag{A3.9}$$

where c is also given in Table A3.1. The quantities (a/R^*) and $\tan\beta$ can be regarded as representative contact strains. Their interpretation as mean contact slopes will be returned to later. As they increase, so does p_r.

A3.2.2 Fully plastic contact

Figure A3.3 shows a wedge-shaped asperity loaded plastically against a softer (left) and a harder (right) counterface, so that it indents or is flattened. The dependence of p_r on asperity slope β and shear flow stress k of the softer material is considered here, by means of slip-line field theory (Appendix 1.2).

In each case, the region ADE is a uniform stress region and the free surface condition along AE requires that $p_1 = k$. Region ABC is also uniformly stressed. Normal force equilibrium across AC gives

$$p_r = p_2 + k \tag{A3.10}$$

Slip-line EDBC is an α-line, so

$$p_2 = p_1 + 2k\psi \equiv k(1 + 2\psi) \tag{A3.11}$$

The angle ψ is chosen to conserve the volume of the flow: material displaced from the overlap between the flat and the asperity must re-appear in the shoulders of the flow, but for small values of β, $\psi \approx \pi/2$. This, with equations (A3.11) and (A3.10), gives

$$p_r \approx 2k(1 + \pi/2) \approx 5k \tag{A3.12}$$

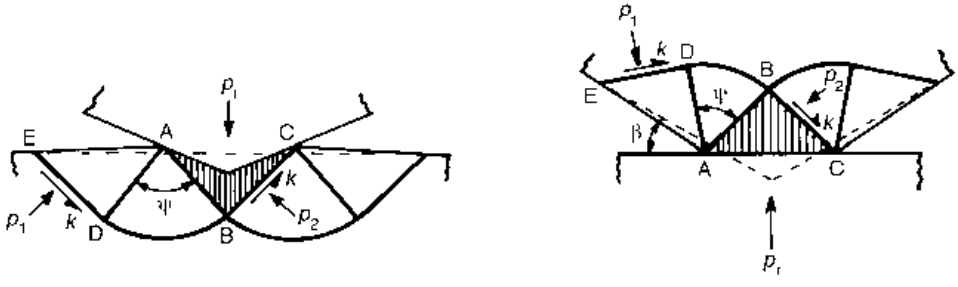


Fig. A3.3 Plastic indenting by, and flattening of, wedge-shaped asperities

A3.2.3 The transition from elastic to plastic contact

The elastic and plastic views of the previous sub-sections are brought together by non-dimensionalizing the contact pressures p_r by k. In Figure A3.4(a), the elastic and plastic model predictions are the dashed lines. The solid line is the actual behaviour. Departure from elastic behaviour first occurs in the range $1 < p_r/k < 2.6$, at values of $(E^*/k)(a/R^*$ or $\tan\beta)$ from 2 to 6.2. The values depend on the asperity shape: they are the last two columns in Table A3.1.

The fully plastic state is developed for $(E^*/k)(a/R^*$ or $\tan\beta)$ greater than about 50. p_r/k continues to increase at larger deformations than this due to strain hardening.

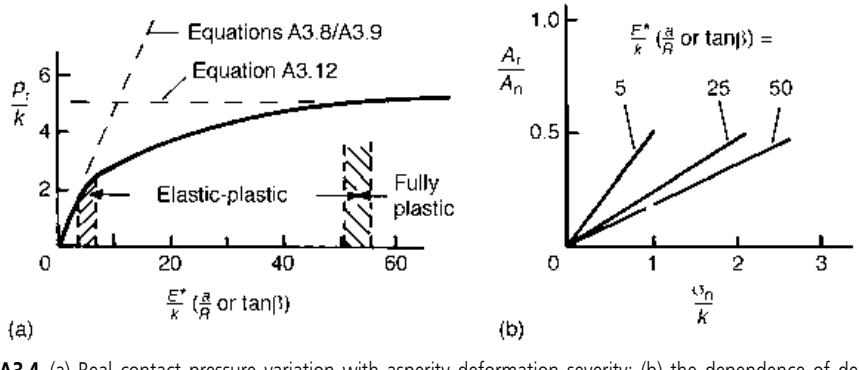


Fig. A3.4 (a) Real contact pressure variation with asperity deformation severity; (b) the dependence of degree of contact on intensity of loading, in the absence of sliding

A real contact area A_r is associated with a surrounding nominal contact area A_n. Figure A3.4(a) can be used with equation (A3.3a), to map how A_r/A_n increases with σ_n/k. The result is shown in Figure A3.4(b) for different values of $(E^*/k)(a/R^*$ or $\tan\beta)$.

A3.3 The normal contact of arrays of asperities on an elastic foundation

In the previous section, loading a single asperity was considered. When all the contacts between two surfaces have the same half-width a, Figure A3.4(b) can be used directly to predict the degree of contact from σ_n/k. However, on real surfaces, asperities have random heights and are not loaded equally. The effect of this on the use of Figure A3.4(b) is the first point considered in this section. In Figure A3.4(b), predictions are only drawn for $A_r/A_n < 0.5$: the second point considered in this section is what happens at higher degrees of contact, when asperity stress fields start to interact.

A3.3.1 Loading of random rough surfaces

Figure A3.5 shows the loading of two rough flat surfaces against a smooth flat counterface. In case (a) all the asperities on the rough surface are identical, imagined as spherical caps of radius R, and are shown in contact with the counterface. In case (b), the same asperities have been shifted in a random manner normal to the surface, so that the peaks have a random distribution σ_s of heights about their mean height. This situation is the most simple that can be pictured, to make the point that an increase of load in case (a) causes the load per contact, the half-width a and the stress severity to increase. However, in case (b), the number of contacts can also increase, so that the load per contact and the severity of stress will increase less slowly with load.

The situation of Figure A3.5(b) was considered by Greenwood and Williamson (1966), supposing the contact stresses to be elastic, and is reproduced in (KLJ Ch. 13). Provided that the number of asperities in contact is a small fraction of the total available (this means in practice that $A_r/A_n \leq 0.5$), the number of contacts grows almost in proportion to the load, so on average the load per contact is almost independent of load. The average real contact pressure $\bar{p}_r$ is

$$\bar{p}_r = (0.3 \text{ to } 0.4)E^*\sqrt{\sigma_s/R} \tag{A3.13a}$$

and is a function only of E^* and the rough surface finish. Compared with equations (A3.8) and (A3.9), $\sqrt{\sigma_s/R}$ is seen as the measure of mean asperity strain or equivalent slope.

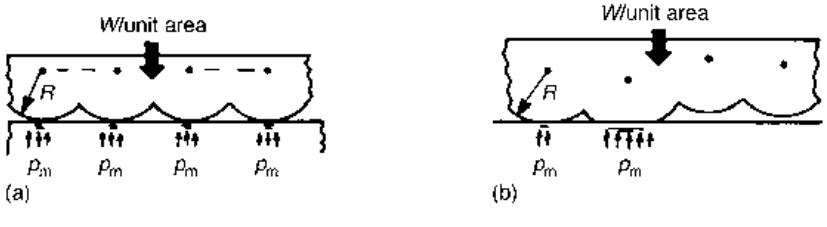


Fig. A3.5 (a) A regular and (b) a random model rough surface loaded on to a flat

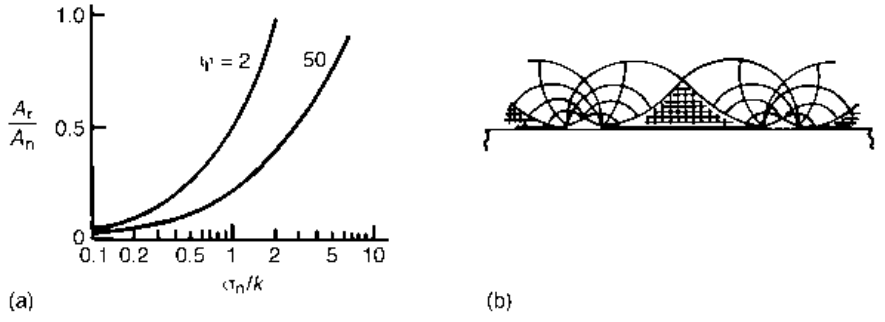


Fig. A3.6 (a) The dependence of A_r/A_n on σ_n/k for different degrees of roughness on an elastic foundation; (b) rigid plastic compression of ridge-shaped asperities

Indeed, later analyses of rough surface elastic contact have replaced $\sqrt{\sigma_s/R}$ by Δ_q, the RMS slope of the rough surface. In non-dimensional form:

$$\frac{\bar{p}_r}{k} = c\,\frac{E^*}{k}\,\Delta_q \tag{A3.13b}$$

The severity index $(E^*/k)\,\Delta_q$, more commonly called the plasticity index Ψ, may be used with Figure A3.4(b) to determine the degree of contact of a rough loaded surface.

A3.3.2 Loading at high degrees of contact

As A_r/A_n increases above 0.5, even for a randomly rough surface, the availability of new contacts becomes exhausted. A load increase will no longer cause a proportional increase in the number of contacts, but will cause increased deformation of existing contacts. A_r/A_n will no longer increase in direct proportion to σ_n/k. Figure A3.6(a) extends Figure A3.4(b) to higher values of σ_n/k and A_r/A_n: note the rescaling of σ_n/k to a log base. At one extreme ($\Psi = 2$), the displacement of material as the surfaces are brought together is taken up by elastic compression. In this example, full contact is reached at $\sigma_n/k = 2$. At the other extreme of fully plastic flow ($\Psi = 50$ or more), material displaced from high spots reappears in the valleys between contacts. Figure A3.6(b) represents a model situation of the plastic crushing of an array of wedge-shaped asperities. The material displaced from the crests of the array by the counterface is extruded into the ever-diminishing gap between the contacts. Slip-line field modelling suggests that, by the stage that the degree of contact has risen to 0.8, the hydrostatic stress beneath the contacts has risen from $\approx 4k$ (for well separated contacts) to $\approx 9k$: then $\sigma_n/k \approx 8$ (Childs, 1973).

A3.4 Asperities with traction, on an elastic foundation

Section A3.3 considered real contacts' ability to support load in the absence of sliding. When shear stresses due to sliding are added to the stresses due to loading, contacts that under load alone are elastic may become plastic; contacts that are already plastic will be overstressed and collapse. These are the sub-topics of this section.

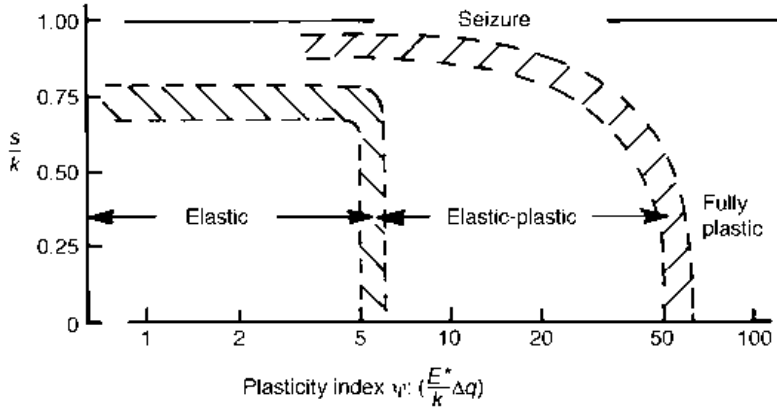


Fig. A3.7 Asperity state of stress dependence on surface shear and plasticity index

A3.4.1 Contact stress regimes under sliding conditions

The stressing of elastic spheres and cylinders loaded against flats, without and with sliding, is reviewed in detail in (KLJ Chs. 4 and 6). Without sliding, the largest shear stress τ_{max} occurs from $0.48a$ to $0.78a$ below the centre of the contact. With sliding, if $\mu_a < 0.25$ to 0.3, τ_{max} is not changed in size by sliding, but the position where it occurs moves towards the surface. For $\mu_a > 0.25$ to 0.3, τ_{max} occurs at the surface and its size rises proportionally to μ_a. The constant of proportionality depends on whether a sphere or a cylinder is being loaded:

$$\tau_{max} = (1.27 \text{ to } 1.5)\mu_a p_r \tag{A3.14}$$

These observations may be applied to the contact of random rough surfaces. Figure A3.7 shows the state of stress (elastic, elastic–plastic or fully plastic) to be expected for different combinations of plasticity index Ψ and s/k. When $s/k = 0$, the transition from elastic to elastic–plastic flow occurs for $\Psi \approx 5$ to 6; fully plastic flow commences for $\Psi \approx 50$ to 60. These values are the same as the transition values of $(E^*/k)(a/R^*)$ shown in Figure A3.4(a).

As s/k increases, the elastic boundary is not altered until s/k reaches 0.67 to 0.78. For larger values, a purely elastic state does not exist. Thus, in Figure A3.7, the elastic region is capped at these values. (They are derived from equation (A3.14), by noting that $\mu_a = s/p_r$ and that plastic flow occurs once $\tau_{max} = k$.)

How the elastic–plastic/fully plastic boundary is influenced by s/k is not well established theoretically. The boundary drawn in Figure A3.7 is a little speculative.

A3.4.2 Junction growth of plastic contacts

A real area of contact A_r, loaded in the absence of sliding, has $\tau_{max} = k$ within it if it is in a plastic state. The addition of a sliding force F to the contact, creating an extra shear stress F/A_r, will, if A_r does not increase, result in an increased τ_{max}. In fact, A_r grows to prevent τ_{max} exceeding k. An alternative view of the cause of this junction growth is that, in the absence of sliding, the material surrounding a plastic contact helps to support the load by

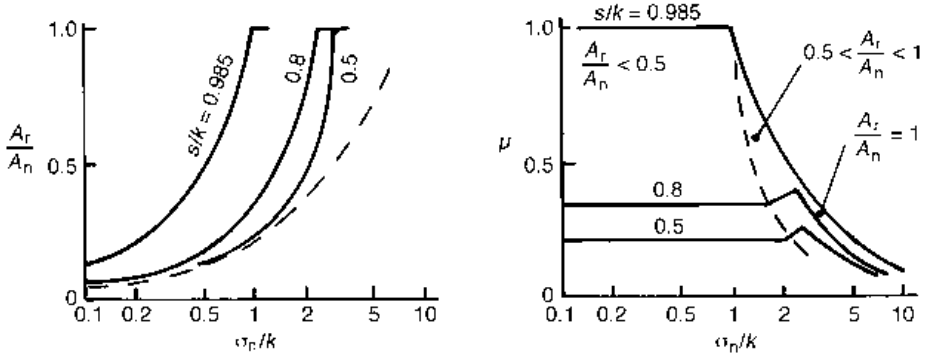


Fig. A3.8 The dependence of A_r/A_n and μ on σ_n/k and s/k, for hard ridges of slope $\beta = 5°$ sliding against a soft flat

imposing a hydrostatic pressure on the deviatoric stress field. Its size, from slip line field modelling, p_2 in equation (A3.11) with $2\psi \approx \pi$, is about $4k$. The addition of surface shear on the contact reduces the surrounding's ability to support the load; in other words, the hydrostatic pressure component supporting the load reduces.

How the slip-line fields of Figure A3.3 become modified by sliding have been studied by Johnson, for the case of a soft asperity on a hard flat (KLJ Ch. 7), and by Oxley (1984) for hard wedges ploughing over a soft flat. The conclusion of both, stemming from the connection of the plastic flow field beneath the contact to the free surface where $p = k$, is that, for s/k close to 1, sliding causes p_r to fall from around $5k$ to $1k$. For a constant load, this causes a fivefold increase in real contact, at least while asperities are sufficiently far apart not to interact with one another.

Figure A3.8 shows Oxley's prediction of how A_r/A_n and μ depend on σ_n/k, for sliding hard wedge-shaped asperities, of slope $\beta = 5°$, over a soft flat, at different levels of surface shear s/k (the dependence of A_r/A_n on σ_n/k in the absence of sliding is shown by the dashed line). It has been chosen because the situation of hard ridges sliding on a soft flat may more realistically represent the condition of a rake face of a cutting tool sliding against a chip than soft asperities sliding on a smooth hard flat. While $A_r/A_n < 0.5$, μ is independent of σ_n/k, but as A_r/A_n approaches 1, μ reduces with increasing σ_n/k.

Although Figure A3.8 is only one example, it illustrates three general points. (1) The hydrostatic pressure within a sliding contact is predicted by slip-line field modelling to be less the larger is s/k, but while it is controlled by the free surface boundary condition, it never becomes less than k. As a result, μ never becomes greater than 1. (2) The reduction in hydrostatic pressure, and hence the junction growth and μ, is very sensitive to s/k when, as in Figure A3.8, s/k is large (close to 1). (3) Once $A_r/A_n = 1$, μ is no longer independent of, but becomes inversely proportional to, load.

A3.5 Bulk yielding

When an asperity is supported on an elastic bulk, the only way to accommodate its plastic distortions is by flow to the free surface. It is this, in the previous section, which ensured that p_r never became less than k. When the bulk is plastic, asperity plastic distortion can be

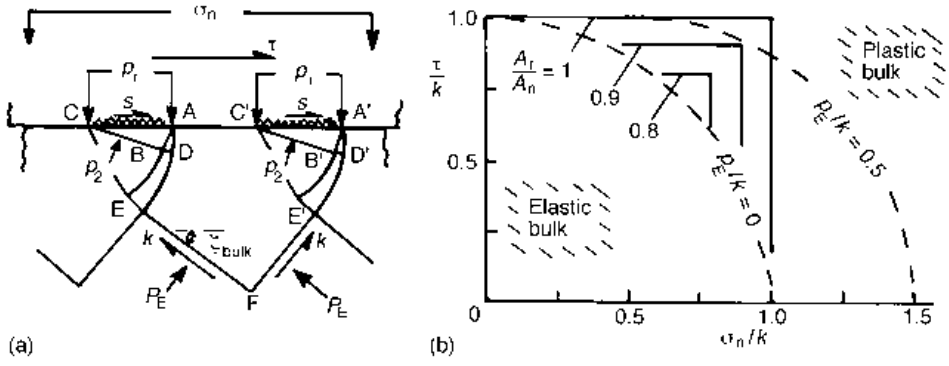


Fig. A3.9 (a) Combined asperity and bulk plastic stressing and (b) derived degrees of contact A_r/A_n, depending on the apparent contact stresses and the bulk field hydrostatic stress level P_E/k, for $s/k = 1$

accommodated by flow into the bulk. Asperity hydrostatic stress and p_r then depend on the state of the bulk flow field: it is possible for p_r to be less than k. If this happens, the degree of contact becomes greater than considered previously. How it depends on the nominal contact stresses σ_n and τ, and on the hydrostatic stress in the bulk field, will now be developed, still for conditions of plane strain and a non-hardening plastic material to which slip-line field theory can be applied.

Figure A3.9(a) is adapted from Sutcliffe (1988). It shows a combined asperity and bulk slip-line field. The bulk field is described by the hydrostatic pressure p_E and slip-lines inclined at ζ_{bulk} to the counterface. The asperity fields ADBC, A′D′B′C′ around the real contacts AC, A′C′ are connected to the bulk by DEC, D′E′C′. It is supposed that p_2, beneath the asperities, has been reduced so much by the influence of p_E that the stresses on AD, A′D′ are no longer sufficient to cause a plastic state to extend to the free surface: region ADEFE′C′ has become rigid. For this to be the case, at least for high values of real surface shear stress s ($s/k \Rightarrow 1$), the results of the previous section suggest $p_r/k < 1$. How p_r/k – and hence (with σ_n/k) A_r/A_n – depends on p_E and ζ_{bulk} can be determined from the slip-line field and its limits of validity.

However, it is simpler to consider the overall force balances between the bulk field, the nominal contact stresses σ_n and τ, and the real contact stresses p_r and s. Force equilibrium between p_E, k, ζ_{bulk} and σ_n and τ creates the relations:

$$\left.\begin{aligned} \tau &= k \cos 2\zeta_{bulk} \\ \sigma_n &= p_E + k \sin 2\zeta_{bulk} \end{aligned}\right\} \qquad \text{(A3.15a)}$$

By elimination of ζ_{bulk}, a relation is formed between p_E and σ_n and τ:

$$\frac{p_E}{k} = \frac{\sigma_n}{k} - \left(1 - \left(\frac{\tau}{k}\right)^2\right)^{1/2} \qquad \text{(A3.15b)}$$

In Figure A3.9(b) the dashed lines show combinations of (τ/k) and (σ_n/k) consistent with asperities existing on a bulk plastic flow in which $p_E/k = 0$ or 0.5. The region marked 'elastic bulk' is that for which (τ/k) and (σ_n/k) are associated with an elastic bulk unless $p_E/k <$ 0. That marked 'plastic bulk' is plastic if $p_E/k > 0.5$.

Equilibrium between the forces of the nominal and real contact stresses gives

$$\frac{\tau}{k} = \frac{A_r}{A_n}\frac{s}{k} \tag{A3.16a}$$

$$\frac{\sigma_n}{k} = \frac{A_r}{A_n}\frac{p_r}{k} \tag{A3.16b}$$

According to equation (A3.16a), $A_r/A_n \Rightarrow \tau/k$ as $s/k \Rightarrow 1$. From equation (A3.16b), if $A_r/A_n < (\sigma_n/k)$, $p_r/k > 1$. However, the slip-line field is not valid if $p_r/k > 1$: flow will break through to the free surface and p_r/k will be limited to 1. Thus, for a plastic asperity on a plastic foundation, when s/k is close to 1, A_r/A_n will equal (τ/k) when $(\tau/k) > (\sigma_n/k)$, and (σ_n/k) when $(\sigma_n/k) > (\tau/k)$, up to its maximum possible value of 1. Contours of $A_r/A_n = 1$, 0.9, 0.8, satisfying this, are added to Figure A3.9(b).

Figure A3.9(b) shows, firstly, the levels of dimensionless hydrostatic stress, p_E/k, needed for a combination of (τ/k) and (σ_n/k) to be associated with a bulk plastic flow. If there is bulk plasticity, it then shows how degrees of contact much larger than when the bulk remains elastic (Figure A3.8a) can be generated at values of $(\sigma_n/k) < 1$. In these conditions the ratio of friction to normal stress (the friction coefficient) becomes greater than 1.

A3.6 Friction coefficients greater than unity

In metal machining, and elsewhere, friction coefficients > 1 have been measured in conditions in which asperities have been plastic but (τ/k) and (σ_n/k) have been too low for bulk plastic flow to be a possibility. What could account for this, that has not been considered in the previous sections?

Work hardening offers two possibilities. First, in the same way as it changes the hydrostatic pressure distribution along the primary shear plane in metal machining (Figures 2.11 and 6.9(b)), it can modify the pressure within a deforming asperity to reduce the mean value of p_r to a value less than k. However, there is likely only to be a small effect with the rake face asperities in machining, already work hardened by previous deformations. A second possibility imagines a little work hardening and high adhesion conditions, leading to the interface becoming stronger than the body of the asperity. Unstable asperity flow, with contact area growth larger than expected for non-hardening materials, has been observed by Bay and Wanheim (1976).

There is a second type of possibility. In the previous sections it has been assumed that an asperity is loaded by an amount W by contact with a counterface and that W does not change as sliding starts. For example, in Section 3.4.2 on junction growth of plastic contacts, it is written that the addition of a sliding force F to a real contact area creates an extra shear stress F/A_r which, if A_r does not increase, will cause τ_{max} to increase. This assumes that the stress W/A_r does not decrease.

If W is constant, the extra force F causes the two sliding surfaces to come closer to one another: it is this that enables A_r to grow. Green (1955) pointed out that, in a steady state of sliding (between two flat surfaces), the surfaces must be displaced parallel to one another. In that case any one junction must go through a load cycle. Figure A3.10 is based on Green's work. With increasing tangential displacement, asperities make contact, deform

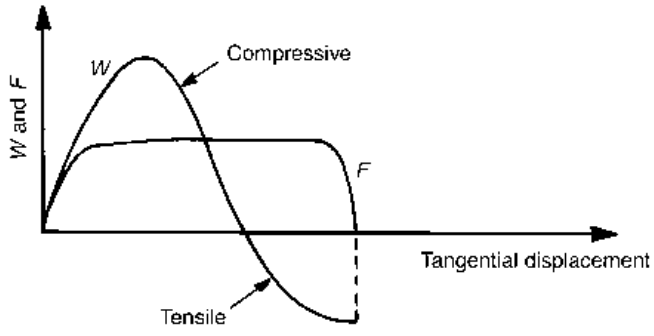


Fig. A3.10 Qualitative junction load history for zero normal displacement

and break. The load rises, passes through a maximum and falls, but the friction force rises and stays constant until failure. If, at any one time, there are many contacts in place, each at a random point in its life cycle, an average friction coefficient will be observed that is obtained from the areas under the curves of Figure A3.10, up to the point of failure. Green argued that when conditions were such that junctions failed when the load dropped to zero, the friction coefficient would be unity. Higher coefficients require junctions to be able to withstand tensile forces, as shown. The exact value of the friction coefficient will depend on the exact specification of how the surfaces come together and move apart; and on the junctions' tensile failure laws. Quantitative predictions do not exist.

References

Bay, N. and Wanheim, T. (1976) Real area of contact and friction stress at high pressure sliding contact. *Wear* **38**, 201–209.

Childs, T. H. C. (1973) The persistence of asperities in indentation experiments. *Wear* **25**, 3–16.

Green, A. P. (1955) Friction between unlubricated metals: a theoretical analysis of the junction model. *Proc. Roy. Soc. Lond.* **A228**, 191–204.

Greenwood, J. A. and Williamson, J. B. P. (1966) Contact of nominally flat surfaces. *Proc. Roy. Soc. Lond.* **A295**, 300–319.

Johnson, K. L. (1985) *Contact Mechanics*. Cambridge: Cambridge University Press.

Oxley, P. L. B. (1984) A slip line field analysis of the transition from local asperity contact to full contact in metallic sliding friction. *Wear* **100**, 171–193.

Sutcliffe, M. P. (1988) Surface asperity deformation in metal forming processes. *Int. J. Mech. Sci.* **30**, 847–868.

Appendix 4

Work material: typical mechanical and thermal behaviours

This appendix holds data that support Chapters 3 and 7, in the first instance. In Chapter 3, reference is made to yield and strain hardening behaviours of aluminium, copper, iron, nickel and titanium alloys, as determined by room-temperature, low strain rate, compression testing. Information on this is given in Section A4.1. The thermal conductivity, heat capacity and diffusivity ranges of these alloys, and their variations with temperature – also used in Chapter 3 to estimate temperature rises during machining – are tabulated in Section A4.2. In Chapter 7 the idea was developed that it is not the strain hardening behaviour of the work materials at room temperature and low strain rates that is needed. What is important for predicting chip formation in machining is the strain hardening behaviour at the temperatures and strain rates that actually occur. Data on this are presented in Section A4.3. This appendix is also a source for applications studies such as are described after Chapter 7.

A4.1 Work material: room temperature, low strain rate, strain hardening behaviours

Figures A4.1 to A4.3 contain representative strain hardening data for commercially pure samples of aluminium, copper, iron, nickel and titanium, and their alloys. The data have been obtained either from plane strain compression tests or from measuring the dependence of yield stress of sheet samples upon reduction of their thickness through cold rolling. In every case, the variation of shear stress, k, with shear strain, γ, is shown. k has been calculated from $\bar{\sigma}/\sqrt{3}$ and γ from $\bar{\varepsilon}\sqrt{3}$. The following is a brief commentary on the figures.

Copper and aluminium alloys (Figure A4.1)

The copper and copper alloys (left-hand panel) are all initially in the annealed state. They show the low initial yield and large amount of strain hardening typical of these face centred cubic metals. The aluminium and aluminium alloys (right-hand panel) show a similar behaviour, but generally at a lower level of stress. Some aluminium alloys can be hardened

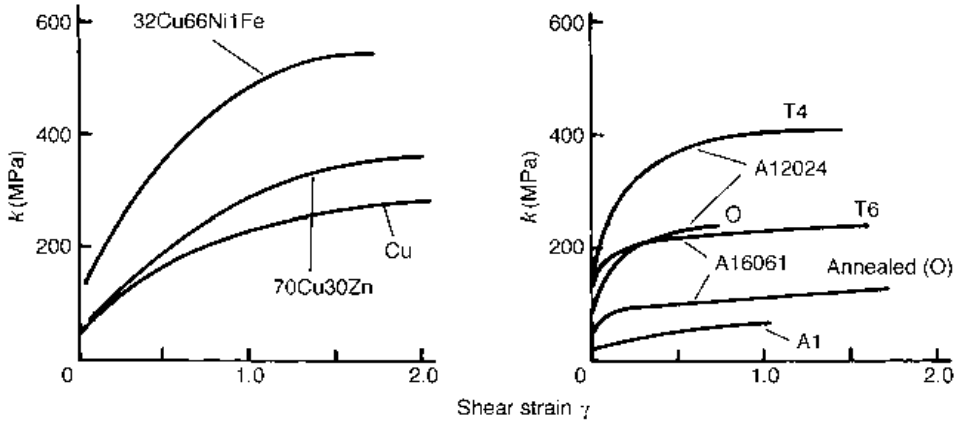


Fig. A4.1 Shear stress-strain behaviours of some copper and aluminium alloys

by ageing, either at room temperature (T4 temper) or above room temperature (T6). The examples of Al2024 (an alloy with 4Cu) and Al6061 (an alloy with 0.5Mg0.5Si) show the extent of hardening by this means. It could be argued that the 32Cu–66Ni alloy shown in the figure is more properly a nickel alloy: it is included here because Figure A4.3, on nickel alloys, is concerned more with Ni–Cr heat resistant alloys.

Ferrous alloys (Figure A4.2)

The left-hand panel contains data for carbon and low alloy steels as received from the hot rolling process. In this state their microstructure is a mixture of ferrite and pearlite (or, for the high carbon steel, pearlite and cementite). In contrast with the copper and aluminium alloys, these body centred cubic materials show a large variation in initial yield stress and, relative to the initial yield, less strain hardening. The right-hand panel shows two austenitic steels, a stainless steel (18Cr8Ni) and a high manganese steel (18Mn5Cr). These face centred cubic alloys show high strain hardening, both absolutely and relative to the body centred steels.

Nickel and titanium alloys (Figure A4.3)

All the nickel alloys (left-hand panel) shown in this figure are for high temperature, creep resistant, use. Commercially they are known as Inconel or Nimonic alloys. They are face centred cubic, with initial yield stress larger than copper alloys and large amounts of strain hardening. The titanium alloys (right-hand panel) are hexagonal close packed (h.c.p.) or mixtures of h.c.p. and body centred cubic. Their initial yield and strain hardening behaviours are intermediate between the face centred and body centred cubic materials.

Further elementary reading on metal alloys, their mechanical properties and uses can be found in Rollason (1973), Cottrell (1975) and Ashby and Jones (1986).

A4.2 Work material: thermal properties

Tables A4.1 to A4.3 contain information on the variation with temperature of the thermal conductivity, heat capacity and diffusivity of a range of work materials. The main single

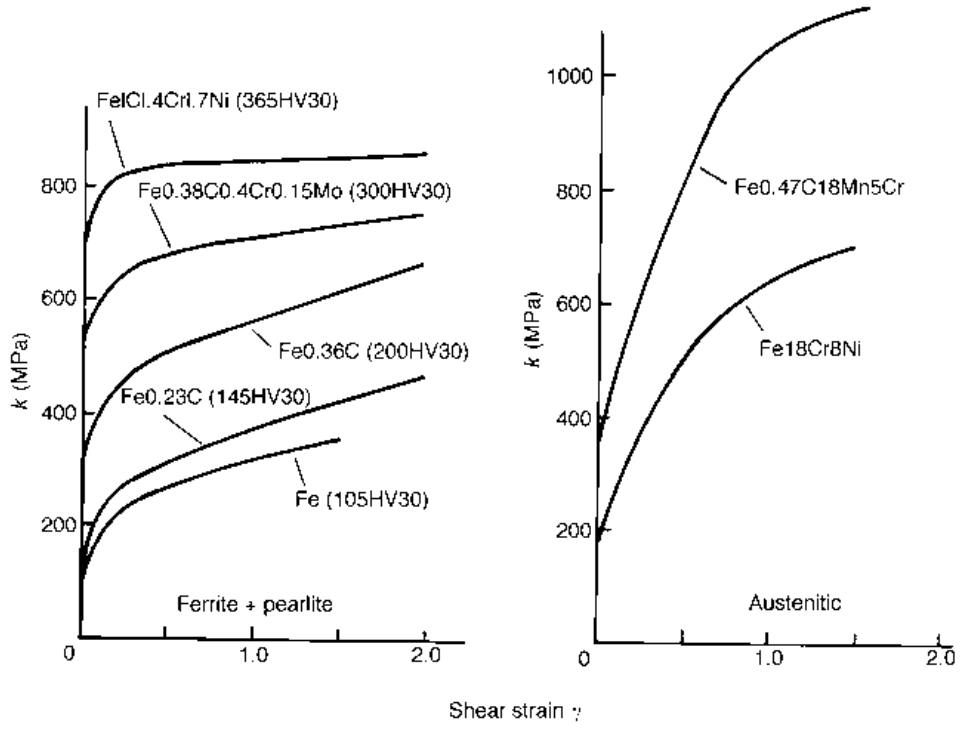


Fig. A4.2 Shear stress-strain behaviours of some ferritic/pearlitic and austenitic steels

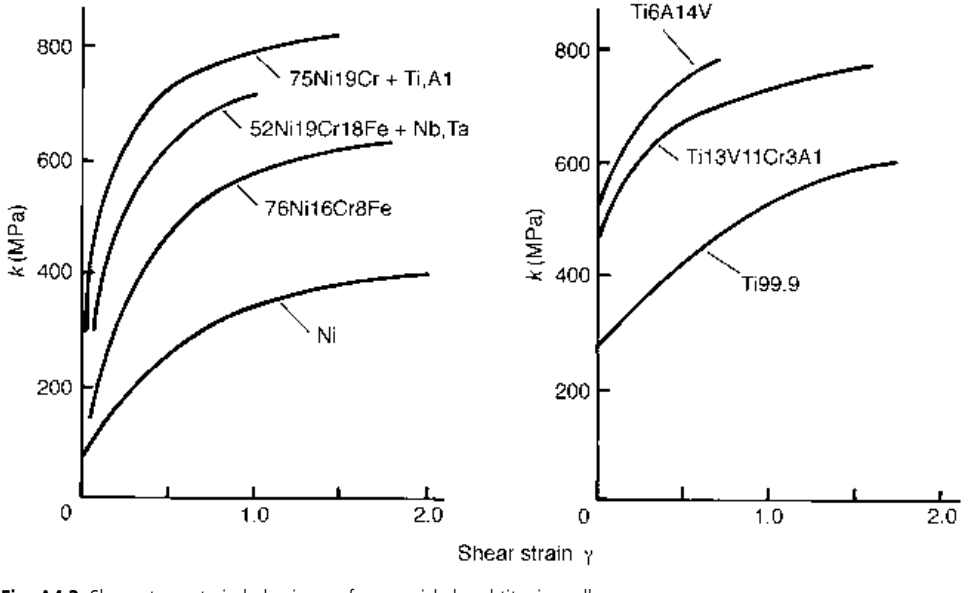


Fig. A4.3 Shear stress-strain behaviours of some nickel and titanium alloys

Table A4.1 Thermal conductivity [W/mK] of some work material groups

Alloy system	Temperature [°C]				
	0	200	400	600	800
Iron and steel					
pure iron*	85	64	50	38	31
0.04–0.25C	52–60	48–54	42–45	35–37	29–30
0.25–0.8C	51–52	46–48	39–42	33–35	29–30
0.8–1.2C	45–51	42–46	37–39	32–33	27–29
low alloy	25–49	30–45	32–40	30–35	27–30
ferritic stainless	21–24	22–25	23–26	24–27	25–29
austenitic stainless	14–17	15–18	17–21	22–25	26–29
high manganese	14	16	19	21	22
Aluminium					
pure, 1000 series	200–240	200–230	200–225	190–220	–
2000 to 7000 series	120–190	160–200	170–210		
Al-Si cast alloys	170–190	170–190	–	–	–
Copper					
pure copper*	380–400	375–395	350–380	320–360	270–330
60/70Cu–40/30Zn	90–120	90–140	110–150	120–150	–
90/95Cu–10/5Sn	50–80	70–100	90–120	–	–
60/90Cu–40/10Ni	20–50	30–70	45–90	60–110	–
Nickel					
pure nickel*	88–94	66–73	54–62	59–67	65–74
70Ni–30Cu	22	28	34	40	46
Superalloys**	11–12.5	11–14	13–16	16–20	20–24
Titanium					
pure titanium*	22	21	21	21	–
α, α–β, β alloys	5.5–8	8–12	10–17	12.5–21	15–25
Ti-6Al-4V	6.6–6.8	8.5–9.1	10.5–12.5	13–16	16–19

*: high and commercial purity; **: including cobalt- and ferrous-base superalloys.

Table A4.2 Heat capacity (MJ/m^3) of some work material groups

Alloy system	Temperature [°C]				
	0	200	400	600	800
Iron and steel					
pure iron, C, low alloy	3.5–3.8	4.1–4.3	4.7–5.0	5.6–5.9	6.7–7.1
ferritic stainless	3.5–4.1	3.8–4.3	4.3–5.0	6.0–7.1	5.8–6.2
austenitic stainless	3.5–4.5	4.2–4.7	4.5–4.8	5.2–5.5	5.6–5.9
high manganese	3.9	4.6	–	–	–
Aluminium					
pure, 1000 series	2.4–2.7	2.6–2.8	2.6–2.9	2.6–2.9	–
2000 to 7000 series	2.1–2.8	2.5–3.1	3.2–3.4	–	–
Al-Si cast alloys	2.3	2.6	2.8	–	–
Copper					
pure copper*	3.3	3.7	3.9	–	–
Zn, Sn, Ni alloys	3.2–3.4	3.6–3.8	3.8–4.0	–	–
Nickel					
pure nickel*	4.1	4.3	4.5	4.9	5.5
70Ni–30Cu	3.8	4.0	4.2	4.6	5.2
Superalloys**	3.3–3.5	3.5–3.6	3.7–3.8	4.2–4.3	4.5–4.8
Titanium					
pure Ti*, α, α–β, β alloys	2.3–2.5	2.55–2.75	2.75–3.05	3.0–3.4	3.3–3.8

*: high and commercial purity; **: including cobalt- and ferrous-base superalloys.

Table A4.3 Diffusivity (mm^2/s) of some work material groups

Alloy system	Temperature [°C] 0	200	400	600	800
Iron and steel					
pure iron*	23	15	10	6.5	4.5
0.04–0.25C	14–16	11–13	8.6–9.3	6.1–6.4	4.2–4.3
0.25–0.8C	14–15	11–12	8.1–8.7	5.7–6.1	4.2–4.3
0.8–1.2C	12–14	10–11	7.6–8.1	5.6–5.7	3.9–4.2
low alloy	7–13	7–11	6.6–8.2	5.2–6.1	3.9–4.3
ferritic stainless	5.1–6.8	5.1–6.5	4.6–6.0	3.4–4.5	4.0–5.0
austenitic stainless	3.2–3.7	3.5–4.0	3.8–4.4	4.0–4.8	4.4–5.2
high manganese	3.6	3.5	–	–	–
Aluminium					
pure, 1000 series	78–100	75–90	70–80	73–76	–
2000 to 7000 series	52–75	55–72	50–65	–	–
Al-Si cast alloys	75–85	65–75	–	–	–
Copper					
pure copper*	115–120	100–110	90–100	–	–
60/70Cu–40/30Zn	25–35	28–35	27–33	25–30	–
90/95Cu–10/5Sn	15–25	20–25	23–30	–	–
60/90Cu–40/10Ni	6–15	8–18	12–22	–	–
Nickel					
pure nickel*	21–23	15–17	12–14	12–14	12–14
70Ni–30Cu	7.4	7.0	8.1	8.7	8.9
Superalloys**	2.8–3.8	3.1–3.9	3.5–4.2	3.8–4.7	4.2–5.3
Titanium					
pure titanium*	9.5	7.6	6.8	–	–
α, α–β, β alloys	2.2–5.0	2.7–5.5	3.2–6.0	3.7–6.4	3.7–6.6
Ti-6Al-4V	2.2–3.0	2.7–3.5	3.2–3.8	3.8–4.2	3.8–4.7

*: high and commercial purity; **: including cobalt- and ferrous-base superalloys.

source of information has been the ASM (1990) *Metals Handbook* but it has been necessary also to gather information from a range of other data sheets.

A4.3 Work material: strain hardening behaviours at high strain rates and temperatures

Published data from interrupted high strain and heating rate Hopkinson bar testing (Chapter 7.4) are gathered here. Stress units are MPa and temperatures T are °C. Strain rates are s^{-1}.

A4.3.1 Non-ferrous face centred cubic metals

For T from 20°C to 300°C, strain rates from 20 s^{-1} to 2000 s^{-1} and strains from 0 to 1, the following form of empirical equation for flow stress, including strain path dependence, has been established (Usui and Shirakashi, 1982)

$$\bar{\sigma} = A\left(e^{-\frac{B}{T+273}}\right)\left(\frac{\dot{\bar{\varepsilon}}}{1000}\right)^{M}\left(\int_{\text{strain path}}\left(\frac{\dot{\bar{\varepsilon}}}{1000}\right)^{m} d\bar{\varepsilon}\right)^{N} \qquad \text{(A4.1a)}$$

For the special case of straining at constant strain rate, this simplifies to

$$\bar{\sigma} = A\left(\mathrm{e}^{-\frac{\mathrm{B}}{T+273}}\right)\left(\frac{\dot{\bar{\varepsilon}}}{1000}\right)^{M+mN}\bar{\varepsilon}^{N} \tag{A4.1b}$$

Coefficients A, B, M, m and N for the following annealed metals are as follows.

Metal	A	B	M	m	N
Aluminium	107	153	0.057	0.064	0.3
α-brass	720	56.7	0.024	0.06	0.5

A4.3.2 Pearlitic carbon and low alloy steels

In early studies, an equation similar to equation (A4.1a) was used but for a changed exponential temperature term and a term dependent on temperature within the strain path integral. Later, this was developed to

$$\sigma = A\left(\frac{\dot{\bar{\varepsilon}}}{1000}\right)^{M}\mathrm{e}^{aT}\left(\frac{\dot{\bar{\varepsilon}}}{1000}\right)^{m}\left(\int_{\text{strain path}}\mathrm{e}^{-aT/N}\left(\frac{\dot{\bar{\varepsilon}}}{1000}\right)^{-m/N}\mathrm{d}\bar{\varepsilon}\right)^{N} \tag{A4.2a}$$

to give a particularly simple form in constant strain rate and temperature conditions:

$$\bar{\sigma} = A\left(\frac{\dot{\bar{\varepsilon}}}{1000}\right)^{M}\bar{\varepsilon}^{N} \tag{A4.2b}$$

A range of measured coefficients is given in Table A4.4, valid for T from 20˚C to 720˚C, strain rates up to 2000 s^{-1} and strains up to 1.

Table A4.4 Flow stress data for annealed or normalized carbon and low alloy steels

Steel	Coefficients of equation (A4.2)
0.1C [1]*	$A = 880\mathrm{e}^{-0.0011T} + 167\mathrm{e}^{-0.00007(T-150)^2} + 108\mathrm{e}^{-0.00002(T-350)^2} + 78\mathrm{e}^{-0.0001(T-650)^2}$ $M = 0.0323 + 0.000014T$ $N = 0.185\mathrm{e}^{-0.0007T} + 0.055\mathrm{e}^{-0.000015(T-370)^2}$ $a = 0.00024$ $m = 0.0019$
0.45C [2]*	$A = 1350\mathrm{e}^{-0.0011T} + 167\mathrm{e}^{-0.00006(T-275)^2}$ $M = 0.036$ $N = 0.17\mathrm{e}^{-0.001T} + 0.09\mathrm{e}^{-0.000015(T-340)^2}$ $a = 0.00014$ $m = 0.0024$
0.38C –Cr–Mo [3]*	$A = 1460\mathrm{e}^{-0.0013T} + 196\mathrm{e}^{-0.000015(T-400)^2} - 39\mathrm{e}^{-0.01(T-100)^2}$ $M = 0.047$ $N = 0.162\mathrm{e}^{-0.001T} + 0.092\mathrm{e}^{-0.0003(T-380)^2}$ $a = 0.000065$ $m = 0.0039$
0.33C –Mn–B [3]*	$A = 1400\mathrm{e}^{-0.0012T} + 177\mathrm{e}^{-0.000030(T-360)^2} - 107\mathrm{e}^{-0.001(T-100)^2}$ $M = 0.0375 + 0.000044\mathrm{T}$ $N = 0.18\mathrm{e}^{-0.0012T} + 0.098\mathrm{e}^{-0.0002(T-440)^2}$ $a = 0.000065$ $m = 0.00039$
0.36C –Cr–Mo Ni[4]*	$A = 1500\mathrm{e}^{-0.0018T} + 380\mathrm{e}^{-0.00001(T-445)^2} + 160\mathrm{e}^{-0.0002(T-570)^2}$ $M = 0.017 + 0.000068\mathrm{T}$ $N = 0.136\mathrm{e}^{-0.0012T} + 0.07\mathrm{e}^{-0.0002(T-465)^2}$ $a = 0.00006$ $m = 0.0025$

*[1] Maekawa *et al.* (1991); [2] Maekawa (1998); [3] Maekawa *et al.* (1996); [4] Childs *et al.* (1990)]

A4.3.3 Other metals

The behaviour of some austenitic steels and titanium alloys has also been studied. An 18%Mn-18%Cr steel's flow stress behaviour has been fitted to equation (A4.2b), with ($\bar{\varepsilon}$/0.3) – replacing $\bar{\varepsilon}$ –, with coefficients (Maekawa *et al.* 1994a)

$$A = 2010e^{-0.0018T}$$

$$M = 0.0047e^{0.0036T} \qquad N = 0.346e^{-0.0008T} + 0.11e^{-0.000032(T-375)^2}$$

A different form has been found appropriate for an austenitic 18%Mn-5%Cr steel, with negligible strain path dependence (Maekawa *et al.* 1993):

$$\bar{\sigma} = 3.02\dot{\bar{\varepsilon}}^{0.00714}[45400/(273 + T) + 58.4 + a(860 - T)\bar{\varepsilon}^{b}]$$

where, for $\bar{\varepsilon} \leq 0.5$ $a + 0.87$, $b = 0.8$; $\bar{\varepsilon} \geq 0.5$ $a = 0.57$, $b = 0.2$

Other forms have been given for a Ti-6Al-4V alloy (Usui *et al.* 1984) and a Ti-6Al-6V-2Sn alloy (Maekawa *et al.* 1994b). For the Ti-6Al-4V alloy:

$$\bar{\sigma} = A(\dot{\bar{\varepsilon}}/1000)^{M}e^{aT}(\dot{\bar{\varepsilon}}/1000)^{m}\left\{c + \left[d + \int_{\text{strain path}} e^{-aT/N}(\dot{\bar{\varepsilon}}/1000)^{-m/N}d\bar{\varepsilon}\right]^{N}\right\}$$

with $A = 2280e^{-0.00155T}$

$$M = 0.028 \qquad N = 0.5 \qquad a = 0.0009 \qquad m = -0.015 \qquad c = 0.239 \qquad d = 0.12$$

The data for the Ti-6Al-6V-2Sn alloy were fitted to equation (A4.2a) with

$$A = 2160e^{-0.0013T} + 29e^{-0.00013(T-80)^2} + 7.5e^{-0.00014(T-300)^2} + 47e^{-0.0001(T-700)^2}$$

$$M = 0.026 + 0.0000T \qquad N = 0.18e^{-0.0016T} + 0.015e^{-0.00001(T-700)^2}$$

$$a = 0.00009 \qquad m = 0.0055$$

References

ASM (1990) *Metals Handbook*, 10th edn. Ohio: ASM.

Ashby, M. F. and Jones, D. R. H. (1986) *Engineering Materials*, Vol. 2. Oxford: Pergamon Press.

Childs, T. H. C. and Maekawa, K. (1990) Computer aided simulation and experimental studies of chip flow and tool wear in turning low alloy steels by cemented carbide tools. *Wear* **139**, 235–250.

Cottrell, A. (1975) *An Introduction to Metallurgy*, 2nd edn. London: Edward Arnold.

Maekawa, K., Kitagawa, T. and Childs, T. H. C. (1991) Effects of flow stress and friction characteristics on the machinability of free cutting steels. In: *Proc. 2nd Int. Conf. on Behaviour of Materials in Machining – Inst. Metals London Book 543*, pp. 132–145.

Maekawa, K., Kitagawa, T., Shirakashi, T. and Childs, T. H. C. (1993) Finite element simulation of three-dimensional continuous chip formation processes. In: *Proc. ASPE Annual Meeting*, Seattle, pp. 519–522.

Maekawa, K., Ohhata, H. and Kitagawa, T. (1994a) Simulation analysis of cutting performance of a three-dimensional cut-away tool. In Usui, E. (ed.*), Advancement of Intelligent Production*. Tokyo: Elsevier, pp. 378–383.

Maekawa, K., Ohshima, I., Kubo, K. and Kitagawa, T. (1994b) The effects of cutting speed and feed on chip flow and tool wear in the machining of a titanium alloy. In: *Proc. 3rd Int. Conf. on Behaviour of Materials in Machining*, Warwick, 15–17 November pp. 152–167.

Maekawa, K., Ohhata, T., Kitagawa, T. and Childs, T. H. C. (1996) Simulation analysis of machinability of leaded Cr-Mo and Mn-B structural steels. *J. Matls Proc. Tech.* **62**, 363–369.

Maekawa, K. (1998) private communication.

Rollason, E. C. (1973) *Metallurgy for Engineers*, 4th edn. London: Edward Arnold.

Usui, E. and Shirakashi, T. (1982) Mechanics of machining – from descriptive to predictive theory. *ASME Publication* **PED 7**, 13–35.

Usui, E., Obikawa, T. and Shirakashi, S. (1984) Study on chip segmentation in machining titanium alloy. In: *Proc. 5th Int. Conf. on Production Engineering*, Tokyo, 9–11 July, pp. 235–239.

Appendix 5

Approximate tool yield and fracture analyses

This appendix supports Section 3.2. The material of Section A5.1 is also referred to in Appendix 1.2.4.

A5.1 Tool yielding

The required tool hardnesses to avoid the yielding shown in Figure 3.19 have been obtained by a method due to Hill (1954).The requirement that the tool does not yield at its apex, together with force equilibrium in the tool, limits the difference between the rake face contact stress and the zero stress on the clearance face and hence places a maximum value on the allowable rake face contact stress.

With the cylindrical polar coordinate system shown in Figure A5.1(a), in which the origin is at the tool apex and the angular variable θ varies from 0 on the rake face to β on the clearance face, and in which the stresses σ_r, σ_θ and τ are positive as shown, the radial and circumferential equilibrium equations are

$$r\frac{\mathrm{d}\sigma_r}{\mathrm{d}r} + (\sigma_r - \sigma_\theta) + \frac{\mathrm{d}\tau}{\mathrm{d}\theta} = 0 \qquad \text{(A5.1a)}$$

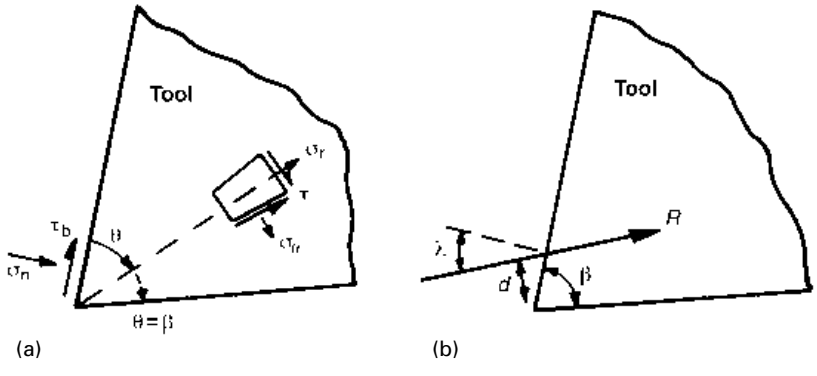


Fig. A5.1 Coordinate systems and definitions for the analysis of tool (a) yielding and (b) fracture

$$r\frac{\mathrm{d}\tau}{\mathrm{d}r} + 2\tau + \frac{\mathrm{d}\sigma_\theta}{\mathrm{d}\theta} = 0 \tag{A5.2a}$$

At the apex, where $r = 0$, these become

$$(\sigma_r - \sigma_\theta) + \frac{\mathrm{d}\tau}{\mathrm{d}\theta} = 0 \tag{A5.1b}$$

$$2\tau + \frac{\mathrm{d}\sigma_\theta}{\mathrm{d}\theta} = 0 \tag{A5.2b}$$

To avoid yielding of the tool, the shear yield stress of which is k_t,

$$\frac{1}{4}(\sigma_\theta - \sigma_r)^2 + \tau^2 < k_t^2 \tag{A5.3}$$

If τ is written as a fraction of k_t

$$\tau = k_t \sin 2\phi \tag{A5.4}$$

where ϕ varies between $\pm \pi/4$, it may be shown, after substituting $(\sigma_r - \sigma_\theta)$ from equation (A5.1b) into equation (A5.3), that equation (A5.3) leads to a limitation of the rate of change of ϕ with θ

$$\left| \frac{\mathrm{d}\phi}{\mathrm{d}\theta} \right| < 1 \tag{A5.5}$$

Furthermore, an expression for the contact stress σ_n on the rake face relative to the value zero on the clearance face is found by integrating equation (A5.2b). After dividing both sides of equation A5.2b by k_{work}

$$\frac{\sigma_n}{k_{work}} = -2\frac{k_t}{k_{work}}\int_0^\beta \sin 2\phi \mathrm{d}\theta \tag{A5.6}$$

The largest value of σ_n/k_{work} is obtained when the integral takes its largest negative value. Figure (A5.2) shows the variation of ϕ with θ that gives that largest negative value: at $\theta = \beta$, $\phi = 0$; and at $\theta = 0$, ϕ is determined by the friction contact stress on the rake face. In Chapter 3 (Figure 3.18) extreme examples of friction stress were considered, up to k_{work} during steady chip creation, but zero at the start of a cut:

$$\phi \equiv \phi_0 = 0, \qquad \tau_f = 0 \tag{A5.7a}$$

$$\phi \equiv \phi_0 = -\frac{1}{2}\sin^{-1}\left(\frac{k_{work}}{k_t}\right), \qquad \tau_f = k_{work} \tag{A5.7b}$$

For $0 < \theta < \beta$, ϕ takes the smallest values allowed by equation (A5.5). Figure A5.2(a) is for the case $\beta > \pi/2 - \phi_0$. If $\beta < \pi/2 - \phi_0$, Figure A5.2(b) applies. Integration of equation (A5.6) for the dependence of ϕ on θ shown in Figure A5.2(a) or (b) as appropriate gives a maximum value of σ_n/k_{work} depending on k_t/k_{work} and β. Inversely, for a specified σ_n, for example

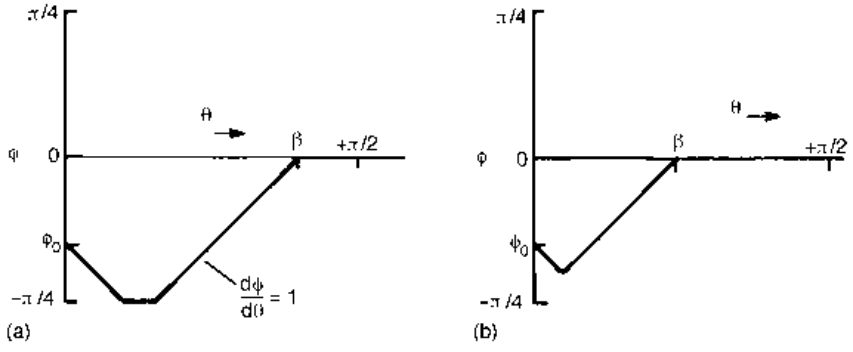


Fig. A5.2 Variations of ϕ with θ that maximize σ_n/k_{work}

$5k_{work}$ or $2.5k_{work}$, a minimum ratio of tool to work shear yield stress to avoid yield can be derived. Taking the tool's Vickers Hardness HV to equal $5k_t$, relations between tool hardness, k_{work} and β to avoid tool yielding can be derived. Thus, the HV/β relations dependent on k_{work} shown in Figure 3.19 are obtained.

A5.2 Tool fracture

Figure A5.1(b) shows a wedge-shaped tool with a line force R per unit length acting at a friction angle λ at a distance d from the apex of the wedge. This force is equivalent to a force R acting at the apex, with a moment $M = Rd$. A classical result of stressing a wedge (Coker and Filon, 1931) is that on the rake face the tensile stress at a distance r from the apex is

$$\sigma_r = -\frac{2R}{r}\left[\frac{\cos\frac{\beta}{2}\sin\left(\lambda+\frac{\beta}{2}\right)}{\beta+\sin\beta} - \frac{\sin\frac{\beta}{2}\cos\left(\lambda+\frac{\beta}{2}\right)}{\beta-\sin\beta}\right] + \frac{2M}{r^2}\frac{\sin\beta}{\beta\cos\beta-\sin\beta} \tag{A5.8}$$

The sizes of tool transverse rupture stress (TRS) relative to the k_{work} required to avoid failure, and which are presented in Figure 3.19, have been obtained by replacing the distributed tool rake face contact stresses by their equivalent line force and moment at the apex, substituting these in equation (A5.8) and differentiating with respect to r to obtain the position and hence the value of the maximum tensile stress. It is supposed that a tool will fracture when the maximum tensile stress is the TRS. The results presented in Figure 3.19 are for the case of a tool entering a cut, assuming that $\tau_f = 0$ and σ_n is constant and equal to $5k_{work}$ over the contact length l between the work and tool. It is found for this example that the maximum tensile stress occurs at $r \approx l$. To replace the distributed stress by the equivalent line force and moment is only marginally justifiable: the treatment is only approximate.

References

Coker, E. G. and Filon, L. N. G. (1931) *A Treatise on Photoelasticity*. London: Cambridge University Press, pp. 328, 367.

Hill, R. (1954) On the limits set by plastic yielding to the intensity of singularities of stress. *J. Mech. Phys. Solids*, **2**, 278–285.

Appendix 6

Tool material properties

More detail is given here than in Chapter 3 of the materials that make up the main tool groupings.

A6.1 High speed steels

The high speed steels are alloy steels with about 0.75% to 1.5% carbon (C), 4% to 4.5% chromium (Cr), between 10% and 20% tungsten (W) and molybdenum (Mo); they can also have vanadium (V), up to 5%, and cobalt (Co), up to 12%. They are strengthened by heating to high temperature (around 1150 to 1250˚C), just below the solidus; then quenching in two stages (to avoid thermal cracking) – to the range 500˚C to 600˚C and then to room temperature; and then tempering typically between 500˚C and 560˚C. Tempering causes hardening by the precipitation of fine carbides. More details may be found in metallurgical texts such as those by Trent (1991) and Hoyle (1988).

There are two series of materials, the T series which is based on W (with no Mo), and the M series which substitutes Mo for some of the W. There are no major technical advantages of one series over the other. The choice is one of cost, varying with the availability of these two elements. The basic grades in each series contain 0.75% to 0.85% C and 4% to 4.5% Cr, with a small amount of V (<2%) but no Co. The addition of extra V, with extra C as well, results in the formation of hard vanadium carbides on tempering. These increase the alloy's room temperature hardness and abrasion resistance but at the expense slightly of its toughness. The addition of Co improves hot hardness, also at the expense of toughness. Table A6.1 gives the nominal compositions of a range of grades.

Table A6.1 Sample compositions of some high speed steels

Grade	Composition (wt. %, balance Fe)					
	C	Cr	W	Mo	V	Co
T1	0.75	4	18	–	1	–
M2	0.85	4	6.5	5	2	–
T6	0.8	4	20.5	–	1.5	12
T15	1.5	4.5	13	–	5	5
M42	1.05	4	1.5	9.5	1	8

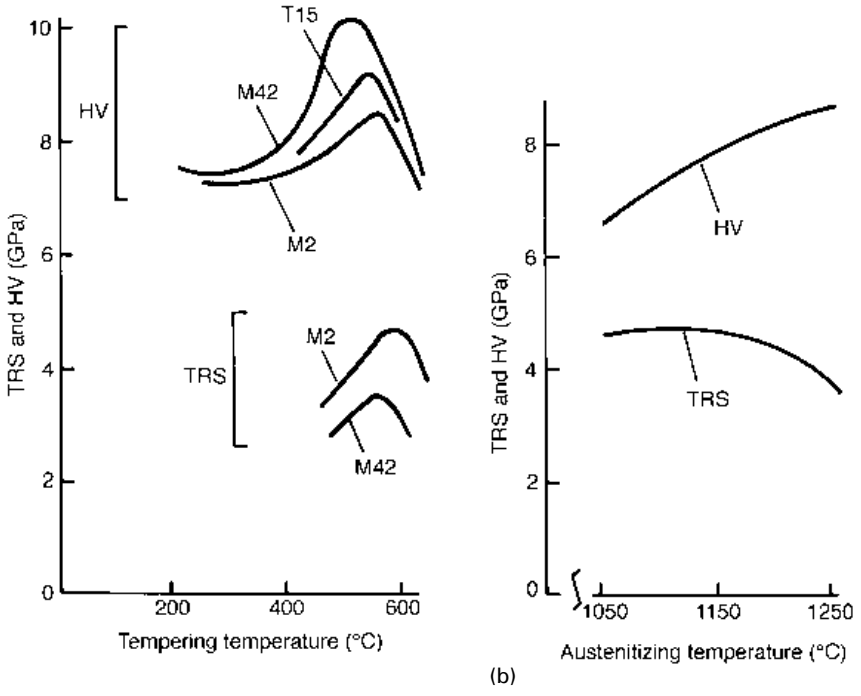


Fig. A6.1 Variations of room temperature HV and TRS with (a) tempering and (b) austenitizing temperature, for a range of high speed steels as indicated

Figure A6.1(a) shows how the room temperature Vickers Hardness (HV) of M2, T15 and M42, and the room temperature tensile rupture stress (TRS) of M2 and M42, typically vary with tempering temperature after quenching from the recommended austenitizing temperatures for these alloys. Figure A6.1(b) shows, for M2, how HV and TRS vary with austenitizing temperature after tempering at 560°C. The data have been derived mainly from Hoyle (1988), converting from Rockwell to Vickers Hardness, with additional data from other sources. The data are presented to show the sensitivity of mechanical properties to composition and heat treatment.

Traditionally, high speed steels have been shaped by hot working. Now, powder metallurgy technology is used to make high speed steel indexable inserts. HV and TRS values are not much changed but there is evidence that fracture toughness (K_{IC} values) can be higher for powder metallurgy than wrought products. Sheldon and Wronski (1987) give K_{IC} at room temperature for sintered T6 as 30 MP $m^{1/2}$ whereas wrought T6 heat treated in the same way has K_{IC} = 15 to 20 MP $m^{1/2}$. This paper also gives the temperature dependence of TRS quoted in Chapter 3 (Figure 3.22).

A6.2 Cemented carbides and cermets

Cemented carbide and cermet cutting tools consist of hard carbide (or carbo-nitride) grains, bonded or cemented together by up to around 20% by weight of cobalt or nickel,

with minor additions of other metals (such as molybdenum or chromium) possible. The hardness of the tools reduces and the toughness increases as the proportion of the metal binder phase is increased.

Cemented carbides and cermets are manufactured by sintering. The reactions that take place during sintering are extremely complex and the creation of good cutting tool grades requires a close attention to detail. A comprehensive monograph has been published (Schwarzkopf and Keiffer, 1960) and since then research reviews have appeared at regular intervals (Exner, 1979; Gurland, 1988). However, from a user's point of view, the elements of cemented carbide tool development are quite clear.

The earliest cemented carbides, developed in the 1920s, consisted of tungsten carbide (WC) cemented together by cobalt (Co). It soon became clear that this material was not suitable for machining steels at high cutting speeds. The WC dissolved in the steel at the temperatures generated by cutting, leading to rapid cratering of the rake face of the cutting tool. It was found that the system titanium carbide (TiC)-Co was more chemically resistant to steel, although cemented carbides based on TiC alone were more brittle than WC-Co. Toughness could be recovered by adding tantalum carbide (TaC). During the 1930s, cemented carbides based on WC-TiC-TaC-Co started to be developed. Tools based on WC-Co, suitable for cutting non-ferrous metals (and also cast iron, which does not get hot enough in machining to trigger rapid dissolution of WC, so tool life remains determined by flank wear) are now known as K-type carbides and those based on WC-TiC-TaC-Co, for steel cutting, as P-type. (In practice, the tantalum carbide often includes niobium; one should then refer to Ta(Nb)C.) During the 1950s, an alternative system for steel cutting began to be studied, based on TiC cemented mainly by nickel (Ni). These have developed to titanium carbo-nitrides (Ti(C,N)) bonded by Ni (with minor amounts of WC and Co), and are known as cermets. Much more detailed data are available on the composition and properties of the K- and P-type carbides (and M-type as well – see later) than on the cermets. The remainder of this section will concentrate mainly on the carbide grades.

The description K-, P- and M-type carbides, although it closely relates to carbide composition, in fact refers not to composition but to performance. An international Standard (ISO 513, 1991) classifies cemented carbide cutting tools by type and grade. Type refers to suitability for steel cutting (P) or non-ferrous materials (K) or to a compromise between the two (M). Grade refers to whether the tool material's mechanical properties have been optimized for hardness and hence abrasive wear resistance, or for toughness. Wear resistance is more important than toughness for low feed, finishing cuts. Toughness is more important for high feed, roughing or interrupted cuts. Grades run from 01 to 50, as properties change from hard to tough.

Different manufacturers achieve a particular tool performance by minor differences of the processing route, so that there is not a one-to-one relation between a tool's type and grade on the one hand and its composition on the other. This is illustrated in Figure A6.2. Each row of the figure presents data on composition, hardness and transverse rupture stress (at room temperature) for one manufacturer's range of tool materials, according to information published by Brookes (1992). The first row is data from a German manufacturer, the second is from a major international company and the third is from a Japanese producer. Each data point in the left hand column represents the TiC-TaC and Co weight % of one tool material (the balance is WC). What type and grade is assigned to the material is indicated by the solid and dashed lines. The ranges of compositions giving P-,

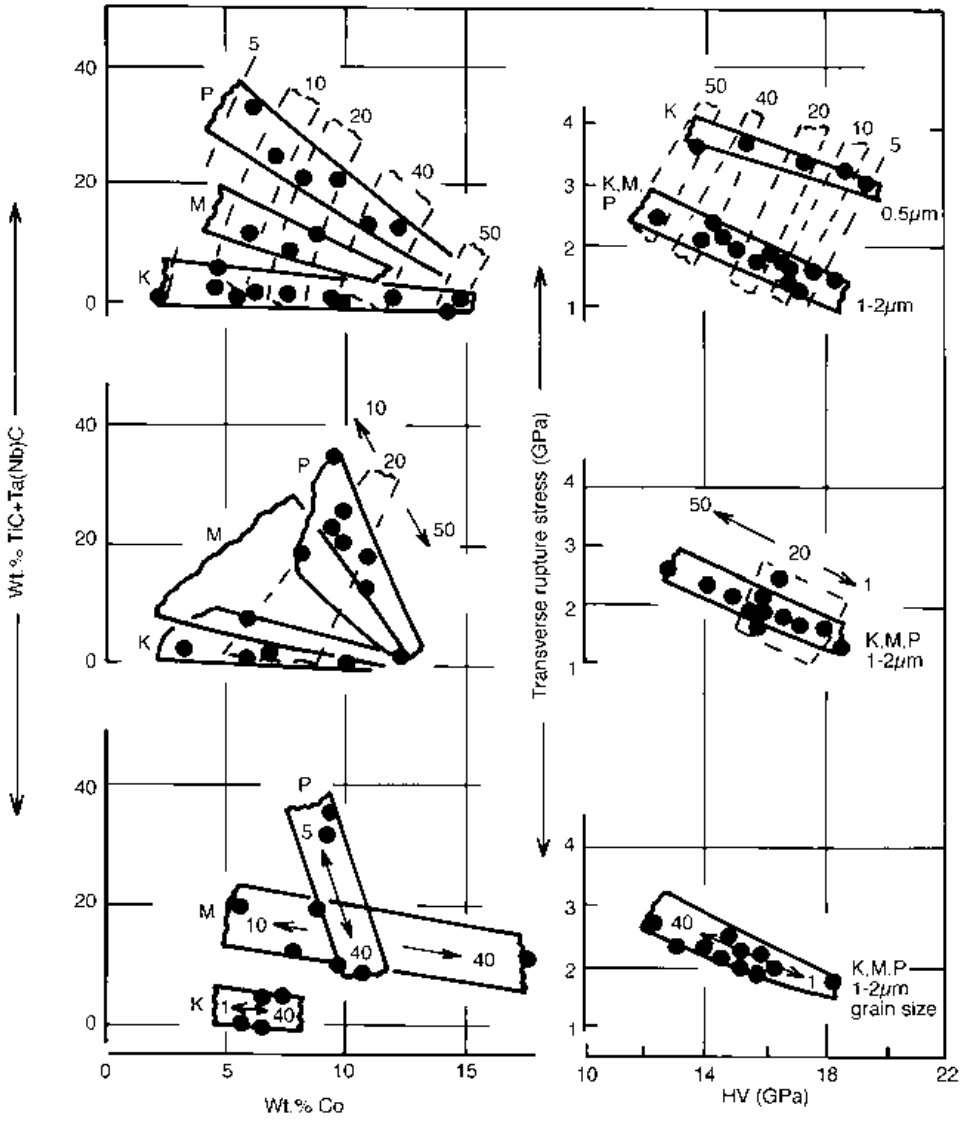


Fig. A6.2 Composition and mechanical property differences of cemented carbide cutting tools classified according to ISO 513 (1991) by three different manufacturers

M- and K-types are slightly different for each producer. So are the ranges of compositions giving the different grades.

The right-hand column shows the relation between transverse rupture stress and hardness for all the grades. It can be seen that the relation depends on the carbide grain size. All three manufacturers produce tool materials of 1 to 2 μm grain size. These have the same relation between transverse rupture stress and hardness, independent of K-, M- and P-type. However, one set of data, in the first row, is for material of sub-micrometre

grain size: it shows a greater transverse rupture stress for a given hardness than the coarser grained material. Such a fine grain size is only achievable with WC-Co (K-type) materials.

The mechanical and physical properties of commercial cemented carbide cutting tools broadly depend on the wt. % of Co, the wt. % of TiC-TaC and the grain size of the material. Rather than describe the material by type and grade, the remainder of this section will describe it by these quantities. For convenience, the classification by amount of TiC-TaC will be by whether the amount of this by weight is in the range 0–3%, 8–15% or 19–35%. The data presented in Brookes (1992) show that very few cutting tool materials have amounts of TiC-TaC outside these ranges.

Figure A6.3 shows that the room temperature hardness of a cemented carbide depends mainly on cobalt content and grain size. Figure A6.4 shows that quantities such as thermal conductivity, K, heat capacity, ρC, thermal expansion coefficient, α_e, Young's modulus, E, and thermal shock resistance, $(\text{TRS}.K)/(E\alpha_e)$, are most influenced by the type of carbide present. Figures A6.2 to A6.4 are the main source of information for the cemented carbide data presented in Chapter 3.

Such detailed information on the properties of cermets is not available in the open literature. Table A6.2 presents data for one manufacturer's products. TiC and TiN are the major hard phase, with WC as a minor part. Ni is the major binder metal, with Co as a minor part. Less complete or differently presented data from other manufacturers, extracted from Brookes (1992) are gathered in Table A6.3.

The densities of the cermets are almost half those of the cemented carbides (the densities of which, because of the high specific weight of tungsten, are around 14 000 to 15 000 kg/m^3 for the WC-Co types and 10 000 to 13 000 kg/m^3 for the high TiC-TaC-Co types). The cermets are mainly described as P-types, although some manufacturers also recommend them as K-types, but because of their limited toughness (TRS < 2.5 GPa, compared with up to 4 GPa for fine grained WC-Co materials), none of them are recommended for heavy duty use, above 30-grade.

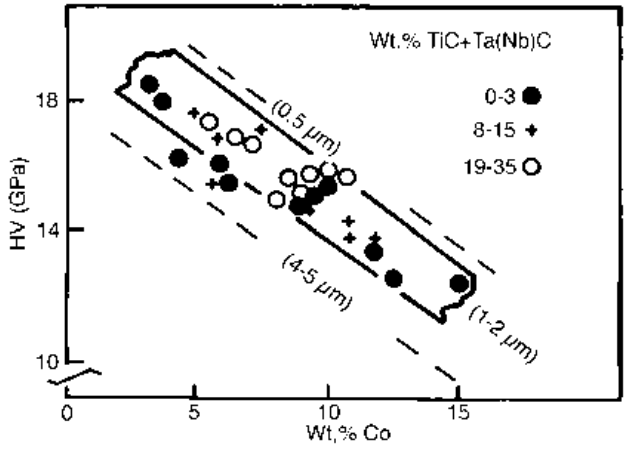


Fig. A6.3 Hardness dependence on % Co and grain size, for cemented carbides

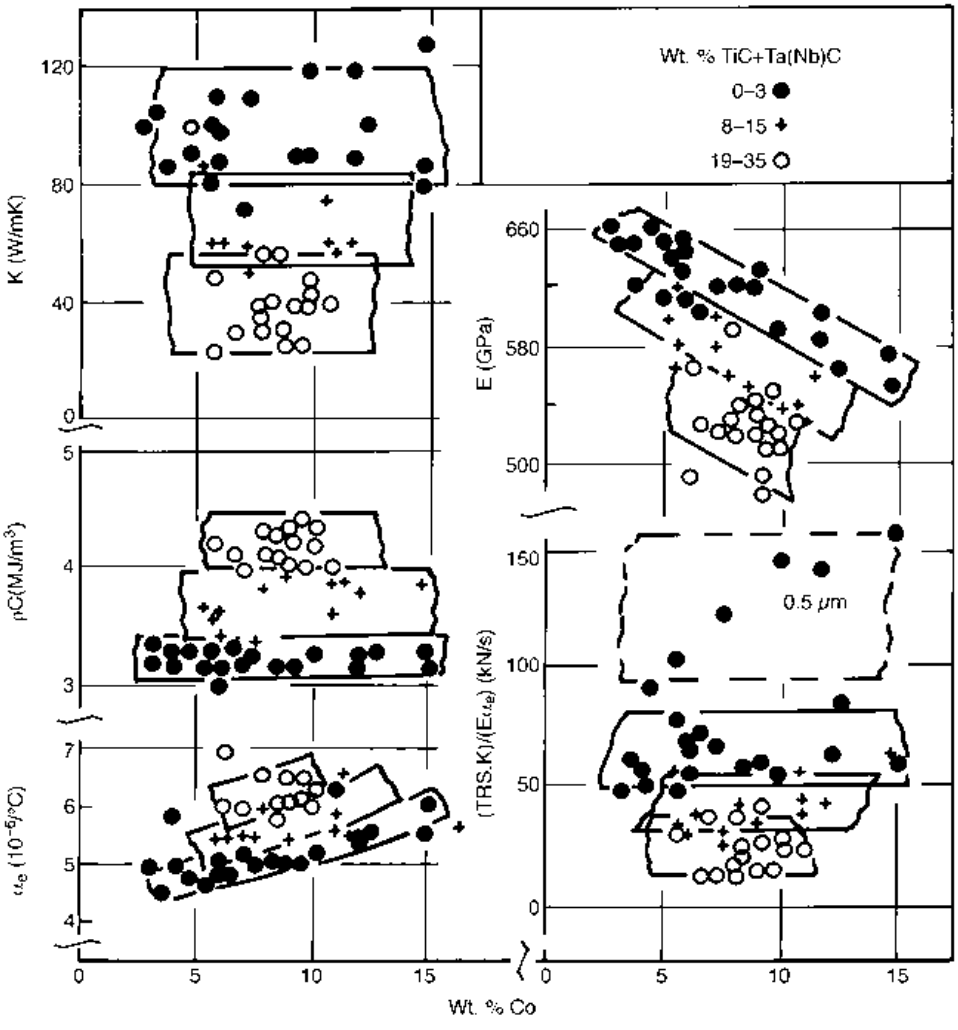


Fig. A6.4 Composition dependence of some properties of cemented carbides

Table A6.2 One manufacturer's range of cermet tool materials

ISO code	Wt. % Ti(C,N) + WC	Wt. % Ni + Co	Grain size [μm]	ρ [kg/m^3]	*HV* [GPa]	*TRS* [GPa]	*K* [W/mK]	*E* [GPa]	α_e [10^{-6} K^{-1}]
P/K01–05	95	5	1	6800	18.1	1.3	11	410	6.7
P10–P15	86	14	1	7100	15.5	1.65	12	400	7.2
P/K05–15	89	11	<1	7000	16.5	1.65	14	410	7.6
P10–P25	85	15	<1	7000	15.2	2.0	19	390	7.4

Table A6.3 Cermet tool materials' data from a range of other manufacturers

ISO code	Wt. % Ti(C,N)	WC	Other carbide	Ni	Co	Grain size [μm]	ρ [kg/m^3]	*HV* [GPa]	*TRS* [GPa]	*K* [W/mK]
P01–10	50	16	20	6	8	< 2	6900	16.2	1.2	20
P05–25	49	16	15	8	12	< 2	7000	14.2	1.8	20
P01–15	48	16	20	5	11	–*	7000	15.7	–*	20
P05	Total carbide: 94		Total metal: 6			–*	6100	17.2	1.8	–*
P10	Total carbide: 86		Total metal: 14			–*	7000	15.7	2.3	11
P20	Total carbide: 82		Total metal: 18			–*	7000	14.2	2.5	16
P01–20	Total carbide: 87		Total metal: 13			2	6600	16.7	1.5	25
P10–30	Total carbide: 83		Total metal: 17			2	7000	15.2	1.8	27
P10–30	Not given					–*	7400	16.0	1.9	29

*: data not provided.

A6.3 Ceramics and superhard materials

Even less systematically detailed information than for cermet tools is available for the composition and properties of ceramic and superhard materials.

Data for tools based on alumina, extracted from Brookes (1992), are gathered in Table A6.4. There are three sub-groups of material. The first, called white alumina because of its colour, is pure alumina together with minor additions (headed 'other' in the table) to promote sintering. These sintering aids can be either magnesium oxide (MgO) or zirconia (ZrO_2): for tool grade aluminas, ZrO_2 is predominantly used. The second group is the black aluminas: alumina to which is added TiC. The third group is SiC whisker reinforced alumina. The data demonstrate that the black aluminas are harder but no tougher than the white aluminas. Silicon carbide whisker reinforcement increases toughness without improving hardness, relative to the black aluminas. All the materials are developed, according to their ISO classification, for finishing duties.

The data in Table A6.4 were all collected before 1992. Recently, a new handbook has appeared which uprates the maximum toughness of whisker reinforced aluminas to 1.2 GPa (*Japanese Carbide Manufacturers Handbook*, 1998). Manufacturers' data in the authors' possession also show maximum hardness of the black aluminas has been enhanced up to 22 GPa; and other information suggests room temperature thermal conductivity can be higher than given, up to 35 W/m K. These extended ranges of data have been included in the construction of Figures 3.20 and 3.21.

Data for silicon nitride based tools, also from Brookes (1992), are collected in Table A6.5. The fact that there is less information for these than for alumina tools reflects the more recent development of these materials for cutting. There are two groups: straight silicon nitrides and sialons. Silicon nitride, without modifications, requires hot pressing for its manufacture. It is also susceptible to contamination by silica (SiO_2). This may segregate at grain boundaries to form silicates which soften at around 1000°C. This is fatal to the performance of cutting tools. One way to prevent these glassy grain boundary phases is by the addition of yttria (Y_2O_3). Thus, almost all silicon nitride based cutting tools have some

Table A6.4 Compositions and properties (pre-1992) of alumina based tool materials

ISO code	Composition, Wt. % Major $A1_2O_3$	TiC	SiC(wh.)	Other	ρ [kg/m^3]	*HV* [GPa]	*TRS* [GPa]	*K* [W/mK]	*E* [GPa]	α_e [10^{-6} K^{-1}]
–	97			3	4000	16.2	0.55	8.7	380	–
K10	96.5			3.5	4020	17.0	0.7	–	–	–
P/K01–10	96			4	4000	17.0	0.7	–	–	–
–	√*			√*	4000	18.6	0.8	–	–	–
P01–15	80	10		0	4150	16.7	0.8	–	–	–
–	71	28		1	4300	17.7	0.55	–	–	–
K05	70	30:Ti(C,N)		0	4250	18.9	0.62	–	–	–
–	√*	√*		√*	4300	19.6	0.9	–	–	–
–	√*	√*		√*	4400	18.6	0.95	–	–	–
K15	75		17	8	3900	14.6	1.0	–	–	–
–	75		25	0	3700	19.6	0.9	6	390	–
K05–15	√*		√*	√*	3700	23.5	0.98	17	410	6.8

*: material present, but composition not given.

Table A6.5 Compositions and properties (pre-1992) of Si_3N_4 based tool materials

ISO code	Composition, Wt. % Si_3N_4	Y_2O_3	Al_2O_3	Other	ρ [kg/m^3]	*HV* [GPa]	*TRS* [GPa]	*K* [W/mK]	*E* [GPa]	α_e [10^{-6} K^{-1}]
K01–10	√*	√*		√*	3250	94.5[1]	0.9	–	–	–
K20	96		2	2	3160	15.7	0.9	–	–	–
K05–20	√*	√*			3250	93.2[1]	0.85	–	–	–
–	√*	√*		√*	3300	93.9[1]	1.1	–	–	–
–	91	0	1	8	3200	14.2	0.8			
K20	90.5	6	3.5		3260	13.7	1.0	–	–	–
K05–20	√*	√*	√*	√*	3300	15.4	0.9	17	280	3.0
K01–30	√*	√*	√*		3300	15.7	0.8	–	–	–
–	80	6	4	10	3200	15.7	0.65	–	300	–

*: material present, but composition not given; [1]: HRA.

addition of Y_2O_3. If Y_2O_3 is added in greater quantities, and also alumina and/or aluminium nitride, an alloy of Si, Al, O and N (sialon) is formed, also containing yttrium. The benefit is that this material can be manufactured by pressureless sintering and maintains its mechanical properties in use up to about 1300˚C. The table shows that the benefits of one group over the other are entirely in the ease of manufacture. There is little to choose between their room temperature mechanical properties (although the sialon materials are likely to have a more reliable high temperature strength). As with the alumina materials, there has been some materials development over the last 10 years. More recent transverse rupture stress data are more commonly in the range 0.95 to 1.2 GPa (*Japanese Carbide Manufacturers' Handbook*, 1998).

Finally, Table A6.6 summarizes the small amount of available information on PcBN and PCD tools. These tools are manufactured in a two-stage process. First, synthetic diamond or cubic boron nitride grits are created at high temperature and pressure. These are then cemented together by binders. Each class of tool has two types of binder, ceramic-based

Table A6.6 Compositions and properties of super hard tool materials

ISO code	PcBN or PCD	Binder materials	ρ [kg/m^3]	*HV* [GPa]	*TRS* [GPa]
P/K01–10	PcBN	ceramic*	3600	38	–
		ceramic*	–	41	0.8
		cermet**	4000	34	–
		cermet**	3900	33	–
		–	–	49	0.6
K01–10	PCD	SiC	–	–	–
		Co	3700	69	–
		Co (18%)	3900	38	–
		–	–	88	1.5
		–	–	88	2.0
		–	–	54	0.6–1.2

*ceramic = Al_2O_3 base; **cermet = carbo-nitrides – Co/WC/AlN up to 18%wt.

for ultimate hardness or metal-based for toughness. For PcBN, the ceramic base is Al_2O_3 and the metal base is sintered carbide or cermet. For PCD, the ceramic is based on SiC and the metal on Co.

References

Brookes, K. J. A (1992)*World Directory and Handbook of Hardmetals and Hard Materials*, 5th edn. East Barnet, UK: International Carbide Data.

Exner, H. E. (1979) Physical and chemical nature of cemented carbides. *Int. Metals Revs*, **24**, 149–173.

Gurland, J. (1988) New scientific approaches to development of tool materials. *Int. Mats Revs*, **33**, 151–166.

Handbook (1998) *Japanese Cemented Carbide Manufacturers' Handbook*. Tokyo: Japanese Cemented Carbide Tool Manufacturers' Association.

Hoyle, G. (1988) *High Speed Steels*. London: Butterworths.

ISO 513 (1991) *Classification of Carbides According to Use*. Geneva: International Standards Organisation.

Schwarzkopf, P. and Keiffer, R. (1960) *Cemented Carbides*. New York: MacMillan.

Shelton, P. W. and Wronski, A. S. (1987) Strength, toughness and stiffness of wrought and directly sintered T6 high speed steel at 20–600˚C. *Mats Sci. Technol.* **3**, 260–267.

Trent, E. M. (1991) *Metal Cutting*, 3rd edn. London: Butterworths.

Appendix 7

Fuzzy logic

This appendix supports Chapter 9 in which fuzzy sets and their operations are introduced to help the optimization of cutting conditions and tool selection. More complete descriptions are given in many textbooks (e.g. Zimmermann, 1991). Applications of fuzzy logic to machining may be found in journals and handbooks (e.g. Dreier *et al.*, 1996).

A7.1 Fuzzy sets

Fuzzy sets were first introduced to represent vagueness in everyday life, especially in natural language. They are not special, but a generalized representation of conventional sets. Five causes of vagueness are generally recognized: incompleteness, non-determinism, multiple meanings, (statistical) uncertainty and non-statistical uncertainty. Fuzziness is non-statistical uncertainty and fuzzy logic deals with it.

Before considering what fuzzy sets are, consider what are conventional, or crisp, sets. As an example, to be used throughout this Appendix, consider the sets of 'ordinary cutting speed' S_o, 'high cutting speed' S_h and 'ultra high cutting speed' S_u. Conventionally, or crisply, they may be defined as

$$S_o = \{V \mid V < V_1\} \tag{A7.1a}$$

$$S_h = \{V \mid V_1 \le V < V_2\} \tag{A7.1b}$$

$$S_u = \{V \mid V_2 \le V\} \tag{A7.1c}$$

where V_1 and V_2 are constants. They have the meaning that if $V = V_1$ or more, the cutting speed is high, but if the cutting speed decreases by only a small value ΔV below V_1, i.e. $V = V_1 - \Delta V$, the cutting speed becomes ordinary. These sets can be represented by membership functions that map all the real elements of the set onto the two points $\{0, 1\}$, e.g. for the set of high cutting speed S_h,

$$\mu_{Sh}(V) = \begin{cases} 1 & V_1 \le V < V_2 \\ 0 & \text{otherwise} \end{cases} \tag{A7.2}$$

Figure A7.1(a) shows the membership functions of three sets $\mu_{So}(V)$, $\mu_{Sh}(V)$ and $\mu_{Su}(V)$. The value of the membership function is called its membership.

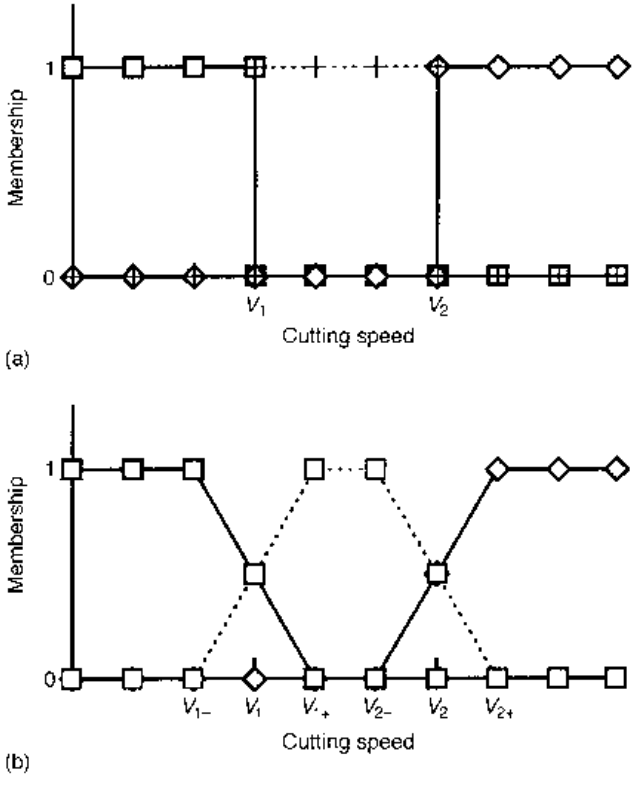


Fig. A7.1 Comparison between (a) crisp and (b) fuzzy sets

However, the sudden transitions between (crispness of) these sets of domains of cutting speed do not satisfy the language needs of machinists and tool engineers. They feel that there must be some transitional region, of significant width, between the domains of ordinary and high (and high and ultra high) cutting speeds. In other words, the membership should be able to change gradually from 0 to 1 or 1 to 0 between the domains.

A fuzzy set is always defined as a membership function, the membership of which has a value in the range [0, 1]. Unlike crisp sets, the membership of fuzzy sets can be fractional. Using this characteristic of fuzzy sets, the domains of cutting speed can be represented by membership functions according to the subjective measure of machinists and tool engineers:

$$\mu_{\tilde{\mathrm{So}}}(V) = 1 - LF(V, V_{1-}, V_{1+}) \tag{A7.3a}$$

$$\mu_{\tilde{\mathrm{Sh}}}(V) = \begin{cases} LF(V, V_{1-}, V_{1+}) & V < V_{1+} \\ 1 & V_{1+} \le V < V_{2-} \\ 1 - LF(V, V_{2-}, V_{2+}) & V_{2-} \le V \end{cases} \tag{A7.3b}$$

$$\mu_{\tilde{\mathrm{Su}}}(V) = LF(V, V_{2-}, V_{2+}) \tag{A7.3c}$$

where V_{1-}, V_{1+}, V_{2-} and V_{2+} are constants and the linear function LF is defined as follows:

$$LF(x, a_1, a_2) = \begin{cases} 0 & x < a_1 \\ \dfrac{x - a_1}{a_2 - a_1} & a_1 \le x < a_2 \\ 1 & a_2 \le x \end{cases} \tag{A7.4a}$$

where x is the variable and a_1 and a_2 are constants.

Figure A7.1(b) shows the membership functions of three fuzzy sets $\mu_{\tilde{S}o}(V)$, $\mu_{\tilde{S}h}(V)$ and $\mu_{\tilde{S}u}(V)$ that result from these definitions: they would usually be drawn on one graph. In a transitional region, for example $[V_{1-}, V_{1+}]$, the membership function $\mu_{\tilde{S}h}(V)$ gradually increases from 0 to 1 as the membership function $\mu_{\tilde{S}o}(V)$ gradually decreases from 1 to 0.

A fuzzy set need not be described by a linear function. Although a triangular function, obtained by letting $V_{1+} = V_{2-}$ in equation (A7.3b), is often used for fuzzy modelling, others may be used. A square function, SF, is used in Section 9.3.3, and is defined as

$$\mathrm{SF}(x, a_1, a_2) = \begin{cases} 0 & x < a_1 \\ \dfrac{2(x - a_1)^2}{(a_2 - a_1)^2} & a_1 \le x < \dfrac{a_1 + a_2}{2} \\ 1 - \dfrac{2(x - a_2)^2}{(a_2 - a_1)^2} & \dfrac{a_1 + a_2}{2} \le x < a_2 \\ 1 & a_2 \le x \end{cases} \tag{A7.4b}$$

When a set of cutting speeds has a finite number of elements, fuzzy sets S_o or S_h, for example, are written as follows:

$$S_o = \mu_{o1}/V_1 + \mu_{o2}/V_2 + \mu_{o3}/V_3 + \ldots + \mu_{on}/V_n \equiv \sum_{i=1}^{n} \mu_{oi}/V_i \tag{A7.5a}$$

$$S_h = \mu_{h1}/V_1 + \mu_{h2}/V_2 + \mu_{h3}/V_3 + \ldots + \mu_{hn}/V_n \equiv \sum_{i=1}^{n} \mu_{hi}/V_i \tag{A7.5b}$$

where each term μ_{oi}/V_i or μ_{hi}/V_i represents the membership $\mu_{\tilde{S}o}(V)$ or $\mu_{\tilde{S}h}(V)$ at speed V_i. The operator '+' means the assembly of elements, not the summation of elements.

A7.2 Fuzzy operations

Among all the fuzzy operations, only two operations, the maximum operation and minimum operation, are described here. The maximum and minimum operations are simply defined as follows: for two memberships μ_1 and μ_2,

$$\mu_1 \mathbf{V} \mu_2 = \begin{cases} \mu_1 & \mu_1 > \mu_2 \\ \mu_2 & \text{otherwise} \end{cases} \tag{A7.6a}$$

$$\mu_1 \Lambda \mu_2 = \begin{cases} \mu_1 & \mu_1 \le \mu_2 \\ \mu_2 & \text{otherwise} \end{cases} \tag{A7.6b}$$

where **V** and **Λ** are the maximum and minimum operators.

The union and intersection of the membership of two fuzzy sets $\mu_{\tilde{S}o}(V)$ and $\mu_{\tilde{S}h}(V)$ at any cutting speed V are respectively defined as, and are given by applying the maximum and minimum operations:

$$\mu_{\tilde{S}o \cup \tilde{S}h}(V) = \mu_{\tilde{S}o}(V) \mathbf{V} \mu_{\tilde{S}h}(V)$$

$$= \begin{cases} 1 - LF(V, V_{1-}, V_{1+}) & V < (V_{1-} + V_{1+})/2 \\ LF(V, V_{1-}, V_{1+}) & (V_{1-} + V_{1+})/2 \le V < V_{1+} \\ 1 & V_{1+} \le V < V_{2-} \\ 1 - LF(V, V_{2-}, V_{2+}) & V_{2-} \le V \end{cases} \tag{A7.7a}$$

$$\mu_{\tilde{S}o \cap \tilde{S}h}(V) = \mu_{\tilde{S}o}(V) \Lambda \mu_{\tilde{S}h}(V)$$

$$= \begin{cases} LF(V, V_{1-}, V_{1+}) & V < (V_{1-} + V_{1+})/2 \\ 1 - LF(V, V_{1-}, V_{1+}) & (V_{1-} + V_{1+})/2 \le V \end{cases} \tag{A7.7b}$$

Figure A7.2 shows the union and intersection of fuzzy sets as defined above.

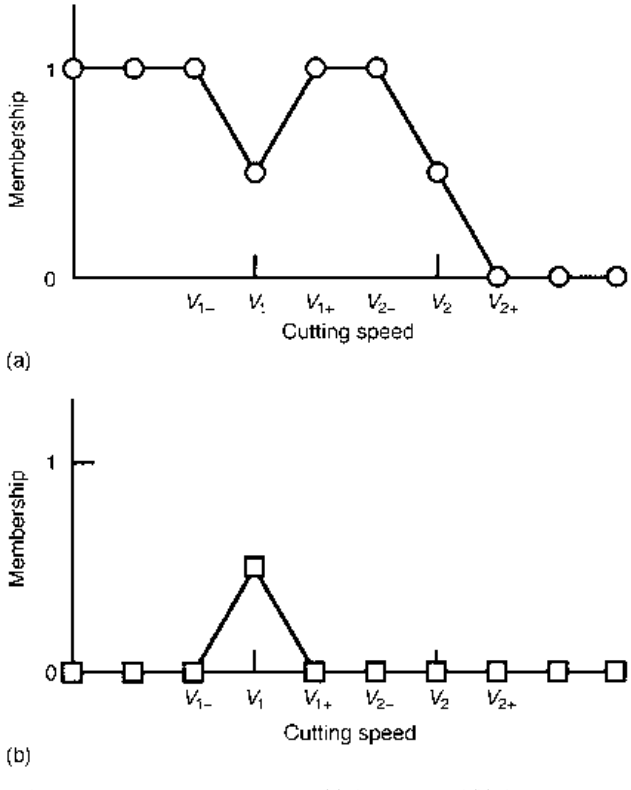


Fig. A7.2 Maximum and minimum operations representing (a) the union and (b) the intersection of two fuzzy sets

Similarly, the union and intersection of the two fuzzy sets S_o and S_h in equations (A7.5a) and (A7.5b) are given as follows:

$$\begin{aligned} S_o \cup S_h &= (\mu_{o1} \mathbf{V} \mu_{h1})/V_1 + (\mu_{o1} \mathbf{V} \mu_{h1})/V_2 + \ldots + (\mu_{o1} \mathbf{V} \mu_{h1})/V_n \\ &\equiv \sum_{i=1}^{n} (\mu_{oi} \mathbf{V} \mu_{hi})/V_i \end{aligned} \tag{A7.8a}$$

$$\begin{aligned} S_o \cap S_h &= (\mu_{o1} \mathbf{\Lambda} \mu_{h1})/V_1 + (\mu_{o1} \mathbf{\Lambda} \mu_{h1})/V_2 + \ldots + (\mu_{o1} \mathbf{\Lambda} \mu_{h1})/V_n \\ &\equiv \sum_{i=1}^{n} (\mu_{oi} \mathbf{\Lambda} \mu_{hi})/V_i \end{aligned} \tag{A7.8b}$$

References

Dreier, M. E., McKeown, W. L. and Scott, H. W. (1996) A fuzzy logic controller to drill small holes. In Chen, C. H. (ed.), *Fuzzy Logic and Neural Network Handbook*. New York: McGraw-Hill, pp. 22.1–22.8.

Zimmermann, H. J. (1991) *Fuzzy Set Theory and Applications*. Boston: Kluwer.

Index