Žydrūnas Vagnorius

Reliability of metal cutting tools:

Stochastic tool life modelling and optimization of tool replacement time

Thesis for the degree of Philosophiae Doctor

Trondheim, October 2010

Norwegian University of Science and Technology
Faculty of Engineering Science and Technology
Department of Production and Quality Engineering

NTNU – Trondheim
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Engineering Science and Technology
Department of Production and Quality Engineering

ISBN 978-82-471-2328-7 (printed ver.)
ISBN 978-82-471-2329-4 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2010:177

Printed by NTNU-trykk

Preface

Conducting a PhD study is an important step in becoming a member of the scientific community. This has been on my mind for a long time. From the young age I was interested in how things worked and why they were as they were. This curiosity followed me through my engineering studies, my early industrial career and eventually impelled me to search for a PhD position.

My pursuit of becoming a researcher began to materialise in March 2007 when I was accepted to PhD studies in the Department of Production and Quality Engineering at the Norwegian University of Science and Technology (NTNU). This way I was given a unique chance to work with scientist and to prove that I was worth being a member of their community. Now after three years of hard work I can reflect on this period and present the results that I have achieved.

My studies were a part of the AVROPROS project. The name is the acronym of the Norwegian title "Avanserte verktøy og robust prosessoptimalisering" meaning "Advanced Tools and Robust Process Optimization". This project was a cooperation between NTNU, SINTEF and Volvo Aero Norge AS and was sponsored by the Research Council of Norway. Its main goal was to increase the level of automation in aerospace industry via development of new, robust machining processes and new advanced tools and via implementation of advanced process monitoring and control systems.

The main focus of the PhD project was robustness and efficiency of machining processes. While these topics are not new, there is still much work to be done. A particular issue is the reliability of cutting tools. The industry today is still relying on deterministic models developed over a century ago, despite the fact that variability is an inherent feature of all machining processes. Therefore in this thesis we develop a stochastic tool life model and propose new approaches for determining the most economic tool replacement time. Besides that we investigate high pressure cooling and show that this modern technique can help to make the machining process more efficient and more robust. Moreover, we demonstrate how statistical tools can be used to perform a macro reliability analysis and spot the "low hanging fruits", i.e., easy improvement opportunities. These issues are elaborated in Part I of this thesis and are covered in the six scientific articles written during this three year period.

The papers are included in Part II of this thesis and are arranged in the chronological order of writing. Articles 1, 2, 4 and 5 have been presented in international conferences. All of these papers have been peer-reviewed and appear in conference proceedings. Article 3 has been published and Article 6 has been accepted for publication in international journals. Moreover, Articles 2 and 5 have been selected by the organizers of the corresponding conferences for publication (subject to further review) in international journals.

In addition to the main articles a number of other documents including meeting minutes, presentation slides, internal reports and term papers have been produced during the study period. Even though these documents were not indented for publication, I believe that some of the material given there might have a certain value to both academics and practitioners. Therefore one literature review article, two term papers and one report are included in Part III of this thesis.

Writing these papers and the thesis was an unique experience, as was the whole period of PhD studies. It has given me a taste of the scientific work, has helped me to develop my research skills and has allowed me to establish valuable contacts for my future carrier. I also hope that the knowledge developed during these PhD studies will give rise to new research ideas and at least to some extent will be applied in practice, this way contributing to improvement of manufacturing processes.

Trondheim,
September 2010

Žydrūnas Vagnorius

Acknowledgements

Over the last few years a number of people, companies and institutions have contributed directly or indirectly to this PhD thesis. Without their support I would not have started the studies in the first place. Later on their encouragement, professional advices and physical help have enabled me to carry out the research, which forms the bases of this thesis.

I would first of all like to thank the Research Council of Norway for providing the funding through the AVROPROS project. Their scholarship has made it possible for me to come to study to Norway. Moreover the funding was also critical for visiting our industrial partners in order to collect data, for purchasing of test materials and tools and finally for disseminating of the research results through journal publications and participation in international conferences.

Next, I would like to express my gratitude to my supervisor Professor Knut Sørby. Thanks to his trust in my knowledge and abilities I was the one to be offered the PhD position and could commence the studies. My supervisor did his best to make this start as smooth and comfortable as possible. Later on he was my teacher and advisor and has contributed to the papers that I have written during these three years. Finally, Professor Knut Sørby has become a friend whom I could occasionally meet outside the office.

I would also like to give a big thanks to Professor Marvin Rausand, whose contribution and expertise in maintenance models is reflected in two of the articles included in this document. Moreover he has given me a lot of practical and technical advices, which have facilitated the writing of the remaining articles as well as this PhD thesis.

One more person that I am very grateful to is Dr. Jan Erik Torjusen, a now retired Volvo Aero Norge AS engineer. Dr. Torjusen was the main link between us and the industrial partners. He has helped to establish contacts with the key people, has arranged a number of important meetings and has helped to resolve all organisational issues.

At this point I must thank Volvo Aero Norge AS for providing us with valuable test pieces and giving access to sensitive information, such as machine tool maintenance records. Extracting the latter data would not have been possible without the cooperation of Kongsberg Terotech AS and especially their engineer Mr. Lars Ingebrigtsen.

Finally, and probably the most importantly, I would like to thank my wife Inesa Vagnorienė. These three years were very difficult for both of us. Being apart for such a long time and meeting only once every month was a huge challenge. Despite that she has always supported my desire to become a researcher and has always encouraged me to continue the studies. Thanks to her patience and love we have lived through this period. I would therefore like to dedicate this PhD thesis to her.

Summary

This PhD thesis is based on six articles and proposes new approaches for modelling of the life of cutting tools and for determining the optimal tool replacement time. These issues are very closely related and play a critical role in machining economics. Replacing a tool too early means wasting of its potential and leads to high costs and reduced productivity. Late replacement poses a risk of wear-out and other types of tool failures, which can damage the component being produced and can cause expensive equipment downtimes. Therefore a lot of work has been done to develop models for predicting the life of a tool and to optimize its replacement time. Probably the best known of them is the Taylor's tool life equation.

Developed in 1906 Taylor's equation expresses the tool life in terms of the cutting speed. Despite being over a century old, this model is still widely used in practice. However, Taylor's equation has a few drawbacks. For example, it ignores the effect of other, though less important, process parameters such as the depth of cut and the feed. To walk around this issue several extensions of Taylor's equation have been proposed and are discussed in this thesis. Nevertheless all these models share another common flaw. They assume that tool life is deterministic, i.e., that given the process parameters the exact time to wear-out can be calculated. Unfortunately, in real machining processes there are a lot of sources of variation that affect the rate of tool wear and influence its life. As a result, deterministic models rarely give accurate estimates and are only valid as approximations.

To improve tool life predictions and assist process planners in choosing the optimal replacement time this PhD thesis proposes new methods. The underlying assumption is that tool life is a stochastic quantity and follows a certain probability distribution. With this in mind the reliability function is derived. Based on the physical analysis of machining processes it is assumed that a tool can fail due to the three main causes: (i) wear, (ii) internal defects and (iii) external stresses.

Tool wear depends on a number of factors, including the characteristics of the tool itself, such as its material, geometry and coating, properties of the workpiece material, cutting parameters, rigidity of the machine tool and the efficiency of the cooling process. This last factor is particularly important as most of the tool wear mechanisms depend on temperature. Therefore in this PhD thesis a lot of attention is given to high pressure cooling, which is an effective way to reduce the temperature in the cutting zone.

Internal defects are micro voids and cracks that develop inside the tool material during its manufacturing process or as a result of inappropriate handling. They act as stress concentrators and lead to shorter than normal tool life. External stresses are severe overloads that cause immediate tool failure regardless of its quality. They are random in nature and may originate from machine operator errors, failure of supporting equipment or some other external sources.

Considering all three failure modes total tool reliability function is found. It is assumed that in a given batch a certain percentage of tools are “bad”, i.e., they contain internal defects, while the rest are “good”. The life of the normal tools is modelled by a two-parameter Weibull distribution. Failures due to internal defects are also accounted for by the Weibull distribution, but with different parameters. Then the life of a tool chosen at random is predicted by the mixture model. In addition, tools of both types can fail due to external stresses, the occurrence of which is model by a homogeneous Poisson process.

The derived tool reliability function is used to determine the replacement time. Two models are proposed for this purpose. The first one is called the minimum acceptable reliability approach. The idea is to select such a replacement period that the reliability of the tool during it would not fall below a certain minimum level. We show that, when the reliability function is known, this can be done by using a simple graphical procedure.

The second model is based on the age replacement policy, which attempts to balance the costs of preventive and failure provoked tool changes. To solve this optimization problem the total time on test (TTT) transform of the reliability function is introduced, and a method for estimating it form the experimental data is proposed. Then, as in case of the first model, the replacement time is found by employing a simple graphical procedure.

For the above approach to be used in practice the expected costs of preventive and failure provoked replacements need to be known. It is shown that the former one can be determined by applying traditional formulas found in machining economics handbooks. The penalty cost, on the other hand, is not so well defined, and no good estimation models are available. Therefore, a new, probability tree-based approach is developed in this thesis.

The relevance and the applicability of the proposed models is tested in a few experimental and case studies described in the appended articles. In Article 1 reliability of machining systems as a whole is investigated, and the stochastic nature of the processes involved is clearly shown. In Article 2 it is demonstrated that a two-parameter Weibull distribution can be used to model the tool life, and a simple replacement model based on the reliability function is proposed. In Article 3 a more generic tool life model is developed, but a two-parameter

Weibull distribution is still found to be a good approximation. The replacement time is than found by employing an optimization procedure based on the age replacement policy. In Article 4 an approach for estimating the penalty cost, which is a key input to the age replacement model, is developed. Finally in Articles 5 and 6 it is shown that high pressure cooling can help to extend the tool life and possibly to reduce its variation, which is the main reason why probabilistic models are needed.

Based on this experimental work and case studies the thesis concludes that stochastic approaches for tool life modelling and for determination of replacement time are relevant and applicable in practice. Therefore further work needs to be done to extend the use of these methods beyond the set-ups and conditions tested throughout the research described in this PhD thesis.

List of main articles

Article 1

Z. Vagnorius, K. Sørby, Reliability of machine tool systems in aircraft industry, in: Proceedings of the Eighth International Conference on Advanced Manufacturing Systems and Technology, CISM, Udine, Italy, 2008, pp. 293–304.

Synopsis. This article presents a case study carried out in a plant manufacturing jet engine components. It demonstrates how statistical methods could be used to analyse the reliability of machine tools and proposes ideas for modelling the occurrence of various types of failures and the time to repair them. In addition some practical issues concerning the maintenance of machine tools are discussed.

Article 2

Z. Vagnorius, K. Sørby, M. Rausand, Probabilistic model for determination of tool replacement time, in: Proceedings of the 12th CIRP Conference on Modelling of Machining Operations, Mondragon Unibertsitateko Zerbitzu Editoriala, Donostia-San Sebastián, Spain, 2009, vol. II, pp. 757–764.

Selected by the scientific committee of the conference for publication (subject to further review) in the Journal of Machining Science and Technology.

Synopsis. In this article a stochastic approach for determining the replacement time for a cutting tool is presented. It is showed that tool life is a random variable and could adequately be modelled by a two-parameter Weibull distribution. Given that, it is demonstrated that the replacement time for a cutting tool could be found from its reliability function. This approach is tested in an experimental study on machining of Inconel 718 with coated carbide inserts.

Article 3

Z. Vagnorius, M. Rausand, K. Sørby, Determining optimal replacement time for metal cutting tools, European Journal of Operational Research 206 (2) (2010) 407–416.

Synopsis. This article presents a model for optimizing the replacement time for a cutting tool. Tool life is assumed to be a random variable. Based on that, a reliability function, including three possible tool failure modes, is developed. The function is then plugged into the age replacement model, which balances the cost of preventive and failure provoked tool replacements. To solve this optimization problem it is proposed to make use of the total time on test (TTT) transform of the tool life distribution function. This approach is tested in an experimental study on machining of Inconel 718 with inserts based on cubic boron nitride (CBN).

Article 4

Z. Vagnorius, K. Sørby, Estimation of cutting tool failure costs, in: 2009 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM 2009), IEEE, Hong Kong, China, 2009, pp. 262–266.

Synopsis. In this article an approach for estimating the expected cost of a cutting tool failure is proposed. The model is based on a probability tree and considers various possible scenarios associated with a tool failure. It is assumed that in each case a certain penalty cost is incured. By multiplying these costs by the likelihood of a corresponding failure scenario the expected cost of a cutting tool failure is found.

Article 5

Z. Vagnorius, K. Sørby, Effect of high pressure cooling on life of SiAlON-based ceramic cutting tools, in: Proceedings of the 5th International Conference on Advances in Production Engineering (APE'2010), Warsaw University of Technology, Warsaw, Poland, 2010, pp. 352–362.

Selected by the scientific committee of the conference for publication (subject to further review) in the Journal of Engineering Manufacture (Part B of the Proceedings of the Institution of Mechanical Engineers).

Synopsis. This article presents a study on application of high pressure cooling in machining of Inconel 718 with SiAlON ceramic inserts. It shows that this technique significantly improves chip breaking, but if not used carefully can lead to shorter life of SiAlON tools. The study demonstrates that in this particular application design and configuration of the high pressure cooling system are critical. Other possible reasons for shorter tool life are also discussed.

Article 6

Z. Vagnorius, K. Sørby, Effect of high pressure cooling on life of SiAlON tools in machining of Inconel 718, *accepted for publication in:* International Journal of Advanced Manufacturing Technology.

Synopsis. This article presents the second study on the application of high pressure cooling in machining of Inconel 718 with SiAlON ceramic inserts. In this case tools made in SiAlON with improved resistance to notching were used, and a special attention was given to system design and configuration. As a result, significantly longer tool life compared to conventional cooling was achieved. Variation of the tool life appeared to be slightly reduced, and chip breaking was considerably improved.

List of abbreviations

BUE	Built-up edge
CBN	Cubic boron nitride
ERP	Enterprise resource planning
FMS	Flexible manufacturing system
HSS	High speed steel
MTBR	Mean time between replacements
MTTF	Mean time to failure
NDM	Near-dry machining
TTT	Total time on test

List of symbols

Symbol	Description	Typical units
α	shape parameter of the Weibull distribution	
α_b	shape parameter of the Weibull distribution in the reliability function of tools with internal defects	
α_g	shape parameter of the Weibull distribution in the reliability function of tools with no internal defects	
λ	scale parameter of the Weibull distribution	min^{-1}
λ_b	scale parameter of the Weibull distribution in the reliability function of tools with internal defects	min^{-1}
λ_g	scale parameter of the Weibull distribution in the reliability function of tools with no internal defects	min^{-1}
λ_s	rate of external stresses	min^{-1}
$\mathcal{T}(T_{(i)})$	total time on test (TTT) at the instant of failure of the ith tool	min
$\phi_F(v)$	scaled TTT transform of the tool life distribution function $F(t)$	
a_p	depth of cut	mm
c	cost of preventively replacing a cutting tool	€
C	constant in Taylor's tool life equation	m/min
$C(t_0)$	rate of cutting tool costs during a replacement period of length t_0	€/min
$C1$	value of a component at the start of a machining process	€
$C2$	cost of machining one component in the penalty cost estimation model	€
$C3$	cost of resetting the cutting tool after a failure	€
$C4$	cost of reworking a component after a tool failure	€
$C5$	value of scrap material	€
C_t	average cost of providing a sharp cutting tool (sharp cutting edge)	€
$C_T(t_0)$	expected tool cost per a replacement period of length t_0	€
C_{pr}	cost of machining one component in cutting speed optimization model	€
D	workpiece diameter	mm
f_n	feed rate	mm/rev

$F(t)$	tool life distribution (tool failure) function	
$F^{-1}(v)$	inverse of the tool life distribution function $F(t)$	
$H_F^{-1}(v)$	TTT transform of the tool life distribution function $F(t)$	
k	penalty cost	€
l	machining distance	mm
M	machine and operator rate	€/min
n	exponent in Taylor's tool life equation	
N_b	number of tools with internal defects in a batch of tools	
N_{bp}	size of a production batch	
N_s	number of tools that failed due to external stresses	
N_t	number of cutting tools needed to machine a batch of components	
p	proportion of cutting tools with internal defects	
$p1$	probability that no rework will be required after a tool failure	
$p2$	expected part of work completed up to the tool failure	
$p3$	probability of successfully reworking the component after a tool failure	
P_r	profit rate	€/min
$R(t)$	total tool reliability function	
$R_b(t)$	reliability function of tools with internal defects	
$R_g(t)$	reliability function of tools with no internal defects	
$R_s(t)$	reliability function of tools with respect to external stresses	
$R_w(t)$	reliability function of a randomly selected tool (either with or without internal defects) that is subjected to wear mechanisms	
$R_{\min}$	minimum acceptable tool reliability	
S	amount of money received by the machine shop for a component	€
t	time	min
t_0	tool replacement time	min
t_l	time to load and unload a component and to return the cutting tool to the beginning of the cut	min
t_m	machining time per component	min
t_{c_t}	tool changing time	min
t_{pr}	production time per component	min
T	tool life	min
T_{ef}	tool life for maximum profit rate	min
v_c	cutting speed	m/min
$v_{c_{oc}}$	cutting speed for minimum cost per piece	m/min
$v_{c_{ot}}$	cutting speed for minimum production time	m/min
VB	average width of the flank wear land	mm
$VB\,\mathrm{max}$	maximum width of the flank wear land	mm
$z(t)$	tool failure rate function	

Contents

Part I Report

Part II Main articles

Part III Other articles and reports

Part I

Report

1

Introduction

Cutting tools lie at the core of machining operations, which are considered to be the most versatile manufacturing techniques for production of highly accurate parts. In the old days potters used their hands, strings, sticks and various chips to shape a piece of clay while rotating it on a foot-driven wheel – the earliest form of a machine tool [1]. Carpenters employed sharp stones and metal chisels in primitive lathes (see Fig. 1.1) to turn wood. The history of metal cutting started in 1774 when the first real boring mill was built [2]. Since that time a number of different purpose machine tools have been constructed, and their efficiency has increased a hundred-fold. According to Benhabib [2], the primary reason for that is the advancement in materials used in cutting tools.

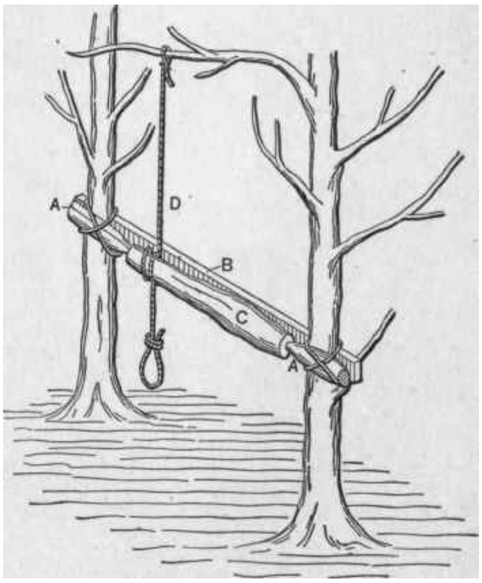


Fig. 1.1: Ancient lathe [1] with (**A**) workpiece holding pivots, (**B**) tool support, (**C**) workpiece and (**D**) power transmission cord

The earliest metal cutting tools were made in hardened carbon steels. They were inexpensive, but softened at just 250°C and were therefore limited to low cutting speeds [3]. The first major leap in the tool making technology was the invention of the Taylor-White hardening process at the beginning of the 20^{th} century. This led to development of high speed steel (HSS) tools [4], which had a better wear resistance and could be used at higher cutting speeds. Despite this improvement the productivity of machining processes was still rather low as the properties of HSS tools deteriorated rapidly at elevated temperatures. A real breakthrough occurred in 1930s when cemented carbide tools were introduced. The supreme properties of these materials, such as high hot hardness, high elastic modulus, perfect thermal conductivity and low thermal expansion [5], made them so versatile that over 50% of the tools used world wide today are estimated to be based on cemented carbides [3]. Other important tool materials are ceramics, cubic boron nitride (CBN) and diamond. All these substances are extremely hard and have an exceptional wear resistance.

The performance of cutting tools, particularly those made in carbides, can be further improved by coating them with multiple protective layers of titanium nitride (TiN), titanium carbide (TiC), titanium carbonitride (TiCN) and/or aluminium oxide (Al_2O_3). This reduces the friction on the surface of the tools, increases their toughness and further improves wear resistance.

Despite these technological improvements tool wear and failure are still the major concerns in all machining processes. Tool wear, such as flank and nose wear, or built-up edge, can negatively affect the finish of the produced surfaces and can cause costly rework [5, 6]. Chipping, or gross fracture can lead to scrapping of the part being machined, while catastrophic tool failure can idle expensive equipment and can even bring down the whole production line, resulting in delayed shipments and loss of customer's good will [7, 8].

A lot efforts have been invested by the researchers to develop models that would allow to predict the tool life and would help to avoid the above-mentioned consequences. However, in most cases variation, which is an inherent property of machining processes, was disregarded. As a result the estimates of the tool life were rarely accurate. Therefore a common practice used in industry today is to replace the tools well before the end of their useful lifetime. Unfortunately, such a conservative strategy leads to increased costs.

In general cost of the cutting tools is said to make only 2–4% of the total manufacturing cost [5]. This estimate is disputed in several studies. Jeang [9] claims that tool costs in metal cutting constitute about one third of the unit production costs. Sharit and Elhence [10] estimate that tool related activities in flexible manufacturing systems (FMS) stand for 25% of the operating costs. This opinion is supported by Gray et al. [11], who review several papers providing industrial data, and conclude that tooling can account for 25–30% of both fixed and variable production costs in an automated machining environment. A similar estimate can be achieved by studying and reprocessing the data presented by Hong [12]. This shows that in conventional machining of Ti–6Al–4V, a typical material used in the aerospace industry, tool costs make up from 9% to over 35% of unit production expenses.

The above estimates show that tool costs are already high in some manufacturing systems. Moreover, the ever-present need for higher productivity, new and often difficult to cut workpiece materials, and the emerging trend of dry or near dry machining (NDM) [3, 5, 13] necessitate the use of more advanced, hence more expensive, tools. This suggests that economic implications of conservative tool replacement policies will become more significant in the near future.

From this discussion it follows that there is a need for new models for predicting the life of cutting tools and for new methods to optimize their replacement time. The proposed approaches should take into account the effects of variability in machining processes and should seek for a balance between the traditional tool costs and the expenses of possible tool failures. These issues are addressed in this PhD thesis.

1.1 Background

The decision when to replace a cutting tool depends on its expected life, which is largely determined by the choice of the process parameters. Cutting speed v_c usually has the biggest effect. Therefore, as discussed in the following sections, the prevailing industry practice is to control this variable in order to adjust the tool life so that the cost-, production time- or profit-related objectives of the machining process are achieved.

1.1.1 Tool life definition

Definition of the tool life varies among different authors and often depends on the intended application. In this thesis I will use ISO 3685 [14] as my main reference. This standard emphasizes that the main function of a cutting tool is to produce workpieces of the desired size and surface quality. Thus in the ideal case it should be replaced as soon as the quality of the produced parts becomes unsatisfactory (unacceptable). In practice, however, in-process monitoring of workpiece quality is difficult. Therefore it must be predicted indirectly from the *tool wear*, which is defined by ISO 3685 as a change of shape of the tool from its original shape, during cutting, resulting from the gradual loss of tool material or deformation. This is a complex process involving several mechanisms and will be discussed in more detail in Section 2.1.

Besides gradually wearing out a cutting tool can fail abruptly. This makes it physically incapable of further cutting and can be caused by the defects introduced during the manufacture of the tool or by various random overloads. Such events are referred to as the occurrence of a phenomenon by ISO 3685 and are discussed in more detail in Sections 2.2 and 2.3.

Considering these failure modes ISO 3685 gives the following definition:

> *tool life* T is the cutting time required to reach a tool life criterion

and the tool life criterion is a predetermined threshold value of a tool wear measure, i.e., the dimension to be measured to evaluate the amount of wear, or the occurrence of a phenomenon.

This formal definition of the tool life is well suited for laboratory tests, where the cutting process can be stopped at predefined intervals, and the tool can be inspected for wear with the help of an optical microscope[1]. Such technique was used extensively in the experimental work presented in this PhD thesis.

1.1.2 Current tool life modelling practice

The most popular empirical model used today is the so-called Taylor's tool life equation [4]. Based on his 26 years of experimental work, Taylor determined that, in rough machining of steel forgings with HSS tools, the cutting speed v_c and the tool life T were related through the equation $v_c = \text{Constant} \times T^{-1/8}$. This relationship has been generalised and is usually written as

$$v_c T^n = C \tag{1.1}$$

where n is an exponent, which depends primarily on the tool material [3, 24], and C is a constant determined by the combination of the tool- and work materials, as well as the geometry of the tool [24].

The physical meaning of the above parameters is illustrated in Fig. 1.2. The experimental data here is taken from Article 2. It is plotted on a double-logarithmic scale and the best fit line is drawn. The exponent n is then the *inverse* of the slope of this curve and shows how sensitive the life of a tool made in a particular material is to the changes in the cutting speed. HSS tools, for example, have a low n (0.10–0.17), which means that a small increase in machining velocity leads to a significant reduction in their life. Ceramic inserts, on the other hand, have a high n (0.40–0.60), thus they can be used over a wider range of speeds. In this particular example n is approximately 0.32, which is a typical value for coated carbide inserts [24].

Constant C shows the cutting speed that would result in the tool life of one minute and can be determined by extending the fitted line until it intersects the

---

[1] On the shop floor direct measurements of tool wear can be difficult or impractical. Therefore indirect methods are sometimes employed (e.g, see reference [15] for a general overview of tool condition monitoring techniques). In this case tool wear is estimated by observing the changes in the characteristics of the cutting process due to the tool wear. Examples of such characteristics are cutting forces [16–19], spindle torque and power [17, 20] or acoustic emissions [21–23]. Some of these methods have already found application in industry, and their use will probably become more widespread in the near future.

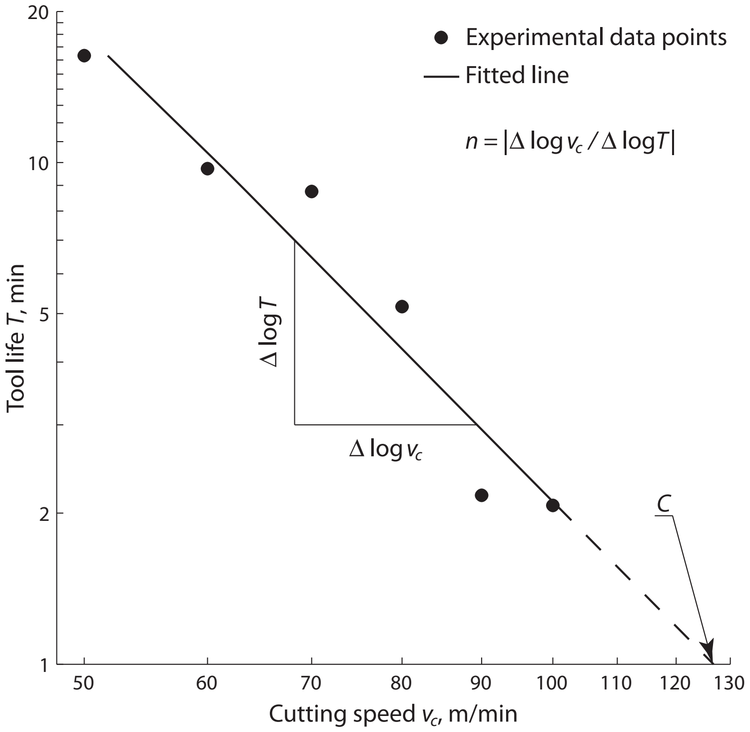


Fig. 1.2: Taylor's tool life curve

abscissa axis. In the above example C is approximately equal to 126 m/min. By substituting it together with the value of n into (1.1) we find that the Taylor's equation for this particular case is $v_c T^{0.32} = 126$.

As shown in Fig. 1.2, when the data is plotted on a double logarithmic scale, the relationship between the tool life and the cutting speed becomes roughly linear. This makes it easier to draw a fitted line, which was probably the main reason for using the logarithmic scale at the time when computers were not available. The linear relationship is generally true for machining of traditional materials, such as steels and cast irons, within the range of typical cutting speeds, where the dominant failure mechanism is flank wear. In case of heat resistant alloys and other high strength materials, however, the straight line relationship does not always apply [25]. One of the reasons is the extreme temperature generated when machining these materials. Narutaki et al. [26] and Kitagawa et al. [27], for example, have shown that in turning of Inconel 718, a heat resistant nickel based alloy used in aerospace industry, temperature on the rake face of the tool can reach 1300°C. Such conditions promote diffusion and therefore lead to increased crater wear. This phenomenon is also common in high speed machining. As a consequence Taylor's equation is least accurate here [24]. The straight line relationship also fails at low cutting speeds, where built-up edge, followed by plucking out of the tool material, typically occurs [3].

Another limitation of equation (1.1) is the fact that it relates the tool life only with the main variable, i.e., the cutting speed, but ignores other, though less important, factors, such as the feed f_n and the depth of cut a_p. To account for the effect of these parameters several extensions of the original formula have been proposed.

Taylor [4] himself studied the influence of the feed on the tool life. The general form of the equation that he has derived can be written as $v_c T^n f_n^y = C$, where y is an experimentally determined exponent.

Woldman and Gibbons [28] went one step further. They observed that constant C in equation (1.1) varied with both the feed and the depth of cut and could be expressed as $C = C_1/(a_p^x f_n^y)$. By substituting this expression into (1.1) they obtained what became known as the extended Taylor's tool life equation

$$v_c T^n a_p^x f_n^y = K \tag{1.2}$$

Constant K and the exponents x and y here depend on the tool- and work materials. Moreover, K varies considerably with the rake angle of the tool [24].

A number of other empirical models, including further extensions of Taylor's equation, have been proposed to account for the effect of workpiece hardness, tool geometry, cutting temperature and other factors on the tool life. A brief overview of these models can be found in [29] and [13].

1.1.3 Current tool life optimization practice

Since tool life is usually the most important practical consideration in selecting cutting conditions [24], tool life prediction equations are an important input to machining process optimization models. These approaches attempt to define a set of cutting parameters that enable to reach the cost or productivity objectives. In general the goal of a machining process is to do maximum possible amount of work in the shortest possible time at the lowest possible cost. These objectives, however, are not always compatible. For example, to maximise the output rate (the amount of work done during a certain time) it might sound intuitive to set the cutting parameters at their highest values. However, this would lead to very short tool life and therefore tool costs would become very high. Moreover, due to frequent replacements a lot of productive time would be lost, thus the output rate would drop. On the other hand, machining at very low cutting speeds, feeds and depths of cut would not do any good either. Even though tool-related costs would decrease, machining time would become very long leading to low output rate. This also implies that to do the same amount of work one would need more man- and machine-hours. Thus the operating costs would increase. In addition to these trade-offs a number of other constraints such the quality of the produced parts, rigidity of the machine tool and the available power need to be taken into account when choosing the machining parameters. Considering all these aspects allows determining the optimal set of cutting conditions.

Choice of feed and depth of cut

As mentioned in Section 1.1.2, feed and depth of cut have less significant influence on the tool life then the cutting speed. However, these two parameters play other important roles. The choice of the feed and the depth of cut defines the cross section of the chip being removed, which in turn affects the metal removal rate and the chip breaking process.

Increasing feed and/or depths of cut leads to higher metal removal rate, hence better productivity. However, at the same time forces, hence power consumption, required to remove the chip also become higher. As a consequence cutting tool is deflected and the whole machine is subjected to more elastic deformation. Therefore the quality of the produced part decreases. Moreover, increasing the feed produces rougher surfaces. Hence, the choice of the depth of cut and the feed depends on the required product quality. When it is low or not critical, as in the case of roughing operations, highest possible feeds and depths of cut can be used. Their maximum values will then be determined by the power available on the machine tool. In finishing operations, though, lower depths of cut and lighter feeds must be used to increase the geometrical accuracy of the part and to achieve smother surfaces.

As mentioned above, the choice of the feed and the depth of cut (as well as the cutting speed) will affect the formation of chips. At light cuts they tend to be thin, flexible and therefore continuous. Even though such chips usually lead to good surface finish [3, 5] they are undesirable due to several reasons. First, continuous chips tend to tangle around the tool holder, the fixtures and the workpiece. Thus they can begin to rub against the machined surface and can damage it. Second, long chips can jam the chip disposal equipment. Such blocks need to be cleared away manually, which for continuous chips are especially undesirable in automated manufacturing. Moreover this poses a safety risk, because chips have sharp edges and can injure the operator who is attempting to remove them. Hence, improving chip breaking is an important consideration when designing a machining process. Besides other measures this can be achieved by increasing the depth of cut and the feed. As a result chips become wider and thicker, thus they are less flexible and are more likely to break into smaller segments. On the other hand, large depth of cut increases the tendency to form built-up edge [5], which is one of the principle factors adversely affecting the finish of the produced surface.

To summarise this discussion we can conclude that high feeds and large depths of cut are desirable, because the metal removal rate is increased and chip breaking can be improved. However, this increases cutting forces, hence power consumption, and can lead to poor quality of the produced part. Therefore the choice needs to be made by taking into account the specific requirements of a particular operation.

Choice of cutting speed

Once the feed and the depth of cut are selected the cutting speed can be determined. As mentioned previously, this parameter not only affects the metal removal rate, but has a substantial influence on the tool wear and life. Therefore, neither very low nor very high speeds yield good results. Thus, methodologies for determining the optimal cutting velocity in terms of machining cost and output rate have been developed. As an example let us discuss the approach proposed by Boothroyd and Knight [3].

Let us say we have a batch of N_{bp} parts that needs to be produced. Then the average cost of machining one component can be found as

$$C_{pr} = M \cdot t_l + M \cdot t_m + M \frac{N_t}{N_{bp}} t_{c_t} + \frac{N_t}{N_{bp}} C_t \tag{1.3}$$

In this equation M is the total machine and operator rate including the overheads, N_t is the number of cutting tools needed to machine the batch of components, C_t is the cost of providing a sharp cutting tool or a sharp cutting edge (when the tool has more than one usable cutting edges) to the machine tool, t_{c_t} is the tool changing time, t_l is the time taken to load and unload a component and to return the tool to the beginning of the cut and t_m is the machining time per component.

The first item in equation (1.3) is the non-productive cost of preparing the component for the operation, removing it from the machine and returning the tool to its initial position. This cost depends on a number of variables, such as the complexity, the size and the weight of the component, the type of the machine tool, the design of the work holding fixtures and other. For a particular operation, where the same type of components is produced, this cost will be constant. The remaining terms of equation (1.3) will depend on the choice of the cutting parameters.

As mentioned previously, machining time t_m decreases as the cutting speed, the feed and the depth of cut are increased. For example, for a cylindrical turning operation $t_m = (\pi l D)/(f_n \cdot v_c)$. Here l is the length to be machined and D is the diameter of the workpiece. Since these parameters, as well as the feed, remain constant during each tool path[2], machining time is only a function of the cutting speed.

The ratio N_t/N_{bp} appearing in the two last terms of equation (1.3) shows the number of tools consumed for machining one component. For simple parts

---

[2] In many occasions more than one path of the tool is needed to complete an operation. For example, several roughing cuts would usually be followed by a finishing cut and the feed in each case might need to be adjusted to achieve the required surface quality. Moreover the total number of paths, hence the overall machining length, would depend on the depth-of-cut settings, and the diameter of the workpiece would also decrease after each cut. Finally, at the end of the path the tool would normally have to be retracted to the starting position, thus the non-productive costs would also become variable. Therefore, in general the optimal cutting speed needs to be determined separately for each tool path.

it would be less than one, while in case of complex components, such as large jet engine parts, several tools would be used, thus N_t/N_{bp} would be big. For a particular type of a component this ratio is equivalent to the ratio between the total amount of work to be done and the tool life, i.e., t_m/T. Machining time is found in the same way as explained previously, and the tool life can be expressed from equation (1.1) (alternatively equation (1.2) could be used) as

$$T = \left(\frac{C}{v_c}\right)^{1/n} \tag{1.4}$$

Thus, since both the machining time and the tool life are expressed in terms of the cutting speed, t_m/T, hence N_t/N_{bp}, is also a function of the speed.

Coming out from the above discussion we can conclude that, given a particular component and the operation that produces it, the first element in equation (1.3) is constant, while the remaining ones mainly depend on the cutting speed. Then taking the derivative $\frac{\mathrm{d}}{\mathrm{d}v_c}C_{pr}$ and equating it to zero gives the speed that minimizes the cost per piece

$$v_{c_{oc}} = C\left(\frac{n}{1-n}\frac{M}{M \cdot t_{c_t} + C_t}\right)^n \tag{1.5}$$

Cost minimization is intuitive and is therefore embedded deep into the minds of managers. While this is obviously an important consideration, in today's volatile markets time often becomes more critical [30]. Thus let us take a look how machining processes can contribute to this new objective.

Examining equation (1.3) we can find out that the production time per component is

$$t_{pr} = t_l + t_m + \frac{N_t}{N_{bp}} t_{c_t} \tag{1.6}$$

Taking the derivative of this equation with respect to the cutting speed and equating it to zero, we find that

$$v_{c_{ot}} = C\left(\frac{n}{1-n}\frac{1}{t_{c_t}}\right)^n \tag{1.7}$$

is the cutting velocity that will yield the minimum production time.

Profit maximization

Substituting equations (1.5) and (1.7) into equation (1.4) gives the tool replacement time that would yield the minimum cost per piece and the shortest production time respectively. It should be noted, however, that being able to produce at low cost or in a short time are not the goals of a company. These are only the measures that under certain market conditions can give a competitive advantage. As emphasized by Goldratt and Cox [31], the true goal of a company is to earn money. Thus the ultimate objective is to maximize the profit.

Boothroyd and Knight's [3] approach for maximizing the profit is a follows. Let S be the amount of money that a machine shop receives for a finished component. The profit rate is then

$$P_r = \frac{S - C_{pr}}{t_{pr}} \tag{1.8}$$

Substituting equations (1.3) and (1.6) for C_{pr} and t_{pr} respectively, finding the derivative with respect to the cutting speed and equating it to zero gives the cutting speed for the maximum profit rate. Then entering the determined speed into equation (1.1) yields the optimum tool replacement time. For the case of cylindrical turning this can be written as

$$T_{ef} = \frac{1-n}{n}\left(t_{c_t} + \frac{t_l \cdot C_t}{S}\right) + \frac{\pi l D \cdot C_t}{nf \cdot SC} T_{ef}^n \tag{1.9}$$

Even though the tool life in this equation is expressed implicitly, Boothroyd and Knight [3] state that certain approximations can be used to find the solution.

To conclude this section I must note that this PhD thesis does not focus on the traditional tool life modelling and optimization methods. Therefore only the main principles were presented here. For more details the reader should consult general literature on machining economics (e.g., see references [3] and [5]).

1.2 Drawback of traditional approaches

Modelling and optimization approaches discussed in Sections 1.1.2 and 1.1.3 are based on the assumption that tool life is *deterministic*. This implies that when the input parameters, such as the cutting speed, the feed and the depth of cut, are given, the exact tool life can be found. Unfortunately, real machining processes are dynamic. This means that, rather than being constant, cutting parameters fluctuate around their pre-set values. Moreover, there are many other factors that cannot be completely controlled. Consequently tool wear and tool life also vary.

In his experiments Taylor [4] was trying to control all the parameters in order to keep them at their fixed values. The magnitude and the cost of his efforts can be appreciated by considering the fact that the pace of all machine tools (they were driven by the same steam engine) in the whole factory often had to be slowed down for lengthy periods in order to allow him to run the tests at a selected constant speed. Despite that, Taylor observed that variation still existed and caused inconsistency in the observed tool life. For example, it is mentioned that even minor deviations from the so-called "standard speed" led to a significant spread of time until wear-out of "identical" tools.

In fact, no two cutting tools are really identical. The differences might lie in their geometry and material composition or might arise from micro defects introduced during the manufacturing process. As a result “identical” tools do not behave the same even when used for the same machining operation.

Besides the variation in the cutting conditions, differences in the shape and the internal structure of the tools, other possible sources of variation are non-uniformity of workpiece properties, such as hardness, micro-structure, composition and surface characteristics, variation in coolant concentration and changes in its condition [24]. To this list one could add the characteristics of the machine tool that promote or limit vibrations, behaviour of fixtures and other supporting equipment when subjected to high cutting forces, actions of the operators and other factors.

Due to the process variation the actual tool life rarely matches the predictions of the deterministic models [32]. As shown by Fenton and Joseph [33], the cost, the productivity and the profit rate estimates given by these approaches are too optimistic. This poses a risk of early wear-outs and high failure costs. To buffer against such uncertainty conservative tool replacement strategies are employed. It is estimated that only 50–80% of the expected tool life is typically used [34]. This means that tool costs are unnecessarily high. Moreover, due to frequent replacements a lot of productive time is wasted, thus the rate of output is lower than possible. In addition to these issues, Noël et al. [35] demonstrate that ignoring process variation leads to inefficient planning of tool supply and hence low availability of minimal manned machining systems.

1.3 Objectives

The analysis of the available models and industrial practices shows that, due to the process variation, tool life estimates are inaccurate, and the choice of the replacement time can be suboptimal. Therefore

the main objective of the PhD project and this thesis is to propose new models that would enable predicting the life of metal cutting tools more accurately and would aid process planners in choosing the most economic tool replacement time.

This objective can be translated into the following tasks:

1. Performing a physical analysis of machining processes
2. Identification of the main factors that determine the life of a cutting tool
3. Proposing new ways to model the effect of these factors on the tool life
4. Developing methods for determining the optimal tool replacement time that would use the new tool life model as an input

Since this PhD project is carried out in a close cooperation with an industrial partner, practical applicability of the proposed approaches is important. Other ways, apart from the improved modelling, to make the tool life predictions more accurate are also welcome. Therefore additional objectives are:

5. Testing of the developed tool life modelling and optimization approaches in a realistic manufacturing environment
6. Proposing other, practical ways for improving the accuracy of the tool life predictions

1.4 Scope

Cutting tools are employed in a wide range of machining operations. This PhD thesis does not intend to cover all of them. Due to its simplicity turning operation was used for analysis. Nevertheless the results could be extended to other processes, for example milling, where similar tools (inserts) can be used, but care should be taken of the specifics of the process.

It should also be mentioned that in this PhD project we have closely cooperated with a manufacturer of jet engine components. Therefore many aspects of the work discussed in this thesis are first of all relevant for the aerospace industry. For example, a typical material that the jet engine components are made in is Inconel 718. Therefore this nickel based alloy was used exclusively in the experiments described in this thesis. The machining parameters, cost estimates and other data are also typical for the aerospace industry. Nevertheless, the proposed approaches are generic and could be adopted to specifics of machining processes used in other types of companies.

1.5 Research approach

The activities performed in this PhD project can be classified as *applied research*. Even though the methods used in each particular case differed slightly, the basic approach was similar and included the following steps:

1. Identification of the need for new knowledge
2. Formulation of hypotheses
3. Performing of experiments or collection of industrial data
4. Data analysis
5. Testing of hypothesis based on the data analysis
6. Dissemination of the new knowledge

The reason for carrying out any kind of research activities is the actual or foreseeable need for new knowledge. Different methods are available for identifying it. In my case this was done through elaborate discussions with my supervisor and our industrial partners. Some additional needs emerged while working on particular issues. This can be seen in Fig. 1.3, which shows how the

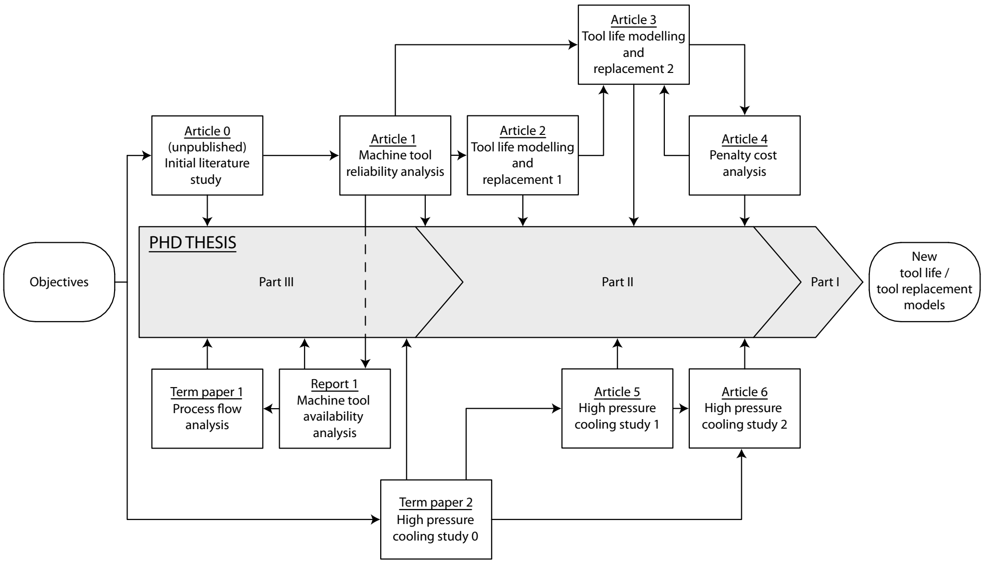


Fig. 1.3: Relationship between research activities

major pieces of work done during these PhD studies build on each other and how they contribute to the thesis as a whole.

Considering the identified gaps in the existing knowledge, hypotheses about possible ways of filling them were formulated. This meant development of new models or proposing of process improvements and claiming that this might solve the problem. To test these hypotheses experiments were performed or real industrial data were collected and were analysed by making use of statistical and mathematical tools. Based on this, conclusions concerning the trueness of the postulated hypotheses were drawn.

The collected new knowledge was disseminated through publications in international journals, presentations in international conferences and meetings with industrial partners. This was done regardless of whether the hypotheses were confirmed or rejected, because both results contribute to the understanding of the phenomena studied and therefore yield new knowledge. In total six articles were published or accepted for publication in journals and conference proceedings. The main methods used in each case are listed in Table 1.1.

This research approach includes a number of possible feedback loops. While developing a model or another solution more insights about the problem to be solved are gained. At the same time it might turn out that additional inputs are required or that the needs for new knowledge have to be clarified. Hypothesis testing might reveal that the collected data is insufficient and that additional experiments need to be carried out. Finally, when disseminating the results, feedback from peers and other parties involved is collected. This way the models and approaches are fine-tuned and turned into new knowledge.

Table 1.1: Use of statistical and mathematical tools

Tool	Main articles					
	1	2	3	4	5	6
Linear interpolation		×	×		×	×
Nelson-Aalen plot	×					
t-test	×					
Distribution fitting	×	×	×			×
TTT transform		×	×			
Probability tree				×		
F-test						×
Levene’s test						×
Wilcoxon (Mann-Whitney) test						×

1.6 Structure of this thesis

This PhD thesis has three parts. Part I contains the main report and is subdivided into six chapters, including the current one. In Chapter 2 the main causes of a cutting tool failure, such as wear, internal defects and external stresses, are discussed. Modelling of the tool life with respect to these causes and derivation of the total tool reliability function is covered in Chapter 3. In Chapter 4 it is showed how the reliability function can be used for determining of the optimal tool replacement time. A method for solving the optimization problem is suggested in Chapter 5. The relevance and practical applicability of the proposed approaches is discussed in Chapter 6. Copies of the six main articles, arranged in the chronological order of writing, are enclosed in Part II of this thesis. Other significant documents, including one literature review article, two term papers and one internal report, are presented in Part III.

2

Causes of cutting tool failure

The ISO 3685 [14] definition given in Section 1.1.1 states that a cutting tool reaches the end of its life when it develops a critical level of wear or when a certain phenomenon, such as a catastrophic failure, occurs. In this section I will discuss these causes in more detail. Since in practice wear-out is the main issue, I will analyse this type of a tool failure first. Then I will discuss two kinds of phenomena that can occur – failure due to internal defects and failure caused by external stresses. As will be seen later, these two modes are different in nature and therefore require a different modelling approach.

2.1 Wear

Tool wear is the change of the shape of the tool from its original shape, during cutting, resulting from the gradual loss of tool material or deformation [14]. The main mechanisms causing this process are: abrasion, adhesion and diffusion [3]. In addition to that, tools suffer from oxidation and corrosion [24, 36, 37].

As the name suggests, abrasion is a process when hard particles that are present in the workpiece or that form during the cutting rub against the tool and mechanically erode it. Adhesion occurs when a chip flows over the tool or when the tool itself slides over a newly produced surface of the component. Since these interfaces are oxide-free, workpiece material can easily weld to the tool, and when the junctions are fractured by the sliding action, small fragments of the tool material are plucked out. Diffusion occurs at high temperature generated during cutting and is characterised by the flow of atoms from the workpiece to the lower concentration zones in the tool material and vice versa. This process changes the structure of the tool material and weakens it. Oxidation takes place when hot tool surfaces are exposed to the atmospheric gasses. As a result hard particles are produced, which promotes abrasion. Corrosion is caused by water that is the main constituent of cutting fluids and leads to development of soft compounds that are easily removed by abrasive forces. The combined action of these mechanisms produces several types of wear on the faces of the tool.

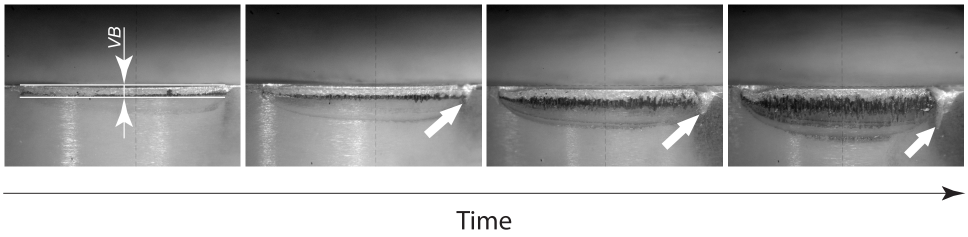


Fig. 2.1: Development of flank wear land and depth-of-cut notch

2.1.1 Types of tool wear

The most common and most broadly studied type of tool damage is the formation of a wear land on its flank face. This damage is called the *flank wear* and occurs when the relief face of the tool rubs against the newly produced surface of the workpiece. Due to this abrasive action a wear land is formed and its width (usually denoted by VB) increases as machining continues. This can be seen in Fig. 2.1, which shows the development of the flank wear land on a ceramic tool used in experiments described in Article 6.

The rate of the flank wear throughout the life of a tool is not constant. A typical example, based on the experimental data from Article 5, is shown in Fig. 2.2. As can be seen here, a tool would usually undergo a period of rapid initial degradation, called the burn-in, which would be followed by a steady wear phase, and eventually the wear-out would start. During this last period, the land produced by the flank wear gets so big that the damage done by the rubbing action on the newly produced workpiece surface becomes unacceptable. Moreover, large frictional forces increase the deflection of the tool, which causes a reduction in the dimensional accuracy of the component and poses an increasing risk of a catastrophic tool failure. Therefore the tool needs to be replaced before it enters the wear-out phase. In practice this is usually done when the average width of the flank wear land (VB) is expected to reach 0.3 mm or when the maximum width ($VB\,\mathrm{max}$) is around 0.6 mm.

On the right side of the flank wear lands, shown in Fig. 2.1, a small *notch* can be seen. This type of damage typically occurs at the boundary of tool's contact with the workpiece and is therefore sometimes called the depth-of-cut notch wear. Usually it develops when the tool cuts through the scale on the surface of the machined part. This hard layer could be the result of work-hardening during the previous path of the tool or could have formed during the fabrication of the workpiece. It causes severe abrasion and leaves a “scar” on both the major flank face and the adjacent area of the rake face of the tool. Such notch acts as a stress concentrator and can lead to tool fracture. As discussed in Articles 5 and 6, this can be a serious issue in machining of nickel-based materials with ceramic tools.

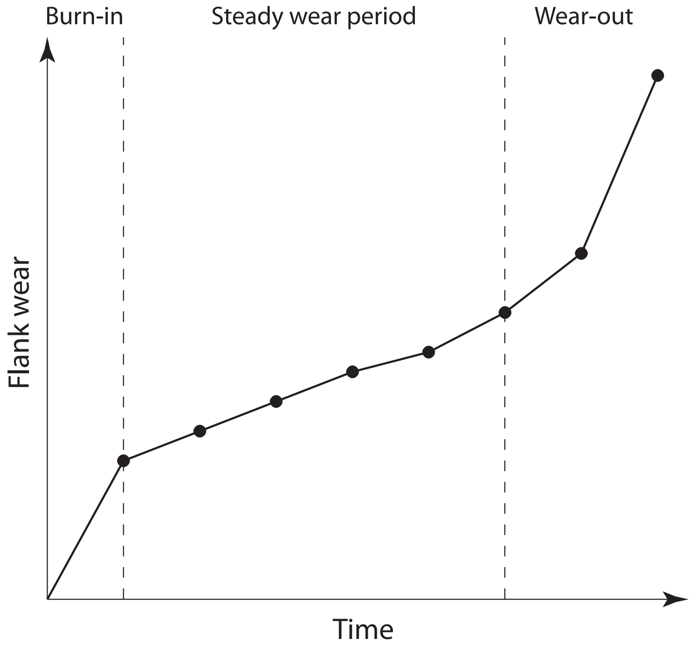


Fig. 2.2: Typical flank wear pattern

In addition to the depth-of-cut notch, in some of the experiments described in Article 5, we observed a notch on the minor flank face of the tool[3]. This kind of damage is sometimes called *groove wear* [36] or *nose radius wear* [24] and is shown in Fig. 2.3. Since it occurs on the trailing side of the tool, which has only a limited contact with the machined surface of the workpiece, groove wear cannot be explained by the abrasion alone. It is likely to be initiated by the corroding action of the water in the coolant [36] or by oxidation, which takes place as the hot surfaces of the tool are exposed to oxygen and nitrogen present in the surrounding atmosphere [37]. In both cases hard oxides can be built and can cause abrasion. The same wear mechanism could also be acting in the depth-of-cut region.

Flank and notch wear were studied explicitly in this PhD project. Other common types of tool degradation are crater wear and edge build-up. *Crater wear* is the formation of a pit on the rake face of the tool. As mentioned in Article 3, a small crater can be beneficial because it acts as a chip breaker. However, when its depth increases, the tool gets weaker and can eventually break. Since crater wear is mainly governed by diffusion [3, 5, 24], this can be a problem when machining at high cutting speeds, or in turning of heat resistant alloys, where even at low speeds temperatures can exceed 1000°C [26, 27].

---

[3] In the experiments under discussion the tool was a round insert. In such case there is no clear limit between the major and the minor flank faces. What is meant here is the side of tool's flank which was in contact with the machined surface of the part.

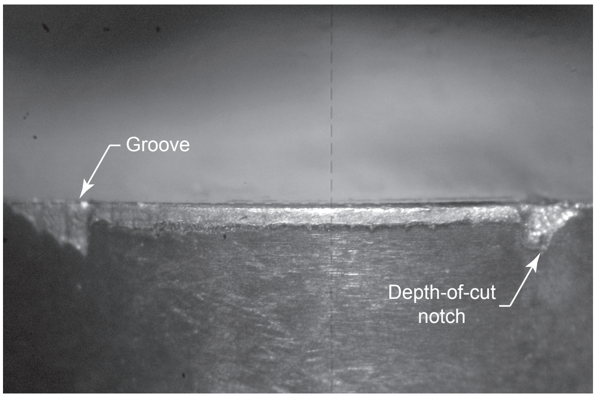


Fig. 2.3: Groove wear

Edge build-up is a process when several layers of work material adhere to the tool and is caused by the high friction between the rake face of the tool and the chip. Presence of the welded material further increases the friction and leads to the formation of new layers [3]. The resulting pile of material is known as the built-up edge (BUE). As it grows larger, BUE breaks apart and is carried away by the chip. This process repeats in a cyclic manner and can sometimes be desirable, because a thin BUE protects the tool and reduces its wear [5]. However, the deposited material blunts the cutting edge, which causes an increase in the cutting forces [3] and has a negative affect on the quality of the produced surface [3, 5, 24].

2.1.2 Factors affecting tool wear

Wear of cutting tools depends on a number of factors. They include the properties of a particular tool, such as its material, geometry and applied coatings, characteristics of the workpiece material being machined, cutting parameters, efficiency of the cooling process and the specifics of the machine tool.

Effect of tool material and geometry

As discussed in Section 1, development of more advanced materials has increased the wear resistance of cutting tools significantly. Carbides were said to be the most versatile of all. Moreover, their characteristics can be further enhanced by applying different coatings, which reduce friction, improve resistance to abrasion and act as a diffusion barrier. Nevertheless, in some applications, tools made in other materials perform better. For example, in machining of heat resistant aerospace alloys with ceramics, several times higher metal removal rate can be achieved [38]. Obviously, ceramics also have limitations. A particular issue is their sensitivity to thermal shocks. Consequently cooling, which in general significantly reduces wear, has to be used with caution or cannot be applied at all [3]. Tools made in other materials, even the newest ones, have their own weaknesses. Therefore in order to achieve the best results it is important to find the right tool for the right application.

Appropriate choice of tool geometry and chip breakers also helps to maximize the performance. For example, use of positive rake angles leads to lower cutting forces and in general reduces tool wear [24]. However, in the case of brittle tools, negative rake angles and high nose radii are needed to increase their toughness. Chip breakers allow to control the flow of the chips, which otherwise might hinder the cooling process and might cause alternating heating and cooling of the cutting edge [3]. This is especially important, when machining materials, which tend to form continuous chips.

Effect of workpiece properties

Tendency to form long chips and other properties of the work material, such as hardness, heat conductivity, chemical affinity to the tool material and proneness to work-hardening, have a substantial affect on the tool wear. However, they usually cannot be changed, because this would affect the properties of the component being produced.

Effect of machine tool properties

Modifying the characteristics of the machine tool, such as the rigidity and vibration damping capacity, is even more problematic. This might take a long time and might require big investments, which can be difficult to justify.

Effect of cutting parameters

Unlike the properties of the workpiece and the characteristics of the machine tool, cutting parameters are at the disposition of the process planner. The effect of cutting speed, feed and depth of cut was discussed previously in this thesis. In general, milder cutting conditions lead to lower tool wear. However this reduces the metal removal rate, thus it should be considered only as the last resort in most operations [24].

Effect of cooling

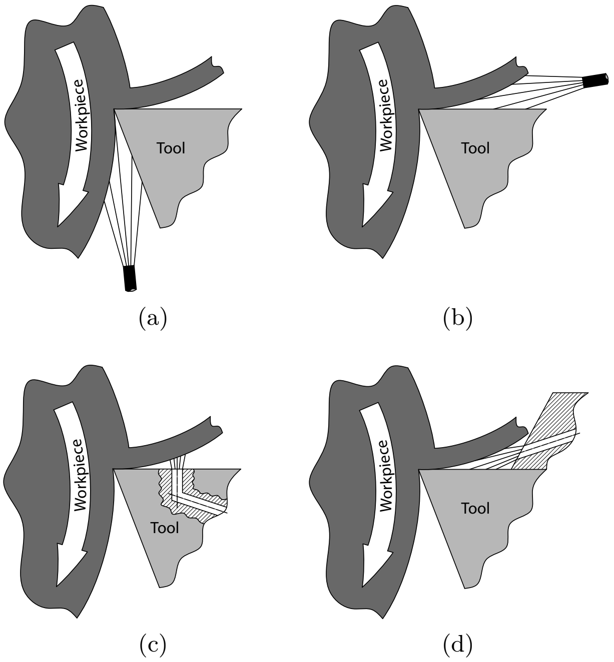
To conclude this section we will discuss the effect of cooling on tool wear. Temperature generated in metal cutting processes can exceed 1000°C and is the major factor governing all types of wear mechanisms [5, 36]. At high temperature tool materials soften and can be easier eroded by abrasive forces. Heat generated as a result of friction between the tool and the chip or between the tool and the workpiece causes adhesion. Rate of diffusion wear is directly proportional to temperature, therefore the location of the maximum depth of a crater (crater wear is mainly driven by diffusion) coincides with the position of the hottest area of the tool [5]. Finally, the intensity of oxidation and corrosion also gets higher when temperature rises because this increases the speed of chemical reactions.

The above discussion implies that reducing temperature generated during the cutting process is an important means of minimizing tool wear. Traditionally this has been achieved by pouring copious quantities of water-based fluids onto the tool. This method (also called conventional cooling) has its limitations. Pigott and Colwell [39] state that conventional cooling is inefficient, because fluids are poured onto the chip, thus cutting tool is cooled indirectly. Moreover at high temperatures coolants are rapidly evaporated. As a result a steam "blanket" is created, which stops the fresh coolant from reaching the tool-chip interface thus rendering conventional flushing ineffective. A few alternative techniques, including internal chilling of the insert, cryogenic-, CO_2- and high pressure cooling, have been tested. The latter method seems to be particularly promising and was studied extensively in this PhD project (see Articles 5 and 6).

The principle behind the high pressure cooling is to supply the cutting fluid in the form of a small jet as it is illustrated in Fig. 2.4. Already the early experiments [39] showed that, in rough turning of aircraft exhaust valves made in a nickel-based alloy, this method could increase the output per tool by over 18 times. Similar enhancement in the life of carbide inserts were obtained in turning of Inconel 718 [40, 41], Waspaloy [42] and Ti–6Al–4V [43–47], milling of titanium [48] and grooving of Inconel 718 [49] and Ti–6Al–4V [50].

The reason behind the above-mentioned improvements is the ability of the high pressure jet to penetrate deeper between the tool and the chip or between the tool and the workpiece. This way cooling takes place closer to the highest temperature zone and is therefore more efficient [39, 51, 52]. Another important advantage to mention is improved chip breaking. Moreover, high pressure cooling can reduce friction [53, 54] and can consequently lead to lower cutting forces [46, 48, 51, 54] and reduced tendency to form built-up edge [39]. It should be mentioned, though, that in some cases application of high pressure cooling requires a special care. A good example is machining with ceramic tools.

Due to the exceptional hot hardness and very good abrasion resistance tools made in ceramics outperform carbides, especially in machining of heat resistant alloys, where very high temperatures are generated [38]. Unfortunately, ceramic materials are sensitive to thermal stresses and have been reported to perform poorly under high pressure cooling [55–58]. The experiments described in Article 5 revealed that one of the reasons for this could be inappropriate system design and configuration leading to an unstable cooling process. Another issue to consider is the sensitivity of ceramic tools to notch wear, which is usually accelerated when high pressure cooling is applied. As we have demonstrated in the study described in Article 6, when these issues are taken care of, significantly improved performance could be achieved. In this particular study a rigid system with a minimal jet impingement angle was designed, and a special



(a) (b)

(c) (d)

Fig. 2.4: Types of high pressure cooling systems

grade of tools with improved resistance to notching was chosen. As a result, notch wear became insignificant, while flank wear was much lower under high pressure cooling as compared to conventional flushing. Consequently a significantly longer tool life was achieved. Moreover there was some evidence that the wear process was more stable. Therefore tool life variability was slightly lower.

2.2 Internal defects

Most modern cutting tools are produced by blending hard particles with a metallic or a ceramic binder and sintering the mixture at a high temperature. The resulting heterogeneous structure can contain various internal defects, such as pores, micro cracks, impurities, unsintered powder and agglomerates of one of the constituents [2, 59, 60]. Additional flaws can appear as a result of socks incurred during transportation and handling of the tools [24].

Internal defects act as stress concentrators and can initiate a fracture mechanism. When a tool is subjected to cyclic loads, which can result from interrupted cutting, vibrations or repeated heating and cooling cycles, cleavage along the stressed zones begins. This weakens the material, thus the crack starts to propagate. Eventually a piece of the tool breaks away. When this takes place close to the cutting edge the consequence is *edge chipping*. As a result, the tool becomes dull, produces surfaces of poor quality and therefore has to be replaced.

If the crack propagates deeper, big portion of the tool can break apart. This phenomenon is called *gross fracture* and leads to immediate tool failure.

2.3 External stresses

External stresses are severe overloads of the tool that arise from external sources, i.e., other than the machining conditions, workpiece properties or the state of the tool itself. Such stresses are random in nature, but are usually so great that they lead to immediate tool failure, irrespective of the level of wear and regardless of whether the tool contains some internal defects or not. A typical analogy is a human being hit by a car or by a bomb blast. In such case the age or the physical condition of the person makes little difference, as the sock exerted by the car or the blast is overwhelming.

One of the causes of external stresses is various human errors. For example, the operator might crash the tool into the workpiece, the programmer might specify wrong positions or cutting parameters in the part program, the tool handler might not clamp the insert properly, and so on. Such failures are a serious concern in machining processes. At the same time this is a very sensitive issue. Therefore it is seldom that the officially reported cause of a failure is a human error. Below is an example.

During the case study described in Article 1 we have learned that after a breakdown the operator would often manipulate the machine manually in order to return the tool to a “safe” position and to restart the automatic cutting cycle. Unfortunately “safe” positions are rarely really safe, and new failures, especially tool breakages, often happen. Taken straightforward such occurrences are human failures. However, the operator was actually forced to run the machine manually, therefore it is unfair to blame him or her. For such reasons collecting data for analysing human failures is difficult. Finding a certain law helping to predict them is even more complicated due to individuality and often unpredictability of the human behaviour. Some of these issues are discussed in Article 1. However, some other interesting findings had to be left out due to the sensitivity of the topics involved.

Besides the human failures, external stresses could result from malfunctioning of sensors, failure of fixtures, tool holders, clamps and other supporting equipment, instabilities in the cooling process, or various other sources. When working on the case study mentioned before, we found a number of notes about machine crashes, likely followed by catastrophic tool failures, which occurred when limit switches got contaminated or otherwise did not function properly. Part fixtures, tool holders, and insert clamps can give way under extreme loads acting in metal cutting. Cooling system can be clogged by small chip fragments or can come to a halt due to a pump failure, which would lead to a rapid overheating and failure of the tool. In addition, various other random and unpredictable events can happen and can cause a catastrophic tool failure.

3

Stochastic tool life modelling

Process variation and different failure modes make the tool life difficult to predict. To deal with this issue, Ermer [61] proposed to periodically update the constants of Taylor's tool life equation on the basis of real production data. Maropoulos and Alamin [62] reported that in practice, such an approach could improve the predictability of the tool life. Others argued that a more natural way of incorporating variation into the modelling was to treat the tool life as a random variable. To the best of our knowledge, this idea was first proposed by Taylor [63], and its feasibility was demonstrated by Wager and Barash [32], who carried out several hundreds of machining tests and found that the real life of HSS tools differed significantly from the deterministic estimates. Therefore

in this PhD thesis it will be assumed that tool life is a *stochastic* quantity and a *probabilistic* model to predict it will be proposed.

3.1 Probability distributions for tool life modelling

Starting with the study carried out by Taylor [63], different probability distributions, including the normal (Gaussian) [32, 64–66], the lognormal [63, 67–70], the Gamma [68], the inverse Gaussian [71], the Bernstein [72–74], the exponential [67, 75] and the Weibull [32, 70, 76–78], have been proposed for modelling the tool life. As discussed in Articles 2 and 3, some of these distributions are not very adequate for this particular application, while others are more realistic.

The normal distribution is simple to use and has symmetrical properties. However, it allows obtaining negative tool lives, which obviously makes it a non-realistic model. The lognormal distribution has a strange failure rate function, which increases for a certain time, but then starts decreasing and approaches zero [79]. While the expected rate of tool failures might be decreasing in some special situations (see Articles 2 and 3), it is strictly increasing in the long run, meaning that the tool is more and more likely to fail as the cutting continues.

For this reason the lognormal distribution is not a realistic tool life model either. The failure rate function of the inverse Gaussian distribution and the Bernstein distributions, which can be reduced to the inverse Gaussian [72], is similar to that of the lognormal distribution. However, instead of decreasing to zero, it approaches a certain non-zero value when the time increases [79]. The problem with the Bernstein and the inverse Gaussian distributions, however, is the difficulty of parameter estimation. Due to this reason, they are not very easy to use in practice. The Weibull distribution, on the other hand, is rather simple to apply and is very flexible, thus it can be adopted for modelling of several shapes, such as decreasing, constant, and increasing, of the failure rate function (e.g., see [79].)

Flexibility and the closed form of the reliability (survivor) and failure rate functions of the Weibull distribution were emphasized by Liu [78] and El Wardany and Elbestawi [70]. Ramalingam and Watson [76] have showed that the Weibull distribution was an appropriate model when the tool failed after the first incurred shock. Following them von Turkovich and Henderer [77] applied Weibull distribution with the shape parameter $\alpha < 1$, i.e., with a decreasing failure rate, to model the life of HSS taps, which primarily failed due to edge chipping. Similarly Rossetto and Levi [67] assumed that the exponential distribution, which is a special case of the Weibull distribution, could be used to model the tool life with respect to fractures. Pandit [75] suggested that the exponential distribution was also a suitable tool life approximation, when high cutting speeds were used, while El Wardany and Elbestawi [70] stated that the Weibull distribution was appropriate, when it was not possible to distinguish between the different failure modes.

Suitability of the Weibull distribution was also verified in our own studies described in Articles 1, 2 and 3. In the case study discussed in Article 1 it appeared to be the most appropriate model for various types of machine tool failures. Unfortunately, we could not check if this distribution was suitable for predicting the wear-outs and breakages of the cutting tools, because very few such failures were mentioned in the maintenance reports. The reason for that was the fact that worn and broken tools were usually replaced by the machine operators. Thus no maintenance intervention was required unless a tool failure occurred as a results of a more serious breakdown. In studies described in Articles 2 and 3, on the other hand, the Weibull distribution turned out to be a fairly good fit for our experimental tool life data. This was especially the case of the study described in Article 2 where a set of coated carbide inserts was tested. In experiments described in Article 3 the tools were based on CBN. In this case a few inserts wore out earlier than predicted by the Weibull model, but the fit to the remaining data points was very good.

From this discussion it follows that the Weibull distribution is very flexible and can be adopted for various applications, including tool life prediction. Therefore the Weibull distribution will used extensively in the models described in the following sections.

3.2 Tool reliability function

In Section 2 of this thesis I have discussed three possible causes of a tool failure: (i) wear, (ii) internal (hidden) tool material defects and (iii) external stresses. Imagine that we have purchased a batch of tools (inserts). We expect that a certain part of them are “bad”, i.e., they contain internal defects and will probably fail early in the cutting process. The rest are “good” tools and will have a normal life. Regardless of their quality, all tools can fail due to external stresses. Modelling of these three scenarios is presented in this section. Then the total tool reliability function, which in general is defined as $R(t) = \Pr(T > t)$ for $t > 0$ and shows the probability that an item will survive to and will still be functioning at a certain time t, is derived.

3.2.1 Modelling of tool wear

As discussed in Section 2.1, tool wear is a gradual, progressive process indicating that the failure rate function should be an increasing function of time. The Weibull distribution has been shown to be an adequate life distribution for wear processes [79] and is so flexible that it can be adapted to most practical cases. The two-parameter Weibull distribution is hence an obvious choice of model for the wear-life of the tool. Then the reliability (survivor) function of the “good” tools can be written as

$$R_g(t) = \exp\left(-(\lambda_g t)^{\alpha_g}\right) \tag{3.1}$$

where t is the time from the start-up of the tool, $\lambda_g > 0$ is a scale parameter, and α_g is a shape parameter. The shape parameter is an indicator of the progressive effect of the wear, while the scale parameter is an indicator of the speed of the wear process. For a fixed value of α_g, the mean time to tool failure will be inversely proportional to the scale parameter λ_g.

3.2.2 Modelling of failures due to internal defects

“Bad” tools are attacked by the same wear mechanisms. Thus, the reliability of these tools can also be modelled by a two-parameter Weibull distribution and can hence be written as

$$R_b(t) = \exp\left(-(\lambda_b t)^{\alpha_b}\right) \tag{3.2}$$

where α_b is the shape and λ_b is the scale parameter for the “bad” tools. Since “bad” inserts contain internal defects, they will generally fail earlier than the “good” ones. Therefore the scale parameter for the “bad” tools must be higher, i.e., $\lambda_b > \lambda_g$.

3.2.3 Modelling of external stresses

As discussed previously, external stresses are random by nature and are usually overwhelming. Thus it is reasonable to assume that the very first occurrence of such a stress causes a complete tool failure, regardless of the level of wear at that instant of time and regardless of whether the tool is “good” or “bad”. On the other hand, such failures could introduce micro cracks in the tool and could theoretically accelerate the wear of the rest of its edges. However, in practice the whole insert is usually discarded immediately after a catastrophic failure. Therefore possible correlation between the external stresses and the rest of the tool failure modes can be disregarded. From this discussion it follows that the external stresses are random events, the very first occurrence of which results in a tool failure, and that other failure modes are independent of the external stresses. Such situations can be adequately modelled by a homogeneous Poisson process [79]. Then the reliability of the cutting tools with respect to external stresses can be expressed as

$$R_s(t) = \exp\left(-\lambda_s t\right) \tag{3.3}$$

where λ_s is the rate (also called the frequency) of external stresses.

3.3 Total tool reliability function

Considering all three types of failure modes the total tool reliability function can be derived. Let us imagine that we pick at random a tool from our batch. Then there is a certain probability, say p, that it is a “bad” tool. This implies that the probability that the tool is “good” is $(1-p)$. The reliability function of a tool that is picked at random from a sample and is subjected to wear mechanisms is hence

$$R_w(t) = (1-p)\cdot\exp\left(-(\lambda_g t)^{\alpha_g}\right) + p\cdot\exp\left(-(\lambda_b t)^{\alpha_b}\right) \tag{3.4}$$

This is the so-called mixed model, which is often used to model failures due to internal defects [80].

If the cutting process was under complete control, as it often happens in laboratory experiments, equation (3.4) could be used to model the tool life. In real production systems, though, the chance that something unpredictable will happen and that the tool will undergo a random overload needs to be taken into account. Then it can be assumed that the tool will fail either due to the combined effect of the wear and the internal defects, modelled by equation (3.4), or due to an external stress, modelled by a homogeneous Poisson process with frequency λ_s, whichever comes first. Such a situation can be modelled as a series system of two virtual components.

A series connected system fails as soon as one of its components breaks down. As an example let us take a simple electrical circuit illustrated in Fig. 3.1. It

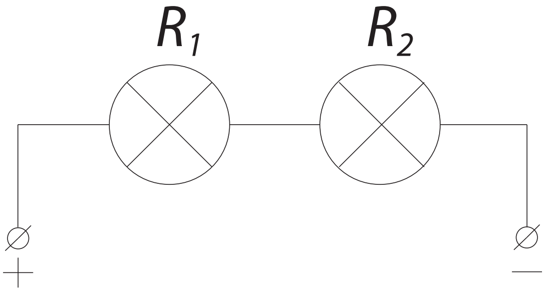


Fig. 3.1: Series-connected system

consists of two lamps connected in series, similarly as in some old Christmas-tree lights. Such a circuit functions as long as all the lamps (in this case two) are in tact. However, as soon as one of them burns, the current stops flowing and the light goes off.

The reliability of such a system can be represented as the product of the reliability functions of the individual components, i.e., $R_1 \cdot R_2$, [79]. Following this analogy, the total reliability function of the tool can be written as

$$R(t) = [(1-p) \cdot \exp\left(-(\lambda_g t)^{\alpha_g}\right) + p \cdot \exp\left(-(\lambda_b t)^{\alpha_b}\right)] \cdot \exp\left(-\lambda_s t\right) \tag{3.5}$$

The parameters entering this equation can be found experimentally. Lets us assume that we have tested n tools and in each case have determined the time until wear-out or other kind of failure, i.e., the tool life $T_{(i)}$, where $i = 1, 2, ..., n$. Let us also assume that some of the tools, say $N_b < n$, appeared to have significantly shorter life then the rest. Then it is reasonable to suspect that these tools contained internal defects. Given that we can estimate p by N_b/n.

The shape and the scale parameters of the Weibull distributions describing the life of both “good” and “bad” tools can be found by applying the maximum likelihood approach to the corresponding data sets. This can be done by using statistical software packages, for example Minitab. As it is showed in Articles 2 and 3, other method could also be applied.

Finding the rate of external stresses experimentally can be difficult, unless n is very large. If this is the case we could estimate λ_s by $N_s / \sum_{j=1}^{n} T_{(j)}$, where N_s is the number of tools that failed due to external stresses and $\sum_{j=1}^{n} T_{(j)}$ is the total observation time. In most cases, however, carrying out a large number of experiments is not practical. Therefore a smarter solution for estimating the rate of external stresses might be finding information about similar processes run elsewhere. A good example is the OREDA database that contains data about the reliability of offshore equipment. If, however, such data is not available, one could start with an educated guess and then improve the estimate as soon as the real production data gets collected.

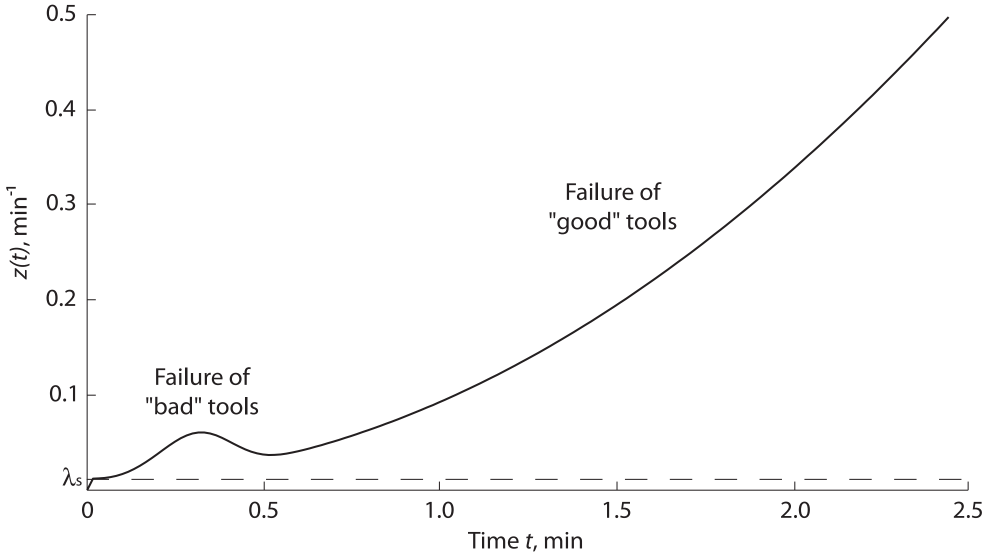


Fig. 3.2: The failure rate function of a tool for $\alpha_g = 3$, $\lambda_g = 0.3\,\mathrm{min}^{-1}$, $\alpha_b = 3.5$, $\lambda_b = 3\,\mathrm{min}^{-1}$, $\lambda_s = 1.2 \times 10^{-2}\,\mathrm{min}^{-1}$, $p = 0.01$

3.4 Tool failure rate function

Equation (3.5) can be used to determine the tool failure rate function $z(t)$, which is the conditional probability that a tool, that has been used and has survived up to a certain time t, will fail in a next short instance of time Δt. Hence, the failure rate function can be expressed as

$$\Pr(t < T \leq t + \Delta t \mid T > t) \approx z(t) \cdot \Delta t \tag{3.6}$$

and can be found as $z(t) = -\frac{\mathrm{d}}{\mathrm{d}t} \ln R(t)$ (e.g., see [79]).

For our case, the shape of $z(t)$ is shown in Fig. 3.2 for selected values of the parameters. This graph illustrates the expected behaviour of the tool. Since both the “good” and the “bad” tools are modelled by Weibull distributions with increasing failure rate functions, the probability of a breakdown during the first few seconds should be negligible. However, due to the external stresses, it is not equal to zero, and the tool can fail even at the very start of the cutting process. Since external stresses are random, we do not know when they are more likely to strike. Therefore, the probability of their occurrence remains the same throughout the process. In the example shown in Fig. 3.2 it is represented by the dashed line and is equal to 0.012 failures per minute, which is equivalent to one failure occurring each 83 minutes.

In addition to the constant risk of sustaining an external stress, the tool soon starts to wear. If we, by chance, have got a “bad” insert, it will tend to chip off or fracture and eventually fail early in the cut. Thus we will have an increasing probability of a failure. If, however, nothing will happen during some time, we will start believing that we have actually got a “good” tool, which will have a normal life. These tools would be unlikely to fail very early, thus for a certain time the probability of a failure would be decreasing.

Good examples of other systems that have a similar failure rate function are Roman bridges. When they were built there were probably some concerns about possible miscalculations, mistakes made during the construction, or internal flaws in the stones used. In spite of that, many of the Roman bridges were long lasting, some of them, like the Pons Fabricius in Italy, have survived for over two millennia. In cases like this our confidence in the system grows – probably the construction work was done properly, and the quality of the stones was most likely good. After all, if nothing has happened during so many years, why should something go wrong now? Thus in our eyes the probability of a failure is decreasing as the time goes.

Unfortunately, all physical systems have a limited life. Erosion and other destructive forces cause degradation even of such reliable structures as the Roman bridges. Thus, if no maintenance was done, sooner or later they would fail. The same is true for the cutting tools. Therefore after some time (approximately after half a minute in the example in Fig. 3.2) the failure rate starts to grow as we realise that the tool is wearing and is more and more likely to fail soon.

4

Models for determining tool replacement time

The replacement time for a cutting tool can be determined directly from its total reliability function. A simple procedure for that was developed in Article 2. It is based on the minimum acceptable level of reliability and is illustrated in the following section. Then it will be shown that the reliability function could also serve as an input to more advanced tool replacement time optimization models. One such approach was developed in Article 3. It is based on the age replacement policy, which seeks for the minimum cost during a replacement cycle, and will be introduced in this chapter.

4.1 Minimum acceptable reliability approach

In the study described in Article 2 we have performed machining experiments and found the tool reliability function $R(t)$. Since no external stresses occurred and all tools seemed to be “good”, it was appropriately modelled by a two-parameter Weibull distribution, the general form of which is

$$R(t) = \exp\left(-(\lambda t)^{\alpha}\right) \quad \text{for } t > 0$$

In such case the replacement time for the tool could be found as follows. Let us say that the *minimum* reliability that would be *acceptable* in a given situation is $R_{\min}$. This implies that the maximum risk that we are willing to take on is $(1 - R_{\min})$. Then, substituting $R_{\min}$ for $R(t)$ in the above equation and solving it for t_0 yields

$$t_0 = \frac{1}{\lambda}\left(\ln\frac{1}{R_{\min}}\right)^{\frac{1}{\alpha}}$$

This expression implies that t_0 is the longest cutting time during which the reliability of a cutting tool would be at least $R_{\min}$. Beyond this limit the risk of a failure would be grater than what we have defined as acceptable. Therefore the tool needs to be replaced no later than at time t_0.

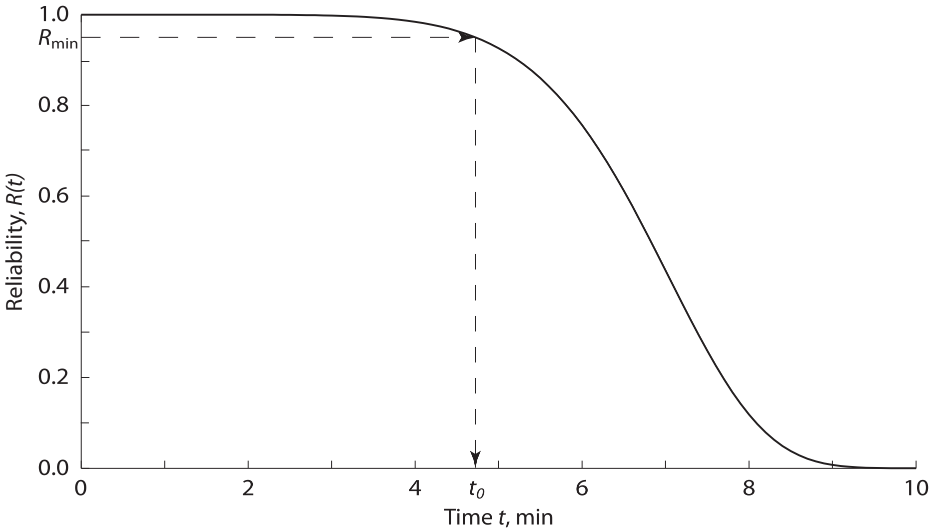


Fig. 4.1: Determining tool replacement time from reliability function

In more general cases, where neither the internal defects nor the external stresses can be ignored, expressing t_0 from the reliability function could be difficult. Thus in Article 2 we have proposed a graphical method, which is illustrated in Fig. 4.1. In this example $R_{\min} = 0.95$ and the resulting replacement time is $t_0 \approx 4.7\,\mathrm{min}$.

4.2 Minimum cost rate approach

The advantage of the above approach is its simplicity, which makes it easy to apply in practice. The drawback, though, is the fact that the choice of the acceptable level of reliability is arbitrary, that is, it is not supported by the cost calculations. In real situations the perception of what risk is acceptable does depend on the expected consequences and costs of a failure. Therefore taking these factors into account is important.

Let us assume that the cost of a tool failure is very low. A possible example could be mass production, where individual components made are relatively inexpensive and equipment often has excessive capacity. In such cases it might be reasonable to take a bigger risk in order to reduce the frequency of replacements. In the extreme case the tool would be changed only upon failure. This way the consumption of tools and the time lost for replacements would be minimized.

A different example is the aerospace industry. The production costs here are high and materials used are expensive. Therefore reworking or scrapping a component due to the damage inflicted by a tool failure is very costly. For such reasons risk tolerance in aerospace industry is low. Consequently cutting tools are replaced more frequently. This, however, means that more productive time is lost for replacements and the consumption of tools is high. Hence, in order to minimize the overall expenses a certain balance between the failure and other related costs needs to be found. Two common strategies to achieve that are scheduled and preventive tool replacements [6, 9, 73, 81–84].

Under the scheduled replacement strategy, also called the block replacement policy, the tool would be changed at fixed time intervals regardless of its age. This resembles the practices used in real machining processes, where the tool is replaced at pre-programmed instances of time or when it fails. However, as stated by Rausand and Høyland [79], the drawback of the block replacement strategy is its wastefulness. Imagine that we perform multiple cuts with the same type of tools, and the scheduled replacement time is at the end of each cut. Then assume that the tool has failed just before the planned replacement. In such case it would be replaced manually and the cut would be completed. Immediately after that the tool would have to be replaced again at the pre-programmed moment. Otherwise it might fail, because the cutting parameters were probably determined so that the useful life of the tool would be consumed during one cut. Thus a nearly new tool would be thrown away.

Under the preventive replacement strategy, also called the age replacement policy, the tool would be changed when it had been used for a certain time or upon failure, and the "clock" would be restarted after each replacement. Thus, the newly installed tool in the previous example would not be changed after completing the cut, but would continue to be used until its age would reach the predetermined limit. This strategy was further elaborated in Article 3, where an optimization model for the tool replacement time was developed.

4.2.1 Age replacement model

Let us imagine that we are running a machining process. We have installed a new tool at time $t = 0$ and are planning to replace it at time t_0, as it is illustrated in Fig. 4.2. In such case we would incur a cost c, which must cover the price of the tool, the expenses of time lost for replacing it and some other elements (see Section 4.2.2). However, the fact that tool life is a stochastic quantity implies that a failure can occur before the planned replacement. In such case the tool would have to be replaced earlier. The costs of such an unplanned tool change would be c, plus a *penalty cost* k, which must cover all extra expenses that are incurred as a result of a tool failure (see Section 4.2.2).

From this discussion it follows that the cost c is incurred each time a tool is replaced, while the penalty k needs to be paid only if the tool fails before the planned replacement. Then the expected tool cost for a given preventive replacement time t_0 is

$$C_T(t_0) = c + k \cdot F(t_0) \tag{4.1}$$

where $F(t_0) = \Pr(T < t_0)$, i.e., the probability that the tool will fail before the selected replacement time t_0.

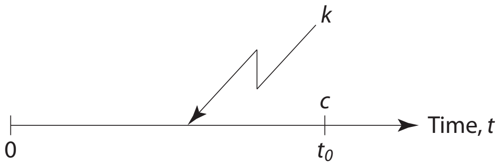


Fig. 4.2: Age replacement policy

Using the same reasoning the time between replacements will be equal to t_0 if the tool does not fail and equal to any value $0 < t < t_0$ if a failure occurs. Then the mean time between replacements (MTBR) can be found as

$$\begin{aligned}\mathrm{MTBR}(t_0) &= \int_0^{t_0} t \cdot f(t)dt + t_0 \cdot \Pr(T \geq t_0) \\ &= \int_0^{t_0} t \cdot f(t)dt + t_0 \cdot R(t_0) \\ &= \int_0^{t_0} t \cdot f(t)dt + t_0 \cdot (1 - F(t_0))\end{aligned}$$

Integrating by parts and simplifying yields

$$\mathrm{MTBR}(t_0) = \int_0^{t_0} (1 - F(t))dt \tag{4.2}$$

Note that in machining operations where tool replacement time can usually not be ignored it is more appropriate to call this quantity the expected cutting time until replacement.

Dividing the expression in equation (4.1) by the expression in equation (4.2) gives the rate of tool costs during the selected replacement period

$$C(t_0) = \frac{c + k \cdot F(t_0)}{\int_0^{t_0} (1 - F(t))\, dt} \tag{4.3}$$

Obviously the optimal replacement time t_0 is the one which minimizes the cost rate. This is called an *age replacement* policy [79].

4.2.2 Estimation of tool replacement costs

To determine the optimal tool replacement time by equation (4.3), the tool failure function $F(t)$, the preventive replacement cost c and the penalty cost k need to be known. The failure function can be determined as $F(t) = 1 - R(t)$, where $R(t)$ is the reliability function given by equation (3.5). The preventive tool replacement and the penalty costs are unknown yet and need to be determined.

Estimation of preventive tool replacement cost

Examining equation (1.3) given in Section 1.1.3 we can observe that the cost of preventively replacing a tool[4] can be found as

$$c = C_t + M \cdot t_{c_t} \tag{4.4}$$

where C_t is the cost of providing a sharp cutting tool or edge to the machine tool, M is the machine and operator rate and t_{c_t} is the tool replacement time.

These cost elements will obviously depend on the type of the cutting tool, machine and operator rate, tool magazine replenishment strategy and other factors. In general we could say that the cost of providing a sharp cutting edge would include the price of the tool, the cost of purchasing, storing and performing other internal activities, the costs of mounting the tool on a tool holder, delivering it to the machine tool and installing it into the tool magazine. After using-up all available edges the worn tool would have to be disposed as a result of which certain additional expenses would be incurred. Then if these costs are accumulated over a long period of time, dividing them over a total number of cutting edges used during the same time would give the average cost of providing a sharp cutting edge.

The machine and operator rate includes the wage of the operator per time unit and the rate of machine depreciation cost. On the top of that an overhead cost needs to be added to cover the administrative expenses of running the machining and the supporting processes.

The average replacement time could be determined by carrying out a time study. It would depend on both the operator (or the tool handler) and the type of the machine tool. In simple systems the operator would need to index the insert or to unclamp the tool and to install a new one manually. Most of these activities would be done while the machine would remain idle. In modern machining systems several identical copies of a tool would be inserted into the magazine and the replacement would be done automatically. The operator or the tool handler would still need to remove the tool from the magazine, index the insert and re-install the tool back. The average time per tool (or edge) to perform these activities would have to be added to the replacement time. It should be mentioned, though, that in some cases removing the tool from the magazine, indexing the insert and re-installing the tool back could be done in parallel without interrupting the machining process. Thus the time spent for performing these activities would not add to the replacement time. The associated costs could then be included in the cost of providing a sharp edge.

---

[4] Note that deterministic approaches disregard tool failures. Therefore all costs in equation (1.3) are preventive replacement costs.

Estimation of penalty cost

Determining the penalty cost is more difficult. As suggested by a few literature sources, a tool failure might cause rework or could lead to scrapping of the component being machined [85, 86], might bring down key equipment causing a loss of productive time and a reduction in throughput [82, 86, 87], and in the worst case could even lead to a loss of customer's good will [8]. To account for such costs an approach was developed in Article 4.

The basic idea of the proposed model is illustrated in Fig. 4.3. Here it is assumed that a workpiece, the value of which is $C1$, has been delivered for machining at a cost $C2$. However, after completing a certain part of the job, say $p2$ (here $0 \leq p2 \leq 1$), a tool failure has occurred. The most forgiving case is when this happens during the early stages of the cutting process. Then the damage inflicted by the failure might be possible to repair by the next operation. Let us say that the probability that this is the case is $p1$. The penalty cost then would be rather low and would mainly arise from the need to return the tool to a safe position before the point where the failure occurred and to restart the process. Let this cost be $C3$.

A more serious situation occurs when the tool fails during the finishing operations. In some of these cases it might still be possible to save the component via rework. Let the probability of this happening be $p3$. The penalty cost $C4$ in this case would be higher, i.e., $C4 > C3$.

In the worst case, the part would be scrapped. Then the penalty cost would be all its value, including the price of the workpiece, plus the cost of the work done up to the failure. Part of this value, say $C5$, could be salvage by selling the component as scrap.

Given these scenarios the expected penalty cost can be found as

$$k = p1 \cdot C3 + (1 - p1) \cdot p3 \cdot C4 + (1 - p1) \cdot (1 - p3)\,(C1 + C2 \cdot p2 - C5) \tag{4.5}$$

Costs $C1$ and $C2$ could be extracted directly from the enterprise resource planning (ERP) system. $C3$ and $C4$ would need to be estimated on the bases of the previous experience, and $C5$ would depend on the market price of the material. Finding probabilities $p1$ through $p3$ would be more tricky. A possible solution could be to follow the approach proposed for estimating the rate of external stresses, i.e., to start with a qualified guess in each case and to improve the estimates as soon as real failure data gets collected.

Equation (4.5) can be generalized and can be written as

$$k = \sum_{i=1}^{n} \left(\prod_{j=1}^{m_i} p_{ij} \right) \cdot k_i \tag{4.6}$$

where $i = 1, 2, ..., n$ is the scenario index, $j = 1, 2, ..., m_i$ is a decision step in the ith scenario, p_{ij} is the probability at the jth decision step of the ith scenario, and k_i is the penalty cost in case of the ith scenario.

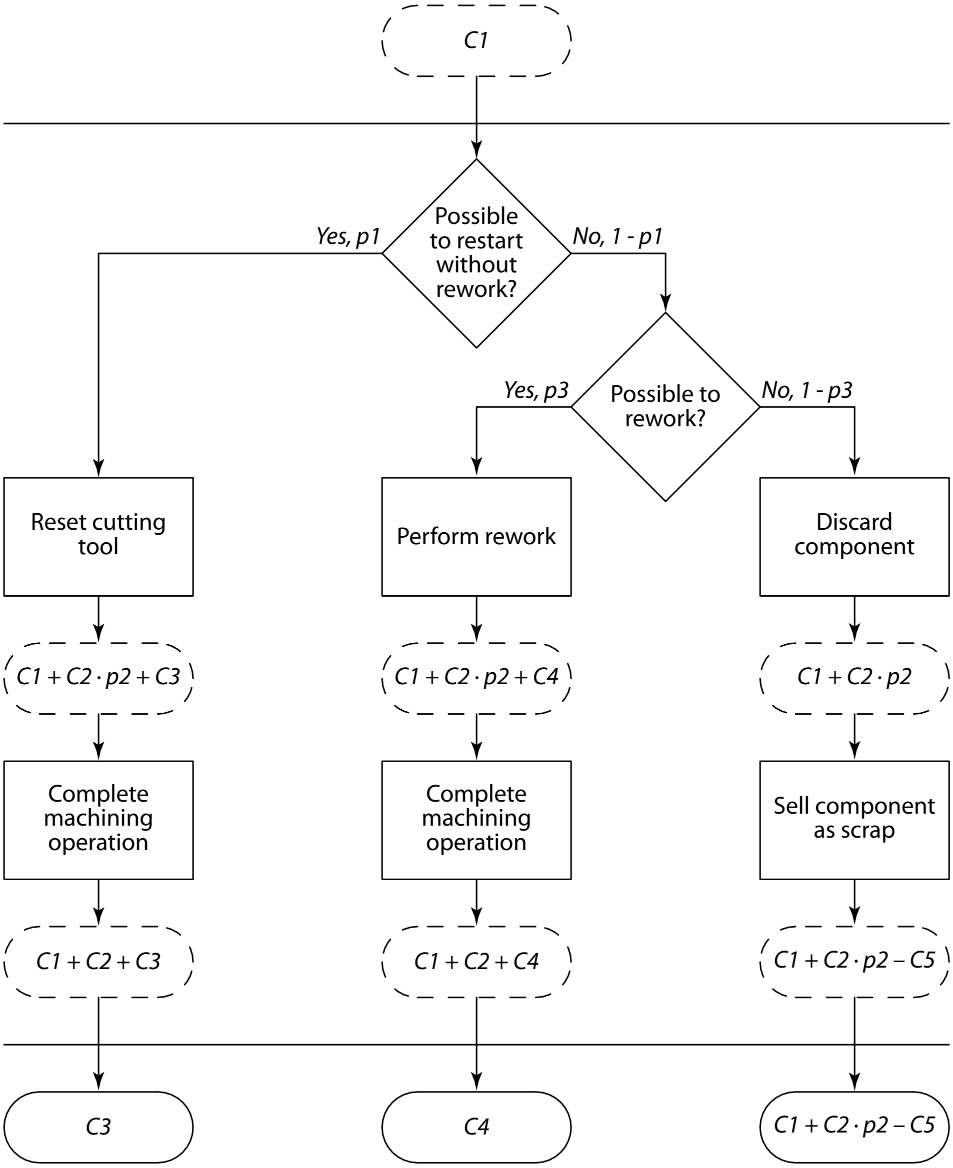


Fig. 4.3: Estimation of tool failure cost

As shown in Article 4, this probability tree-based approach is very flexible and can be extended beyond the machining process, for example, to include the costs of the quality control and customer losses. This gives a more realistic picture of the tool failure costs. However, the model becomes quite complex and requires more parameters to be estimated. Therefore its practical usability reduces. Hence, the most appropriate level of detail for a particular situation has to be chosen.

5

Solving optimization problem

In the previous section we introduced the age replacement model, given by equation (4.3), which enables to balance the costs of the preventive and failure provoked tool replacements, and showed how these costs could be estimated. Given that, the optimal tool replacement time can be found by minimizing equation (4.3). This includes calculating the integral $\int_0^{t_0}(1 - F(t))\,dt$, which can be difficult due to the complexity of the tool reliability function (note that $R(t) = 1 - F(t)$). One solution is to use computer programs, such as Matlab. Another approach is to apply the total time on test (TTT) transform to the tool reliability function. As will be demonstrated next, with the help of this technique the task of finding the optimal tool replacement time can be accomplished by following a simple graphical procedure.

5.1 Total time on test transform

Let us say we know the tool life distribution function $F(t)$. Then

$$H_F^{-1}(v) = \int_0^{F^{-1}(v)} (1 - F(t))\,dt \tag{5.1}$$

is its *TTT transform.* In this equation $F^{-1}(v)$ is the inverse of the probability distribution function $F(t)$. For example, for the two-parameter Weibull distribution $F(t) = 1 - \exp\left(-(\lambda t)^{\alpha}\right)$ for $t \geq 0$, $\lambda > 0$ and $\alpha > 0$. By rearranging this expression and by taking logarithms of both sides we find that the inverse function is $t = \frac{1}{\lambda}(-\ln(1 - F(t)))^{1/\alpha}$. Then by denoting $F(t) = v$, hence $t = F^{-1}(v)$, we can write the inverse function of the two-parameter Weibull distribution as $F^{-1}(v) = \frac{1}{\lambda}(-\ln(1 - v))^{1/\alpha}$.

Note that since $v = F(t)$ it satisfies the condition $0 \leq v \leq 1$. Moreover, when $v \to 1$, $F^{-1}(v) \to \infty$. This implies that

$$H_F^{-1}(1) = \int_0^{F^{-1}(1)} (1 - F(t))\,dt = \int_0^{\infty} (1 - F(t))\,dt = \text{MTTF} \tag{5.2}$$

i.e., $H_F^{-1}(1)$ is the mean time to tool failure. Then the *scaled* TTT transform of the distribution $F(t)$ is given by

$$\phi_F(v) = \frac{H_F^{-1}(v)}{\text{MTTF}} \tag{5.3}$$

The function $\phi_F(v)$ is defined for $0 \leq v \leq 1$ and takes values $0 \leq \phi_F(v) \leq 1$, such that the plot of the function $\phi_F(v)$ is contained in a 1×1 square. This graph is called the *TTT-plot* and is unique for each distribution.

5.2 Application of TTT transform to age replacement model

TTT transform can be used to find the optimal tool replacement time. For this let us return to equation (4.3). Since $F^{-1}(F(t)) = t$, we can express the denominator in terms of the TTT transform as follows

$$\int_0^{t_0} (1 - F(t))\, dt = \int_0^{F^{-1}(F(t_0))} (1 - F(t))\, dt = H_F^{-1}(F(t_0)) \tag{5.4}$$

Moreover equation (5.3) implies that

$$H_F^{-1}(F(t_0)) = \phi_F(F(t_0)) \cdot \text{MTTF} \tag{5.5}$$

Then by combining equations (4.3), (5.4) and (5.5) we get

$$C(t_0) = \frac{1}{\text{MTTF}} \cdot \frac{c + k \cdot F(t_0)}{\phi_F(F(t_0))} \tag{5.6}$$

For the age t_0 to be the optimal tool replacement time, the condition $C'(t_0) = 0$ must be fulfilled. By introducing $v_0 = F(t_0)$, that is, $t_0 = F^{-1}(v_0)$, we can write the derivative as follows

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}v_0} C(v_0) &= \left(\frac{1}{\text{MTTF}} \cdot \frac{c + k \cdot v_0}{\phi_F(v_0)} \right)' \\
&= \frac{1}{\text{MTTF}} \cdot \left(\frac{c + k \cdot v_0}{\phi_F(v_0)} \right)' \\
&= \frac{1}{\text{MTTF}} \cdot \frac{(c + k \cdot v_0)' \cdot \phi_F(v_0) - (c + k \cdot v_0) \cdot \phi'_F(v_0)}{\phi_F^2(v_0)} \\
&= \frac{1}{\text{MTTF}} \cdot \frac{k \cdot \phi_F(v_0) - (c + k \cdot v_0) \cdot \phi'_F(v_0)}{\phi_F^2(v_0)}
\end{aligned}$$

Then the optimality condition can be written as

$$\frac{1}{\text{MTTF}} \cdot \frac{k \cdot \phi_F(v_0) - (c + k \cdot v_0) \cdot \phi'_F(v_0)}{\phi_F^2(v_0)} = 0 \tag{5.7}$$

By multiplying both sides of equation (5.7) by MTTF $\cdot\, \phi_F^2(v_0)$ we get

$$k \cdot \phi_F(v_0) - (c + k \cdot v_0) \cdot \phi_F'(v_0) = 0$$

From this we can express $\phi_F'(v_0)$

$$\phi_F'(v_0) = \frac{k \cdot \phi_F(v_0)}{c + k \cdot v_0}$$

Finally by dividing both the numerator and the denominator by k we obtain

$$\phi_F'(v_0) = \frac{\phi_F(v_0)}{c/k + v_0} \tag{5.8}$$

which is the derivative of the scaled TTT transform of the tool life distribution function $F(t)$ at the optimal point v_0.

5.3 Parametric optimization procedure

Equation (5.8) implies that if $t_0 = F^{-1}(v_0)$ is the optimal tool replacement time, then the slope of the tangent line to the TTT transform of the tool life distribution function $F(t)$ at the point v_0 must be equal to $\phi_F(v_0)/(c/k + v_0)$. This is illustrated in Fig. 5.1. By making use of it we can derive the following procedure for determining the optimal tool replacement time:

1. Construct the scaled TTT-plot of the tool life distribution function $F(t)$.
2. Extend the abscissa axis to the left of the TTT-plot and indicate the point $(-c/k, 0)$.
3. Draw a tangent to the TTT-plot from $(-c/k, 0)$.
4. Read the abscissa value v_0 for the tangent point.
5. Calculate the optimal replacement time as $t_0 = F^{-1}(v_0)$.

The scaled TTT-transform of the tool life distribution function can be constructed and the inverse can be found with special computer programs. It is also possible to use general purpose mathematical software packages, such as Matlab or other.

5.4 Non-parametric approach

The optimization approach shown above requires the tool life distribution function $F(t)$ to be known. This is usually not the case. A possible solution is to run a set of tool life test, fit a probability distribution to the data and then

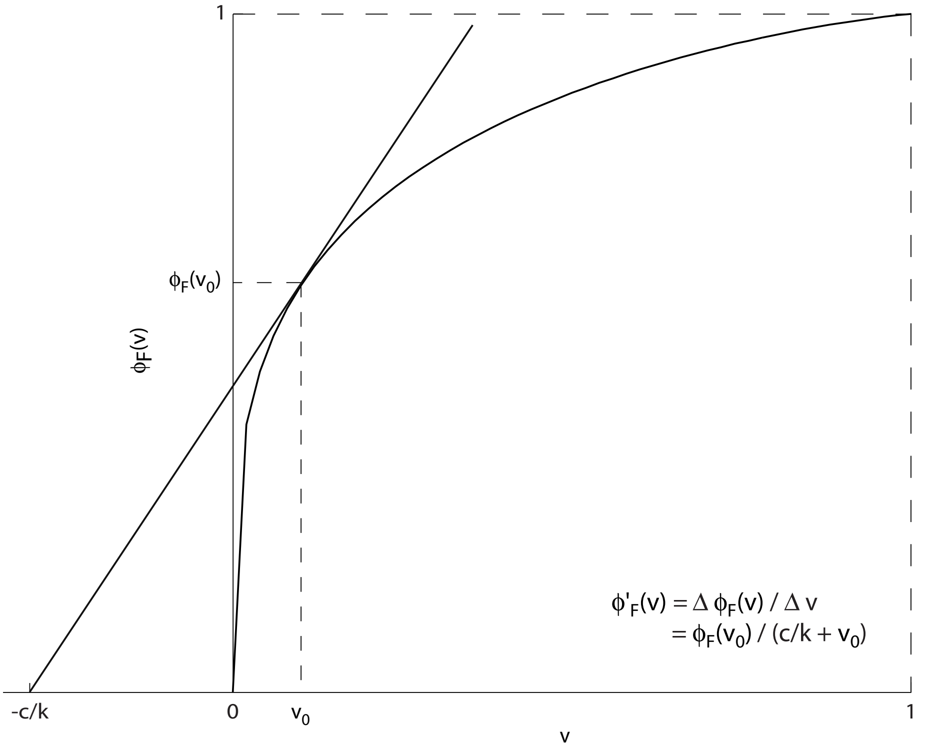


Fig. 5.1: Finding optimal tool replacement time from parametric TTT-plot

apply the above procedure. A more practical approach, though, is to estimate the TTT-plot. A procedure for that is described in [79] and will be illustrated in the following section. Then the estimated TTT-plot will be used to determine the optimal tool replacement time.

5.4.1 Estimation of TTT-plot

To illustrate the estimation of the TTT-plot let use an example. During the experiments described in Article 3 a set of 12 CBN inserts was tested. The recorded tool lives are listed in Table 5.1. Assuming that all 12 test runs were started at the same time $t = 0$, the first failure occurred after 62 seconds. The rest of the tools were still functioning, thus the total time on test (i.e., the cumulated operating time) was $62 + 11 \times 62 = 744$ seconds. The next failure occurred at the 97th second (counting from $t = 0$) and the cumulated operating time was $62 + 97 + 10 \times 97 = 1129$ seconds. Then

$$\mathcal{T}(T_{(i)}) = \sum_{j=1}^{i} T_{(j)} + (n - i)T_{(i)} \tag{5.9}$$

is the total (or accumulated) time on test when the ith tool fails at time $T_{(i)}$.

Table 5.1: Observed tool life data and TTT estimates

i	$T_{(i)}$	$\sum_{j=1}^{i} T_{(j)}$	$\sum_{j=1}^{i} T_{(j)} + (n-i)T_{(i)}$	$=$	$\mathcal{T}(T_{(i)})$	i/n	$\frac{\mathcal{T}(T_{(i)})}{\mathcal{T}(T_{(n)})}$
1	62	62	$62 + 11 \times 62$	$=$	744	0.08	0.374
2	97	159	$159 + 10 \times 97$	$=$	1129	0.17	0.568
3	113	272	$272 + 9 \times 113$	$=$	1289	0.25	0.648
4	149	421	$421 + 8 \times 149$	$=$	1613	0.33	0.811
5	151	572	$572 + 7 \times 151$	$=$	1629	0.42	0.819
6	175	747	$747 + 6 \times 175$	$=$	1797	0.50	0.904
7	182	929	$929 + 5 \times 182$	$=$	1839	0.58	0.925
8	189	1118	$1118 + 4 \times 189$	$=$	1874	0.67	0.943
9	202	1320	$1320 + 3 \times 202$	$=$	1926	0.75	0.969
10	205	1525	$1525 + 2 \times 205$	$=$	1935	0.83	0.973
11	215	1740	$1740 + 1 \times 215$	$=$	1955	0.92	0.983
12	248	1988	$1988 + 0 \times 248$	$=$	1988	1.00	1.000

Correspondingly, $\mathcal{T}(T_{(n)})$ is the total time on test after the last (i.e., nth) tool fails at time $T_{(n)}$. By plotting the $\mathcal{T}(T_{(i)})/\mathcal{T}(T_{(n)})$ versus i/n, as it is shown in Fig. 5.2, we get the empirical (non-parametric) TTT-plot of our data. According to Rausand and Høyland [79], $\mathcal{T}(T_{(i)})/\mathcal{T}(T_{(n)})$ is the natural estimate for the scaled TTT transform $\phi_F(v)$ for $v = i/n$; $i = 0, 1, 2, \ldots, n$.

This plot is very useful. For instance, we could check if a certain probability distribution is an appropriate model for our data. In the example in Fig. 5.2 it was assumed that the two-parameter Weibull distribution with the shape parameter $\alpha = 3.80$ could be used. By comparing its scaled TTT transform against the experimental data points we can conclude that in general the match is quite good, even though some deviation on the left-hand side can be seen.

The empirical TTT-plot also reveals the failure trends in the process under analysis. In general, a concave curve, as the one in Fig. 5.2, signals that the failure rate is increasing, convex shape means that it is decreasing and in case of the constant failure rate the TTT-plot is a straight line [79]. Combination of these patterns is also possible. An S-shaped TTT-plot, for instance, would indicate that the failure rate function has a bathtub form, which is common to many systems. Take for example a machine tool.

As we have learned during the case study described in Article 1, the rate of machine tool failures at the start of its exploitation can be considerably higher than that of similar equipment that has been in production for a longer period. This could be due to manufacturing defects, control software bugs or other problems. As these issues are worked out the number of problems, hence the failure rate, gradually decreases. This burn-in phase is followed by the useful life period, where the rate of failures is quite stable, though slight increasing due to the natural ageing. Eventually equipment enters the wear-out phase, and the number of failures starts increasing, unless serious maintenance and replacement of the worn parts is done.

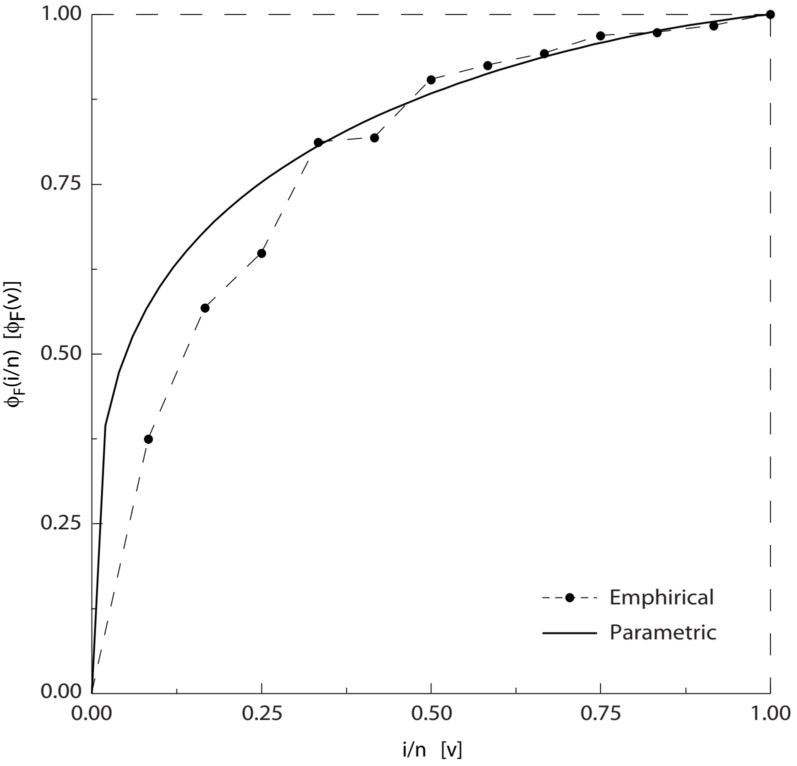


Fig. 5.2: Empirical TTT-plot with overlay parametric TTT-plot for Weibull distribution with $\alpha = 3.80$

5.4.2 Non-parametric optimization procedure

Besides having other useful features, the empirical TTT-plot offers a simple way to find the optimal tool replacement time. The procedure is as follows:

1. Build a data table similar to Table 5.1.
2. Plot $\mathcal{T}(T_{(i)})/\mathcal{T}(T_{(n)})$ versus i/n to construct the empirical TTT-plot.
3. Extend the abscissa axis to the left of the empirical TTT-plot and indicate the point $(-c/k, 0)$.
4. Draw a line through the point $(-c/k, 0)$ and one of the data points making up the TTT-plot, so that all other data points remain under the line.
5. Go back to the data table that was built in Step 1 and find the experimentally observed tool life linked to the data point through which the line passes. This is the optimal tool replacement time.

This procedure is illustrated in Fig. 5.3. Its main advantage is that we do not need to assume any specific distribution for the tool life, i.e., it is non-

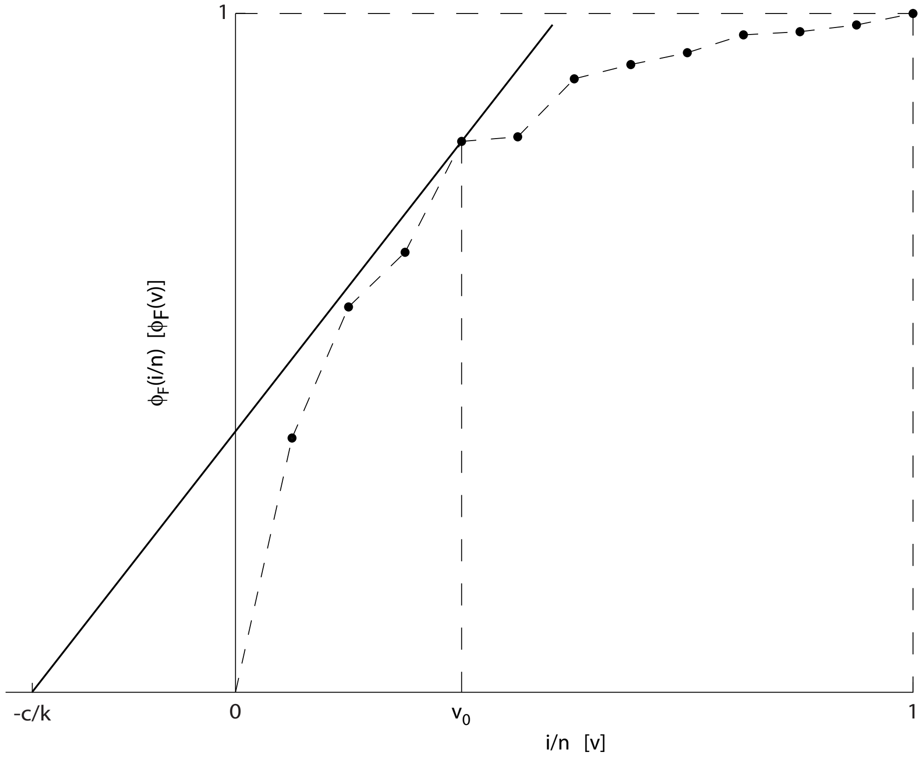


Fig. 5.3: Finding optimal tool replacement time from empirical TTT-plot

parametric. Moreover this method is very simple. All that is needed to construct the empirical TTT-plot, is a data table similar to Table 5.1. Such a table can easily be made with a spreadsheet program or even by hand calculations.

6

Summary and recommendations for future work

In this thesis the results of the work done during a three-year PhD project at the Norwegian University of Science and Technology are presented. Throughout this period some important issues concerning the efficiency of machining processes were studied. A special attention was given to the cutting tool as it plays a critical role in the economics of the process and is one of the main elements determining the reliability of the whole system. It was shown that due to the various sources of uncertainty the existing tool life prediction models gave inaccurate estimates, which lead to conservative tool replacement strategies and reduced the efficiency of the machining processes. To deal with the uncertainty, it was suggested treating the life of a cutting tool as a stochastic variable. Based on this, a set of new models and approaches for determining the optimal replacement time for the tool was proposed. Moreover a few other means, such as high pressure cooling, faster and more accurate reporting about machine tool breakdowns and more thorough analysis of the failure causes, to extend the tool life, reduce its variation and improve the availability and the efficiency of machining processes were studied. Based on this work a number of articles have been written, the copies of which can be found in the remaining two parts of this PhD thesis. The goal of the current part was to present the most important results, and in this concluding section their practical applicability as well as the contribution to the academic society is discussed. At the end a few recommendations for the future work are given.

6.1 Summary and discussion of the main results

The core of the method proposed in this thesis is the tool reliability function, which shows the probability that a tool will survive and will still be functioning at a certain time t. The analysis of the cutting processes revealed that this probability was affected by a few major factors, including wear, presence of internal defects and occurrence of external stresses.

Wear was said to be the most significant issue. Therefore various types of wear, the main physical mechanisms causing them and the means of controlling wear, including the high pressure cooling, which was studied extensively during this PhD project, were discussed. This analysis showed that tool wear was a gradual, progressive process. It was argued that in such cases the two-parameter Weibull distribution with an increasing failure rate function was the most appropriate model. Since tools with internal defects also suffer from wear, their reliability was also modelled by the Weibull distribution. However, the parameters in this case were assumed to be different to account for the fact that the life of the tools with internal defects is usually much shorter than that of the normal ones. Then it was assumed that the expected fraction of tools with internal defects was p, and the life of a tool picked at random was suggested to be modelled by a mixture approach. In addition to that, a tool, regardless of its quality, was assumed to be exposed to random external stresses, like operator errors or failure of auxiliary equipment. Such failures were included in the general tool life model as a homogeneous Poisson process.

Development of this reliability function is an important contribution. This allows us to determine the probability of survival until the selected moment of time. Having such information the planner of a machining process can decide if the risk is acceptable or not. In the latter case he or she could choose to replace the tool earlier, thus increasing the probability of survival. With this principle in mind we have developed two tool replacement models.

The idea of the first approach was to establish the minimum acceptable reliability level and then to choose the longest tool replacement time that would result in the selected probability of survival. The main advantage of this method is its simplicity, which makes it easy to implement in practice. The drawback, though, is that the lower reliability bound is established subjectively, i.e., without explicit consideration of various tool costs. Nevertheless this model could still be applied in practice, especially in those cases where the cost of a tool failure is difficult to estimate.

The second approach is based on the same principle, but in this case the decision represents a balance between the costs of preventive and failure provoked tool replacements. For this purpose an age replacement model was proposed. Based on it, the optimal tool replacement time is such that minimizes the mean total cost per a replacement period. To solve this optimization problem it was suggested to employ the TTT transform of the tool life distribution function. It was demonstrated that in this case the optimal tool replacement time could be determined graphically. The advantage of such approach is its simplicity, which again makes it relatively easy to apply in practice. A technical challenge, though, would be keeping track of the actual age of the tool. In modern manufacturing systems, this can be accomplished through the use of identification chips attached to tool holders. Another issue would be determining the cost of a tool failure, which is an essential input to the proposed approach.

Studying of the available literature showed that there were no good methods for calculating the costs of a cutting tool failure. Therefore a new model was developed and presented in this thesis. The approach is based on the probability-tree, which makes it very flexible and allows including various possible consequences of a tool failure. In general, a higher level of detail and a wider perspective give a more realistic picture of the real costs. However, this makes the model more complex. Therefore it was recommended for the user to choose the most appropriate level of detail according to his/her particular needs.

Besides considering the tool replacement problem we have investigated the effect of high pressure cooling on the tool life. This technique provides several benefits, such as prolonged tool life and improved chip breaking. During this PhD project we have demonstrated that these advantages also applied to ceramic tools, despite the fact that in most other studies an opposite result was achieved. This is an important contribution, because it means that the efficiency of the processes employing ceramic tools, for example rough machining of heat resistant alloys, could be improved with the help of high pressure cooling. Moreover we have found some evidence that the variability of the tool life could be reduced, which would make the tool life more predictable.

6.2 Recommendations for further work

The goal of this PhD thesis was to demonstrate how stochastic approaches could be used to model and predict the life of a cutting tool and to determine its replacement time. The validity of the concept was tested by publishing the results in peer reviewed journals and conferences and was verified in experimental and case studies. The next step is to extend the applicability of these ideas. For this some additional work needs to be done.

One particular issue of interest is testing of the proposed approaches under different sets of machining conditions. In the experiments conducted during this PhD project it was assumed that the cutting speed, the feed and the depth of cut were fixed. Changing one of these parameters would inevitably affect the reliability of the tool, hence the optimal replacement time. Therefore, to extend the applicability of the approach beyond the conditions tested during this PhD project, a law relating the parameters of the tool life distribution function to the cutting conditions would have to be found. This would mainly concern the “good” tools. Hence the functions of interest are $\alpha_g = f(v_c, f_n, a_p)$ and $\lambda_g = f(v_c, f_n, a_p)$. The life of the “bad” tools is first of all determined by the number of defects that they contain and only then by the wear. Therefore the effect of the process parameters on the life of these tools could be ignored. The rate of external stresses would not be affected either. This was illustrated in this thesis by using the car accident and the bomb blast analogies.

The above extension would allow to use the proposed concept for optimization of the total cost in a machining process. As it was discussed in this thesis, the choice of the cutting conditions affects the machining time, hence the productivity and the costs. This factor is not accounted for by the proposed models. It was assumed that the cutting parameters were determined in advance, and given that the balance between the preventive and the failure provoked tool replacement costs was found. Knowing functions $\alpha_g = f(v_c, f_n, a_p)$ and $\lambda_g = f(v_c, f_n, a_p)$, on the other hand, would allow finding the replacement time and the expected tool costs under different machining conditions. Then an iterative procedure for determining the optimal combination of the machining conditions and the tool replacement time could be developed.

Another issue that requires some additional consideration is the estimation of the tool failure costs. The model presented in this report contained only a few probabilities and cost elements which could be found by using simple procedures. In a more general case, like the one shown in Article 4, the model becomes more complex and the number of probabilities and related costs increases. In such case methods for estimation of these parameters would have to be developed.

The last recommendation for future work is to study the ways for reducing the process variation. As discussed in this thesis, process variation is the cause of poor tool life predictability and cannot be completely eliminated, which is the main reason for using stochastic tool life modelling methods. This however does not mean that variation cannot be reduced. In this PhD project we have studied high pressure cooling and showed that this technique could possibly make the process more predictable. These experiments should be extended to include different cutting conditions, other types of tools and workpiece materials. Besides the high pressure cooling other solutions for reducing process variation and making the life of cutting tools more predictable might also be feasible and need to be investigated.

References

[1] C.O.E. Perrigo, Lathe Design, Construction and Operation, The Norman W. Henley Publishing Co., New York, 1919.

[2] B. Benhabib, Manufacturing: Design, Production, Automation and Integration, Marcel Dekker, New York, 2003.

[3] G. Boothroyd, W.A. Knight, Fundamentals of Machining and Machine Tools, Taylor & Francis, Boca Raton, 2006.

[4] F.W. Taylor, On the art of cutting metals, Proceedings of the American Society of Mechanical Engineers 28 (3) (1906) 31–350.

[5] S. Kalpakjian, S.R. Schmid, Manufacturing Engineering and Technology, Prentice Hall, Singapore, 2006.

[6] Y.V. Hui, L.C. Leung, R. Linn, Optimal machining conditions with costs of quality and tool maintenance for turning, International Journal of Production Research 39 (4) (2001) 647–665.

[7] M.L. Vineyard, J.R. Meredith, Effect of maintenance policies on FMS failures, International Journal of Production Research 30 (11) (1992) 2647–2657.

[8] Y. Hui, L. Leung, Optimal economic tool regrinding with Taguchi's quality loss function, Engineering Economist 39 (4) (1994) 313–331.

[9] A. Jeang, Reliable tool replacement policy for quality and cost, European Journal of Operational Research 108 (2) (1998) 334–344.

[10] J. Sharit, S. Elhence, Computerization of tool-replacement decision making in flexible manufacturing systems: a human-systems perspective, International Journal of Production Research 27 (12) (1989) 2027–2039.

[11] A.E. Gray, A. Seidmann, K.E. Stecke, Synthesis of decision models for tool management in automated manufacturing, Management Science 39 (5) (1993) 549–567.

[12] S.Y. Hong, Economical and ecological cryogenic machining, Journal of Manufacturing Science and Engineering 123 (2) (2001) 331–338.

[13] P.W. Marksberry, I.S. Jawahir, A comprehensive tool-wear/tool-life performance model in the evaluation of NDM (near dry machining) for sustainable manufacturing, International Journal of Machine Tools and Manufacture 48 (7–8) (2008) 878–886.

[14] ISO 3685, Tool-life testing with single-point turning tools, International Organization for Standardization, Geneva, 1993.

[15] G. Byrne, D. Dornfeld, I. Inasaki, G. Ketteler, W. Konig, R. Teti, Tool condition monitoring (TCM) – the status of research and industrial application, CIRP Annals – Manufacturing Technology 44 (2) (1995) 541–567.

[16] K. Glass, R. Colbaugh, Real-time tool wear estimation using cutting force measurements, in: IEEE International Conference on Robotics and Automation, 1996, vol. 4, pp. 3067–3072.

[17] H.M. Ertunc, K.A. Loparo, H. Ocak, Tool wear condition monitoring in drilling operations using hidden Markov models (HMMs), International Journal of Machine Tools and Manufacture 41 (9) (2001) 1363–1384.

[18] S.E. Oraby, D.R. Hayhurst, Tool life determination based on the measurement of wear and tool force ratio variation, International Journal of Machine Tools and Manufacture 44 (12–13) (2004).

[19] K. Zhu, Y.S. Wong, G.S. Hong, Multi-category micro-milling tool wear monitoring with continuous hidden Markov models, Mechanical Systems and Signal Processing, In Press, Corrected Proof (2008) 1–14.

[20] G. Dini, F. Tognazzi, Tool condition monitoring in end milling using a torque-based sensorized toolholder, Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture 221 (1) (2007) 11–23.

[21] K. Iwata, T. Moriwaki, An application of acoustic emission measurement to in-process sensing of tool wear, CIRP Annals – Manufacturing Technology 25 (1) (1977) 21–26.

[22] X. Li, A brief review: Acoustic emission method for tool wear monitoring during turning, International Journal of Machine Tools and Manufacture 42 (2) (2002) 157–165.

[23] X. Mu, C. Xu, Tool wear monitoring of acoustic emission signals from milling processes, in: Proceedings of the 1st International Workshop on Education Technology and Computer Science, ETCS 2009, IEEE, Wuhan, Hubei, China, 2009, vol. 1, pp. 431–435.

[24] D.A. Stephenson, J.S. Agapiou, Metal Cutting Theory and Practice, second ed., Taylor & Francis, Boca Raton, 2006.

[25] M. Kronenberg, Replacing the Taylor formula by a new tool life equation, International Journal of Machine Tool Design and Research 10 (2) (1970) 193–202.

[26] N. Narutaki, Y. Yamane, K. Hayashi, T. Kitagawa, K. Uehara, High-speed machining of Inconel 718 with ceramic tools, CIRP Annals – Manufacturing Technology 42 (1) (1993) 103–106.

[27] T. Kitagawa, A. Kubo, K. Maekawa, Temperature and wear of cutting tools in high-speed machining of Inconel 718 and Ti–6Al–6V–2Sn, Wear 202 (2) (1997) 142–148.

[28] N.E. Woldman, R.C. Gibbons, Machinability and Machining of Metals, McGraw-Hill, New York, 1951.

[29] I.S. Jawahir, R. Ghosh, X.D. Fang, P.X. Li, An investigation of the effects of chip flow on tool-wear in machining with complex grooved tools, Wear 184 (2) (1995) 145–154.

[30] M. Christopher, Logistics and Supply Chain Management: Strategies for Reducing Cost and Improving Service, second ed., Financial Times/Prentice Hall, London, 1998.

[31] E.M. Goldratt, J. Cox, The goal, revised ed., Gower Publishing Ltd, Aldershot, 1989.

[32] J.G. Wager, M.M. Barash, Study of the distribution of the life of HSS tools, Journal of Engineering for Industry, Transactions of the ASME 93 (1971) 1044–1050.

[33] R.G. Fenton, N.D. Joseph, Effects of the statistical nature of tool-life on the economics of machining, International Journal of Machine Tool Design & Research 19 (1) (1979) 43–50.

[34] H. Wiklund, Bayesian and regression approaches to on-line prediction of residual tool life, Quality and Reliability Engineering International 14 (5) (1998) 303–309.

[35] M. Noël, M.S. Sodhi, B.F. Lamond, Tool planning for a lights-out machining system, Journal of Manufacturing Systems 26 (3–4) (2007) 161–166.

[36] T. Kurimoto, G. Barrow, B. Davies, The influence of aqueous fluids on the wear characteristics and life of carbide cutting tools, CIRP Annals – Manufacturing Technology 31 (1) (1982) 19–23.

[37] A.S. Coromant, Modern Skskärande Bearbetning: En Praktisk Handbok, AB Sandvik Coromant, Sandviken, 1994.

[38] J. Vigneau, J. Boulanger, F. Le Maître, Behaviour of ceramic tools during the machining of nickel base alloys, CIRP Annals – Manufacturing Technology 31 (1) (1982) 35–39.

[39] R. Pigott, A. Colwell, Hi-jet system for increasing tool life, Society of Automotive Engineers – Transactions 6 (3) (1952) 547–564.

[40] E.O. Ezugwu, J. Bonney, Effect of high-pressure coolant supply when machining nickel-base, Inconel 718, alloy with coated carbide tools, Journal of Materials Processing Technology 153–154 (2004) 1045–1050.

[41] E.O. Ezugwu, J. Bonney, Finish machining of nickel-base inconel 718 alloy with coated carbide tool under conventional and high-pressure coolant supplies, Tribology Transactions 48 (1) (2005) 76–81.

[42] K. Sørby, M.S. Bjørnstad, K. Tønnessen, The effect of high-pressure cutting fluid on chip breaking and tool life in turning of nickel-base alloys, in: CIRP 2nd International Conference on High Performance Cutting, 2006.

[43] A.R. Machado, J. Wallbank, I. Pashby, E.O. Ezugwu, Tool performance and chip control when machining Ti6Al4V and Inconel 901 using high pressure coolant supply, Machining Science and Technology 2 (1) (1998) 1–12.

[44] E.O. Ezugwu, J. Bonney, R.B. Da Silva, A.R. Machado, E. Ugwoha, High productivity rough turning of Ti–6Al–4V alloy, with flood and high-pressure cooling, Tribology Transactions 52 (3) (2009) 395–400.

[45] A. Nandy, S. Paul, Effect of coolant pressure, nozzle diameter, impingemnt angle and spot distance in high pressure cooling with neat oil in turning Ti–6AL–4V, Machining Science and Technology 12 (4) (2008) 445–473.

[46] A. Nandy, M. Gowrishankar, S. Paul, Some studies on high-pressure cooling in turning of Ti–6AL–4V, International Journal of Machine Tools and Manufacture 49 (2) (2009) 182–198.

[47] E.O. Ezugwu, R.B. Da Silva, J. Bonney, A.R. Machado, Evaluation of the performance of CBN tools when turning Ti–6AL–4V alloy with high pressure coolant supplies, International Journal of Machine Tools and Manufacture 45 (9) (2005) 1009–1014.

[48] R. Kovacevic, C. Cherukuthota, M. Mazurkiewicz, High pressure waterjet cooling/lubrication to improve machining efficiency in milling, International Journal of Machine Tools and Manufacture 35 (10) (1995) 1459–1473.

[49] R. Wertheim, J. Rotberg, A. Ber, Influence of high-pressure flushing through the rake face of the cutting tool, CIRP Annals – Manufacturing Technology 41 (1) (1992) 101–106.

[50] K. Sørby, K. Tønnessen, High-pressure cooling of face-grooving operations in Ti6Al4V, Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture 220 (B10) (2006) 1621–1627.

[51] B.K. Nagpal, C.S. Sharma, Cutting fluids performance – Part 1 – Optimization of pressure for "Hi-Jet" method of cutting fluid application, Journal of Engineering for Industry, Transactions of the ASME 95 (1973) 881–889.

[52] J. Kaminski, B. Alvelid, Temperature reduction in the cutting zone in water-jet assisted turning, Journal of Materials Processing Technology 106 (1–3) (2000) 68–73.

[53] C.S. Sharma, W.B. Rice, R. Salmon, Some effects of injecting cutting fluids directly into the chip-tool interface, Journal of Engineering for Industry, Transactions of the ASME 93 (1971) 441–444.

[54] M. Mazurkiewicz, Z. Kubala, J. Chow, Metal machining with high-pressure water-jet cooling assistance – a new possibility, Journal of Engineering for Industry, Transactions of the ASME 111 (1989) 7–12.

[55] E.O. Ezugwu, A.R. Machado, I.R. Pashby, J. Wallbank, The effect of high-pressure coolant supply when machining a heat-resistant nickel-based superalloy, Lubrication Engineering 47 (9) (1991) 751–757.

[56] K.M.C. Öjmertz, H.-B. Oskarson, Wear on SiC-whiskers reinforced ceramic inserts when cutting Inconel with waterjet assistance, Tribology Transactions 42 (3) (1999) 471–478.

[57] E.O. Ezugwu, J. Bonney, Effect of high-pressure coolant supplies when machining nickel-base, Inconel 718, alloy with ceramic tools, Tribology Transactions 46 (4) (2003) 580–584.

[58] E.O. Ezugwu, J. Bonney, D. Fadare, W. Sales, Machining of nickel-base, Inconel 718, alloy with ceramic tools under finishing conditions with various coolant supply pressures, Journal of Materials Processing Technology 162–163 (2005) 609–614.

[59] J. Dusza, L. Parilak, M. Slesar, Fracture characteristics of ceramic and cermet cutting tools, Ceramics International 13 (3) (1987) 133–137.

[60] M. Zeren, S. Karagoz, Defect characterization in the diamond cutting tools, Materials Characterization 57 (2) (2006) 111–114.

[61] D.D. Ermer, A Bayesian model of machining economics for optimization by adaptive control, Journal of Engineering for Industry, Transactions of the ASME 95 (1970) 628–632.

[62] P.G. Maropoulos, B. Alamin, Integrated tool life prediction and management for an intelligent tool selection system, Journal of Materials Processing Technology 61 (1) (1996) 225–230.

[63] J. Taylor, Carbide cutting tool variance and breakage: unknown factors in machining economics, in: Advances in Machine Tool Design and Research 1967: Proceedings of the 8th International M.T.D.R. Conference, 1968, pp. 487–504.

[64] C. Zhou, R.A. Wysk, Integrated system for selecting optimum cutting speeds and tool replacement times, International Journal of Machine Tools & Manufacture 32 (5) (1992) 695–707.

[65] W.T. Kwon, J.S. Park, S. Kang, Effect of group IV elements on the cutting characteristics of Ti(C,N) cermet tools and reliability analysis, Journal of Materials Processing Technology 166 (1) (2005) 9–14.

[66] P. Dašić, A. Natsis, G. Petropoulos, Models of reliability for cutting tools: Examples in manufacturing and agricultural engineering, Strojniški Vestnik – Journal of Mechanical Engineering 54 (2) (2008) 122–130.

[67] S. Rossetto, R. Levi, Fracture and wear as factors affecting stochastic tool-life models and machining economics, Journal of Engineering for Industry, Transactions of the ASME 99 (1) (1977) 281–287.

[68] S.S. Ramalingam, Tool-life distributions Part 2: Multiple-injury tool-life model, Journal of Engineering for Industry, Transactions of the ASME 99 (3) (1977) 523–531.

[69] K. Hitomi, N. Nakamura, S. Inoue, Reliability analysis of cutting tools, Journal of Engineering for Industry, Transactions of the ASME 101 (2) (1979) 185–190.

[70] T.E. El Wardany, M.A. Elbestawi, Prediction of tool failure rate in turning hardened steels, International Journal of Advanced Manufacturing Technology 13 (1) (1997) 1–16.

[71] G. Galante, A. Lobardo, A. Passannanti, Tool-life modelling as a stochastic process, International Journal of Machine Tools & Manufacture 38 (10) (1998) 1361–1369.

[72] M. Ahmad, A.K. Sheikh, Bernstein reliability model: Derivation and estimation of parameters, Reliability Engineering 8 (3) (1984) 131–148.

[73] A. Jeang, Tool replacement policy for probabilistic tool life and random wear process, Quality and Reliability Engineering International 15 (3) (1999) 205–212.

[74] M. Zhang, Z. He, Z. Liu, Tool replacement method study based on process capability and cost, in: Industrial Electronics and Applications, IEEE, Harbin, China, 2007, pp. 516–521.

[75] S.M. Pandit, Data dependent systems approach to stochastic tool life and reliability, Journal of Engineering for Industry, Transactions of the ASME 100 (3) (1978) 318–322.

[76] S.S. Ramalingam, J.D. Watson, Tool-life distributions Part 1: Single-injury tool-life model, Journal of Engineering for Industry, Transactions of the ASME 99 (3) (1977) 519–522.

[77] B. von Turkovich, W. Henderer, On the tool life of high speed steel tools, CIRP Annals – Manufacturing Technology 27 (1) (1978) 35–38.

[78] H.H. Liu, Cutting-tool reliability assessment in variable machining conditions, IEEE Transactions on Reliability 45 (4) (1996) 573–581.

[79] M. Rausand, A. Høyland, System Reliability Theory; Models, Statistical Methods, and Applications, Wiley, Hoboken, NJ, 2004.

[80] D.N.P. Murthy, M. Rausand, T. Østerås, Product Reliability; Specification and Performance, Springer, London, 2008.

[81] A.K. Sheikh, L.A. Kendall, S.M. Pandit, Probabilistic optimization of multitool machining operations, Journal of Engineering for Industry, Transactions of the ASME 102 (3) (1980) 239–246.

[82] U. La Commare, S.N. La Diega, A. Passannanti, Optimum tool replacement policies with penalty cost for unforeseen tool failure, International Journal of Machine Tool Design and Research 23 (4) (1983) 237–243.

[83] E.E. Iakovou, C. Koulamas, Adaptive tool replacement policies in machining economics, Journal of Manufacturing Science and Engineering 118 (4) (1996) 658–663.

[84] S. Schall, M.J. Chandra, Optimal tool inspection intervals using a process control approach, International Journal of Production Research 28 (5) (1990) 841–851.

[85] C. Koulamas, B. Lambert, M. Smith, Optimal machining conditions and buffer space size for the two-stage case, International Journal of Production Research 25 (3) (1987) 327–336.

[86] W.J. Zdeblick, R. De Vor, J.F. Kahles, A Comprehensive Machining Cost Model and Optimization Technique, CIRP Annals – Manufacturing Technology 30 (1) (1981) 405–408.

[87] E. Iakovou, C. Ip, C. Koulamas, Optimal solutions for the machining economics problem with stochastically distributed tool lives, European Journal of Operational Research 92 (1) (1996) 63–68.

Part II

Main articles

Article 1

Zydrunas Vagnorius[*] and Knut Sørby[†]

Reliability of machine tool systems in aircraft industry

In: Proceedings of the Eighth International Conference on Advanced Manufacturing Systems and Technology, CISM, Udine, Italy, 2008, pp. 293–304

---

[*] Zydrunas Vagnorius performed data analysis and wrote the entire text.

[†] Professor Knut Sørby established the contact with and arranged the trips to the company, where the data for the analysis was collected. He also contributed to the writing process by coming up with comments that helped to improve the readability and the overall quality of the text.

Is not included due to copyright

Article 2

Zydrunas Vagnorius[*], Knut Sørby[†] and Marvin Rausand[‡]

Probabilistic model for determination of tool replacement time

In: Proceedings of the 12th CIRP Conference on Modelling of Machining Operations, Mondragon Unibertsitateko Zerbitzu Editoriala, Donostia-San Sebastián, Spain, 2009, vol. II, pp. 757–764

This article was selected by the scientific committee of the conference for publication (subject to further review) in the Journal of Machining Science and Technology

---

[*] Zydrunas Vagnorius developed the tool replacement model, performed data analysis and wrote the whole text, except for one paragraph in Section 2. He also ran the machining experiments, during which the data for testing of the proposed approach was collected, and performed tool wear measurements.

[†] Professor Knut Sørby proposed the basic modelling idea. He also supervised the machining experiments and contributed to the writing process by coming up with comments that helped to improve the readability and the overall quality of the text.

[‡] Professor Marvin Rausand wrote the paragraph about the inverse Gaussian distribution in Section 2. He also contributed to the writing of the remaining parts by coming up with comments that helped to improve the readability and the overall quality of the text.

Is not included due to copyright

Article 3

Zydrunas Vagnorius*, Marvin Rausand† and Knut Sørby‡

Determining optimal replacement time for metal cutting tools

European Journal of Operational Research 206 (2) (2010) 407–416

---

* Zydrunas Vagnorius wrote the whole text, except for Section 4, drew all tables and figures, revised the article a few times, including the Section 4, and produced the final manuscript. He contributed to the development of the tool reliability function and the tool replacement model. He also ran the machining experiments, during which the data for testing of the proposed approaches was collected, performed tool wear measurements and conducted data analysis.

† Professor Marvin Rausand came up with the idea for modelling the reliability of the cutting tool and developed the approach for determining the optimal replacement time. He wrote most of the Section 4, sketched Figure 1 in Section 4 and Figures 5 and 6 in Section 5 and edited the rest of the article a few times, which has significantly improved the readability and the overall quality of the text

‡ Professor Knut Sørby participated in the development of the tool reliability function and the tool replacement model. His practical experience with metal cutting processes contributed significantly to the realism of these models. Then Professor Sørby supervised the machining experiments. He also contributed to the writing process by coming up with comments that helped to improve the readability and the overall quality of the text.

Is not included due to copyright

Article 4

Zydrunas Vagnorius* and Knut Sørby†

Estimation of cutting tool failure costs

In: 2009 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM 2009), IEEE, Hong Kong, China, 2009, pp. 262–266

---

* Zydrunas Vagnorius developed the model for estimating the expected cost of a cutting tool failure and wrote the entire text.

† Professor Knut Sørby contributed to the writing process by coming up with comments that helped to improve the readability and the overall quality of the text.

Estimation of Cutting Tool Failure Costs

Z. Vagnorius, K. Sørby
Department of Production and Quality Engineering, Norwegian University of Science and Technology,
Trondheim, Norway
(zydrunas.vagnorius@ntnu.no)

Abstract – **Probabilistic cutting tool replacement models assume that tool life is stochastic in nature. This implies that a tool can wear out before the planned replacement, as a result of which penalty costs are incurred. If these costs, as well as the tool failure function and the cost of scheduled replacement, are known, optimal tool replacement time can be found. While many researchers have focused on the latter two elements, there are very few articles explaining what penalty costs are and how they should be calculated. Therefore this article presents an approach for estimating the costs of tool failure for a one-stage machining operation.**

Keywords – **cost, failure, machining, tool**

I. INTRODUCTION

In one of our earlier studies [1] we have showed that due to variation in machining processes it is most appropriate to treat tool life as a random variable. This implies that the real tool life can be either longer or shorter than expected. In the former case tools are replaced too early, thus the replacement cost, including the losses resulting from reduced throughput, is incurred more frequently. On the other hand, when a tool fails earlier than expected there is a risk of damaging the workpiece and incurring other extra costs, which are collectively called *penalty costs*.

The two types of costs are inversely related, because replacing tools earlier, thus increasing the replacement costs, usually reduces the risk of incurring the penalty costs and vice versa. Therefore, statistical approaches, such as the age replacement model [2-5], seek for a tool replacement time that balances the two costs, so that the total cost is minimized. To apply these ideas practically, however, requires the values of the costs to be known. This is usually not a problem when it comes to traditional tool replacement costs. Calculation of the penalty costs, though, appears to be difficult.

Literature on this topic is rather scarce. Taylor [6] states that unit machining costs are likely to increase by 100% as a result of a tool failure. Similar estimates are given by Koulamas et al. [7]. They assume that while some tool failures might not affect the quality of the product, others might lead to rework and 50% increase in unit cost, or scrapping of the part and the loss of 100% of its value. The same consequences are studied by Zdeblick et al. [8], who employ a Markov chain to determine the machining costs. They suggest that when a part is scrapped penalty costs are even higher than the money spent for making it, because a replacement part needs to be produced to maintain the required output. This opinion is supported by La Commare et al. [9]. They assume that upon scrapping a part total penalty cost is equal to unit production expenses plus an extra charge equivalent to the cost of 2.5 – 100 labor minutes. Iakovou et al. [10] point to another important issue – the loss of productive time, and suggest that penalty costs be proportional to the downtime caused by the failure. Hui and Leung [4], who employ Taguchi’s quadratic loss function to study the effect of tool condition on product quality, go even further. They claim that a poor condition of the tool can not only lead to immediate scrapping of the part, but can also result in wastage of subsequent production resources, and can even cause a loss of customers good will.

Apart from the above examples, most other papers dealing with tool replacement focus on the model itself, while failure costs are not explicitly analyzed. Thus, if stochastic tool replacement models are to be applied in practice, generalized approaches for calculating these expenses are needed. To start with we propose a model for calculating the expected cost of a tool failure for a one-stage machining operation.

II. TOOL LIFE AND FAILURE

According to ISO 3685:1993(E) [11], tool life t_0 is the cutting time required to reach a tool life criterion. This can be illustrated graphically as shown in Fig. 1. Since tool life criteria are generally conservative, replacing a tool slightly after t_0 does not necessarily result in workpiece damage or other extra costs. At the time t_{cr}, though, tool degradation reaches a critical limit where it becomes detrimental to the quality of the produced part. We will call this Type 1 tool failure.

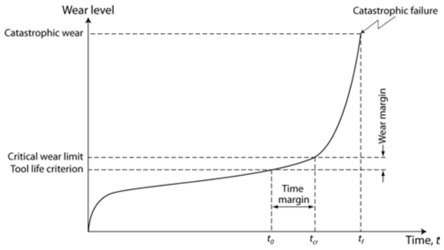


Fig. 1. Typical tool wear curve.

978-1-4244-4870-8/09/$26.00 ©2009 IEEE

Continuing machining with a worn tool leads to a rapid increase in cutting forces. Eventually the load becomes unbearable, as a result of which the tool breaks at the time t_f and needs to be replaced immediately. We will call this Type 2 or a catastrophic tool failure.

If machining processes were deterministic all tools would be replaced before the wear-out. In practice, however, tool wear varies significantly. Therefore at the selected replacement time the wear can have exceeded the critical limit. In some other cases the wear rate is so high that the tool breaks before being replaced. In addition, catastrophic tool failures can occur independently from gradual wear. They can result from external stresses causing severe tool overloads, or micro defects, which weaken the tool and lead to its breakage.

III. EFFECT OF TOOL FAILURE

The distinction between the two types of failures is important to keep in mind, because the consequences in the two cases are different. To analyze them let us consider a production process shown in Fig. 2.

A supplier (either external or internal) delivers a part for finish machining. The operation is supervised by an operator or a monitoring system. In both cases Type 2 tool failures, can be detected and the process can be stopped. Detecting Type 1 failures is technically also possible [12-16], but is more complicated and is not commonly applied yet. If the damage caused by the failure can be repaired, the process is restarted. Otherwise the part is scrapped. The penalty for delivering a defective part is high, thus all but the scrapped parts are inspected for quality. Defective parts can be sent back for rework or must be scrapped. Good parts are delivered to a customer (either external or internal), who also checks the product and can return it.

Let us now assume that the tool life distribution function is known, thus the probability of Type 1 and Type 2 failures can be determined. Possible consequences of such events are illustrated in Fig. 3. The notations used in the figure are defined in Table I and will only be commented in the text when necessary.

If a failure has occurred the probability that it is of Type 2 is $p1$. As mentioned before, these failures can be detected. In practice, though, monitoring process is not completely reliable. Therefore, it is reasonable to introduce $p2$ as the probability of detecting a Type 2 failure. Let us assume that the failure has been detected and machining has been stopped. The operator would then replace the broken tool, retract it to a "safe" position and try to restart the process. Unfortunately, "safe" positions are not always safe, thus manual restarts can result in new tool failures. The breakage could also occur during the last path of the tool. Then there would be little room for correcting possible damage on the part, and it might eventually need to be scrapped. Penalty cost in such case would have to cover the money invested in making the part, which at the moment of the failure is equal to parts initial value plus the cost of the work completed before the occurrence of the failure, i.e., $C1+C2\cdot p3$ (see dashed boxes). To this it would have to be added the restart cost $C3$, covering the costs resulting from increased tool consumption and wasted productive time. Some of these losses could be recovered by selling the rejected part as scrap at price $C4$. Thus, the penalty cost would be $C1+C2\cdot p3+C3-C4$.

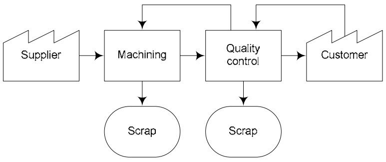


Fig. 2. Production process.

Another scenario would occur if the restart was successful. Then the work on the part would be completed and the part would be sent for quality control. Since tool failures can introduce surface distortions and cause other problems, the likelihood of such part failing the quality control would be increased (note that cases when a part fails the inspection due to reasons not directly attributable to tool failure should be disregarded here). If this happened, the part would be reworked. Failure of the repair would lead to scrapping of the part, the value of which by then would be increased by the cost of the quality control $C5$ and the cost of the rework $C6$.

Successfully reworked parts would be delivered to the customer, where they would undergo further inspection and could still be rejected. Penalties in such case could differ significantly. Some customers would return the product in exchange for a replacement part. The rejected unit could be reworked again or would have to be scrapped. The losses would then include the value of the part, as well as the overhead, delivery and the return costs. Other customers might refuse to take the replacement product and would request money refund, which in addition would mean the lost of actual and possibly future sales. Since such costs are difficult to quantify, we will not go further into this discussion, but will assume that the penalty here is the average cost of handling customer complains per a returned part.

When the customer does not complain the penalty is the sum of all other extra expenses incurred as a result of the tool failure. In this example these are $C3$ and $C6$.

Now let us return to the quality control and check what happens when the part passes the inspection. Such product could be immediately delivered to the customer. If the customer would reject it, the cost would be the same as in the previous example. However, if the part was accepted, the penalty cost would be only $C3$, as no rework is done on the parts passing the quality control.

Going even further back we assume that a Type 2 failure has occurred, but was not detected. This would usually cause severe damage, so it would be very unlikely that the part would pass the quality control. The chance of

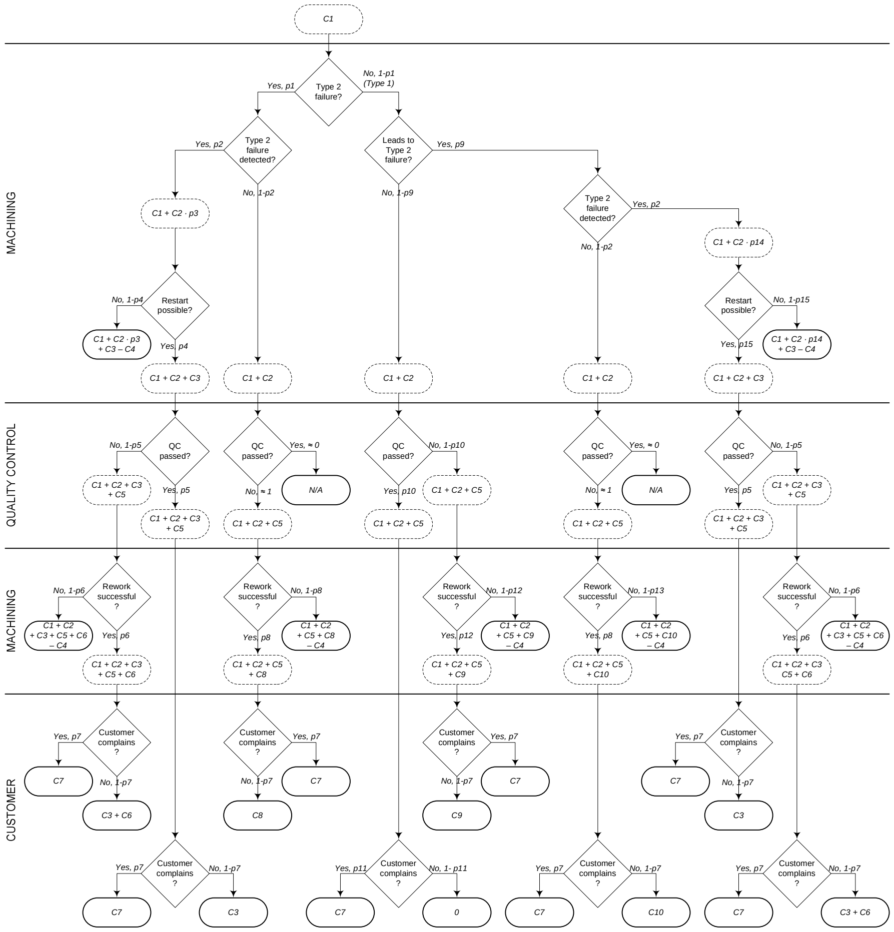


Fig. 3. Tool failure scenarios and resulting costs.

a successful rework would also be much lower than previously, i.e. $p8 << p6$. The remaining logic is the same as in the previous scenario.

Now let us have a look at Type 1 tool failures. As mentioned before, in-process detection of these failures is difficult. Thus, we will assume that Type 1 failures can only be detected if they lead to Type 2 failures.

The damage caused by a Type 1 tool failure can be small, so it is possible that such a defective part would slip through the quality control and would be delivered to the customer. If the customer rejected such a part, the cost would be the same as in previously discussed cases. The probability of the reject, though, would be grater here, i.e., $p11 > p7$. On the other hand, it is also possible that the damage would be so insignificant that the customer would not complain about it and no penalties would be incurred.

If the defect was detected during the quality control, the part would be reworked. The probability of success $p12$ here would generally be grater than in case of rework after a Type 2 failure.

The scenario when a Type 1 failure develops into a Type 2 failure is similar to the one when a Type 2 failure happens at once. We would expect, however, that the time to failure would be longer, thus the part would be closer to completion, i.e., $p14 > p3$. This implies that the chance of successful restart would be lower here, i.e. $p15 < p4$.

TABLE I
NOMENCLATURE

Symbol	Description
$C1$	initial part (workpiece) value
$C2$	machining cost per part
$C3$	expected restart cost, given that Type 2 tool failure was detected
$C4$	scrap value of a part
$C5$	quality inspection cost per part
$C6$	expected rework cost, given that the part has sustained a detected Type 2 tool failure, but has failed quality control
$C7$	average cost of handling customer complaints per defective part
$C8$	expected rework cost, given that the part has sustained an undetected Type 2 tool failure
$C9$	expected rework cost, given that the part has sustained an undetected Type 1 tool failure
$C10$	expected rework cost, given that the part has sustained a Type 1 followed by an undetected Type 2 tool failure
$p1$	probability that the tool failure is of Type 2
$p2$	probability of detecting Type 2 tool failure
$p3$	mean time to Type 2 tool failure, expressed as a fraction of the operation time
$p4$	probability of successful restart upon detecting a Type 2 tool failure
$p5$	probability of passing quality control, given that the part has sustained a detected Type 2 tool failure
$p6$	probability of successful rework, given that the part has sustained a detected Type 2 tool failure, but has failed quality control
$p7$	probability that a customer will reject a part, given that the part has sustained a tool failure, but has met the quality requirements
$p8$	probability of successful rework, given that the part has sustained an undetected Type 2 tool failure
$p9$	probability that a Type 1 failure will lead to a Type 2 tool failure
$p10$	probability of passing quality control, given that the part has sustained an undetected Type 1 tool failure
$p11$	probability that a customer will reject a defective part
$p12$	probability of successful rework, given that the part has sustained an undetected Type 1 tool failure
$p13$	probability of successful rework, given that the part has sustained a Type 1 followed by an undetected Type 2 tool failure
$p14$	mean time to Type 1 followed by Type 2 tool failure, expressed as a fraction of the operation time
$p15$	probability of successful restart, given that the part has sustained an undetected Type 1 followed by a detected Type 2 tool failure

IV. ESTIMATION OF EXPECTED PENALTY COST

We have discussed possible scenarios and costs resulting from Type 1 and Type 2 tool failures. When selecting the replacement time, though, we do not know which of the scenarios will occur. Therefore, we need to determine the expected cost of a tool failure.

The starting point here is to estimate the values of the inputs listed in Table I. Calculation of costs would usually not be a problem. $C1$, $C2$ and $C5$ can be extracted directly from the ERP. Rework and complaints handling expenses are often "accumulated" in separate cost centers, thus they can also be found. $C4$ can be obtained from a local scrap dealer, while $C3$ can be calculated based on the average downtime caused by a Type 2 tool failure. Finding various probabilities, on the other hand, is more difficult. Some data can be extracted from historical maintenance or quality control records. These records, however, are often collected for specific purposes and can therefore be difficult to use for other purposes. In such case one should start with the best estimates of experienced personnel. Later on a system for collecting the required data should be put in place and the initial estimates updated.

Once the input parameters are estimated the expected penalty cost can be determined by calculating the probability of each scenario, multiplying it by the corresponding cost, and summing the resulting products over all scenarios. For example, the probability that the failure will be of Type 2 and will be detected, but the restart will fail is $p1 \cdot p2 \cdot (1-p4)$. The corresponding penalty cost is $C1 + C2 \cdot p3 + C3 - C4$. If we repeat these calculations for all scenarios and sum the resulting products we will find the expected penalty cost. Mathematically this can be written as:

$$k = \sum_{i=1}^{n} \left(\prod_{j=1}^{m_i} p_{ij} \right) \cdot k_i \qquad (1)$$

where $i = 1, 2, ..., n$ is scenario index, $j = 1, 2, ..., m_i$ is a decision step in the i-th scenario, p_{ij} is the probability at the j-th decision step of the i-th scenario, and k_i is the penalty cost in case of i-th scenario.

Equation (1) can be plugged into the age replacement or other statistical models to find the optimal tool replacement time. Discussion about these models is out of the scope of this short article. Therefore the reader should consult general literature on replacement models.

V. DISCUSSION

The model presented in this paper is based on a simple manufacturing process. Real processes are usually more complex and may consist of several machining operations carried out in multiple roughing and finishing steps. Moreover, products often undergo other processing before being checked for quality. This increases unit production costs, thus scrapping a part becomes costlier.

The supply network could be more complex as well. A product might be sold to a wholesaler or a dealer and only then reach the consumer. The defect made by the tool failure could thus propagate all the way down the supply chain, and the cost could become huge.

Another issue is constrained resources. In traditional accounting rework immediately increases the production cost. Obviously, power consumed by the machine tool or its depreciation need to be taken into account. If, however, rework is performed by operators with excessive capacity, then their work does not cost anything extra, unless they need to work overtime. Situation becomes different when machines and operators are constrained resources. Then every rejected piece or time lost for rework reduces throughput, hence sales.

Yet another point to consider is tool room's capacity. Tool failures lead to increased demand for cutting tools, thus more load on the tool room. If their capacity was limited, then frequent tool failures could hinder the supply of tools to all processes and have a negative affect on the smooth operation and the output of the whole system.

On the other hand, the model can sometimes be simplified. For example, we have mentioned that the penalty for delivering a defective part to a consumer could be very high. Such costs, however, are very difficult to quantify. Moreover, the product would usually undergo several inspections before reaching the consumer. This means that the probability of incurring penalties would be decreasing rapidly when going down the supply chain, and the contribution to the expected penalty cost would be low. It is therefore reasonable to stop the calculations at a certain level, for example at the first-tie customers.

Other simplifications are possible too. For example, if the monitoring process is absolutely reliable, the probability of a catastrophic tool failure propagating further can be set to zero, thus disregarding all related consequences. It is also likely that after rework parts would undergo scrutinized quality control, so the probability of complains would become negligible.

VI. CONCLUSIONS

In this paper we have developed an approach for estimation of expected cutting tool failure cost. Within the model a process is defined by its customer and supplier. Both of them can be generalized and can mean a sub-process or even a single cut. The approach can therefore be used to study various machining processes on different levels of detail. What is needed is to split the process into manageable steps, estimate the probabilities and costs of a tool failure in these steps and then build a probability tree. Having built the probability tree calculation of expected tool failure cost becomes straight forward.

The expected tool failure cost is a necessary input to statistical tool replacement models. These models have been shown to be superior to deterministic techniques. Therefore, the development of this approach should facilitate the implementation of statistical replacement models and this way would contribute to reduction of cutting tool related costs.

The issue of course is the amount of the required input data, collecting which might be cumbersome. Nevertheless, we believe that the expected gains, especially in the industries where cutting tool costs are high, would outweigh the required efforts.

ACKNOWLEDGMENT

The authors thank The Research Council of Norway and Volvo Aero Norge AS for supporting this work through the AVROPROS project.

REFERENCES

[1] Z. Vagnorius, K. Sørby and M. Rausand, "Probabilistic model for determination of tool replacement time," in *Proc. 12th CIRP Conference on Modelling of Machining Operations*, Donostia-San Sebastian, Spain, pp. 757–764.

[2] A. K. Sheikh, L. A. Kendall and S. M. Pandit, "Probabilistic optimization of multitool machining operations," *Journal of Engineering for Industry, Transactions of the ASME*, vol. 102, no. 3, pp. 239–246, 1980.

[3] E. E. Iakovou and C. Koulamas, "Adaptive tool replacement policies in machining economics," *Journal of Manufacturing Science and Engineering*, vol. 118, no. 4, pp. 658–663, 1996.

[4] Y. V. Hui and L. C. Leung, "Optimal economic tool regrinding with Taguchi's quality loss function," *Engineering Economist*, vol. 39, no. 4, pp. 313–331, 1994.

[5] A. Jeang, "Reliable tool replacement policy for quality and cost," *European Journal of Operational Research*, vol. 108, no. 2, pp. 334–344, 1998.

[6] J. Taylor, "Carbide cutting tool variance and breakage: unknown factors in machining economics," in *Proc. 8th International MTDR Conference*, Manchester, UK, pp. 487–504.

[7] C. Koulamas, B. Lambert and M. Smith, "Optimal machining conditions and buffer space size for the two-stage case," *International Journal of Production Research*, vol. 25, no. 3, pp. 327–336, 1987.

[8] W. J. Zdeblick, R. E. De Vor and J. F. Kahles, "A comprehensive machining cost model and optimization technique," *CIRP Annals – Manufacturing Technology*, vol. 30, no. 1, pp. 405–408, 1981.

[9] U. La Commare, S. Noto La Diega and A. Passannanti, "Optimum tool replacement policies with penalty cost for unforeseen tool failure," *International Journal of Machine Tool Design and Research*, vol. 23, no. 4, pp. 237–243, 1983.

[10] E. Iakovou, C. M. Ip and C. Koulamas, "Optimal solutions for the machining economics problem with stochastically distributed tool lives," *European Journal of Operational Research*, vol. 92, no. 1, pp. 63–68, 1996.

[11] ISO 3685, "Tool-life testing with single-point turning tools," International Organization for Standardization, 1993.

[12] S. Kalpakjian and S. R. Schmid, *Manufacturing Engineering and Technology*. Singapore: Prentice Hall, 2006, pp. 634–635.

[13] H. M. Ertunc, K. A. Loparo and H. Ocak, "Tool wear condition monitoring in drilling operations using hidden Markov models (HMMs)," *International Journal of Machine Tools and Manufacture*, vol. 41, no. 9, pp. 1363–1384, 2001.

[14] S. E. Oraby and D. R. Hayhurst, "Tool life determination based on the measurement of wear and tool force ratio variation," *International Journal of Machine Tools and Manufacture*, vol. 44, no. 12–13, pp. 1261–1269, 2004.

[15] G. Dini and F. Tognazzi, "Tool condition monitoring in end milling using a torque-based sensorized toolholder," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, vol. 221, no. 1, pp. 11–23, 2007.

[16] W. W. Luggen, *Flexible Manufacturing Cells and Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1991, pp. 191–192.

Article 5

Zydrunas Vagnorius[*] and Knut Sørby[†]

Effect of high pressure cooling on life of SiAlON-based ceramic cutting tools

In: Proceedings of the 5th International Conference on Advances in Production Engineering (APE'2010), Warsaw University of Technology, Warsaw, Poland, 2010, pp. 352–362

The scientific committee of the conference has invited us to submit an extended version of this paper to the Journal of Engineering Manufacture (Part B of the Proceedings of the Institution of Mechanical Engineers)

---

[*] Zydrunas Vagnorius performed data analysis and wrote the entire text. He ran the machining experiments, during which the data for evaluating the effect of the high pressure cooling on the life of ceramic tools was collected, and performed tool wear measurements.

[†] Professor Knut Sørby supervised the machining experiments and contributed to the writing process by coming up with comments that helped to improve the readability and the overall quality of the text.

Is not included due to copyright

Article 6

Zydrunas Vagnorius[*] and Knut Sørby[†]

Effect of high pressure cooling on life of SiAlON tools in machining of Inconel 718

Accepted for publication in International Journal of Advanced Manufacturing Technology

---

[*] Zydrunas Vagnorius performed data analysis and wrote the entire text. He also ran the machining experiments, during which the data for evaluating the effect of the high pressure cooling on the life of ceramic tools was collected, and performed tool wear measurements.

[†] Professor Knut Sørby supervised the machining experiments. He also contributed to the writing process by coming up with comments that helped to improve the readability and the overall quality of the text.

Int J Adv Manuf Technol manuscript No.
(will be inserted by the editor)

Effect of high pressure cooling on life of SiAlON tools in machining of Inconel 718

Zydrunas Vagnorius · **Knut Sørby**

Received: date / Accepted: date

Abstract High pressure cooling has proven to be very effective when machining with carbide inserts. Longer tool life and improved chip breaking are among the most commonly mentioned advantages. Nevertheless this cooling method has been reported to reduce the life of ceramic tools in machining of heat resistant alloys. The main reason for that is said to be the accelerated notch wear. Therefore in this study SiAlON ceramic inserts with improved resistance to notching were tested in machining of Inconel 718 under high pressure cooling. The results were compared to conventional cooling. It turned out that while notch wear was still slightly increased when high pressure cooling was applied, it was no longer critical for the tool life. Flank wear, on the other hand, was reduced, which led to significantly longer tool life. The variation of the tool life appeared to be slightly less and chip breaking was considerably improved. This shows that when used properly high pressure cooling can help to increase the productivity in machining of heat-resistant alloys with ceramic tools.

Keywords High pressure cooling · SiAlON cutting tools · Inconel 718 · Tool life

1 Introduction

Machining of heat resistant aerospace materials, such as the nickel-based alloy Inconel 718, is characterised by low cutting speeds and therefore poor productivity. The main reasons for that are high hardness and low thermal conductivity of these materials. As a result a very high temperature is generated in the cutting zone. Narutaki et al. [1] and Kitagawa et al. [2] have experimentally shown that, in turning of Inconel 718 under conventional cooling, temperature on the rake face of ceramic inserts can reach 1300°C. At such temperature cutting tools soften significantly, thus they can be easier eroded by abrasion. In addition, heat promotes diffusion wear and can cause thermal shocks and fatigue. Therefore, to achieve a reasonable tool life heat resistant alloys are often machined at speeds as low as 30 to 100 m/min [2].

One way to raise the efficiency in machining of heat resistant alloys is to use more advanced cutting tool materials. A good example is ceramics. As shown by Vigneau et al. [3], when turning Inconel 718 with alumina, cermet and silicon nitride-based inserts, metal removal rate can be increased up to four times as compared to carbides. The reason for this is the exceptional hardness and abrasion resistance of these tool materials. Moreover, ceramics have a very high melting point, thus they remain stable and retain their supreme properties at elevated temperatures.

Besides the desirable properties ceramic tools also have some weaknesses. A particular concern is their sensitivity to thermal stresses. Owing to this drawback it is sometimes recommended to use no or very small quantities of coolant when working with ceramic tools [4]. However, due to extreme temperatures generated when machining heat resistant alloys, dry cutting can only be performed at relatively low speeds [5] and is therefore inefficient [6]. Thus, despite the improved heat resistance of ceramics tools, measures to reduce temperature in the cutting zone need to be taken.

Traditionally large quantities of fluids have been poured onto the tool to extract the heat. This technique has proven to be effective in machining of steels and other materials, but, as shown by Kitagawa et al. [2], provides insufficient cooling in cutting of heat resistant alloys. The issue is that in the range of temperatures developed in machining of these ma-

---

Zydrunas Vagnorius (✉) · Knut Sørby
Department of Production and Quality Engineering,
Norwegian University of Science and Technology,
S.P. Andersens v. 5, NO-7491 Trondheim, Norway
Tel.: +47-735-93919
Fax.: +47-735-97117
E-mail: zydrunas.vagnorius@ntnu.no

terials coolants are rapidly evaporated. As a result a steam "blanket" is created, which stops the coolant from reaching the tool-chip interface thus rendering conventional flushing ineffective. A few alternative techniques, including internal chilling of the insert, cryogenic-, CO_2- and high pressure cooling, have been tested. The latter method seems to be particularly promising.

In high pressure cooling cutting fluid is supplied in the form of a small jet. This "pushes" the coolant closer to the cutting edge, hence cooling becomes more effective. Already the early experiments performed by Pigott and Colwell [7] showed that, in rough turning of aircraft exhaust valves made in a nickel-based material, this method could increase the output per carbide tool by over 18 times. Later applications of high pressure cooling in various machining operations confirmed the effectiveness of this method.

From the above discussion it follows that high pressure cooling is an effective way to reduce the temperature in the cutting zone and therefore leads to improved tool life. Then, if the same advantage also applied to ceramic inserts, productivity in machining of heat resistant alloys could be increased significantly. As mentioned previously, these tool materials are generally sensitive to temperature variations. Nevertheless certain sorts of ceramics, such as alumina reinforced with SiC whiskers and SiAlON, have improved resistance to thermal shocks [8, 9]. It is therefore interesting to investigate whether these tools could be used under high pressure cooling. A few studies involving SiC-whiskers reinforced inserts have already been carried out and are reviewed in Section 3. In the present work the focus was placed on SiAlON tools, which were used in machining of a heat resistant alloy, Inconel 718, under high pressure cooling.

2 Effect of high pressure cooling on machining performance

The effect of high pressure cooling has been studied from various perspectives, including its influence on friction [10, 11], cutting forces [11–16], surface finish [12–18] and surface integrity [19]. The most significant advantages of using high pressure cooling, though, are considerably improved chip breaking and reduced temperature in the cutting zone, leading to longer tool life.

2.1 Effect on chip breaking

The most noticeable benefit of using high pressure cooling is very efficient chip breaking. In nearly all studies that will be reviewed next, high pressure cooling produced short chips. Under conventional cooling the same test conditions resulted in long continuous chips, which are undesirable, especially in automated machining. Mazurkiewicz et al. [11] suggest that this improvement is due to the hydro-wedge, which is created as the focussed coolant jet penetrates between tool's rake face and the chip. This wedge acts as a regular chip breaker, i.e., it lifts the chip up and reduces its curl radius. Eventually the chip is broken down and is flushed away by the powerful jet.

2.2 Effect on temperature and tool life

Applying cutting fluids at high pressure significantly improves the efficiency of the cooling process. As shown by Nagpal and Sharma [12], this way an up to 45% reduction in tool-chip interface temperature can be achieved. Similar observations have been made by Kaminski and Alvelid [20]. Due to these improvements high pressure cooling usually leads to significantly longer tool life.

Wertheim et al. [21] applied through-tool high pressure cooling in grooving of Inconel 718 and a few more materials. As result both crater and flank wear of carbide inserts was considerably reduced. Consequently tool life increased.

Ezugwu and Bonney [15, 18] applied high pressure cooling in rough and finish turning of Inconel 718 with coated carbide tools. They used a number of rejection criteria, such as the level of tool wear (flank, nose and notch), cutting edge failure and workpiece's surface roughness, and showed that under correct choice of machining and system parameters a considerably longer tool life can be achieved.

The effect of high pressure cooling on the performance of carbide tools has also been investigate in machining of other aerospace materials. Sørby et al. [22] used high pressure cooling in turning of Waspaloy. This reduced the flank wear and resulted in less edge chipping. Similar effects were observed in grooving of Ti–6Al–4V [23]. Application of high pressure cooling in turning of Ti–6Al–4V has been investigated by Machado et al. [24], Nandy and Paul [25], Nandy et al. [16] and Ezugwu et al. [26]. In all of these cases significant reduction in tool wear, hence improvement in tool life, was observed.

To explain the above-discussed improvements in tool life under high pressure cooling a few theories have been proposed. Kaminski and Alvelid [20] suggested that the key was the ability of the jet to break the steam barrier, which builds up when coolant gets evaporated. As a result fresh coolant can reach the tool and can carry away the heat.

Another critical factor is the ability of the pressurized fluid to penetrate deeper into the interface between tool's flank face and the workpiece (in flank face cooling) or between the rake face and the chip (in rake face cooling). In the former case the jet is not obstructed by the chip. Therefore coolant can be pushed closer to the cutting edge. In rake face cooling the chip is in the way of the jet. However, according to Mazurkiewicz et al. [11], in this case a hydro-wedge is created. As a result the chip is lifted up giving access for the coolant to the cutting edge.

Lifting up of the chip has another important effect. Measurements of the width of the worn area on the rake face of the tool show that the length of the contact between the tool and the chip is reduced in high pressure cooling [8, 16, 27]. Meanwhile Sadik and Lindström [28] have demonstrated that reducing the chip contact length leads to a decrease in tool temperature and consequently to lower flank wear. This suggests that the drop in temperature, followed by improved tool life, in high pressure cooling is at least partially due to the mechanisms provoked by the shorter chip contact length. On the other hand, Sadik and Lindström [28] observed that, when the chip contact length was reduced beyond a certain limit, both the temperature and the flank wear increased substantially. Sadik and Lindström explain that in this case forces act on a very small area, thus compressive stress is increased. Moreover, the reduction in the chip contact length means that the highest temperature region is "pushed" closer to the cutting edge. This causes elastic deformation of the tool. Consequently the area of contact between it and the workpiece is enlarged leading to increased flank wear.

Reduction in the chip contact length, followed by the concentration of stresses and shift of the highest temperature zone closer to the cutting edge, is a possible explanation for those cases, where high pressure cooling led to shorter tool life. Machado et al. [24] and Ezugwu et al. [8], for example, observed reduced life of uncoated carbide tools in turning of Inconel 901 (though some improvement was achieved at the highest cutting speed). Results presented by Sharman et al. [19] show that, in machining of Inconel 718 under high pressure cooling, tool life was worse than, or at best equivalent to, that obtained under conventional flushing. Under some test conditions shorter tool life in turning of Inconel 718 was also observed by Ezugwu and Bonney [15, 18].

3 Effect of high pressure cooling on performance of ceramic tools

As discussed in the previous section, high pressure cooling leads to more efficient chip breaking and usually extends the tool life. Despite these improvements the number of reported studies on the application of this technique when machining with ceramic tools is scarce, and, as shown by the examples below, the results can be mixed.

Ezugwu et al. [8] experimented with high pressure cooling in turning of Inconel 901 with SiC-whiskers reinforced ceramic inserts. They observed that cooling at a pressure of 14 MPa enhanced chip breaking. However, it generally led to reduced tool life as compared to conventional flushing. The reason for this was said to be the accelerated notch wear.

Analogous results were achieved by Öjmertz and Oskarson [31], who tested rough turning of Inconel 718 with SiC-whiskers reinforced ceramic tools. They observed that high pressure cooling led to better chip control, reduced tendency to built-up edge formation and therefore better surface quality as compared to dry machining. However, a clear tendency towards increasing depth-of-cut notch wear was observed as the pressure was raised from 80 to 360 MPa.

A similar work was done by Ezugwu and Bonney [29], who applied coolant at a pressure of 11–20 MPa in rough turning of Inconel 718 with SiC-whiskers reinforced ceramic tools. Despite the lower pressure they also observed that jet cooling caused severe notching and therefore led to shorter tool life as compared to conventional flushing.

Ezugwu et al. [30] achieved slightly more promising results under finishing conditions. In general the tool life improved at coolant pressures of 11 and 15 MPa. At 20 MPa, though, it dropped significantly due to the accelerated notch wear. Tool life was also shorter at 11 MPa when the speed was increased to 300 m/min.

According to Ezugwu and Bonney [29], the reduction in the life of ceramic tools that has been observed when high pressure cooling was applied could be caused by hydrodynamic erosion. They suggest that when the jet hits the tool it comes to a sudden rest and builds a stagnation pressure. To release it coolant tries to escape through the depth-of-cut region. This way small abrasive particles caught in the fluid are flushed away at a high velocity, which causes severe wear in this region. Öjmertz and Oskarson [31] on the other hand, suggest that cooling at a high pressure could reduce the temperature of the workpiece below a certain threshold. This would increase its strength and would result in a higher tool contact pressure. Consequently the wear would intensify leading to shorter tool life.

4 Hypotheses

The review of the previous work on high pressure cooling shows that this technique is very effective when machining with carbide tools, but in general has a negative influence on the performance of ceramic inserts. It should be emphasized, however, that in the studies reported so far only ceramics based on alumina reinforced with SiC-whiskers have been used under high pressure cooling, while tools made in SiAlON have not been tested. Therefore performance of SiAlON inserts in machining of Inconel 718 under high pressure cooling will be investigated in this study.

Since tool life is one of the main considerations in machining economics, the main goal of this study is to check whether the application of high pressure cooling could prolong the life of SiAlON tools. Thus, the central hypothesis to be tested is:

$$(1) \quad H_{0_1}\colon \mu_{hpc} = \mu_{conv} \quad \text{versus} \quad H_{1_1}\colon \mu_{hpc} > \mu_{conv}$$

where μ is the mean tool life and indexes hpc and $conv$ stand for high pressure- and conventional cooling respectively.

In addition it is important to note that it is now commonly accepted that tool life is a stochastic rather than a deterministic quantity (e.g., see [32]). As a consequence the actual tool life rarely matches the predicted values. This leads to conservative replacement strategies. According to Wiklund [33], only 50–80% of the expected life is typically used. As a result tool consumption and related replacement costs are higher than necessary, as are the losses in terms of the productive time. For these reasons it is important to investigate not only the mean, but also the variance of the tool life. Demonstrating that it could be reduced by applying high pressure cooling, would mean that this way the service length of cutting tools would become more predictable, which can be expected to have a substantial economical effect. Such outcome would be reasonable considering the fact that a focussed high pressure jet is more stable than a low pressure stream and should therefore result in a more stable cooling process. Improved chip breaking should also add stability to the cooling process as it would not be obstructed by long chips. Hence the second hypotheses to be tested in this study is:

$$(2) \quad H_{0_2}\colon \sigma^2_{hpc} = \sigma^2_{conv} \quad \text{versus} \quad H_{1_2}\colon \sigma^2_{hpc} < \sigma^2_{conv}$$

where σ^2 is the variance of the tool life, while indexes hpc and $conv$ stand for high pressure- and conventional cooling respectively.

5 Experimental work

In order to test the above hypotheses machining experiments were performed. For this purpose 20 SiAlON ceramic inserts and an Inconel 718 workpiece were prepared. As mentioned in Section 1, very high temperatures have been recorded when machining this material. Therefore it was decided that, in order to maximize the extraction of heat, high pressure cooling should be used in combination with conventional flushing. This technique will be referred to as high pressure-assisted cooling in the following text. The physical configuration of the experimental cooling system and the test conditions are described next.

5.1 Experimental set-up

Cutting experiments were carried out with a Hessapp DV80 lathe (see Fig. 1). The machine is equipped with an auxiliary pump, which delivers pressurised (up to 40 MPa) cutting fluid to the outlet on the tool turret. From this point coolant is transported via a copper tube to a custom-made insert clamp with internal channels (see Fig. 2). Such system is very rigid, thus the direction and the target point of the jet do not change as a result of the reactive forces. This was expected to add stability to the cooling process, hence to minimize thermal variation, which ceramic tools are known to be sensitive to.

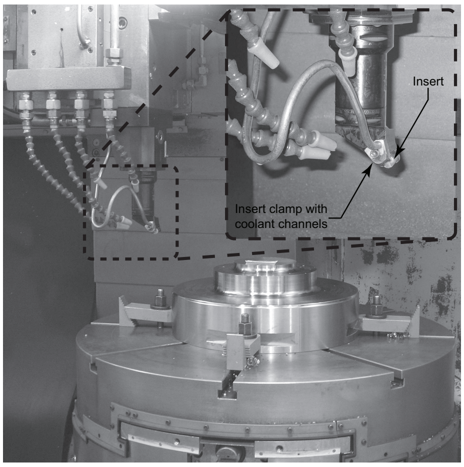


Fig. 1: Experimental set-up

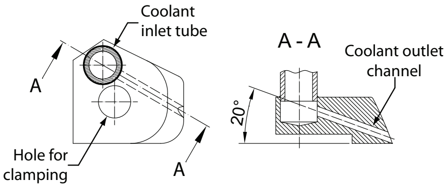


Fig. 2: Insert clamp with coolant channels

Since ceramic tools are also known to be brittle, special measures were taken to minimize the occurrence of mechanical shocks. Workpiece was securely clamped on the pallet, and a few millimetres of material were removed from its sides to compensate for the centring error. Moreover, the tool was programmed to follow an arc-shaped trajectory at the start and the end of each cut. This way a smooth entrance to and exit from the workpiece was achieved.

Cutting conditions used in these experiments are typical for semi-rough machining of Inconel 718 with ceramic tools. They were kept constant and were the same during both high pressure-assisted cooling and the control tests, where only conventional cooling was applied. This was done in order to assure that the only source of possible difference in the observed results was the cooling method. For further experimental details please see Table 1.

Table 1: Experimental conditions

Workpiece	Material	Inconel 718
	Hardness	33 HRC
	Shape	Solid cylinder
	Diameter	400 mm
	Height	150 mm
Tool	Material	SiAlON
	Insert type	RCGX 120700 E
	Holder type	PCLN
Cutting data	Operation	Facing
	Speed, v_c	300 m/min
	Feed, f_n	0.2 mm/rev
	Depth of cut, a_p	1.0 mm
Cooling	Coolant	Emulsion, 4%
	Control method	Flooding at 0.7 MPa
	Test method	Flooding + High pressure cooling at 20 MPa

5.2 Procedure

Test inserts were labelled with numbers from 1 to 20. Half of them were selected for high pressure-assisted cooling tests. The rest were controls and were used for machining under conventional cooling alone.

The experiment was carried out in multiple tool paths divided into several cuts. To make sure that the conditions in each of the test cuts were approximately the same, each path of the tool was executed in three stages.

Stage I. The previous experience has showed that a burr tends to form on the outer edge of the cylindrical workpiece. As a result the tool sustains a shock at the beginning of the next path. To avoid this, a dummy tool was used to remove the burr.

Stage II. At this stage the test tools were used. Due to the size of the workpiece each path was split into three parts resulting into three cuts of equal duration. In order to avoid any possible bias due to inhomogeneity of workpiece material or scale that could have formed on its surface during the previous tool path, experiments were *randomized*. To accomplish that, a number from 1 to 20 was drawn to select the tool. After completing the cut, which on average took 35 seconds, the insert was removed from the tool holder and the wear was measured with a Mitutoyo toolmaker's microscope. Then a new random number was drawn to determine the tool to be used in the next cut.

Stage III. After completing the third cut a dummy insert was used again to remove a few millimetres of material. The reason for this last operation was to avoid the contact between the test tool and the core that formed in the centre of the workpiece.

Having finished one path, the tool was lowered by the amount of the depth of cut, i.e., by 1 mm, and the three-stage procedure was repeated again. This work was continued until each insert had performed four cuts. By that time the wear on all test tools had reached the limit to be defined in the following section.

5.3 Tool life criterion

According to ISO 3685 [34], the most common life measures for tools of ceramics are the average and the maximum width of the flank wear land. Depth-of-cut notch wear, which is often mentioned to be an issue when machining nickel-based alloys with ceramic tools, is said to depend on the accuracy of repeated depth settings and must therefore be excluded from the flank wear measurements. Another common problem with ceramic tools is edge chipping. According to ISO 3685 [34], to a certain extent this type of wear is taken into account by the maximum width of the flank wear land, which for the latter is the recommended measure when edge chipping is expected. Thus the maximum width of the flank wear land *VB* max of 0.6 mm was chosen as the tool life criterion in this study.

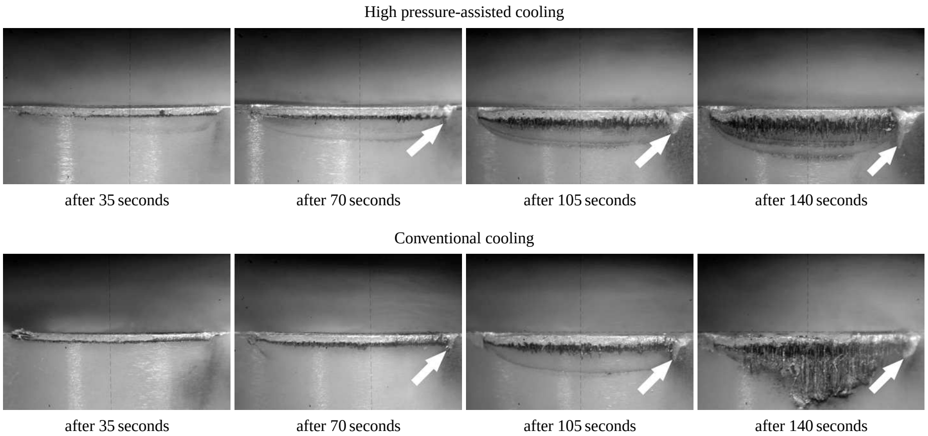
6 Results

This section presents the results of the experiments. It starts with a discussion about the observed wear of SiAlON inserts under conventional and high pressure-assisted cooling. Coming out from this, tool lives are derived and the hypothesis postulated in Section 4 are tested. In addition the observed effect of high pressure cooling on chip breaking is briefly discussed.

6.1 Tool wear

As mentioned in Section 3, notch wear at the depth-of-cut region is usually the most serious issue when machining nickel-based alloys with ceramic tools under high pressure cooling. In our experiments notch wear only became significant at later stages of the cutting process (see Fig. 3). In general it was more intense under high pressure than under conventional cooling. In the latter case the maximum length of the depth-of-cut mark was 0.59 mm and in case of high pressure cooling it was 0.82 mm. Such level of notch wear was considered to be within reasonable limits, hence not critical for the tool life.

Flank wear was clearly visible from the first cuts (see Fig. 3) and was increasing steadily as machining continued. As illustrated in Fig. 4, its rate was slightly higher under


High pressure-assisted cooling

after 35 seconds | after 70 seconds | after 105 seconds | after 140 seconds

Conventional cooling

after 35 seconds | after 70 seconds | after 105 seconds | after 140 seconds

Fig. 3: Tool wear under high pressure-assisted and conventional cooling

conventional cooling, and the gap was growing as the cutting progressed further. Moreover, in case of conventional cooling there was more work material being welded to the tool. This attached layer was carefully removed in order to expose the flank face of the tool and to measure the true amount of wear.

6.2 Tool life

In Section 5.3 the critical width of the maximum flank wear land was set at 0.6 mm. The time when this limit was reached was determined by interpolating between the two nearest experimental points. The results together with the basic sample statistics are shown in Table 2 (note that the tool lives seen here are typical for semi-rough machining of Inconel 718 with ceramic inserts). This data will now be used to test the hypotheses postulated in Section 4. Since most of the commonly used statistical methods are based on the assumption of equal variances we will test H_{0_2} first, and will then proceed with H_{0_1}.

6.2.1 Comparison of variances

The data shown in Table 2 suggest that the variance was lower when high pressure cooling was applied. However, this difference could be due to random rather than systematic causes. To check this let us assume that the two data samples came from two normal distributions. Then we can verify the validity of hypothesis H_{0_2} by applying the F-test.

Given our experimental data, the statistic of the test is $f_0 = s^2_{hpc}/s^2_{conv} = 0.6449$. We would reject H_{0_2} if

$$f_0 < f_{1-\alpha,n_{hpc}-1,n_{conv}-1} \tag{1}$$

where n is the sample size and α is the significance level. In our case $n_{hpc} = n_{conv} = 10$, and we chose α to be 0.05. Thus we would reject H_{0_2} if $f_0 < f_{0.95,9,9}$. By making use of the identity $f_{1-\alpha,u,v} = 1/f_{\alpha,u,v}$ and by looking up the tables for the F-distribution we find that $f_{0.95,9,9}$ is approximately equal to 0.3145. Since this number is less than our test statistic we cannot reject H_{0_2}, i.e., we do not have enough evidence to claim that the variances under the two types of cooling are different. In fact the p-value in this case is 0.524. Thus it is very likely that the cause of the observed differences was random variation.

Despite the fact that we cannot reject H_{0_2} it should be mentioned that, based on the collected data, standard deviation of the tool life was approximately 1.25 times larger in case of the conventional cooling. However, this estimate is based on samples of size 10, which are small from a statistical point of view. Given that, we can read from the operating curves for the F-distribution (found in statistical handbooks) that even if the difference revealed by the experimental data represented the true difference between the standard deviations, which would imply that the alternative hypothesis H_{1_2} was true, the probability of accepting H_{0_2} would still be more than 80%.

In the above calculations we have assumed that the data sets were normally distributed. Due to its simplicity and symmetrical shape normal distribution is sometimes used to model the life of cutting tools (e.g., see [32, 35–37]). Nevertheless, there are a couple issues with this assumption. First, the normal distribution allows for negative values and is therefore not a realistic life model. Second, careful examination of the data shows that even if it was a viable model, the normal distribution does not describe the tool lives ob-

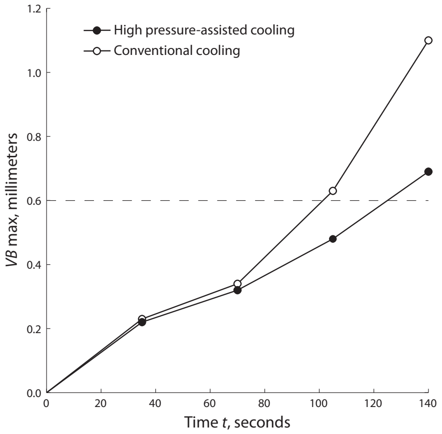


Fig. 4: Development of flank wear land (each point represents the average of 10 observations)

served under the conventional cooling well. Therefore, to check whether the conclusion drawn above is correct, we have dropped the normality assumption and applied the Levene's test.

The calculated test statistic is $W = 0.0216$ (the calculations are cumbersome and therefore not shown here). We would reject H_{0_2} if

$$W > f_{\alpha,k-1,N-k} \tag{2}$$

where N is the total number of data points and k is the number of subgroups. In our case $N = 20$ and $k = 2$. Then if we kept α at 0.05 we would reject H_{0_2} if $W > f_{0.05,1,18}$. Looking up the tables for the F-distribution we find that $f_{0.05,1,18}$ is approximately equal to 4.4139. Since this value is much grater than our test statistic we cannot reject H_{0_2}, i.e., we do not have enough evidence that the variances under the two types of cooling are different. The p-value for this test is 0.885 and is even greater than under the normality assumption. Thus it is very likely that the differences were due to the random variation.

6.2.2 Comparison of means

In the previous section we concluded that the normal distribution was not an appropriate model for our experimental data. Thus we cannot use the t-test to compare the means, i.e., to test the hypothesis H_{0_1}. An alternative solution is to apply the non-parametric Wilcoxon (Mann-Whitney) test for the medians $\widetilde{T}_{hpc}$ and $\widetilde{T}_{conv}$.

The test statistic W_{hpc} is the sum of ranks for the tool lives observed under high pressure-assisted cooling and is equal to 154. The probability of obtaining such a high number, given that $\widetilde{T}_{hpc} = \widetilde{T}_{conv}$, is 0.0001. In other words it is very unlikely that experimentally collected data would show such a difference if the true medians were equal. Indeed, the 95% confidence interval for $\widetilde{T}_{hpc} - \widetilde{T}_{conv}$ is (16.84, 31.57). Based on these calculations we can reject H_{0_1} in favor of H_{1_1}, i.e., we can conclude that the life of SiAlON ceramic tools seems to be longer when high pressure cooling is applied.

Table 2: Observed tool lives (in seconds)

	High pressure-assisted cooling	Conventional cooling
Tool life, T	136	115
	134	110
	126	108
	140	112
	113	101
	131	109
	133	90
	128	83
	117	108
	129	107
Sample mean, $\overline{T}$	129	104
Median, $\widetilde{T}$	129.9	108.2
Sample variance, s^2	69	107

6.3 Chip breaking

Besides affecting the tool wear, high pressure cooling had a considerable influence on chip flow. As can be seen in Fig. 5, the discoloured area, i.e., the region where the chip was in contact with the tool, covers nearly half of the rake face of the insert used under conventional cooling, while on the tool used under high pressure-assisted cooling there is almost no discolouration. This shows that in the latter case the chip was curling up much earlier. As a consequence it was broken into shorter segments. This is illustrated in Fig. 6. As can be seen here, chips produced under high pressure cooling were very short, needle-like, while under conventional flushing they were long tubular.

7 Discussion

In this study machining of a heat-resistant aerospace material, Inconel 718, with SiAlON ceramic inserts under high pressure cooling was tested. The presented overview of earlier research shows that the latter technique usually leads to longer tool life and significantly improves chip breaking. When applied in machining of nickel-based alloys with ceramic tools, however, high pressure cooling has been reported to accelerate notch wear and therefore to lead to re-


Conventional cooling High pressure-assisted cooling

Fig. 5: Effect of cooling method on chip contact length

duced tool life. To overcome this problem, in this study ceramic tools with improved resistance to notching were used, and special measures were taken to minimize the occurrence of thermal and mechanical shocks.

The above set-up proved to be effective. Even though notch wear was still more intense under high pressure-assisted cooling, it remained within reasonable limits and was not critical for the tool life. Flank wear, on the other hand, was reduced as a result of high pressure cooling. Therefore tool life was significantly longer.

The reduction in flank wear and consequently the improvement in tool life is the result of more efficient cooling. As indicated by the observed reduction in chip contact length, when applied at a high pressure, coolant overcomes the resistance of the chip, lifts it up and penetrates closer to the cutting edge, where the highest temperature occurs. Moreover, the speed of the coolant flow is much higher under high pressure cooling, which for the dissipation of heat is more rapid. Combination of these two factors leads to more efficient cooling of the cutting edge. As a consequence, the intensity of wear processes is reduced. On the other hand, rapid cooling can lead to thermal cracking, followed by micro chipping. This effect can be expected to be pronounced in the depth-of-cut region, where coolant has a direct contact with the heated cutting zone. For this reason the rate of cooling, hence the likelihood of thermal cracking and micro chipping, should be particular high here, which would explain the observed increase (though not very significant) in notch wear.

The results of this study also suggest that the variance of the tool life might be reduced by applying high pressure cooling. This would be reasonable considering the fact that in such case cooling process is not obstructed by long chips and is probably more stable. However, we did not have enough statistical evidence to support this claim, despite the fact that the sample sizes that we used in our experiments are rather big for this type of studies.

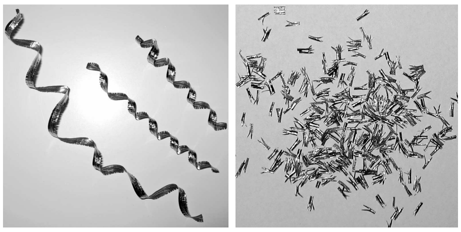

Conventional cooling High pressure-assisted cooling

Fig. 6: Effect of cooling method on chip breaking

8 Conclusions

The results achieved in this study show that, when the machining process is properly designed, high pressure-assisted cooling can help to extend the life of SiAlON ceramic tools, hence reduce the costs in machining of heat resistant alloys. Alternatively the cutting speed could be increased, which would make the process more productive. Moreover, high pressure-assisted cooling significantly improved chip breaking. Since long chips can scratch the workpiece and can block the disposal equipment, this advantage is important, especially in unattended machining.

Acknowledgements We would like to express our gratitude to Volvo Aero Norge AS and The Research Council of Norway for supporting this work through the AVROPROS project. The authors would also like to thank Professor Marvin Rausand at the Norwegian University of Science and Technology for helpful comments.

References

1. Narutaki N, Yamane Y, Hayashi K, Kitagawa T, Uehara K (1993) High-speed machining of Inconel 718 with ceramic tools. CIRP Annals – Manufacturing Technology 42(1):103–106
2. Kitagawa T, Kubo A, Maekawa K (1997) Temperature and wear of cutting tools in high-speed machining of Inconel 718 and Ti–6Al–6V–2Sn. Wear 202(2):142–148
3. Vigneau J, Boulanger J, Le Maitre F (1982) Behaviour of ceramic tools during the machining of nickel base alloys. CIRP Annals – Manufacturing Technology 31(1):35–39
4. Boothroyd G, Knight WA (2006) Fundamentals of Machining and Machine Tools. Taylor & Francis, Boca Raton
5. Li L, He N, Wang M, Wang ZG (2002) High speed cutting of Inconel 718 with coated carbide and ceramic inserts. Journal of Materials Processing Technology 129(1–3):127–130
6. Klocke F, Eisenblätter G (1997) Dry cutting. CIRP Annals – Manufacturing Technology 46(2):519–526
7. Pigott R, Colwell A (1952) Hi-jet system for increasing tool life. Society of Automotive Engineers – Transactions 6(3):547–564
8. Ezugwu EO, Machado AR, Pashby IR, Wallbank J (1991) The effect of high-pressure coolant supply when machining a heat-resistant nickel-based superalloy. Lubrication Engineering 47(9):751–757
9. Kalpakjian S, Schmid SR (2006) Manufacturing Engineering and Technology. Prentice Hall, Singapore

10. Sharma CS, Rice WB, Salmon R (1971) Some effects of injecting cutting fluids directly into the chip-tool interface. Journal of Engineering for Industry, Transactions of the ASME 93:441–444
11. Mazurkiewicz M, Kubala Z, Chow J (1989) Metal machining with high-pressure water-jet cooling assistance – a new possibility. Journal of Engineering for Industry, Transactions of the ASME 111:7–12
12. Nagpal BK, Sharma CS (1973) Cutting fluids performance – Part 1 – Optimization of pressure for "Hi-jet" method of cutting fluid application. Journal of Engineering for Industry, Transactions of the ASME 95:881–889
13. Kovacevic R, Cherukuthota C, Mohan R (1995) Improving milling performance with high pressure waterjet assisted cooling/lubrication. Journal of Engineering for Industry 117(3):331–339
14. Kovacevic R, Cherukuthota C, Mazurkiewicz M (1995) High pressure waterjet cooling/lubrication to improve machining efficiency in milling. International Journal of Machine Tools and Manufacture 35(10):1459–1473
15. Ezugwu EO, Bonney J (2005) Finish machining of nickel-base Inconel 718 alloy with coated carbide tool under conventional and high-pressure coolant supplies. Tribology Transactions 48(1):76–81
16. Nandy A, Gowrishankar M, Paul S (2009) Some studies on high-pressure cooling in turning of Ti–6Al–4V. International Journal of Machine Tools and Manufacture 49(2):182–198
17. Senthil Kumar A, Rahman M, Ng SL (2002) Effect of high-pressure coolant on machining performance. International Journal of Advanced Manufacturing Technology 20(2):83–91
18. Ezugwu EO, Bonney J (2004) Effect of high-pressure coolant supply when machining nickel-base, Inconel 718, alloy with coated carbide tools. Journal of Materials Processing Technology 153–154:1045–1050
19. Sharman A, Hughes J, Ridgway K (2008) Surface integrity and tool life when turning Inconel 718 using ultra-high pressure and flood coolant systems. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture 222(6):653–664
20. Kaminski J, Alvelid B (2000) Temperature reduction in the cutting zone in water-jet assisted turning. Journal of Materials Processing Technology 106(1–3):68–73
21. Wertheim R, Rotberg J, Ber A (1992) Influence of high-pressure flushing through the rake face of the cutting tool. CIRP Annals – Manufacturing Technology 41(1):101–106.
22. Sørby K, Bjørnstad MS, Tønnessen K (2006) The effect of high-pressure cutting fluid on chip breaking and tool life in turning of nickel-base alloys. In: CIRP 2nd International Conference on High Performance Cutting
23. Sørby K, Tønnessen K (2006) High-pressure cooling of face-grooving operations in Ti6Al4V. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture 220(B10):1621–1627
24. Machado AR, Wallbank J, Pashby I, Ezugwu EO (1998) Tool performance and chip control when machining Ti6Al4V and Inconel 901 using high pressure coolant supply. Machining Science and Technology 2(1):1–12
25. Nandy A, Paul S (2008) Effect of coolant pressure, nozzle diameter, impingement angle and spot distance in high pressure cooling with neat oil in turning Ti–6Al–4V. Machining Science and Technology 12(4):445–473
26. Ezugwu EO, Bonney J, Da Silva RB, Machado AR, Ugwoha E (2009) High productivity rough turning of Ti–6Al–4V alloy, with flood and high-pressure cooling. Tribology Transactions 52(3):395–400
27. Kramar D, Krajnik P, Kopac J (2010) Capability of high pressure cooling in the turning of surface hardened piston rods. Journal of Materials Processing Technology 210(2):212–218
28. Sadik MI, Lindström B (1995) The effect of restricted contact length on tool performance. Journal of Materials Processing Technology 48(1–4):275–282
29. Ezugwu EO, Bonney J (2003) Effect of high-pressure coolant supplies when machining nickel-base, Inconel 718, alloy with ceramic tools. Tribology Transactions 46(4):580–584
30. Ezugwu EO, Bonney J, Fadare D, Sales W (2005) Machining of nickel-base, Inconel 718, alloy with ceramic tools under finishing conditions with various coolant supply pressures. Journal of Materials Processing Technology 162–163:609–614
31. Öjmertz K, Oskarson HB (1999) Wear on SiC-whiskers reinforced ceramic inserts when cutting Inconel with waterjet assistance. Tribology Transactions 42(3):471–478
32. Wager JG, Barash MM (1971) Study of the distribution of the life of HSS tools. Journal of Engineering for Industry, Transactions of the ASME 93:1044–1050
33. Wiklund H (1998) Bayesian and regression approaches to on-line prediction of residual tool life. Quality and Reliability Engineering International 14(5):303–309
34. ISO 3685 (1993) Tool-life testing with single-point turning tools. International Organization for Standardization, Geneva
35. Zhou C, Wysk RA (1992) Integrated system for selecting optimum cutting speeds and tool replacement times. International Journal of Machine Tools and Manufacture 32(5):695–707
36. Kwon WT, Park JS, Kang S (2005) Effect of group IV elements on the cutting characteristics of Ti(C,N) cermet tools and reliability analysis. Journal of Materials Processing Technology 166(1):9–14
37. Dašić P, Natsis A, Petropoulos G (2008) Models of reliability for cutting tools: Examples in manufacturing and agricultural engineering. Strojniški Vestnik – Journal of Mechanical Engineering 54(2):122–130

Part III

Other articles and reports

Article 0 (unpublished)

Zydrunas Vagnorius

Reliability of flexible manufacturing systems

This literature review article served as a reference for defining the main issues to be addressed in the PhD project

RELIABILITY OF FLEXIBLE MANUFACTURING SYSTEMS

Effect on System Performance, Failure Modes and Countermeasures

Zydrunas Vagnorius

Keywords: flexible manufacturing systems, reliability, availability, performance, failure.

1 INTRODUCTION

Traditional mass production lines have evolved at the beginning of the twentieth century and were based on big batches and high volumes. Since then markets have changed significantly, so today Henry Ford would hardly be able to sell his Model T car in any colour as long as it was black. Nowadays, demand is shifting towards more and more customized, higher quality products, faster deliveries, while harsh competition in global markets means that individual products are sold in smaller quantities and lower prices. In such environment *flexible manufacturing systems* (FMS) and their subset, called *flexible manufacturing cells* (FMC), are gaining popularity [1, 2].

As the name suggests the main advantage these systems offer is *flexibility*. Flexibility is an umbrella term for ability to respond to changing business conditions. For this reason understanding of what it is depends on the business sector that a company is engaged in. In case of manufacturing enterprises flexibility usually means the ease of switch between different product types, called product mix flexibility, ability to cope with fluctuations in production quantities, called volume flexibility, ability to operate economically in small batch environment, deal with tight schedules, and potential to produce customized goods.

Another important feature of FMSs is *interlinking* of resources through centralized computer control and automated transportation system, which enables coordination of such activities as scheduling, material handling and tool sharing. The benefits of integration are synergetic and include reduction of lead time, predictability of operation and consistency of results [3]. Interlinking also allows using equipment more efficiently. Vineyard and Meredith [4] state that compared to regular machines, which run around 20% of time, FMSs can reach up to 70 – 80% utilization rates. Luggen [1] claims that asset usage in FMS can even be as high as 80 – 90%. These statements are backed by case studies carried out by Hackstein and Budenbender [2], which show average actual utilization of machines in two different companies of 84.8% and 84.9% respectively, and Wiendahl and Springer [5], who report that actual utilization of different resources in two investigated systems varied between 64% and 92%.

Other benefits from implementing FMS include 60 – 80% inventory reduction, 30 – 50% percent direct labour savings and 40 – 50% floor space reduction [1]. Despite of these improvements users are often not satisfied after establishment of FMS [2, 5]. One of the main reasons for the disappointment is said to be unrealistic user expectations [1]. FMS cannot create order out of chaos nor can it solve problems of mismanaged or poorly organised plant. Lack of strategic necessity is another issue. Many FMSs fail because technology is sold for the sake of technology rather than to solve strategic tasks [1]. Moreover, successful implementations require management commitment and close cooperation with system supplier throughout the lifetime of FMS.

While these are quite political reasons having more to do with company's management than FMS itself, there is another and more serious concern. In general we know that complex systems tend to be less reliable than their simpler counterparts and FMS is of course a complex system. Moreover, due to extremely high equipment utilization FMSs will experience bigger amount of wear and tear than traditional machine tools [4]. Eventually, there seems to be a trade-off between the main characteristic features of FMS and *reliability*. These last issues will be our main focus throughout this paper, but before we elaborate on them, let us first discuss two things: the concept of reliability and its place in production systems.

2 THE CONCEPT OF RELIABILITY

Rausand and Høyland [6] say that reliability, as human attribute, has been praised for a very long time, but for technical purposes this concept has not been applied before the end of World War I. Indeed, we can think about reliable ally, reliable spouse, reliable company or reliable information, which refers to the degree we can trust these non-technical "entities". In technical terms, however, *reliability is the ability of an item to perform a required function, under given environmental and operational conditions and for a stated period of time* [6].

Let us consider a car example. Those interested in Formula 1 racing remember last year's Japanese Grand Prix, when engine of Michael Schumacher's bolide failed costing him the eighth world champion title. This example first of all demonstrates how crucial is the ability of a system to perform a required function or in other words how frustrating its failure might be. Another point that this example emphasizes is the role of operational conditions.

Despite their technological superiority Formula 1 bolides break down much more often than regular cars. The reason for this is extreme operating conditions. Racing cars are pushed to the limits of their capabilities, which causes huge over-stresses and accelerated wear. For example, tyres have to be replaced at least once during the race, because, due to aggressive driving and extreme operating temperature, they wear out rapidly. On the other hand an old Lithuanian saying has it that a ploughshare loses the shine (starts to rust) if not used. Similarly, a drop of tyre temperature below a specified level reduces grip and causes graining of rubber, which eventually leads to vibration of the wheel and often damages the suspension. Thus, operating conditions affect reliability of a system. The same can be said about environment. In both cases neither of the extremes is good, so the best is if operating and environmental conditions stay within a certain optimal band.

The concluding part of reliability definition says that an item should perform its function for a stated period of time. In most cases a part or a product is desired to serve as long as possible. Nevertheless it is evident that no item can live forever. This is not possible technically nor might even be desirable, because in some cases extended system lifetime is a drawback rather than an advantage. While this last claim might sound contradictory, below examples show that in certain situations it is true.

One of earlier world champions has once said that an engine of a Formula 1 car should fail immediately after crossing the finish line. If it does not so, than it means that an engine was built too reliable at the expense of extra weight and reduced speed. Therefore, if the engine of Michael Schumacher's bolide would have failed only 17 laps later, nobody would have even noticed that.

Another example comes from the history of wars, which has contributed significantly to the development of reliability. It is said that bearings of German tanks in World War II were so reliable that could have served for years. However, the lifetime of the whole tank in the battlefield was only a few minutes. Under such conditions long lifetime of components did not pay off, but quantity and speed of production did. For this reason Russia, whose tanks were very crude in finish, but therefore could be produced faster and in bigger numbers, had a crucial advantage.

These last examples do not disparage the importance of reliability – without any doubt it is very important. What these examples do say, though, is that reliability, like any other feature of an item, must be coupled to the requirements of a particular system and specific condition of its use. In other words we should think about reliability in terms of the effect on the expected performance of the system as a whole.

3 PLACE OF RELIABILITY IN PRODUCTION SYSTEMS

FMS is a type of production systems. Individual installations may differ from industry to industry, but usually consist of value adding equipment, such as machine tools, automatic painting, welding and assembly stations, and supportive resources, such as coordinate measuring and washing machines. The resources are linked together by material and tool handling systems and jointly controlled by centralized computer system. Failure of any of these subsystems is directly associated with repair costs, including the replacement of expensive units, scrapping of broken tools and debugging of programs, to mention but few. Moreover, in industries, such as aerospace, where raw material costs are substantial, damaged parts due to equipment failure, also count. However, focusing purely on direct failure cost has one clear disadvantage.

Thomas Corbett, inspired by Eliyahu Goldratt, the father of Theory of Constrains (TOC), in his book Throughput Accounting [7] argues that there is a clear bound for any cost reduction efforts – cost can not be less than zero. Therefore an alternative strategy to improve company’s profitability is to maximize output, which is only limited by the ability of the market to consume goods. From this it follows that cost reduction attitude is the right strategy for mature products, the markets of which are saturated or declining, while for any industry that experiences growth maximization of throughput must be of supreme importance.

Back to reliability, failure of a resource does not only mean expensive repairs, it also means that while being down, equipment is not available for further production. This disrupts regularity of the flow, might cause “starvation” of other resources, and eventually leads to reduced throughput and associated income. What is more important, equipment failures might result in missed deadlines and damaged reputation of a reliable supplier, thus reduced future earnings. So, besides direct costs, another important effect of reliability is on availability of resources.

Rausand and Høyland [6] define average *availability* as:

$$A_{av} = \frac{MTTF}{MTTF + MTTR}$$

Here, *MTTF* stands for mean time to failure, denoting the mean functioning time, and *MTTR* is mean time to repair, marking the mean downtime after a failure has occurred. Sometime a third factor, *MWT* – mean waiting time, is added to the denominator of this equation, because repairs might be delayed by missing replacement parts, limited capacity of service personnel or a need for a specialist to come and solve the problem. Expressed this way, availability shows the proportion of the total time that equipment was functional. Unfortunately functional does not mean functioning.

In real manufacturing systems, machines may not be running, even though they are in order. This can be caused by the lack of raw parts or ongoing maintenance activities. Hackstein and

Budenbender [2] refer to disturbances of this type as *organisational*. According to them, technical downtimes include systematic non-productive times, such as run-up of the plant at the beginning of a shift, direct technical problems occurring in the machine itself, such as tool changing faults, and indirect technical disturbances that immerge elsewhere in the system, such as interrupted part supply due to failure of transportation system. As mentioned before, availability of resources is further reduced by non-technical, i.e. organisational, non-productive times such as lack of raw materials or planned maintenance.

Similar classification is advocated by Wiendahl and Springer [5]. They distinguish between technical and organisational failures as well as breakdowns conditioned by linkage and conception. Linkage is a common name used for shop floor logistics system and central computer control, thus downtimes condition by linkage are equivalent to Hackstein and Budenbender's [2] indirect technical failures. Downtimes conditioned by conception are embedded in the system during design phase, thus are similar to systematic disturbances.

A further extension of these ideas is Nakajima's *overall equipment effectiveness* (OEE) model [8], which assumes that system performance is decreased by the so called six big losses:

Downtime:
1. Equipment failures.
2. Setups and adjustments.

Speed losses:
3. Idling and minor stoppages.
4. Reduced speed.

Defects:
5. Process defects (scrap and rework).
6. Reduced yield.

The first two losses reduce the time that machine is available for production, because neither in the failed state nor during the setups machines do produce any output. Thus, *availability* of a resource is calculated as:

$$Availability = \frac{Loading\ time\ \text{-}\ Downtime}{Loading\ time} \times 100\%$$

Loading time is total time during a certain period (day or month) less planned downtime, such as maintenance or daily production meetings. Downtime refers to unexpected problems, such as equipment failures. The difference between loading time and downtime gives operating time that will be used later on. So availability shows the percentage of time, which was available for manufacturing after planned and unplanned breakdowns.

As before, available does not imply running and running does not mean producing at maximum possible speed. First, this is because machines might not work due to organisational issues, such as interrupted material supply, as well as minor technical stoppages caused by sensor contamination and resulting signal errors or by limit switch failure [2, 8]. Next, to achieve the required process capability less reliable machines might need to run at speeds below theoretical values. Speed can also be reduced by partial failures, which do not bring a machine completely down, but limit its capabilities. For any of these reasons real cycle times can be longer than rated ones, therefore actual output will be below estimated amount. This reduction is evaluated by *performance efficiency*:

$$Performance\ efficiency = \frac{Theoretical\ cycle\ time \times Processed\ amount}{Operating\ time} \times 100\%$$

It might sound paradoxical, but there can be found cases, when performance efficiency is more than 100%. This happens if the actual cycle time is shorter than the rated time, so the real output exceeds the estimated amount. Occurrences like that are common in manual operations and relate to the way theoretical cycle time is determined.

One method to estimate cycle time is to use historical data from similar processes. If such information is not available, rated time can be calculated employing scientific methods, based on motion analysis. In the worst case cycle time is "estimated" by guessing. Neither of these methods is accurate, because any estimation is subjected to error. An alternative way is to measure the actual time during the test batch production, but then a human factor comes into play. As the time goes on operators naturally get faster due to learning effect. Moreover, in some companies people are paid for productivity, which is measured as throughput in excess of theoretical output. In such cases people tend to work slower when measurements are taken to make it easier to exceed the theoretical yield and earn a bonus later on.

The following issues can distort efficiency measurements, encourage unreasonable optimism, and what is worse they can reduce the motivation to continue with further improvements. To avoid these pitfalls it is important to update theoretical cycle times periodically and keep the records as accurate as possible. At the same time we should note that machines, as opposed to manual operations, are less influenced by the factors mentioned above. However, rated time should still be updated each time a process is modified.

The last two big losses stand for parts that fail quality inspection at any place in the process and have to be scrapped or reworked. In both cases they are considered as defects, because even if an item can be salvaged through rework it distorts regular flow and consumes extra time that otherwise could be used to make other parts. Then the *rate of quality products is*:

$$Rate\ of\ quality\ products = \frac{Processed\ amount - Defect\ amount}{Processed\ amount} \times 100\%$$

After availability, performance efficiency and rate of quality products are determined, overall effectiveness of a particular machine can be expressed as a product of the three:

$$OEE = Availability \times Performance\ Efficiency \times Rate\ of\ Quality\ Products$$

It is easy to see that OEE is closely linked to reliability and is also in line with earlier discussed models. First, reliability directly influences availability – the more reliable the machines, the less downtimes and the more time available for production. This corresponds to technical availability. Next, by taking into account planned downtimes, availability (in OEE) considers some of the organisational non-productive times. Organisational issues, as well as technical problems, are further taken care of by performance efficiency, because both organisational disturbances and minor technical stoppages mean that less amount will be processed during the same period of operation. Finally, reduced speeds due to technical condition of equipment or for the sake of process capability increase real cycle time and further reduce output.

In addition OEE adds a quality dimension, which in terms of reliability can be justified as follows. If a machine produces a defective part it fails to perform the required function – to make a good product. In the worst case defective part has to be scrapped and the whole time to manufacture it is wasted. Even if the effects of such a failure can be minimized through rework, it still consumes extra time, which otherwise could be used to generate more throughput. Finally, it was already mentioned that a physical failure of a machine can lead to damaged parts. Therefore poor reliability will deteriorate quality and overall equipment efficiency.

Unfortunately, despite its attractiveness, OEE, as well as other models discussed so far, has one inherent problem – it is best suited for individual machines. In practice, however, resources seldom operate alone. In general they are part of a complex network, which on the product level reduces to separate lines. Allcock [9] quotes a representative from Siemens company, who says that OEE does not take account of factors upstream and downstream in the supply chain. As the below example shows, this might be a serious drawback.

Let us consider a simple production process shown in Figure 1. The system consists of three non-identical machining centres MC1, MC2 and MC3 (in practice even machines of the same type and age will rarely exhibit identical performance) with maximum theoretical capacity of 80, 50 and 90 parts per day respectively. For simplicity let us consider only technical availability, which is most sensitive to reliability, i.e. we will assume that setups are negligibly short and process is simple, so no serious organisational troubles occur. Statistics shows that machine MC1 has an average technical availability of 80 percent, machine MC2 is quite new thus is available 90 percent of time and the average technical availability of MC3 is 70 percent.

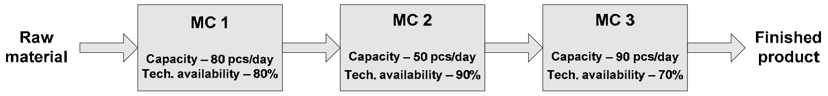


Figure 1. Production system with a constraint.

From the first glance it may seem that due to frequent failures machine MC3 requires the biggest attention, while MC2 is doing best, so is the least important. However, a closer look shows that an improved efficiency of MC3 will lead nowhere, because the line can at best produce $50 \times 0.9 = 45$ finished products, which is the best possible output of MC2 taking into account its reliability. At the same time MC3 with a poor technical availability of 70 percent, can still finish $90 \times 0.7 = 63$ pieces per day. Thus, investment in improved reliability of MC3 can only be justified if the cost savings from less frequent failures are greater than the invested efforts, or if the gained capacity can be used for other operations (e.g. to help MC2). However, as long as the maximisation of the throughput of and profit from the product in question is concerned, main efforts should be devoted to machine MC2, because it is the bottleneck of the line and every minute lost in it means a minute lost in the whole line. From this point of view, the second most important resource is MC1, because it feeds MC2 and the least important is MC3.

This is an oversimplified example of course, because in general we should consider other factors affecting performance of a manufacturing line. Moreover, average values should be replaced with corresponding confidence intervals to have a more accurate picture of the system. Nevertheless, this example shows that focusing on individual machines can be misleading as it may not contribute to the overall improvement. Thus, to avoid sub-optimisations, a different approach, which considers the system as a whole, is needed.

During the research in former Western Germany, Hackstein and Budenbender [2] used earlier introduced model of availability to evaluate the productivity of two complete manufacturing cells. Even though no calculations are shown, from the presented results it appears that performance of the whole system was measured in terms of the performance of machine tools, which on their own are assumed to be independent.

Associating of system performance with the efficiency of machine tools is inline with our example, because machine tools directly contribute to the throughput and, just like MC2, act as a limiting resource that determines the throughput of entire system. Other equipment, such as transportation system, washing or measuring machines, are supportive and their efficiency should be viewed in terms of the affect on the performance of value adding activities. For example, just like a lengthy failure of MC1 can cause starvation of the bottleneck MC2, a long breakdown of automated guided vehicle (AGV) can lead to lack of material at machine tools and therefore will be immediately reflected on reduced output. Short failures of AGV's, however, would remain "unnoticed" as long as they were able to serve the machine tools, just like low reliability of MC3 is not significant, as long as it is able to maintain the required production rate.

The assumption of independency is revealed by the fact that system losses appear to have been evaluated by adding up capacity lost at each individual machining centre. Total non-productive time was then expressed as a percentage of the total available hours, which is again the sum of individual capacities. Such approach is only valid for parallel structures, where resources do not directly depend on each other (as long as they do not have to "compete" for jobs). In parallel arrangements machines can be viewed as common pool, so individual capacities and productivity losses can be added together. In general case, however, parallel structures would be part of longer lines, and efficiency of the whole pool would depend on other resources.

Nachiappan and Anantharaman [10] report that some researches have proposed an extended OEE approach for flow lines, where availability, performance efficiency and rate of quality products are replaced by line availability (*LA*), flow line performance efficiency (*LP*) and flow line quality efficiency (*LQ*) respectively. Then, like OEE, *overall line effectiveness* (OLE) is determined as a product of these three parameters. To calculate the contributing factors themselves two *rules of thumbs* are proposed. According to the first rule *LA*, *LP* and *LQ* are determined as an average of corresponding efficiencies of individual machines. The second rule multiples availabilities, performance efficiencies and rates of quality products of machines to get corresponding efficiencies of the whole line. To examine the validity of these rules let us return to example in Figure 1.

We have already shown that efficiency of MC3 was limited by the output of MC2 and could not be improved, unless excessive capacity was used elsewhere. What is more, efficiency of MC1 also depends on MC2, because running MC1 at the rate exceeding the capability of MC2 would only increase work-in-process inventory (WIP) and associated costs, but the output of the system would remain unchanged. In order to avoid build-up of inventories, the output rate of all resources has to be subordinated to the slowest member of the chain. So in our case all machines will run at the rate of approximately 45 parts per day.

Let us now apply the first rule of thumbs. Average capacity of machines MC1 through MC3 is 73.33 units per day, and average availability, i.e. line availability *LA*, is 80%. Since we have not considered yet any other losses, but availability, this suggests that we should have around 58 parts (73.33 × 0.8) at the end of a day. However, we know that instead the line will produce only 45 finished products per day. This discrepancy arises because simple average ignores the existence of constrains in the line. The issue can be solved by assigning different weights to machines. However, we are going to show that averaging is not suitable for *LQ* either.

Let us assume that 10% of parts are scrapped at each machine, i.e. rate of quality products of MC1 through MC3 is 90%. Then the first machine will make 45 × 0.9 = 40.5 good parts. Of course we should take this reduction into account and release a bit more parts to MC1 to retain the output of 45 parts, so that no bottleneck capacity would be wasted. However, let us ignore this possibility for the sake of simplicity of this example. Then after the parts will have gone through MC2 there will remain 40.5 × 0.9 = 36.5 good parts and finally 36.5 × 0.9 = 32.8 good parts will exit the line. But according to the first rule of thumbs, line quality efficiency *LA* is the average of individual rates of quality parts, i.e. 90%. This for the line should produce 45 × 0.9 = 40.5 good items. This is another demonstration that the first rule of thumbs is not a valid approximation.

In fact to determine the real output of the defect-prone line we have just used the second rule of thumbs. But now let us assume that quality problems have been eliminated, while availability issues remain. Then we shall make 45 parts at each machine and their performance efficiencies will be 56.3% (45 / 80 × 100%), 90% and 50% respectively. Applying the second rule of thumbs we should multiply these figures to determine the performance efficiency of the line. However, in such way we would assume that the throughput of the line would be reduced at each stage. Unfortunately, this is not true, because even in a non-subordinated line the output of MC3 will be the same as MC2 – 45 parts, i.e. the performance will not be reduced at the last stage. Thus, despite that the second rules of thumbs works for line quality efficiency, it is not valid for line performance efficiency, which for neither of the two rules is valid for all three constituents of overall line efficiency metrics.

The conclusion to be drawn here is that in spite of the apparent simplicity, extended OEE and Hackstein and Budenbender's [2] availability models are not suitable for a general case of flow lines. Instead, Nachiappan and Anantharaman [10] have developed an alternative method to evaluate OLE. It is based on the fact that the output of a machine, operating in a continuous line, will be the input for the next resource, just like 45 parts form MC2 were the input for MC3. The authors also argue that since defective parts will not be transferred to the next machine, there is no sense in distinguishing between *LP* and *LQ*. Instead a new parameter, called *line production quality performance efficiency* (*LPQP*), is introduced and OLE is calculated as:

$$OLE = LA \times LPQP$$

LA is the operating time of *n*-th machine (OT_n), expressed as a percentage of loading time (*LT*):

$$LA = \frac{OT_n}{[LT]} \times 100$$

Loading time of the line is total available or calendar time (*CT*) less planned downtime (*PD*), which is the same as loading time in OEE model. However, what is different is that loading time is calculated only once – at the beginning of the line, whereas in extended OEE model it would be determined for each single machine.

$$LT = CT - PD_1$$

Operating time at any machine is determined by subtracting planned and unplanned downtimes (*DT*) of the machine in question from the operating time of the previous machine:

$$OT_i = [OT_{i-1} - PD_i] - DT_i$$

This equation assumes that the output or the operating time of the previous machine is the input or the calendar time for the succeeding resource. We should note, however, that this assumption is not absolutely correct, because even in a well balanced line in short term machines will not run at exactly the same rate, which for some WIP is inevitable and even necessary for smooth operation. Given that, when a machine stops, the next one can still continue running as long as inventory, which was possibly built by the preceding machine before it broke down, is depleted. Moreover, after repair, the first resource might be able to catch up, which way the second machine would not be influenced by the failure of the first one. So the validity of the above assumption largely depends on the degree of line balance, the amount of margin that each resource possesses and the time to repair the broken machine.

If this limitation is ignored, OT_n can be determined by applying the equation downstream the chain, i.e. starting with the first and finishing with the n-th machine. Note that the first work centre is assumed to be independent of other resources, which for the operating time of the first machine is:

$$OT_1 = [CT - PD_1] - DT_1$$

This, however, is not an absolute truth either, because if the first machine is not the bottleneck it will probably not need to run full time to satisfy the demand. To compensate for this problem, the unavoidable idle time of the first machine can be included in its planned downtimes.

So, by dividing the operating time of the last member of the chain by the time that was input at the beginning of the line, *LA* evaluates the total time loses in the line. The second parameter *LPQP* is calculated by multiplying the number of good products made at the n-th machine (G_n) by the cycle time of the bottleneck resource (*CYT*) and dividing it by the operating time of the first machine:

$$LPQP = \frac{(G_n \times CYT)}{OT_1} \times 100$$

Cycle time of the bottleneck resource is used, because it determines the output of all whole line. Thus *LPQP* is the ratio between the total time that was theoretically needed to make the number of good parts actually produces and the time that was available at the beginning of the line.

The number of good parts produced is the maximum possible output (n) of a machine under given conditions (taking into account the effect of constraints) minus the degraded quality (D) and rejected parts (R):

$$G_i = n_i - (D_i + R_i)$$

The first machine is said to be independent of other resources, so its maximum possible output will be equal to its theoretical throughput ($n_1 = N_1$). Note once again that this is not an absolute truth, because even the machines located upstream the bottleneck resource have to be subordinated to it in order to avoid the build-up of inventories. However, since only the final number of good parts matters, you will soon see that this drawback will have no effect on *LPQP*.

In general for any machine theoretical throughput can be calculated as operating time, less setups and other idle times, which the authors call performance reduction time (*PRT*), divided over part cycle time (*cyt*) at a particular machine:

$$N_i = \frac{OT_i \text{ - } PRT_i}{cyt_i}$$

The output of all other machines, but the first one, will depend on previous resources in such way, that maximum possible output will be less or equal to the output of the previous machine. In our example, machine MC3 can in the best case make 50 parts (in case of no failures of MC2), even though it is cable to produce 90 pieces per day. So if the next machine is less capable than the previous one feeding it or if their capacities match, i.e. if $N_i \leq G_{i-1}$, then the machine will at best match its theoretical output, i.e. $n_i = N_i$. In the opposite case, when the next machine is capable to process more than the previous one, i.e. when $N_i > G_{i-1}$, the machine will only produce as much as the previous one will supply, i.e. $n_i = G_{i-1}$.

So, *LPQP* is a joint measure, which takes into account the reduction of line performance due to setups and other minor disturbances as well as quality issues, because only the yield of the good parts is counted at the end of the line.

Applying this approach Nachiappan and Anantharaman [10] have achieved several interesting results. The result of interest for us is the one that confirms the conclusion that we have made using our simple example in Figure 1 – efficiency of the whole line depends on the efficiency of the bottleneck resource. If we now focus on this critical resource, we can observe that higher reliability will help to improve both line availability and production quality performance efficiency.

Similarly like in OEE model effect on line availability efficiency is straightforward. Recall that *LA* is proportional to operating time, which on its turn depends on downtimes ($OT_i = [OT_{i-1} – PD_i] – DT_i$). So the more reliable the machine, the fewer downtimes and the higher time available for production. If there is more time available, then the potential to produce good parts is bigger too. So reliability will also affect *LPQP*. Moreover, due to the fact that machine failures often lead to damaged parts, higher reliability implies less damaged parts, thus more good parts produced at the end of the line ($LPQP = (G_n \times CYT) / OT_1 \times 100$).

Thus, improved reliability of the bottleneck resource pays off immediately. However, quite often reliability is improved at the expense of reduced cutting speed or more conservative tool lives. Unfortunately, this will extend the cycle time of the bottleneck machine and further reduce its output, so as long as the efficiency of the bottleneck is concerned increasing its reliability by reducing cutting speed and going for more conservative tool lifetimes is not a solution, because the net result might be reduced throughput and OLE.

To sum it up, we have analyzed our system requirements and effects of reliability on its performance. We have determined that reliability improvement gives best results if applied on critical resources. Once again, we do not intend to say that reliability of other machines or elements of the system is not important. For example, if each machine produces only good quality products, the *LPQP* will be improved considerable. However, we have demonstrated that the first priority must be assigned to limiting resources.

4 FAILURES IN FMS

Now when we have discussed the role of reliability on the performance of production systems, let us return to our earlier discussion about FMS and look at some particular reliability questions related to this type of production systems.

4.1 Reliability versus Flexibility and Integration

Earlier we have mentioned that there seems to be a trade-off between the main characteristic features of FMS and reliability. Bennett and Jenney [11] have studied 100 machine tools of different complexity and have found evidence to suggest that equipment designed for small batch production, i.e. to provide more flexibility, could be likely to fail more frequently than machines, which run for long periods between setups. This is probably because longer runs mean more stable manufacturing environment, while each setup is associated with potential workpiece positioning, tool changing and other random faults.

Meredith [3] states that integration, the second pillar of FMS, is also a source of most of the problems, because it increases system complexity. This claim is supported by the research carried out by Hackstein and Budenbender [2], who have observed two FMSs for 20 and 15 days respectively. In both cases big drops (compared to the average) in technical availability were observed on a few particular days. In the first FMS this was caused by control software faults, while in the second company problems with pallet transporter were said to be responsible for reduced productivity.

The authors have also analyzed a third case, where a large mechanical engineering concern has reconfigured its stand-alone machines as an FMS. Observations showed that due to interlinking of resources this reorganisation had a significant positive effect on organisational problems and system availability in general, but taking technical availability separately, number of downtimes has increased, as compared to stand alone mode. The increase was attributed mainly to disturbances in transportation system.

The essential role of linkage is further emphasised by Wiendahl and Springer [5]. In one of the FMSs, observed for 83 hours, they have counted seven technical failures that lasted more than one hour. The main cause of these lengthy downtimes was said to be control system failures, which in spite of the low frequency resulted in 50% of the overall time losses. In addition there were 40 transport car faults observed, which in total lasted 2.8 hours, but caused a loss of 6.8 hours of productive time of machine tools.

The above case studies back up the findings of Meredith [3] that interlinking of resources is an important advantage of FMS, but at the same time it is a major source problems. This calls for particular attention to elimination of failures of integrating subsystems, such as automated material handling or centralized computer control, because in the worst case their breakdowns can deteriorate the performance of entire system.

4.2 Vital Few versus Trivial Many

The two FMSs observed by Hackstein and Budenbender [2] experienced a total of 442 stochastically distributed technical (for comparison 444 organisational) problems during 20 days and 159 technical (for comparison 336 organisational) issues during 15 days respectively. Wiendahl and Springer [5] report that over a period of 122 operational hours 620 downtimes (including organisational) with a total duration of 113.6 hours were recorded in one FMS. Another system was observed for 83 hours. During this time 753 non-productive events (including organisational) were observed and accounted for a total of 84.7 hours. In addition 40 technical failures, 2.8 hours in total, of transport cart were registered.

Most of the above failures turned out to be short. As many as 48% of downtimes in the first FMS and 32% of downtimes in the second FMS, observed by Hackstein and Budenbender [2], had lasted up to 10 minutes, but totalled only for 13% and 5% of total non-productive times in the first and second FMS respectively. The same tendency was observed with organisational troubles. Wiendahl and Springer [5] have calculated that two thirds of downtimes in the first FMS were shorter than five minutes, but shared only 12.6% of total duration. In the second case, half of downtimes lasted up to 10 minutes, but amounted for only 3% of total non-productive time.

Other researchers, for example, Vineyard, Meredith and Amoako-Gyampah [12] and Thilander [13], also conclude that firms will usually experience a big number of short failures. Hackstein and Budenbender [2] suggest that a big number of short breakdowns should be interpreted as a sign of operating team’s ability to eliminate problems quickly. Wiendahl and Springer [5] also point to the essential role of operators when dealing with frequent disturbances. However, there are other points to see here.

Frequent, although short, failures distort regularity of the flow and are a threat for unmanned manufacturing [5], because in such case operator is not available to eliminate those problems and they turn into long downtimes. Moreover, the fact that a big number of problems stands for a small portion of total non-productive time means that efforts to avoid those numerous disturbances will have minor positive effect on equipment availability. For comparison let us consider the effect of long breakdowns.

Hackstein and Budenbender [2] report that only 6% of failures have lasted longer than 120 minutes, but in spite of that were responsible for more than 40% of the total technical non-productive time. Wiendahl and Springer [5] mention that in the first case study 3.7% of downtimes were over one hour, but still were responsible for nearly 40% time losses. Similar analysis in the second FMS showed that as few as 7 out of 40 technical failures lasted longer than one hour, but resulted in 50% of technical non-productive time. These ratios show that unlike the frequent failures, a small number of long issues had a major effect on the system. This is a well known Pareto's phenomenon, widely employed in Total Quality Management and other areas. It states that 80% of quality problems are caused by just 20% of factors, so the key point is to determine and eliminate those vital few causes, as Joseph Juran, one of quality gurus, calls them.

The conclusion that we would like to make here is that as long as availability of resources is concerned the focus should be placed on the few long downtimes. This task is much easier if problems are of similar character. For example, Hackstein and Budenbender [2] state that in the second FMS lengthy downtimes were caused by pallet transporter failures, while in the first one the most likely source of long non-productive periods was control software. In general however, problems can be split over different categories, making them difficult to predict and prevent. So identification of main failures modes and categorisation of collected failure data is important step in reliability analysis.

4.3 Failure Modes

Such information, however, seems to be difficult to find in literature. Bennett and Jenney [11] distinguish between mechanical, electrical and hydraulic failures of machine tools. Unfortunately, FMS is more than just machines, thus Hackstein and Budenbender [2], besides sporadic disturbances in electrical subassemblies (failure of limit switches, failure of contacts, signal transmission errors due to sensor contamination), sporadic disturbances in pneumatic and hydraulic subassemblies (mechanical failure of valves, failure of lines) and failure of mechanical subassemblies due to wear outs, speak about software and material handling system failures.

A similar classification is used by Vineyard and Meredith [4] and Vineyard et al. [12]. The two papers present data from a nine-month case study, where an FMS, working on a simple corrective maintenance bases, was observed. Corrective maintenance means that equipment was repaired after failures, but no preventive actions were taken, so the data is assumed to be pure, not influenced by earlier maintenance activities. The authors have come up with the following set of failure types:

1. Mechanical – screws, turrets, bearings, and slides.
2. Hydraulic – pumps, solenoids, and fuses.
3. Electronic – solid state components, circuit boards, drives, and controls.
4. Electrical – motors, transformers, relays, and switches.
5. Software – logic errors in the system or part programs.
6. Human – stoppages due to personnel inaction or inappropriate actions.

All failures had decreasing rates and followed Weibull probability distribution, which is said to be consistent with failure patterns of other complex systems [12]. Apparently in case of FMS it is difficult to control all factors, influencing reliability, at the installation stage. For this reason, most of the problems show up in the beginning of the service, after which system performance stabilizes.

Concerning the frequency, a total of 1310 individual failures were observed, 520 of which were human, 243 software, 213 electrical, 165 mechanical, 107 hydraulic, and 62 electronic. Figure 2 below gives a graphical representation of the data. As can be seen, human failures accounted for nearly 40% of the total number of problems. However, we have previously demonstrated that frequent disturbances do not necessary consume a lot of productive time. So a better measure is length of downtime attributed to each category.

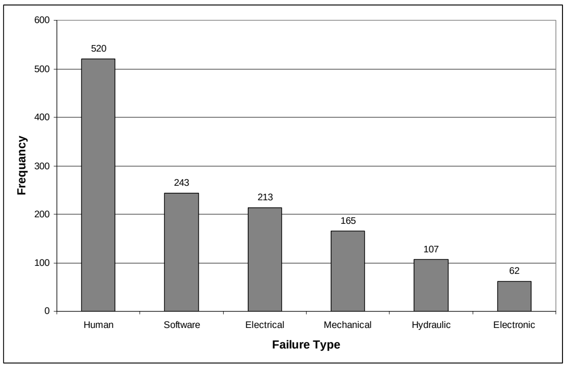


Figure 2. Number of failures [4, 12].

An estimate of downtime per failure category can be obtained by multiplying the above frequencies by average repair times, which were shown to follow the Lognormal probability distribution [12]. Data reveals that the longest to repair were hydraulic problems and there were quite a few of them. This is inline with earlier discussion about Pareto’s Law, according to which we should see a few very long failures. But on the other hand human failures, which were the most frequent, also required long repairs, with total downtime of 470 hours (see Figure 3), which stands for nearly 43% of the total losses.

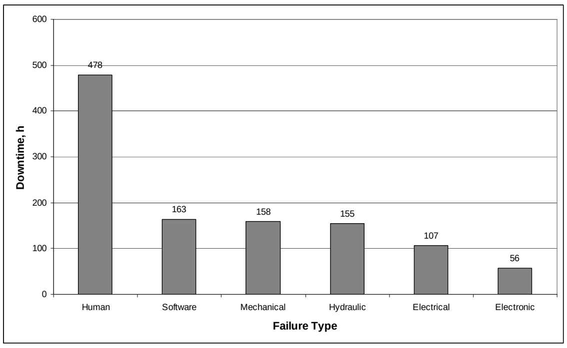


Figure 3. Downtime (*based on data from* [4] *and* [12]).

This contradiction can partially be explained by the way downtimes are assigned to certain categories. For example, operators’ mistakes are very short failures, but often lead to mechanical damage and long downtimes. Employing the root cause principle these breakdowns are classified as human, rather than mechanical. On the other hand the root cause method can be difficult to apply in practice. A particular problem concerns humans, because people fear to admit that their actions or inactions have caused the damage. For this reason it is likely that some of the mechanical failures, in the case study under discussion, should have actually been classified as human failures [12]. This would counterbalance the previous assignment of mechanical breakdowns to human failures and drive us even further from the Pareto's principle.

Another explanation could be possible interaction between individual failures [12], meaning that a small problem can provoke a serious failure. The relationship between human and mechanical failures is one example. Interaction between other types of failures is possible two. If we consider a car analogy, leaking oil (hydraulic failure) can cause poor lubrication and jam the engine (mechanical failure) if not noticed on time. Moreover, new problems can be triggered by maintenance personnel, because when attempting to repair such complex systems as FMS, it is easy to forget small things, like tightening a bolt, which later on can develop into a new failure [4, 12].

So failures do not occur in isolation, they interact, therefore care must be taken when collecting and interpreting the data. In the case under discussion, results contradict with the Pareto's Law in the sense that most frequent disturbances also took long time to eliminate. But on the other hand the results obey the law, because the above analysis shows that there is one clear group of problems, which stands for a vast amount – 43% or even bigger (if we assume that some mechanical failures were actually human failures) share of downtime. Thus from the above data it looks that human factor is the most crucial one affecting the reliability of FMS.

4.4 Dealing with Human Failures

Frequent failures and long downtimes caused by human mistakes suggest that in order to improve the reliability of FMS one should attempt to limit the involvement of people in production process. The first thing that comes to mind is automation. Unfortunately, even in unmanned manufacturing the role of operators and other personnel remains important. Such activities as fixturing and unclamping of workpieces, packaging and other operations, especially those requiring subtitle judgement, like various visual inspections or initialization of rework, still need to be done manually. What is more important, Nakajima [8] states that maintenance is difficult to automate. Therefore, in spite that humans cause lots of downtimes, it looks that production systems will have to live with it for the nearest future. Thus, if people can not be completely eliminated, some measures need to be taken in order to reduce the negative effects of their activity.

4.4.1 Training

Human failures were said to be stoppages due to personnel inaction or inappropriate actions. Very few people do this deliberately. Instead, in many cases they just do not know how to deal with a particular situation and what actions to take. Or, even if they know what to do, it takes too long time to accomplish and problems, which could have been avoided, evolve into big breakdown. Such issues can be prevented through instructions and training.

As a proverb says, nobody was born educated. All the skills are gained through hard work and experience. Thus, through training people first of all learn how to do their job right. Next, they find out what the critical things, which should be taken care of, are. In addition people learn the procedures that need to be followed and actions that need to be taken if things go wrong. Finally, by training people develop these skills, so it takes less time to act when necessary.

As will be discussed later, Vineyard and Meredith [4] have proven that such preventive measures have a big positive effect on human failures. Moreover, training does not only prevent human mistakes, but also helps to eliminate failures of other types. Wiendahl and Springer [5] report that nearly 60% of the downtimes in one of their case studies were possible to eliminate by operators. Hackstein and Budenbender [2] second their words by saying that the greater the number of faults, which can be eliminated by the operators, the shorter the non-productive time, since there will be no need to wait for the repair and maintenance staff. Finally, Vineyard and Meredith [4] wrap it up stating that by eliminating operators automation has also removed a valuable maintenance resource.

4.4.2 Loyalty and Favourable Work Environment

Training is an important way to attack human failures. However employee turnover reduces its efficiency and increases cost, because new employees need to be trained for the same things for which previous workers had already been trained. If this work is omitted, earlier problems will reappear. So employee turnover should be reduced if a reliable process is to be achieved. Measures to do this include pay system and other incentives, promotion of employee loyalty, through company picnics, open door days, and other.

Unfortunately even well trained and experienced people do make mistakes. The reasons are fatigue, stress, the so called slips [14], when intended action does not occur as planned, and other physical and psychological problems. To minimize the effect of these soft factors companies should create favourable working conditions.

A good example is Volvo engine plant in Skövde, Sweden. Local told once that in the old days working conditions in the factory were so bad, that parents used to intimidate their children that if they were not going to improve at school they would have to work in Volvo. Today however, things are very different. Factory looks like a big park with corridors full of natural trees and flowers, benches to take a rest during the breaks. In the workshops machines are painted in green, because this colour is believed to have a positive relaxing effect. We think that this is a good example how stress at work can be addressed in order to reduce the number of human errors.

4.4.3 Poka-yoke

Effective training and creation of favourable atmosphere can considerably reduce the number of human errors. However, as the time goes, old problems reappear, because people tend to forget things, especially in repetitive operations. In addition new errors occur for different random causes. This makes human errors difficult to predict and control. Moreover, some of people’s mistakes are not necessarily their fault, as poorly-designed processes that require a great deal of attention can contribute severely to problems [15]. A system that both encourages the development or reengineering of processes in such a way that a probability of error occurrence is minimized and helps to prevent them from happening during the operation is called *poka-yoke*.

Poka-yoke is a Japanese word meaning mistake-proofing. In fact the original term introduced in 1961 by Shigeo Shingo, a Japanese quality guru, was baka-yoke, meaning fool-proofing. The two English version are now used interchangeable, but mistake-proofing is preferred because fool-proofing turned out to be dishonourable and offensive [16].

The first poka-yoke device that Shigeo Shingo proposed was based on the idea of a checklist. Just like an unchecked box informs us that something has not been done, a spring remaining on a small dish after assembly was a signal for Yamada Electric workers that a spring had been left out and the assembly had to be corrected [16]. Later on various mistake-proofing techniques, based on limit switches, sensors, gauges, sound and light alarms and complex systems, such as fuzzy logic neural network that detects tool breakage automatically and stops the machine [14], were developed. From this point of view poka-yoke is closely related to monitoring systems, which will be discussed later in this article. However, the reason for considering poka-yoke here is the fact that it is not limited to physical devices. Instead, poka-yoke is a more general term meaning any design technique or a device that make it impossible or at least very difficult to make a mistake.

Shingo's first mistake-proofing attempt, that we have presented, was actually a redesign of a push-button assembly process. Other examples of inexpensive process modification, given by Shingo [16], are arrangement of insulating tapes into groups of 10 rather than into continuous line to ensure that all 10 tapes have been applied, or redesign of a bridge so that left-hand and right-hand parts could not be mixed. Concerning the FMS, these ideas can be applied in fixture design to safeguard correct positioning of workpiece. For instance, if orientation of a part maters, a simple pin on the jig and a corresponding hole in the workpiece will allow to clamp it only one way. Or, if parts are asymmetrical, the asymmetry should be increased, which will both allow the positioning only in the desired orientation and will make it easier for operator to notice shape irregularity and insert the workpiece into the fixture quicker.

4.5 Debugging

Software appears to be the second most vulnerable element of FMS. As mentioned before it is an important feature of the system. Software is at the heart of contemporary numerical control (NC). Moreover it enables to link all resources and coordinate their actions. However, due to the same essential role, software failure can bring the whole FMS down.

The origin of software problems is human in nature. They are "embedded" into the package during programming. The good thing about software problems, however, is that unlike human mistakes once eliminated they seldom return [12]. Thus an effective debugging during the installation and test runs of FMS should prevent software problems from appearing during production. However, all possible cases are not possible to simulate at the initial stage. Moreover, design and part program changes as well as modifications of the control software will introduce new errors. This emphasizes the role of cooperation between programmer, who might not be a specialist of the cutting process, and process designers.

Eventually, some of the problems will still pop up during production. Thus it is a smart idea to have a backup of part programs and a few software storage options. It is even more important to leave a possibility to run machines in a stand alone mode open, which will quickly pay-off in case of FMS's control system failure [1].

4.6 Maintenance

As discussed earlier equipment in FMS configuration is utilized much heavier compared to conventional machines, thus is more prone to failures. This emphasises the importance of maintenance. Approaches developed in this area, such as Preventive Maintenance, Productive Maintenance and Total Productive Maintenance, focus on preventive maintenance. Nakajima [8] uses the analogy of healthcare. Just like daily hygiene, periodic health checks and early treatment prevent illnesses and extend human life, daily maintenance, inspections and preventive repairs prevent failures of machines and prolong their useful lifetime.

However, preventive maintenance alone is not sufficient. This is because equipment failure mechanisms change throughout different stages of its use. Thus, trial runs during start-up period, simple tidiness, promoted by procedures like 5S, and delegation of maintenance tasks to operators, this way showing trust and improving employee participation, are needed. Such principles should become part of company's daily live and must be guided by overall maintenance policy. These issues are universal and are well explained in Nakajima's book on TPM [8], so will not be discussed in more detail. Instead we shall have a look at some specific FMS's maintenance problems, because FMS is a very complex system and behaviour of equipment was said to be very different here.

Vineyard and Meredith [4] have used the empirical data, presented in chapter 4.3, to model the affects of different maintenance policies on the performance of FMS. Total number of maintenance tasks was chosen as a parameter. Analysis of simulation results showed that the choice of the maintenance policy had a significant effect on the total number of maintenance tasks. Surprisingly, simple corrective approach, i.e. applying no preventive actions, but repairing broken machines, turned out to be the best, resulting in least amount of maintenance work required. The only policy, which could not statistically be proven to be worse, was a hybrid 30-day opportunistic on failure approach. The latter policy means that system is run on the same corrective bases for 30 days, after which it switches to an on failure opportunistic policy. The latter approach means that maintenance is triggered by a failure, after which not only the broken unit is repaired, but also preventive maintenance on other parts of the machine is performed. Then, policy is reset back to corrective for another 30 days and the cycle repeats.

At the same time, applied alone the on failure opportunistic policy caused the biggest total number of maintenance tasks. These controversial findings are explainable by the extreme complexity of FMS. As mentioned before, new failures can be triggered when attempting to repair broken machines. As Nakajima [8] states, maintenance is still difficult to automate, so even in unmanned operation, maintenance will have to be done manually. Unfortunately, human errors were showed to be the most frequent and crucial for performance of FMS. Thus any activity involving human intervention, like maintenance, can be a source of new troubles.

An interesting example to support this assumption comes from aviation history. It has been reported that on British Airways' flight No.5390 from Birmingham to Malaga a big part of aircrafts windshield blew out nearly sucking out the captain. The investigation revealed that this failure was caused by the maintenance staff that had replaced a worn out window, but by mistakenly used wrong size screws.

Comparison of the effect of maintenance policies on separate failure categories provided no evidence that the choice would have a significant influence on any failure type, except for human and electronic failures. While the number of electronic failures was increased significantly when on failure opportunistic maintenance policy was applied, the same policy, as well as 30-day opportunistic on failure policy, has minimised the average number of human errors. The result was explained by the nature of the two types of failures. Because of the difficulty to measure the state of components and because of high infant mortality of new elements, testing and replacement of functioning electronic items for preventive reasons was said to be infeasible. On the other hand, as discussed earlier, humans tend to respond well to preventive measures, such as training.

The final point that Vineyard and Meredith [4] mention is that their study did not consider the long term effects of preventive maintenance. As suggested by Nakajima[8], preventive maintenance is the health care for equipment, thus extends its lifetime. Therefore, 30-day opportunistic on failure policy was proposed for FMSs, as statistically it was not worse than corrective policy in terms of total number of maintenance tasks, it was as good as other policies in case of electronic failures, and included preventive maintenance, which kept the number of human mistakes low and was most likely to offer payoffs in the long term.

A similar study was carried out by Savsar [17]. However, he has come to a completely opposite conclusions, stating that opportunity triggered maintenance policy, which seems to be identical to on failure opportunistic policy, would lead to the best system performance, while corrective policy would do worst. In fact, Savsar [17] has used production rate, over number of maintenance tasks, as a performance measure. This makes the comparison of the results a bit difficult.

In general we believe that system output is a better indicator than the number of maintenance tasks, because maximisation of productivity and associated income must be the ultimate target of any company. Next, it is normal that any policy involving preventive maintenance will lead to more maintenance task, because in addition to breakdowns, every preventive activity also counts, as it consumes some of the productive time. Moreover, this drawback can be overcome by scheduling maintenance for off-work hours, for example during the third shift. On the other hand, however, in unmanned operation company might want to run machines all three shifts, so then preventive maintenance, whenever performed, would still cause non-productive periods.

Another reason discouraging the use of number of maintenance tasks as a parameter is earlier discussion about the duration of different failures and corresponding frequencies. We have found enough evidence to state that numerous failures, therefore the corresponding maintenance tasks, do not necessarily consume most of the productive time. But at the same time we have also discussed that short and frequent disturbances is a real obstacle for unattended manufacturing, in which case counting the number of required maintenance tasks makes sense.

What is more, Vineyard and Meredith's [4] research has a solid empirical backup, which seems to be missing in Savsar's [17] model. Moreover, in the latter case it is assumed that all failures due to wear-outs can be eliminated by preventive measures. While this might seem to be reasonable in theory, we think that it is difficult to achieve in real life situations, because FMS consists of a huge number of components and keeping the wear of each part under full control is complicated. Finally, Savsar [17] does not consider the human factor, i.e. the possibility of new failures being introduced during maintenance.

After all it looks that comparison of maintenance policies is sensitive to the choice of performance measures and different assumptions allowed in the models. Nevertheless conclusions drawn by Vineyard and Meredith [4] seem to be more solid. However, contradictory findings of Savsar [17] do not allow to point towards one best maintenance policy for FMS. This echoes the opinion of Nakajima's [8] that preventive maintenance, or any other policy, applied alone is not sufficient and should be complemented by other measures that are company, equipment type and age specific.

4.7 Monitoring

Besides other advantages brought by FMS and advanced technology around it is the reduced dependency of production systems on skilled labour. However, elimination of humans from manufacturing process at the same time means elimination a valuable maintenance management resource, because in addition to their daily work operators used to continuously monitor the system [4]. Through long years of experience humans develop such tacit skills as ability to recognize process abnormalities through vision, hearing, smell and even intuition, which are of great importance when identifying emerging disturbances at an early stage [13]. However in FMS, especially in unattended operation, these advantages are gone, so maintenance of the systems must rely purely on sophisticated monitoring system.

Monitoring consists of two tasks: detection and diagnosis [18, 19]. The role of detection is to determine abnormal behaviour of the supervised system by collecting information about representative parameters and comparing it against the established limit values. If any misbehaviour is detected information is sent to diagnosis block. In general diagnosis consists of fault localisation, identification of its origin and prognosis of potential consequences of a particular failure [18]. According to Hu et al. [20] this job can take up to 80% of the whole downtime resulting from the failure, which for rapid fault diagnosis is a pre-requisite to attainment of high plant availability [2].

As we come to the object of monitoring, Ly, Toguyeni and Craye [18] say that it is both production processes and control system. Surveying of the latter is called direct control monitoring, which aims to avoid sending of control orders to the system if some of its components are in inappropriate states. A typical example is said to be assuring that the door of a machine is closed before cutting starts. Thus, direct control monitoring acts like information filter. Diagnosis in this case is restricted to localisation of faults and the main purpose is to improve system security.

Production process monitoring, as the name says, surveys operation of resources. It also aims to improve the security of people and machines, but at the same time it tends to assure maximum availability of resources. This task can be divided into curative and predictive monitoring. The difference between the two approaches is the type of failures addressed: curative monitoring deals with complete failures, such as sudden breakdowns, while predictive monitoring takes into account progressive failures and tries to predict the breakdowns before they happen.

Detection principle in curative monitoring is entirely based on physical sensors [18]. Signals are collected, analyzed and classified. In case of misbehaviour symptoms are generated and reported to diagnosis. The analysis of the symptoms is carried out using knowledge base. According to Hu et al. [20] acquisition of knowledge for such database consists of fault tree analysis, studying of control information and principles, such as electric or hydraulic circuit drawings, acquisition of condition-monitoring knowledge, which reflects the faults or abnormalities from the condition-monitoring point of view, and collection of expertise. Based on the results of analysis immediate actions to be taken to guarantee human and equipment safety are determined.

Detection in principle in predictive monitoring can be either direct or indirect [18, 19]. Direct predictive monitoring, just like curative monitoring, uses physical sensors to detect any signs of component wear, age or other types of progressive failures that might affect the performance of the system and suggest replacing failing parts. For example, an increased load on spindle drive motor, higher feed-forces or excessive acoustic emissions are all signs of tool wear, because a worn tool draws more power than a sharp one, it also needs more force to push it against the workpiece and just before failure of worn tool acoustic emissions increase up to five times [1]. Degradation of other systems can be detected in a similar fashion. For example, heat and vibrations at a joint can signal excessive friction due to poor lubrication, caused by oil leaks.

Such detection approach is very reactive and efficient because the decision to replace a part or a unit can be made on the basis of real-time information about the condition of the item, rather than on the statistical service time. The issue with latter policy is that like any estimate, statistical service time is subjected to error. In one case the real lifetime of a part can turn out to be shorter than estimated, which for such item will break down before the scheduled replacement. Such failure will probably be detected by curative monitoring system and serious damages will be prevented. However, time losses are usually lower, if the replacement is planned rather than forced. On the other extreme, an item may appear to be more durable than expected; this for replacing it too early would mean wasting of the residual life and associated money.

However, besides these advantages direct predictive monitoring has a few disadvantages. First of all direct predictive monitoring system has to be planned in the FMS implementation stage, because later on it might not be technically possible to install the required sensors. Second, such system needs a lot of physical instrumentation, so it can be rather expensive [18]. Third, Vineyard and Meredith [4] have demonstrated that some systems, such as electronics, do not benefit from predictive monitoring. Nevertheless the main issue with direct predictive monitoring is that a big number of sensors increases system complexity and makes it less reliable [19]. This conflict arises because sensors, like any other item, are not 100% reliable; therefore they can physically fail and cause serious issues. An example is temperature sensor in a car, which is supposed to turn on the cooling fan as soon as the cooling liquid temperature reaches 90°C. Fails of such device may easily lead to engine overheating. In order to avoid such problems critical systems are equipped with redundant sensors. This diminishes the effect of single device failure, but on the other hand it increases the probability of false alarms and associated downtimes. For example, a false fire alarm may lead to evacuation of manufacturing facilities and lengthy searches for a possible fire.

For the above reasons Ly et al. [18] have proposed an indirect predictive monitoring approach, which unlike direct predictive monitoring is based on logical sensors. The idea of the method is observation of material flows. Any deviation from the planned throughput is assumed to be either due to self failure of a machine, failure of the transportation system or failure of other resources, which may lead to starvation of the machine in question or blockage of its output due to the saturation of buffers at the succeeding resource. Such deviations are called drift. Drift is analyzed by a diagnostics algorithm, which uses mathematically described operating data to trace the failure back to its source.

In fact the detection of drift still needs a few physical sensors to measure the flows. However, the number of these devices can be minimised by careful analysis of workshop layout. This is a considerable advantage as it simplifies the system. Ly et al. [18] say that other important features of indirect predictive monitoring are its capacity to diagnose multiple failures and ability to control and improve the performance of the manufacturing system. This for this approach is suggested to be used as complementary technique to curative and direct predictive monitoring.

Unfortunately, Toguyeni, a co-author of the above method, and Korbaa [19] state that although innovating on many aspects, the suggested approach has several limitations, especially from the point of view of reactivity. This for they have proposed a new indirect monitoring method for FMS under cyclic planning. Cyclic planning recognizes the existence of bottleneck resources in the process and by building the manufacturing schedule around those critical machines attempts to maximise the output of the system, while at the same time minimizing WIP. Moreover the output of the system becomes predictable, as parts are always made in the same order and finished at roughly predefined moments. For this reason to detect a failure is enough to observe a deviation from the planned output date at the end of the line. On the one hand this does not give a direct answer on which of the resource is failing. Such information needs to be extracted, i.e. diagnosed, by employing resolution trees. But on the other hand, positioning of monitoring sensor only in one place makes the system even simpler, which increases its reliability [19].

Nevertheless, reactivity of this monitoring approach still remains the major drawback. Moreover, this implies that the method is sensitive to existence of margins in the system. As discussed previously, a machine with excessive capacity can easily catch up after own failures. In such case, breakdowns shorter than the margin of corresponding resources will not disturb the regularity of the flow, measured at the end of the line, which for they will remain “hidden”. Thus, in spite of innovative ideas and offered advantages, due to low reactivity indirect predictive monitoring seems to have a limited practical applicability.

4.8 Scheduling of Failures

Training of operators, encouragement of loyalty, reduction of stress at work, poka-yole, debugging maintenance and monitoring are important ways to improve the reliability of FMS. However, it is naïve to think that each single problem can be prevented. Moreover, some machines will be more prone to failures than others, as well as some specific products and processes will be more likely to cause troubles than others. Identification of such risk factors can be especially important for unmanned production, because problems occurring during the day, when maintenance staff is available, will probably cause much less downtime than those happening during unattended night shift, when no operator is available to restore automatic cycle. For this reason, as many problems as possible should be "scheduled" for the first and second shift when operators' and other maintenance staff's availability is maximum.

"Scheduling of failures" involves classifying of products according to complexity. The most complex parts should be produced during the day, while the simplest ones can be left for night shift. The same stands for new products. Even though they might look simple, new products are unfamiliar, thus are more likely to cause troubles. Therefore, the new parts should be made under supervision of an operator, while standard pieces can be machines unattended.

In addition to the above, the possibility of jams, snarled chips and broken tools can be minimized by programming the system to operate at more conservative speeds and feeds with lighter than usual depth of cut [1]. The issue arises when there are one or a few bottleneck resources in the system. Then, as discussed earlier, any time lost at those critical machines will decrease the throughput. To compensate for this reduction, more aggressive process parameters can be used during the day shifts.

Next, due to complexity of FMS and its failure mechanisms, cooperation with equipment provider is important. Sometimes local maintenance staff will need to consult the manufacturer via "hot line" in order to solve the problem [1]. This is yet another reason for scheduling of problems for the day time, as specialist at suppliers office will hardly be available during the night.

4.9 Don't Automate, Obliterate

"Scheduling of problems" for the right time can reduce the time to repair the equipment. However, what to do if the products have been sorted out, leaving only the simple and standard ones for unattended manufacturing, but the process of those selected parts still has to go through a highly unreliable machine? *Don't Automate, Obliterate!* With this slogan Hammer [21] points towards the importance of reengineering. If the process is inefficient, why should one try to improve it? Perhaps it is easier to change it? In case of unreliable machine, perhaps the parts could be routed through other resources, or perhaps the process could be rethought this way that the troublesome operation would no longer be needed?

Hammer [21] presents two case studies, which show that by simplifying business processes outstanding improvements can be achieved. Nachiappan and Anantharaman [10] have experimented with their model by gradually shortening the line, i.e. reducing process complexity, and have observed that OLE was improved significantly. Since we have demonstrated that OLE is directly linked to reliability, the role of reengineering and simplification should not be forgotten if process reliability is to be improved.

5 CONCLUSIONS

In this paper we have analyzed some of FMS's reliability issues. The main conclusions that we have come up with are as follows:

1. Reliability of FMS, like any other feature of a system, must be coupled with the requirements of particular business situation.

2. Reliability in productions systems does not only mean costly repairs, but also affects the availability of resources for further production and influences product quality.

3. Availability of machines is reduced by both technical and organisational disturbances.

4. Reliability must be seen in the context of the system as a whole, because focussing on reliability of individual machines may not contribute to overall improvement.

5. Performance of the system first of all depends on efficiency and reliability of constraining resources. However, improvement of reliability of those machines via more conservative operating parameters is not a solution, as it will reduce the output of the system.

6. From the presented case studies there seems to be an inverse relationship between flexibility and integration features offered by FMS and its own reliability. Especial attention needs to be placed on reliability of central control, material and tool handling systems, as their failures affect operation of entire FMS.

7. Companies will usually experience a big number of short downtimes. Despite their duration, these disturbances are a big threat for unmanned operation, when no operator is available to restore automatic cycle. Nevertheless, short failures consume a minor portion of production time, while very few long problems have major effect. Thus sorting out and prevention of the latter is essential for system performance.

8. Human mistakes appear to be most common and serious failure of FMS. In spite of that it is not possible to eliminate people from production process.

9. Measure to reduce the negative consequences of human activity include training, reducing employee turnover, attacking stress at work and poka-yoke.

10. Software is the second most vulnerable element of FMS. Countermeasures include close cooperation between programmers and process experts, debugging and backup of part programs, as well as possibility to run machines in stand-alone mode.

11. Maintenance is the main remedy for the remaining types of failures. Research in this area demonstrates that the choice of maintenance policy has a significant effect on system performance, but contradictory findings of two similar studies dos not allow to choose one best maintenance approach.

12. In the absence of operators during unattended operation monitoring systems plays an essential role in assuring high availability and performance of FMS.

13. Monitoring must survey both the software and hardware part of the system and can be curative or predictive. Predictive monitoring can be achieved in direct or indirect way.

14. Direct predictive monitoring is very reactive and efficient, but high number of required sensors can eventually reduce system reliability instead of improving it.

15. Indirect predictive monitoring is more simple and reliable, but lacks reactivity, so has limited practical applicability.

16. Reliability of FMS can be further improved by “making” failures happen at the right time, i.e. when maintenance staff and manufacturer support is available. This involves classifying products and processes by complexity and likelihood to fail.

17. Reliability can also be improved by eliminating volatile processes, i.e. by reengineering and simplifying them, o by replacing prone to failure machines by more reliable ones.

REFERENCES

1. Luggen, W.W., *Flexible manufacturing cells and systems*. 1991, Englewood Cliffs, NJ: Prentice-Hall. XV, 461 s.
2. Hackstein, R. and W. Budenbender, *Flexible manufacturing systems in practice -- results of a study of operational use.* Engineering Costs and Production Economics, 1991. **21**(3): p. 259-276.
3. Meredith, J., *Installation of flexible manufacturing systems teaches management lessons in integration, labor, costs, benefits.* Industrial Engineering, 1988. **20**: p. 18-27.
4. Vineyard, M.L. and J.R. Meredith, *Effect of maintenance policies on FMS failures.* International Journal of Production Research, 1992. **30**(11): p. 2647-2657.
5. Wiendahl, H.-P. and G. Springer, International journal of advanced manufacturing technology, 1988. **3**(1): p. 55-69.
6. Rausand, M. and A. Høyland, *System Reliability Theory. Models, Statistical Methods, and Applications*. 2004, New Jersey: John Wiley & Sons Inc.
7. Corbett, T., *Throughput Accounting*. 1998, Great Barrington: North River Press.
8. Nakajima, S., *Introduction to TPM: total productive maintenance*. 1988, Cambridge, Mass.: Productivity Press. XX, 129 s.
9. Allcock, A.A., *Understanding OEE.* Machinery and production engineering, 2004. **162**(4102): p. 16-17.
10. Nachiappan, R.R.M., *Evaluation of overall line effectiveness (OLE) in a continuous product line manufacturing system.* Journal of Manufacturing Technology Management, 2006. **17**(7): p. 987-1008.
11. Bennett, D.J. and B.W. Jenney, *Reliability: Its implications in production systems design.* Omega, 1980. **8**(4): p. 433-440.
12. Vineyard, M., K. Amoako-Gyampah, and J.R. Meredith, *Failure rate distributions for flexible manufacturing systems: An empirical study.* European Journal of Operational Research, 1999. **116**(1): p. 139-155.
13. Thilander, M., *Some Observations of Operation and Maintenance in Two Swedish Firms.* Integrated Manufacturing Systems, 1992. **3**(2): p. 24-31.
14. Grout, J.J.R., *Mistake-proofing production.* Production and inventory management journal, 1997. **38**(3): p. 33-7.
15. Straker, D. *Poka yoke*. 2007 [cited 2007 August 16]; Available from: http://syque.com/improvement/Poka%20yoke.htm.
16. Shingō, S., *Zero quality control: source inspection and the poka-yoke system*. 1986, Cambridge, Mass.: Productivity Press. xxv, 303 s.
17. Savsar, M., *Effects of maintenance policies on the productivity of flexible manufacturing cells.* Omega, 2006. **34**(3): p. 274-282.

18. Ly, F., A.K.A. Toguyeni, and E. Craye, *Indirect predictive monitoring in flexible manufacturing systems.* Robotics and Computer-Integrated Manufacturing, 2000. **16**(5): p. 321-338.
19. Toguyeni, A. and O. Korbaa, *Indirect monitoring of the failures of a flexible manufacturing system under cyclic scheduling.* Robotics and Computer-Integrated Manufacturing, 2005. **21**(1): p. 1-10.
20. Hu, W., et al., *A systematic approach to integrated fault diagnosis of flexible manufacturing systems.* International Journal of Machine Tools and Manufacture, 2000. **40**(11): p. 1587-1602.
21. Hammer, M., *Reengineering Work: Don't Automate, Obliterate.* Harvard Business Review, 1990. **68**(4): p. 104-112.

Term paper 1

Zydrunas Vagnorius

An approach for analysing and improving manufacturing performance in aerospace industry

This article was written during the Phd course “Analysis, Modelling and Design in Logistics Systems” at the Norwegian University of Science and Technology. It is based on a real case study and contains some confidential information. Therefore some parts of the text have been changed or removed.

AN APPROACH FOR ANALYSING AND IMPROVING MANUFACTURING PERFORMANCE IN AEROSPACE INDUSTRY

Z. Vagnorius

Department of Production and Quality Engineering, Norwegian University of Science and Technology,
Trondheim, Norway

KEYWORDS: Aerospace industry, Manufacturing performance, Constraints.

ABSTRACT. Resent changes and trends in aerospace market are encouraging component manufacturers to expand production capacity. At the same time competition, record high fuel prices and other factors force to keep a close eye on costs. Several approaches, like Lean Manufacturing, automation or process optimisation employing mathematical methods provide support to this type of problem. However, these methods are not directly suited for the specifics of aerospace industry. Therefore this paper combines techniques from different approaches to develop an analysis and improvement method adopted for aerospace. The approach was applied in a real company and helped to study the manufacturing system, enabled to identify critical areas, and pointed towards viable solutions. A few of the solutions are presented in the article. In the concluding part possible extensions of the method beyond aerospace are discussed.

1 INTRODUCTION

Aviation is recovering after the crisis caused by the Gulf War and terrorist attacks of 11 September 2001. The International Air Transport Association reports considerable increase in passenger numbers [1, 2]. Freight demand, even though moderately, has also been growing over the last few years. These trends had a positive effect on the aerospace industry. Airbus claims having received a record number of firm orders in 2005 and having delivered more aircraft in 2006 than ever before [3]. Boeing reports having received over 1000 commercial airplane orders for the third consecutive year and having set a company record for total orders in 2007 [4]. Moreover it is estimated that due to aging around 40% of the planes flying today will retire in 20 years [5, 6]. This suggests that demand for new vehicles will continue to rise in the near future.

Growing demand for airplanes immediately transforms into new orders for components manufacturers and further members of the supply chain. To capitalize on this upturn, market players are expanding their capacity by investing in new manufacturing resources [7]. However, Crute et al. [8] state that global aerospace market already has over-capacity, as a result of which profits are declining. Other factors adding to the pressure in this industry are record high jet fuel prices [9] and plans of European Commission to include aviation into emission trading system [10, 11]. Therefore Korane [5] concludes that with production volumes ramping up aerospace manufactures are facing demands to hold the line on costs and improve on other aspects.

To summarise, two market requirements can be identified. The major focus for the players today is to increase the throughput. The second requirement is cost efficiency. Several authors compare this situation in aerospace to challenges in automotive industry a few decades ago. As discussed by Womak et al. [12], western car makers have faced a serious crises, as they struggled to compete against cheaper, but higher quality, Japanese products. This has forced manufacturers to rethink their strategies and focus on the wastes in their value chains. Crute et al. [8] claim that similar challenges will provoke Lean revolution in aerospace.

This opinion is questioned by James-Moore and Gibbons [13]. They have interviewed employees in several aerospace companies and concluded that the driving forces to adopt Lean practices were similar to automotive case. Types of products, though, were found to be significantly different. Aerospace companies typically make highly customised parts at irregular time intervals, which is a complete opposite of stable volume mass products in car industry. Therefore, according to James-Moore and Gibbons [13], in some areas aerospace might need modified or fundamentally different approaches. A contra argument from Crute et al. [8] is that lower volume production is even closer to Lean ideals. They show two examples, where application of Lean philosophy in aerospace has led to capacity gains, reduced scrap rates and other advantages.

Another allusion to automotive industry is automation. According to Webb and Eastwood [14] greater use of automation in aerospace is desirable due to a number of factors including cost, the requirement for greater flexibility, safety legislations and other. This view is paralleled by Korane [5], who suggest that current developments in the market are an excellent opportunity to benefit from advanced technologies. On the other hand, a manufacturing engineer from one aerospace company states that automation is more economically feasible in car business due to very high production rates [15]. Nevertheless, he admits that there are important things to learn from the experience of automotive industry.

Hackstein and Budenbender [16] also emphasise the importance of advanced manufacturing systems under current economic constrains. However, they note that real benefits of such systems are not clear. Therefore they propose a model to analyse operational behaviour of automated manufacturing systems. According to the model, production capacity is determined by technical and organisational downtimes. Authors test the model in two real systems. They demonstrate that automation can slightly increase the number of technical disturbances, but reduces the amount organisational issues and has an overall positive effect.

Vineyard and Meredith [17] suggest that as manufacturing systems get more sophisticated, their performance becomes highly dependant on maintenance. According to them this is especially profound in JIT environment, where unexpected stoppages can disrupt the flow of finished goods, delay shipments and add to intangible, but real, costs of customer good will. Therefore they employed statistical methods to determine the strategy that would minimise the number of different failures in a manufacturing system. A similar approach to the same problem was taken by Savsar [18]. He used a different set of assumptions and therefore came to completely opposite conclusions. For this reason no generalised conclusions can be drawn here.

A few more ways to study and improve manufacturing performance discussed in the literature include optimisation of process parameters [19], viewing production systems as socio technical entities [20], or complete reengineering of processes [21]. The latter approaches, as the ones discussed before, have arguments both supporting and discouraging their use in particular cases. Nevertheless all of them provide useful insights. Therefore in the following sections different ideas will be combined to develop a method for analysis and improvement of a real aerospace manufacturing system.

The remaining part of the paper is structured as follows. First, the underlying approach is chosen and analysis technique is developed. Next, company under study is introduced. After that, data collection process is described. Gathered data is plugged into analysis model to study the production system. Based on the analysis, possible improvement areas are identified. This is followed by a discussion of a few viable solutions. The concluding part is a reflection of the work, giving the main findings and discussing possible application of the model for analysis of other manufacturing systems.

2 ANALYSIS AND IMPROVEMENT APPROACH

Aerospace market demands simultaneous increase in productivity and cost efficiency. This echoes the advantages offered by Lean Manufacturing [12]. However, the author of this article agrees with James-Moore and Gibbons [13] that while some principles, for example waste elimination, and tools, such as mapping techniques, are definitely applicable to aerospace, all aspects of the philosophy cannot be adopted here. The main problem is the difficulty to achieve continuous flow, which is one of the milestones of Lean Manufacturing. The reasons for this are:

- Irregular production rates.
- Bulkiness of equipment, which makes layout optimisation difficult and expensive.
- Strict regulations, which limit the possibilities to change operations and balance the lines.
- Geographical spread of suppliers, producers and customers, which makes piecemeal deliveries impossible.

Another approach suggested in the literature was automation. Full scale solutions, such as flexible manufacturing systems (FMS), have been reported to have a positive effect on both cost and productivity [22-24]. This results from better coordination and improved shop floor logistics. The factor limiting implementation of FMS are high investment costs [16, 22]. Smaller scale automation solutions are cheaper and desirable. However, their primary merit is cost reduction and not necessarily productivity improvement [25]. After all automation needs to be supported by analytical tools, which help to identify critical areas to be addressed.

Previously mentioned attempts to optimize the performance by employing mathematical methods [17, 18] gave contradictory results. This suggests that full scale manufacturing systems might be too complex for purely mathematical analysis. Therefore mathematical methods must be based on experimental data and used in combination with qualitative tools.

The conclusion to be made here is that available methods are of limited use in aerospace industry. Therefore there is a need to develop an approach which is better suited for the specifics of this sector. The author of this article believes that current situation in aerospace industry looks somewhat similar to the case described in Goldratt’s “Goal” [26]. The hypothetical factory in the book has sufficient demand, has installed new resources, but still can not achieve the desired performance. Goldratt [26] argues that the reason for this is existence of bottlenecks in the system. For illustrative purposes consider a simple system in Figure 1.

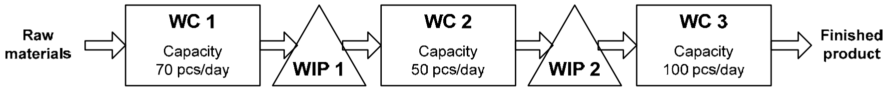


FIGURE 1. Production line with an internal constraint.

Let's assume that there are enough raw materials and sufficient market demand. Then work centre WC1 can produce 70 pieces per day. However WC2 can process only 50 parts. The remaining 20 units accumulate as work in process inventory (WIP). The capacity of WC3 is 100 units, but only 50 pieces are supplied by WC2. Thus the throughput of the whole line is 50 parts, i.e. the capacity of work centre WC2. WC2 is the bottleneck of the system. Therefore its throughput must be increased to raise the output of the line. The production rate of other resources needs to be adjusted (subordinated) to the bottleneck to avoid inventory build-up [26].

This approach has several advantages:

- It enables to identify critical resources.
- Management (elevation) of bottlenecks improves the throughput of the system.
- Subordination of other resources to the bottleneck reduces inventory carrying costs.

The disadvantage is the assumption of infinite supply and demand. This drawback is overcome by extending the concept of the bottleneck. Goldratt [27] says that constraint is any internal (e.g. insufficient capacity) or external (e.g. saturated markets, lack of raw materials) factor that limits the throughput. Any system has one or more constrains. Otherwise throughput and profit would go to infinity. Thus, performance of the system can only be improved by identifying and elevating of constraints. These are the basic ideas of the Theory of Constraints (TOC). The author of this paper thinks that the principles of TOC can be used in response to the issues of aerospace industry. For this reason TOC will be the underlying idea of the approach discussed next.

The methodology that was developed is illustrated by Figure 2. The author suggests to start drilling down to constraints by performing product family analysis. The reasons for that are:

- Individual products, especially in aerospace manufacturing, might be made in relatively low volumes. Therefore focussing on them might not give considerable improvement.
- Not all products are of equal importance for the customer or profitability of the company.

Carrying out product family analysis allows arriving at bigger groups of similar items. This requires information about product range (product codes), routing tables and sales data of main items. For more details see, for example, Bicheno [28].

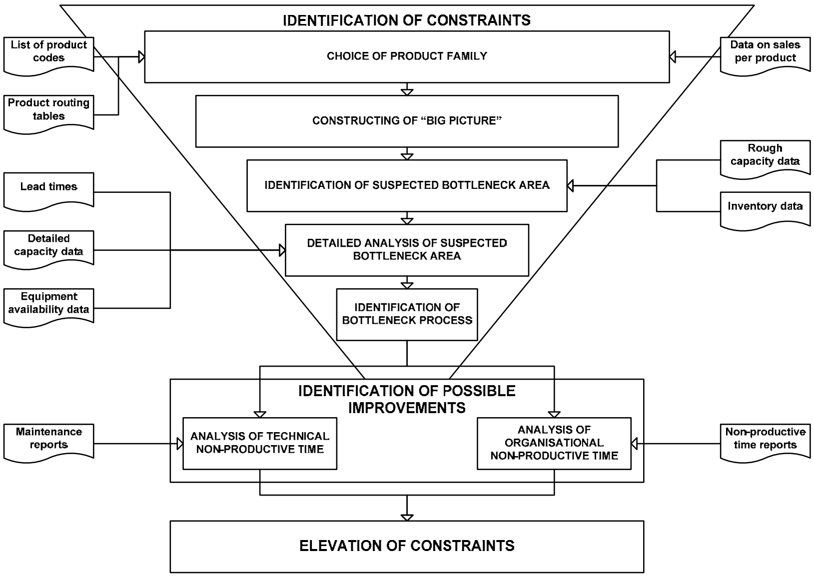


FIGURE 2. Analysis approach.

Next step is drawing the “Big Picture” of the system. Experience of the author of this paper shows that different people within an organisation can have different understanding of what is critical for the manufacturing performance. Thus, starting with a high level study of the whole system allows avoiding bias from the very beginning. This can be accomplished by following the production flow of a chosen product, and drawing symbols for the main steps. Any flow chart symbols are suitable here.

Keeping the “Big Picture” in mind the process needs to be screened for possible constraints. Goldratt [26] suggest that an area with the biggest amount of WIP is a “good” candidate to be a bottleneck. Another way is to perform a rough capacity analysis. Data for it can usually be extracted from ERP systems. An area with the lowest output rate can also be suspected to be a constraint.

“Candidate” area/areas have to be studied in more detail. For each work centre the following information needs to be obtained:

- Lead time (LT). The most reliable way to determine the lead time is to perform a time study. This is because information in company database can be outdated or represent the best possible scenario.
- Full capacity can be calculated as the theoretical total number of hours at each work centre.
- Activated capacity. Due to different amount of work at different operations, resources might work on different schedules during the day. Thus, activated capacity is the total number of hours planned for a work centre per day.
- Availability (technical) can be determined as:

  $$Availability = \frac{Total\ PT - (Total\ DT_{corrective} + Total\ DT_{preventive})}{Total\ PT} \times 100\%$$

  In the above formula PT stands for planned time and DT for down time. Some researchers assign preventive maintenance to organisational downtimes [16]. However earlier work of the author of this paper has shown that in practice the distinction between corrective and preventive repairs in not always clear [29]. Thus it is more reasonable to look at availability as a percentage of time available for production after all kinds of repairs.
- Effective capacity is determined by multiplying activated capacity by availability. Note that resources are not likely to fail when they are not active, so using of full capacity here is not reasonable.
- Production rate can be calculated by dividing the sales volume of all products in the chosen family over the number of working days. The time periods must be the same.
- Capacity requirements are determined by multiplying the production rate by lead time at particular work centre.
- Utilization of effective, activated and full capacity can be calculated by dividing capacity requirements over the corresponding capacity and multiplying it by 100%.

More details on the above calculations can be found in Rother and Shook [30] and Hoop and Spearman [31].

The main input for further analysis is utilization of effective capacity. It shows how heavily a resource is loaded during up time (when it is not failed). A resource with the highest utilization of effective capacity is critical for the system. Following the logic of TOC, such resource must be the focus of further work if the throughput is to be increased.

To searched for possible wasted activities in critical work centre the author of this paper proposes to use the model developed by Hackstein and Budenbender [16]. As mentioned earlier they suggest that utilization is determined by technical downturns and organizational non-productive time.

Extent of technical problems can be estimated from availability values. If availabilities are low this means that there are a big number of technical failures. In such case improvements activities should focus on maintenance. Analysis of maintenance reports can reveal the main types and sources of problems.

If availability is not an issue, but utilization of effective capacity is low, then problems are most likely of organizational nature. Examples of organisational issues are lack of raw materials, cleaning and other [16]. Identification of these problems can be complicated. In some companies employees are asked to report, what caused their non-productive time. Such reports can be biased, but can still provide useful information.

The final step of the approach is elevation of constraints. Proposed solutions must be based on technical and organisational problems in a particular company.

3 CASE STUDY

In this section the developed method is applied to a real manufacturing system. The steps from the Choice of Product Family through Identification of Suspected Bottleneck Area were performed in direct cooperation with the company. Later analysis is mainly the work of the author of this paper.

3.1 COMPANY DESCRIPTION

The company under study is a European manufacturer of jet engine parts, including:

- Low pressure turbine cases (hereafter turbine cases)
- Exhaust cases
- Jet engine shafts

Main customers are suppliers of engines for civil aircraft builders (Boeing, Airbus). Smaller part of the business is manufacturing of components for military planes. In recent years company's production volumes have been steadily rising [32]. To expand the capacity investment in several new machining systems was done. However, the throughput of the system is still said to be below the market demand [7]. At the same time company faces a challenge to reduce cost in order to stay competitive [33].

3.2 DATA COLLECTION

Data necessary to test the approach was collected at different time moments (on as required bases). This process is described next.

- Data on the range of turbine cases (see §3.3 Choice of Product Family) and sales of main models during a period of 22 months was received by courtesy of company's sales department.
- Routing tables of turbine cases were received by courtesy of Process Manager.
- No information about inventories was supplied for analysis.
- Capacity data (detailed). The number of machines, operators and planned working hours in each work centre was recorded during factory tour (courtesy of Process Manager).
- Lead time. The initial idea was to perform a time study and collect primary information about lead times. However, due to existing agreements with local labour union, the author of this article was not allowed to carry out any measurements. Therefore lead times used in later analysis are based on statistical data extracted from company's database (courtesy of Process Manager). They represent long term average time to complete an operation and include setup time, cutting time, tool changeovers and measurements. Tool changeovers are dictated by the wear and are unavoidable part of the cutting process. Measurements need to be done while part is still fixed in the machine, so that minor corrections can be done immediately. The reason why setups are also included is that each part needs a setup at every machine (batch size is one through entire production process), so all setup time is allocated to one piece.
- Maintenance reports for a period of 22 months were received by courtesy of Maintenance Manager. These reports are prepared manually and show: equipment failures, preventive repairs, descriptions of failures and repairs, and associated repair times. The author of this paper used this information to calculate availability of resources and analyse technical problems.
- "Spindle-on" reports for a period of 22 months were received by courtesy of Maintenance Manager. These reports show the time that machine spindle was active (rotating). Data is regularly downloaded (by maintenance engineers) from machine tool controllers. The author of this paper used this information to support the analysis of both technical and organisational problems.
- Non-productive time reports for a period of 22 months were received by courtesy of Planning Manager. These reports showed the reasons and total number of man-hours lost in different cost centres. They were used for analysis of organisational non-productive times. It must be noted that downtimes reported in the reports are related to, but not the same as, machine time losses. For example, if a particular machine is not running due to lack of raw materials, but its operator can do productive work on another work station, then downtime of the first machine is not reflected in the reports. For this reason quantitative data given in the reports was not used.

3.3 CHOICE OF PRODUCT FAMILY

Turbine cases are company's best selling products. Therefore after informal interview with company representatives this group was chosen for further analysis. Turbine cases are made in a few versions: one main model and several low volume options. Routing for all of them is the same. Minor customisations are handled by modifying the control software. Due to these differences cycle times are slightly different too. However, according to the sales data, the main model constituted nearly 95% of the total production volume over the last two years. Thus its cycle times were used throughout the analysis.

3.4 BIG PICTURE

Figure 3 below shows a simplified representation of the production process (drawn with the help of Process Manager). The starting point is the reception of raw material from suppliers. Parts are physically stored in a dedicated section of the central warehouse. Raw materials are then delivered to machining cell where the main job on the part is done. Resources here are flexible turning and machining centres and are dedicated to turbine cases. The remaining operations are mainly supporting and make use of men and general purpose equipment shared between several product families.

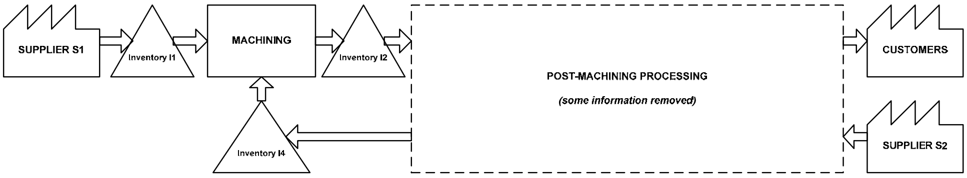


FIGURE 3. Production process of a turbine case.

After machining, turbine cases undergo post-machining processing.

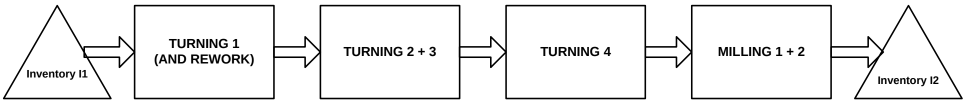
The text describing the post-machining part of the process has been removed.
As will be seen in the next section it was not essential for further analysis.

3.5 ANALYSIS OF MACHINING PROCESSES

After having studied the manufacturing process a discussion was held to identify the areas limiting the throughput of the system. The possibility that market is a constraint was immediately excluded, as customer demand is beyond current system capacity. The other end of the channel – material supply, was mentioned as one source of problems. No data, however, was available to continue the analysis here. A more serious concern was said to be the machining section. Despite resent investments the output of this section is lower than expected, suggesting that it could be the bottleneck of the whole line. Observed amounts of inventories also support this conclusion (as mentioned in §2, detailed data on inventories was not available). For this reason the decision was taken to perform a more detailed analysis in machining section.

As shown in Figure 4, machining consists of turning and milling operations. Some turning operations, as well as some milling operations share machines, so they are represented by a single box and will be analysed as one work centre. First work centre also performs rework on as required bases. The frequency, amount and duration of these corrective jobs are very variable. Moreover, this information is of restricted use. Therefore rework was not considered in later analysis.

As mentioned before no data on inventories was available, so only the incoming and outgoing material storage locations are depicted in the figure. Work in progress inventory (WIP) is not shown.



Number of machines = *	Number of machines = *	Number of machines = *	Number of machines = *
Number of operators = *	Number of operators = *	Number of operators = *	Number of operators = *
LT = * h	LT = * h	LT = * h	LT = * h
Full capacity = * h/day	Full capacity = * h/day	Full capacity = * h/day	Full capacity = * h/day
Activ.capacity = * h/day	Activ.capacity = * h/day	Activ.capacity = * h/day	Activ.capacity = * h/day
Availability = 99.2%	Availability = 94.3%	Availability = 96.6%	Availability = 92.0%
Effect.capacity = * h/day	Effect.capacity = * h/day	Effect.capacity = * h/day	Effect.capacity = * h/day
Prod. rate = * pc/day	Prod. rate = * pc/day	Prod. rate = * pc/day	Prod. rate = * pc/day
Capacity req. = * h/day	Capacity req. = * h/day	Capacity req. = * h/day	Capacity req. = * h/day
Utilization of ef.c.= 25.7%	Utilization of ef.c.= 70.9%	Utilization of ef.c.= 48.6%	Utilization of ef.c.= 39.1%
Utilization of act.c. = *%	Utilization of act.c. = *%	Utilization of act.c. = *%	Utilization of act.c. = *%
Utilization of full c.= *%	Utilization of full c.= *%	Utilization of full c.= *%	Utilization of full c.= *%

Note. Calculations were performed following the procedure described in §2 Analysis and Improvement Approach (data marked with * has been removed).

FIGURE 4. Machining processes.

3.6 IDENTIFICATION OF BOTTLENECK PROCESS

Table 1 summarises the main results of the calculations. With nearly 71% utilisation of effective capacity turning 2 + 3 is the busiest work centre. This makes it the constraint of the machining section and the whole production line.

TABLE 1. Utilization of work centres

WORK CENTRE	UTILISATION OF EFFECTIVE CAPACITY
Turning 2 + 3	70.9%
Turning 4	48.6%
Milling 1 + 2	39.1%
Turning 1	25.7%

The following comments need to be added to the above conclusion. On one hand 71% is quite high utilization, especially taking into account that market demand continues to increase. On the other hand XX hours every day (see Figure 4) are not even activated. The latter point looks strange, because milling 1 + 2 is run full three shifts even though it does not appear to be a critical work centre. Discussion of this issue with Process Manager revealed that milling operations used to be the biggest problem, mainly due to frequent failures. Our earlier study on reliability of machine tools [29] confirms that milling machines have highest rates of occurrence of failures (ROCOF). Moreover, as will be discussed later, real availability and effective capacity of machines is probably lower due to waiting time for service. Because milling operations have highest ROCOF, it is likely that the effect of this waiting is also most substantial here. In that case milling 1 + 2 might become the true capacity constrain of the system. Unfortunately, there was no data to confirm this hypothesis.

3.7 NON-PRODUCTIVE TIME ANALYSIS

To search for any wasted potential in the identified critical work centre, the author of this article studied the use of resources belonging to it. Figure 5 shows an example of the best machine. The time of interest here is spindle utilization. It shows the proportion of time during which equipment was cutting metal, i.e. adding value to the product. In this example average utilisation is around 40%. Later on it was found out that spindle normally rotates when tool is approaching the workpiece, as well as during some tool measurements. These are clearly non-value adding activities, thus the true value adding time is lower.

TECHNICAL NON-PRODUCTIVE TIME. From Figure 5 the effect of equipment failures and preventive repairs appears to be insignificant. This observation is paralleled by earlier results (recall Figure 4). Availability of turning 2 + 3, as well as other work centres, is fairly high and looks too optimistic even for aerospace industry. More thorough consideration of this question revealed that maintenance reports, that were the bases for availability analysis, did not include waiting time for maintenance assistance. No statistics is currently available about these waiting times. The company admitted that this was a serious problem. If waiting times were substantial, availability, and therefore effective capacity, of machines could be considerably lower.

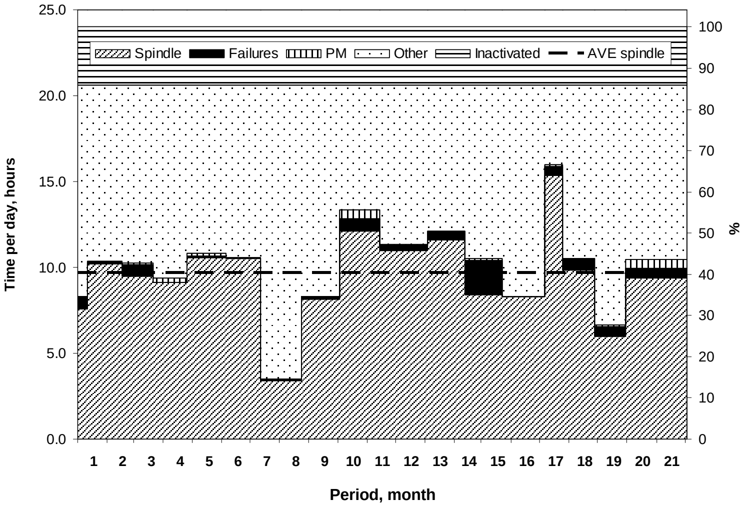


FIGURE 5. Distribution of manufacturing time.

Part of the remaining time loses could be attributed to setups, tool changeovers and measurements. These activities were included in the lead time (recall the discussion about data collection in §3.2), but none are productive. Thus, even though setups, tool changeovers or measurements cannot be completely eliminated, the time associated with them should be reduced as much as possible.

ORGANISATIONAL PROBLEMS. While various setups and measurements are obvious causes for non-productive time, they can not explain all remaining losses. Following the research of Hackstein and Budenbender [16], most of the wastes labelled “Other” in Figure 5 should be linked to organisational problems. To see what kind of problems these could be, the author has analysed non-productive time reports. The following main categories of non-productive time were identified:

- Training
- Miscellaneous side tasks
- Waiting for materials
- Training provoked by material shortages
- Waiting for tools

Analysis showed that the most significant time “losses” were due to training. This can be explained by expanding production and a big number of new operators needed to run the equipment. In fact, it is increasingly challenging to find people with required qualification (according to statistical data, unemployment in the country, where the company is based, stood at 2.4% in March 2008 [34]). As a consequence more time needs to be spent for training of operators before they can do their job.

Miscellaneous side tasks mean that operators were given jobs different from their main responsibility. While all machines are automatic to some extent, in most cases operator’s presence is required to monitor the process. Thus if an operator is busy with another task, in most cases machine is not running.

Waiting for materials was the third biggest category of time losses. To it could be added training provoked by material shortages. These problems are directly related, but not limited, to raw material delivery troubles. Currently there are very few raw material suppliers in the world. Under growing demand for aircraft components they are struggling to supply enough raw materials required to satisfy this demand. In addition raw material received from some of the suppliers has been observed to behave very differently compared to raw parts from other sources. These variations in material behaviour are causing serious quality problems and are said to be the main reason for the discussions with customers. Material shortages could also result from logistical problems in the factory itself.

Waiting for tools could be linked to lack of capacity in the tool presetting room. On the other hand it could be due to too frequent tool replacements, or internal logistics problems.

4 ELEVATION OF CONSTRAINTS

In the previous section the analysis model was used to identify problems in the production system. This section discussed how these issues could be handled in order to improve the productivity. The main focus is on turning 2 + 3. However, it is important to remember that as this constrain gets elevated, other resources might become critical [26]. Therefore some of the improvement ideas will be more general and applicable to all work centres. It must also be mention that part of the solutions discussed next were developed by the company and are already adopted in production. These will be presented as practical examples.

Calculations showed that turning 2 + 3 is the capacity constraint of the system. However it is not a real bottleneck, because there is still room to increase its throughput. The easiest way to do that is activating of the remaining capacity. The reason, why it has not been done yet is lack of man power. On the other hand, milling 1 + 2 is run full three shifts even though it does not appear to be a critical work centre. Thus, if additional men power cannot be found it might be smart to reallocate the time of available operators from non-critical machines to the critical ones.

Another way to improve productivity of the system is to raise reliability of critical resources. This would have several positive effects, including improved availability, hence increased throughput, less time lost for rework and better product quality. On the one hand, availability of resources is already high, so it might not be easy to raise it higher. Moreover, preventive maintenance was included in availability calculations. Reduction of this time is hardly possible and could have adverse consequences. On the other hand, waiting time for maintenance assistance is manageable. The first thing to do is to start continuously observing time spent for waiting. Today it is believed that the main cause for delayed repairs is insufficient maintenance capacity, which is again related to availability of qualified engineers in the labour market. Even so, it is possible to review the priorities of actual maintenance engineers making it clear that whenever possible critical machines should be serviced first, while other ones should be repaired during the remaining time.

Next improvement typically recommended to save bottleneck time is setup reduction. Today every part needs a setup at the beginning of each operation. This includes delicate positioning of a turbine case in the fixture, clamping and attachment of part-fixture unit to a specific pallet before it can be sent to machine. The need for this last setup operation arises because of different pallet systems employed by different machine tool manufacturers. To reduce time losses here, company has developed a new fixture with an adaptation layer. This solves the problem of different pallet systems and eliminates manual attachment of fixture to a pallet, saving a lot of time. Yet another advantage is simplified measurements.

As mentioned before, measurements used to be done while a part was still fixed in the machine, to make it possible to do required correction immediately. Using the new design, fixture can be removed from the machine in a short time. Machine can start working on a new part, while measurements are performed in parallel. If measurements show that minor corrections are needed, part can be sent back to the machine without loosing too much time.

Besides setups and measurements, other time losses are related to tool changeovers. Traditional way to determine tool life is by using Taylor’s Law [19]. Detailed discussion of this approach is out of the scope of the article, therefore only points important for understanding of implication on productivity will be mentioned. Taylor’s equation uses two empirical constants to determine tool life under given cutting speed. Since it is not practical to perform big number of experiments constants are generalized to remain valid for wider range of applications. For this reason they tend to be conservative, i.e. tool life determined by Taylor’s formula is shorter than the real one. Moreover, aircraft industry itself is known to be conservative, so manufacturers add their own safety margin. As a result typical time between tool replacements in manufacturing of aircraft components is only four to five minutes. At the same time changeover, which involves tool measurement, is a machine constant and takes around two minutes. The consequence of such conservativeness is a lot of productive time wasted for changeovers.

The straightforward solution to this problem would be to extend the tool life and free some of the capacity. The concern, though, is that such approach would have negative effect on process and product reliability. Thus, to choose optimal tool life, productivity gains due to extended tool life should be traded against the increased risk of failures. We are quite convinced that such approach would extend the tool life and have a positive effect on the productivity.

The solutions discussed so far were directed towards technical issues. Further ones will address organisational troubles. One such issue was the dependence of the manufacturing system on manpower. To walk round these problems company is investing in unattended manufacturing. The concept has been proven by automating a few processes in other manufacturing cells within the factory. However, there is still a lot of work to be done. As an example issues of process monitoring and tool delivery are discussed.

Process monitoring is one of the main functions of an operator today. His/her responsibility is to observe the production process and take necessary actions if something goes wrong. This is a subtle work and needs highly developed skills. Unfortunately, due to complicated situation in labour market, skilled operators are difficult to find. Thus the challenge in solving this problem is to develop an automated system, which would replace operator's eyes, ears and nose, so that production process could be monitored effectively.

Tool delivery problems were previously mentioned to be one of the reasons for organisational downtimes. Industrial tool delivery systems already exist. Their feasibility was studied in the company. On one hand, study has shown that people can do the job more efficiently than automated systems [32]. Unfortunately this only true provided that operators are available. A solution could be a hybrid system, where tool delivery would be automated, but could be switched to manual mode if necessary.

The last organisational problem to be discussed is raw material shortages. It was found out that these were mainly due to lack of suppliers. In addition some of the existing suppliers were facing problems to supply raw parts with consistent properties. The problem is believed to be technological process. To improve the situation, suppliers need to be assisted in process development, so that it results in more uniform raw material behaviour.

5 CONCLUDING REMARKS

In this paper some of the latest developments in aerospace industry were discussed. The general trend was growing demand for both passenger and cargo aircrafts, which immediately transformed into increasing demand for aircraft components. This trend emphasised the need to improve productivity. At the same time competition, rocketing fuel prices and legal regulations were forcing to have a tight control over costs. To help to respond to these conflicting demands this paper has proposed an analysis and improvement approach.

The method uses theory of constrains as an underlying idea, but combines it with mapping and performance evaluation tools offered by other analysis techniques. The approach was applied to a real manufacturing system in aerospace industry and had the following benefits:

- It allowed to construct the “Big Picture” of the system, thus helped to avoid sub optimisations.
- It enabled to identify critical resources.
- It pointed towards possible solutions.

Moreover the method proved to be useful as a learning tool. Both during the high level and detailed process analysis important facts about the system were uncovered.

On the other hand the fact that the approach was developed to address the specific manufacturing problems of one industry sector can also be seen as its drawback. However, the author thinks that main ideas, such as the principle of constrained resources, are universal. Therefore the approach and some of the suggested improvements could be equally applicable to other types of industries. For instance, the problem of setups is an important issue in all manufacturing companies. The solution that was discussed is an example of “conversion” of internal setups to external ones and this way saving of valuable production time.

In some cases, of course, extending the use of the approach might require some modifications. For example, in simpler systems two steps to dig to core problems might not be needed. In other companies it might be desirable to penetrate even deeper, for example operation level. Then the procedure could be continued by splitting critical operations into subroutines and identifying the “slow mowers”.

Another possible weakness of the method is big amount of required data. The study has showed that some of these data might be sensitive and therefore difficult to get (e.g. lead times or data on inventories). Other data might not be accurate enough (e.g. maintenance or non-productive time reports). However, it was discussed that models not based on real data can give confusing recommendations. Thus despite the difficulties, data needs to be collected if reliable analysis is to be performed.

As a concluding remark the author must repeat a well known truth, that self analysis and improvement process must be continuous. Once solutions are implemented the approach should be reapplied to check for new “candidates” to be critical resources. If this work is done after significantly long time, during which important changes might have taken place, the whole procedure, starting with the product family analysis might need to be repeated.

REFERENCES

1. IATA, (2006), Annual Report 2006, http://www.iata.org/iata/Sites/agm/file/2006/file/annual_report_06.pdf, Accessed 21.03.2007.
2. IATA, (2007), Annual Report 2007, http://www.iata.org/NR/rdonlyres/9CE110BA-B989-448F-9991-67DF69397A6D/0/ar2007.pdf, Accessed 13.11.2007.
3. Airbus, (2006), Annual Review 2006, http://www.airbus.com/store/mm_repository/pdf/att00008917/media_object_file_Airbus2006AnnualReview.pdf, Accessed 13.11.2007.
4. Boeing, (2008), Boeing Sets Third Consecutive Record for Commercial Airplane Orders in 2007, http://www.boeing.com/commercial/news/2008/q1/080103d_nr.html, Accessed 25.06.2008.
5. Korane, K. J., (2005), Trends in Aerospace Machining, http://machinedesign.com/ContentItem/59982/TrendsInAerospaceMachining.aspx, Accessed 20.03.2007.
6. Boeing, (2007), Current Market Outlook 2007. New Airplanes., http://www.boeing.com/commercial/cmo/new.html, Accessed 13.11.2007.
7. Torjusen, J. E., (2007), Personal Communication in Volvo Aero Norge AS, Kongsberg.
8. Crute, V., Ward, Y., Brown, S., Graves, A., (2003), Implementing Lean in Aerospace--Challenging the Assumptions and Understanding the Challenges, Technovation, Vol 23, 917-928.
9. IATA, (2008), Jet Fuel Price Development, http://www.iata.org/whatwedo/economics/fuel_monitor/price_development.htm, Accessed 25.06.08.
10. Commission, E., (2006), Bulletin Eu 12-2006. Environment (11/15). http://europa.eu/bulletin/en/200612/p121011.htm, Accessed 25.05.08.
11. Tiefensee, W., (2007), Member States Agree on Including Aviation in Emissions Trading, http://www.bmvbs.de/-,2969.998494/Tiefensee-Member-States-agree-.htm, Accessed 25.06.08.
12. Womack, J. P., Jones, D. T., Roos, D., (1990), The Machine That Changed the World, Rawson Associates, New York.
13. James-Moore, S. M., Gibbons, A., (1997), Is Lean Manufacture Universally Relevant? An Investigative Methodology, International journal of operations & production management, Vol 17, 899.
14. Webb, P., Eastwood, S. J., (2004), An Evaluation of the Ti2 Manufacturing System for the Machining of Airframe Subassemblies, Proceedings of the I MECH E Part B Journal of Engineering Manufacture, Vol 218, 819-826.
15. Waurzyniak, P. P., (2006), Modular Automation for the Aerospace Industry, Manufacturing Engineering, Vol 136, 81-90.
16. Hackstein, R., Budenbender, W., (1991), Flexible Manufacturing Systems in Practice -- Results of a Study of Operational Use, Engineering Costs and Production Economics, Vol 21, 259-276.
17. Vineyard, M. L., Meredith, J. R., (1992), Effect of Maintenance Policies on Fms Failures, International Journal of Production Research, Vol 30, 2647-2657.
18. Savsar, M., (2006), Effects of Maintenance Policies on the Productivity of Flexible Manufacturing Cells, Omega, Vol 34, 274-282.
19. Boothroyd, G., Knight, W. A., (2006), Fundamentals of Machining and Machine Tools, Taylor & Francis, Boca Raton.
20. Thilander, M., (1992), Some Observations of Operation and Maintenance in Two Swedish Firms, Integrated Manufacturing Systems, Vol 3, 24-31.
21. Hammer, M., (1990), Reengineering Work: Don't Automate, Obliterate, Harvard Business Review, Vol 68, 104-112.
22. Luggen, W. W., (1991), Flexible Manufacturing Cells and Systems, Prentice-Hall, Englewood Cliffs, NJ.
23. Wiendahl, H.-P., Springer, G., (1988), International journal of advanced manufacturing technology, Vol 3, 55-69.
24. Meredith, J., (1988), Installation of Flexible Manufacturing Systems Teaches Management Lessons in Integration, Labor, Costs, Benefits, Industrial Engineering, Vol 20, 18-27.
25. Torjusen, J. E., (2007), Personal Communication 2 in Volvo Aero Norge AS, Kongsberg.
26. Goldratt, E. M., Cox, J., (1989), The Goal, Revised edn, Gower Publishing Ltd, Aldershot.
27. Goldratt, E. M., (1990), What Is This Thing Called Theory of Constraints and How Should It Be Implemented?, North River Press, Croton-on-Hudson, N.Y.
28. Bicheno, J., (2004), The New Lean Toolbox: Towards Fast, Flexible Flow, PICSIE Books, Buckingham.

29. Vagnorius, Z., Sørby, K., (2008), Reliability of Machine Tool Systems in Aircraft Industry, Paper presented at the Proceedings of 8th International Conference on Advanced Manufacturing Systems and Technology, Udine, Italy.

30. Rother, M. M., (2003), Learning to See : Value Stream Mapping to Create Value and Eliminate Muda / by Mike Rother and John Shook ; Foreword by Jim Womack and Dan Jones.

31. Hopp, W. J. W. J., (1996), Factory Physics : Foundations of Manufacturing Management / Wallace J. Hopp, Mark L. Spearman.

32. Dr. Torjusen, J. E., (2007), Personal Communication in Volvo Aero Norge AS, Kongsberg.

33. AS, V. A. N., (2005), Annul Report 2004, Volvo Aero Norge As, http://www.volvo.com/NR/rdonlyres/24462F17-BB6E-4B4A-B12C-836BE8A7D358/0/VAN2004engelsk.pdf, Accessed 20.03.2007.

34. Norway, S., (2007), Labour Force Survey. Seasonally Adjusted Series, December 2006, http://www.ssb.no/english/subjects/06/01/akumnd_en/, Accessed 20.03.2007.

Term paper 2

Zydrunas Vagnorius[*] and Knut Sørby[†]

Effect of high pressure cooling on lifetime of SiALON-based ceramic cutting tools – Preliminary study

Writing of this article was a part of the final exam at the SFI Norman PhD School organized by the Norwegian University of Science and Technology

---

[*] Zydrunas Vagnorius performed data analysis and wrote the entire text. He also ran the machining experiments, during which the data for evaluating the effect of the high pressure cooling on the life of ceramic tools was collected, and performed tool wear measurements.

[†] Professor Knut Sørby supervised the machining experiments. He also contributed to the writing process by coming up with comments that helped to improve the readability and the overall quality of the text.

Effect of High Pressure Cooling on Lifetime of SiALON-based Ceramic Cutting Tools – Preliminary Study

Z. Vagnorius* and K. Sørby
Norwegian University of Science and Technology, Department of Production and Quality Engineering,
S. P. Andersens v. 5, NO-7491 Trondheim, Norway
*zydrunas.vagnorius@ntnu.no

Abstract

The properties of heat resistant alloys used in aerospace make them difficult to machine. A particular problem is high temperature generated in the cutting zone, which leads to short tool lives and low productivity. One of the solutions to the problem is to use cutting tool materials, such as ceramics, which can tolerate higher temperature. Additional improvements can be achieved through more efficient heat extraction from the cutting zone. High pressure cooling has proven to be very effective for this purpose. Therefore it was tested in machining of a nickel based aerospace material Inconel 718. The results of the study have showed, on one hand, that this technique has a negative effect on the lifetime of SiAlON-based tools. On the other hand a very efficient chip breaking was achieved. Moreover, a potential to improve the tool life and productivity was identified and is discussed in this paper.

1 INTRODUCTION

Machining of heat resistant aerospace materials such as Inconel 718 or Inconel 901 is characterised by low cutting speeds and therefore low productivity and high production costs. The main reasons for that are high hardness and low thermal conductivity of these nickel-based alloys. As a result very high temperature is generated in the cutting zone. Narutaki et al.[1] and Kitagawa et al. [2] have experimentally shown that, in turning of Inconel 718 under conventional cooling, temperature on the rake face of ceramic tools reaches 900°C at the cutting speed of just 30 m/min and climbs to as high as 1300°C at 300 m/min. Such high temperature can have a considerable negative effect on the dimensional accuracy and surface quality of produced parts. Moreover temperature is an important factor governing the tool life. As it increases tool materials soften, thus they can be easier eroded by abrasion. In addition, high temperature promotes diffusion wear and can cause thermal shocks and fatigue. Therefore, to achieve a reasonable tool life and avoid other problems caused by high temperature, heat resistant alloys are often machined at cutting speeds as low as 30 to 100 m/min [2].

One way to raise the efficiency in machining of these materials is to use more advanced cutting tools. Plain and coated carbide inserts are the industry standard today. In machining of nickel base alloys, though, materials such as ceramics perform much better [3]. Ceramics are very hard and have good abrasion resistance. Their melting point is very high, thus hardness is retained even at high temperatures. This for ceramic tools can be used at higher cutting speeds. Moreover, ceramics are chemically more stable than carbides, thus the tendency to form built-up edge is reduced.

Despite the improvements achieved with ceramic tools, productivity in machining of aerospace materials is still relatively low. Therefore additional measures need to be taken to deal with high temperatures. Traditionally copious quantities of fluids have been poured onto the tool to extract the heat. This technique is quite effective in machining of steels and other materials, but, as shown by Kitagawa et al. [2], has a negligible effect in cutting of aerospace alloys. The issue is that at the range of temperatures developed in machining of these materials coolants are rapidly evaporated. As a result a steam “blanket” is created, which stops the fresh coolant from reaching the tool-chip interface thus rendering conventional flushing ineffective. A few alternative techniques, including internal chilling of the insert, cryogenic and CO_2 cooling, or high pressure cooling, have been tested. The latter method seems to be particularly promising.

The principle behind the high pressure cooling is to supply the cutting fluid in the form of a small jet. Already the early experiments showed that, in rough turning of aircraft exhaust valves made in inconel, this method could increase the output per tool grind by over 18 times [4]. Similar results were obtained in machining of steels. The reason behind these impressive improvements is the ability of the high pressure jet to penetrate deeper into the interfaces between the tool and the chip or the workpiece. This way cooling takes place closer to the highest temperature zone and is therefore more efficient. Moreover, high pressure cooling has been reported to enhance chip control, reduce cutting forces, eliminate the tendency to form built-up edge, improve dimensional accuracy and surface integrity.

The impressive results achieved in using high pressure cooling in various applications encourage us to try it in machining of aerospace materials with ceramics. A concern, though, is the sensitivity of these materials to thermal stresses. Due to this weakness it is often recommended to use no or very small quantities of coolant when machining with ceramic tools [5]. However, due to extreme temperatures generated in cutting of aerospace alloys, one or another way to dissipate heat has to be used. Moreover, some ceramics, such as alumina reinforced with SiC whiskers or SiAlON have improved resistance to thermal shocks [6, 7]. It is therefore interesting to investigate whether the performance of ceramic tools, hence the productivity in machining of heat resistant materials, could be further improved by using high pressure cooling. A few attempts to apply this technique when machining with SiC-whiskers reinforced tools have been reported and will be reviewed in later sections. In the present study we will focus on SiAlON. Thus, machining of Inconel 718 with tools based on this material will be performed under high pressure cooling.

The results of our experiments are presented, discussed and summarised in Sections 4, 5 and 6 respectively. In addition an overview of main aspects of high pressure cooling is provided in Section 2, and in Section 3 a few earlier attempts to apply this technique when machining with ceramic tools are discussed.

2 HIGH PRESSURE COOLING

2.1 Principle

According to Vosough and Svenningsson [8], the experiments with high pressure cooling were started in 1938 by O. W. Boston. The credit for inventing the method, however, is usually given to Pigott and Colwell [4]. In the paper published in 1952 they argued that conventional overhead streams were inefficient because cutting tool was cooled indirectly through the chip, and only a very small quantity of the fluid was doing the actual cooling. Pigott and Colwell [4] demonstrated that a much better performance could be achieved if the cutting edge was cooled directly by supplying a small $\phi 0.25$ mm stream of neat oil from below to the wedge formed by tool's flank face and the workpiece. To make the coolant flow upwards a pressure of 0.2 – 4.1 MPa had to be maintained. This concept was called the "Hi-Jet" system and is roughly illustrated in Figure 1a.

Most of the later researchers abandoned Pigott and Colwell's [4] original idea of cooling the flank face or used this method mainly for comparison purposes. It was recognized that an even more efficient cooling could be achieved if fluids were delivered to the interface between the chip and the tool's rake face, where the highest temperature occurred. To reach this zone Sharma et al. [9] injected coolant through a hole in the tool as it is illustrated in Figure 1b. This idea has also been tested in grooving [10], milling [11, 12] and drilling [13].

Another way to deliver the coolant to the tool-chip interface is through channels in the tool holder as shown in Figure 1c. This system is robust, does not required custom-made tools and does not obstruct tool changing, which for it is used in industrial applications of high pressure cooling. It has also been employed in

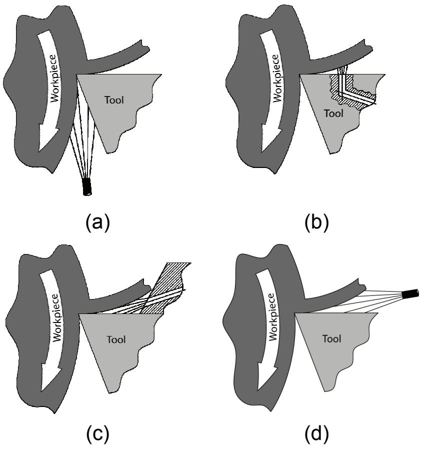


Figure 1: Types of high pressure cooling systems.

research studies done by Ezugwu et al. [7, 14-19], Kaminski and Alvelid [20], Sørby et al. [21, 22] and Sharman et al. [23].

The last system configuration to be discussed is shown in Figure 1d. It employs an external nozzle for coolant delivery. This makes the system simple as standard tools and tool holders can be used. Moreover, the position of the nozzle can be easily adjusted to achieve the desired distance from the tool, required jet impingement angle and the location of the target point. Simplicity and flexibility is probably what makes this design so common in research studies [13, 24-27]. This configuration was also used in the current study.

2.2 Effects of high pressure cooling

Tool temperature

Removing the heat generated during the machining process is one of the main functions of cutting fluids. Therefore water or water soluble oils usually give the best performance. In the original study on high pressure cooling, however, Pigott and Colwell [4] achieved the longest tool life when using neat oil. To explain this strange result the authors suggested that the cooling effect was caused by oil boiling. In this case, however, temperature of the cutting edge was only 24°C lower than in dry machining and only 7°C below that recorded in flooding with neat oil. Therefore it is more likely that the improvements were due to better lubrication rather than improved cooling.

In later studies traditional coolants were usually used. Moreover in most cases fluids were delivered to the tool-chip interface, where the highest temperature occurs. The study performed by Nagpal and Sharma [28] shows that, compared to conventional cooling, application of soluble oil under high pressure can lead to up to 45% reduction in tool-chip interface temperature. Similar results are also reported by Kaminski and Alvelid [20].

There are a few possible explanations to such reduction in temperature. As argued by Pigott and Colwell [4], flood cooling is indirect and hence inefficient. When applied in the form of a jet fluid is delivered directly to the tool's rake or/and flank face. Therefore cooling becomes more effective. Moreover, according to Mazurkiewicz et al. [25], fluid applied at high pressure creates a hydro-wedge, which lifts up the chip. This way coolant gets access to the highest temperature zone, thus cooling is improved.

Lifting up of the chip has another important effect. Measurements of the width of the worn area on the tool's rake face [7, 27] show that the length of the contact between the tool and the chip is reduced in high pressure cooling. Meanwhile Sadik and Lindström [29] have demonstrated that reducing the chip contact length (up to a certain limit) leads to a decrease in tool temperature. This suggests that temperature drop observed in high pressure cooling is at least partially due to shorter chip contact length.

Yet another possible explanation for considerable temperature reduction as a result of high pressure cooling is the ability of the jet to break the steam barrier, which builds up when coolant gets evaporated. Kaminski and Alvelid [20] suggest that the breakthrough occurs when a certain pressure is reached. This hypothesis is supported by the results of a resent study performed by Ezugwu et al. [19]. They report that applying coolant at 7 MPa has doubled the lifetime of uncoated carbide tools in machining of Ti-6Al-4V. However, the advantage was diminishing rapidly as the cutting speed was increased. From these results we can suspect that at higher cutting speeds the amount of heat, and therefore steam, was so great that at a 7 MPa jet was too weak to break through it and reach the highest temperature zone.

Friction

Heat generated during the machining process is at least to some extent due to the rubbing of the workpiece and the chip against the tool. Therefore cooling must be combined with lubrication in order to reduce friction. As discussed previously, the results presented by Pigott and Colwell [4] suggest that in high pressure cooling improvement in lubrication can be so remarkable that it might provide more advantage than cooling. Sharma et al. [9] have also focussed on this issue. They have found out that, when injected through a hole on the tool's rake face, mineral oil with 1.2% of sulphur reduced the friction by 60%. It is suggested that effective lubrication was thanks to the formation of a sulphide film. This film, however, remained stable only up to 800°C, which indicates that using oils as cutting fluids is inefficient at higher temperatures. Mazurkiewicz et al. [25] on the other hand have shown that, when applied at a high pressure, even tap water, which is a good coolant, but a poor lubricant, can reduce the friction coefficient from 0.75 to just 0.4.

Pigott and Colwell [4] also mention that built-up edge was eliminated and suggest that this was a sign of the effectiveness of lubrication. Similarly, Kovacevic et al. [11, 12] have observed that chips produced under high pressure cooling had smoother surface than those collected after conventional flushing. This shows that less seizure between the tool and the chip must have occurred and indicates that lubrication was better.

Cutting forces

Due to reduced friction high pressure cooling usually leads to lower cutting forces and therefore reduced energy consumption. The results presented by Nagpal and Sharma [28] show that both the cutting force and specific energy were generally lower in rough turning of carbon steel under high pressure cooling than in flooding or dry cutting. Mazurkiewicz et al. [25] report having achieved a 23% and 50% reduction in cutting and feed forces respectively. Nandy et al. [27] observed a significant decrease in feed force in turning of Ti-6Al-4V, while Kovacevic et al. [11, 12] have found out that, under various process parameters, cutting force in milling of stainless steel and Ti-6Al-4V was considerably lower under high pressure cooling. On the other hand some researchers [17, 18, 20] have observed that the effect of high pressure cooling on the cutting forces can be either insignificant or even detrimental.

Kaminski and Alvelid [20] have noticed that, compared to dry machining, cutting forces decreased only marginally when high pressure cooling was applied. They suggest that the answer lies in chip contact length. As it is decreased the frictional forces are also reduced. However, in those cases when the chip contact length is already small, it is difficult to reduce it even further. Thus the effect on the cutting forces becomes insignificant.

Another explanation for possible increase in cutting forces is provided by Ezugwu and Bonney [17]. Their data shows that, depending on the process parameters, forces in finish turning of Inconel 718 under high pressure cooling can be either lower or higher than in conventional flushing. However, in the former case both the feed and the cutting forces rose as the speed was increased, while in the latter case they remained quite stable. Ezugwu and Bonney [17] suggest that the increase was due to a reactive force introduced by the high pressure jet.

Tool wear and tool life

One of the most significant advantages offered by high pressure cooling is improved tool life. When discussing tool life, though, care should be taken of how it is defined. Traditionally tool life has been considered to be the time it took to reach the critical width of the flank wear land. In some operations, though, other types of wear such as notch or nose wear become dominant and define the tool life. Therefore it is more precise to speak about reduced wear than improved tool life.

Reduction in tool wear in high pressure cooling is a joint effect of improved cooling, more efficient lubrication and lower cutting forces. Pigott and Colwell [4] have observed that unlike in flooding, tool nose of high speed tools did not wear appreciably in high pressure cooling. Failure was mainly caused by crater wear, which was much more uniform than in flooding. Moreover, it is mentioned that high pressure cooling eliminated built-up edge. As a result of these improvements tool life in turning of carbon, alloy and stainless steels increased more than 17 times.

Wertheim et al. [10] applied through-tool (see Figure 1b) high pressure cooling in grooving. This reduced the crater wear, which was critical for the tool life in machining of alloy steel with of coated carbide inserts. When grooving alloy and stainless steels as well as Inconel 718 with coated carbide tools, flank wear became dominant. In all cases it was considerably reduced when high pressure cooling was applied.

Machado et al. [15] tested high pressure cooling in turning of Ti-6Al-4V with uncoated carbide tools. They have observed that the main failure modes were maximum flank and crater wear. The main mechanism behind these types of wear was diffusion. Since high pressure cooling reduces the temperature, diffusion wear was reduced. Therefore up to 300% increase in tool life was achieved.

Similar improvement in the life time of uncoated carbide tools when turning Ti-6A-4V was achieved by Ezugwu et al. [19] and Nandy et al. [27]. In the former study high pressure cooling resulted in significantly less nose wear, which was the dominant failure mode. In the latter case high pressure cooling reduced flank wear and minimized edge depression. Lower flank wear and edge depression were also observed by Nandy and Paul [26]. In this case, however, the results were compared to dry machining, and the coolant was neat oil.

Improved tool life of uncoated carbide tools as a result of high pressure cooling was also observed by Ezugwu et al. [18]. They have used these results as a baseline to evaluate the performance of cubic boron nitride tools in turning of Ti-6Al-4V. In general these tools performed very poorly due to severe notching and chipping. However, high pressure cooling had a clearly positive effect on the inserts which contained only 50% cubic boron nitride.

Yet another study on application of high pressure cooling in machining of titanium was performed by Kovacevic et al. [12]. The presented results show that at the start of a down milling operation the rate of the flank wear was similar in both high pressure and conventional cooling. After six minutes, though, the wear rate in flooding increased rapidly, while in high pressure cooling it remained stable. Thus, in the latter case longer tool life was achieved.

Ezugwu and Bonney have investigated the effect of high pressure cooling in rough [16] and finish [17] turning of Inconel 718 with coated carbide tools. They claim to have used a complex set of rejection criteria, such as the level of tool wear (flank, nose and notch), workpiece's surface roughness and cutting edge failure. The studies have showed that under correct choice of machining and system parameters a considerably longer tool life can be achieved.

Sørby et al. [21] have tested high pressure cooling in finish turning of Waspaloy (a nickel based aerospace material) with carbide tools. They have observed reduced flank wear and less edge chipping, hence longer tool life.

The above cases show that high pressure cooling can significantly reduce wear and therefore prolong tool life. Nevertheless there is evidence that in some situations high pressure cooling can have a detrimental effect on tool life.

Machado et al. [15] and Ezugwu et al. [7] have observed shorter life of uncoated carbide tools in turning of Inconel 901 (some improvement was observed at the highest cutting speed where cooling is most critical). In both cases the main reason for reduced tool life was increased depth of cut notch wear. Ezugwu and Bonney [14] suggest that the increase in notch wear under high pressure cooling could be caused by hydrodynamic erosion. When the jet hits the tool, it comes to a sudden rest and builds a stagnation pressure. To release it coolant tries to escape through the depth of the cut region. This way small abrasive particles caught in the fluid are flushed away at high velocity and cause severe wear.

Under some conditions reduction in the life of carbide tools when machining nickel based alloys under high pressure cooling has also been observed by Ezugwu and Bonney [16, 17] and Sharman et al. [23]. This was particularly the case at higher coolant pressures. A possible explanation to these results is offered by Öjmertz and Oskarson [24]. They suggest that at high pressure workpiece might be cooled down below a certain threshold temperature. This would increase its strength and result in higher tool contact pressure. Another clue is given by Sadik and Lindström [29], who have investigated the effect of the tool-chip contact length on tool temperature and wear.

As discussed previously, high pressure cooling reduces tool-chip contact length. Sadik and Lindström [29] have experimentally demonstrated that this leads to lower flank face temperature and consequently reduced flank wear. However, when the contact length is reduced beyond certain limit both the temperature and the flank wear increase substantially. Sadik and Lindström [29] explain that in this case forces act on a very small area, thus compressive stress is increased. Moreover, the reduction in the chip contact length means that the highest temperature region is pushed closer to the cutting edge which undergoes elastic deformation. Consequently the contact area between the workpiece and tool's flank face is enlarged leading to increased flank wear. Thus it is likely that at high coolant pressure the chip contact length is reduced so much that the phenomena described by Sadik and Lindström [29] come into action.

Chip breaking

The most noticeable and therefore most often mentioned benefit of high pressure cooling is very efficient chip breaking. In nearly all studies reviewed here high pressure cooling produced short segmented chips. Under conventional cooling the same test conditions resulted in long continuous chips, which are a serious hazard, especially in automated machining. Mazurkiewicz et al. [25] suggests that this improvement is due to the hydro-wedge, which is created as the focussed coolant jet penetrates between the tool's rake face and the chip. This wedge acts as a regular chip breaker, i.e. it lifts the chip up and reduces its' curl radius. Eventually the chip breaks and is flushed away by the powerful jet.

In a few rare cases, though, high pressure cooling did not provide sufficient chip breaking. The explanation here seems to be the properties of the workpiece material and the choice of the system parameters, discussed in Section 2.3. Sharma et al. [9] have applied high pressure cooling in machining of steel tubes. The coolant was injected into the tool-chip interface through a hole on the rake face. As a result chip curl diameter reduced to 1.3 – 2.5 cm compared to 10 cm in dry machining. Nevertheless, in the latter case chips were segmented, while in high pressure cooling they became ribbon-like. Continuous chips in milling of steel under high pressure cooling were also observed by Senthil Kumar et al. [30].

Machado et al. [15] have focussed on the chip control when turning Ti-6Al-4V and Inconel 901. In general good fragmentation was observed. However, at low speeds and feeds continuous chips were produced. It is thought that at low speeds the presence of the built-up edge changed the chip shape, thus high pressure jet could not hit it at a critical point and break. At low feed the chip was very thin and thus too flexible to be broken.

Some issues with chip breaking in machining of nickel based alloys are reported by Scharman et al. [23], Ezugwu and Bonney [17] and Sørby et al. [21]. In all three studies short segmented chips were produced at 15 – 20 MPa. At lower pressure Ezugwu and Bonney [17] observed short tubular chips, while in conventional cooling long tubular chips were produced. Sørby et al. [21] mention that at 10 MPa chip breaking was usually satisfactory. However, in some cases unbroken chips were produced.

Surface finish

Pigott and Colwell [4] observed that in high pressure cooling surface roughness improved from $5 \mu m$ to $1.3 \mu m$. They assumed that this was due to reduced nose wear. Moreover Pigott and Colwell [4] mention that high pressure cooling eliminated built-up edge, which could have also contributed to better surface quality. Kovacevic et al. [11, 12] have analysed surface finish in milling. Their results show that under all combinations of cutting parameters surface roughness was better in high pressure cooling than in flooding. Improved surface roughness is also reported by Kaminski and Alvelid [20]. However, in this case high pressure cooling was compared to dry machining.

In other studies the effect on surface roughness depends on process parameters or the results are inconclusive [19, 23]. Data presented by Nagpal and Sharma [28] shows that at the feed of 0.2 mm/min surface roughness was smaller in high pressure than in conventional cooling. On the other hand, at both lower and higher feeds the results appear too be mixed. Sentil Kumar et al. [30] have focussed on the relationship between the surface finish and workpiece hardness. They have found out that surface in high pressure cooling was always smother than in dry milling. The difference from conventional cooling, on the other hand, became significant only when workpiece hardness exceeded 35 HRC.

Finally there are a few documented cases were surface finish was worse. Nandy et al. [27] report that, compared to conventional cooling, roughness increased under high pressure cooling with both neat and water soluble oil. Similarly, the results presented by Ezugwu and Bonney [16] show that, while at lower coolant pressure surface roughness was similar in both high pressure and conventional cooling, it was clearly worse at 15 MPa and beyond.

A possible reason for poorer surface finish is given by Kaminski and Alvelid [20]. They suggest that due to unsuitable adjustment of the jet's direction the chip could be forced against the already machined part and scratch it. This issue was analysed by Sørby et al. [22]. They have observed a slight increase in surface roughness when high pressure cooling was applied. Moreover, welding of small chip particles to Ti-6Al-4V workpiece and tearing of material out from its surface was detected. Similar observations were made by Ezugwu and Bonney [17], who have detected small pits on the surface of an Inconel 718 part. These were said to have occurred when high pressure jet removed hard particles from the surface.

Surface integrity

Improved cooling, reduced friction, lower cutting forces and tool wear suggest that surfaces should be subjected to less deformation in high pressure cooling. An extensive study on this issue was performed by Sharman et al. [23].

They have found out that lowest residual stresses were induced when a 45 MPa jet was directed to the tool's flank face. While in conventional cooling high tensile stresses were observed, in this case small compressive stresses were detected. However, when the same jet was directed to the rake face, high tensile stresses were developed. Therefore Sharman et al. [23] state that the effect of high pressure cooling on surface integrity was neither positive nor detrimental.

Sharman et al. [23] also mention that high pressure cooling reduced surface micro hardness. A contrary conclusion can be made from the data presented by Ezugwu et al. [19]. The micro hardness here seems to be slightly higher in high pressure cooling. Nevertheless the authors state that high pressure cooling resulted in less plastic deformation. Reduced plastic deformation was also observed by Sørby et al. [22].

Another interesting observation made by Sharman et al. [23] is that at high pressure the jet alone can alter surface integrity. To come to this conclusion they have removed the cutting tool and sprayed the workpiece with coolant at 45 MPa. As a result feed peaks were flattened.

Coolant consumption

According to the review done by Dahlman [31], the costs of purchasing, handling and disposing of cutting fluids account for 17% of the costs in automotive industry. While this number might be slightly different in other industries, it is a clear indication that reduction in coolant consumption would give a big economic advantage. Moreover, this is also critical from the environmental point of view.

In the studies discussed in this paper coolant flow rates in flooding operations ranged between 2.2 and 30 l/min. The flow rate in high pressure cooling depended on the pressure and nozzle diameter. It is reported to be from 0.1 to 15.1 l/min. Therefore, it looks that an efficient high pressure system can help to reduce coolant consumption.

2.3 Effect of system parameters on efficiency of high pressure cooling

The efficiency of high pressure cooling depends on a number of external variables, such as the type of machining operation, workpiece material, tool material, coating and geometry, and cutting parameters. These variables define how severe the machining conditions are and how critical the improved cooling, lubrication and chip breaking is. Other parameters such as the direction of the jet, coolant pressure, jet impingement angle, nozzle diameter and the position of the target point are internal system variables. The effects of these variables on the efficiency of high pressure cooling are discussed next.

Cooling direction

As mentioned in Section 2.1, coolant in a high pressure system can be delivered either from below to the space between the workpiece and tool's flank face or from the top to the interface between the chip and tool's rake face. The mechanisms acting in the two cases are slightly different. Therefore a few studies have been carried out to compare the two approaches.

When delivered to the flank face, coolant has direct access to the cutting edge. In rake face cooling on the other hand, the access to the cutting edge is hindered by the huge pressure exerted by the chip. It can therefore be expected that flank face cooling should be more effective in reducing tool wear. This is partially confirmed by Sharman et al. [23], who have observed reduced flank wear of carbide tools in finish turning of Inconel 718. However, no effect on the rate of notch wear was noticed, thus no increase in tool life was achieved. In rake face cooling, on the other hand, significant reduction in tool life occurred at higher pressure levels. Finally, supplying coolant from both directions did not yield any significant difference in tool life. Sørby et al. [21], on the other hand, have observed that combined flushing from both directions was beneficial as was cooling from both sides separately.

Conflicting conclusions can be made from the observations made by Nandy and Paul [26]. They report that cooling the rake face of carbide inserts with neat oil was more efficient in reducing flank wear than targeting the jet from below on the flank face. Simultaneous cooling from both directions gave intermediate wear results, thus it was concluded that application of flank cooling in addition to rake face cooling was not beneficial.

Sharman et al. [23] have observed a clear difference in the level of residual stresses developed when cooling from the two directions. The lowest deformation was detected in flank face cooling. Even more important is that stresses turned from tensile (observed in conventional cooling) to compressive. In rake face cooling, on the other hand, high tensile stresses were developed.

Another area, where differences appear is chip breaking. In rake face cooling the jet hits the chip directly this way making it curl and eventually break. In flank face cooling the jet hits the tool and has limited contact with the chip. Therefore, as shown by Sharman et al. [23], higher pressure is needed to achieve optimal chip breaking in flank face cooling.

Jet impingement angle

Kaminski and Alvelid [20] have emphasized the importance of nozzle orientation on the chip flow. This topic has been investigated by Nandy and Paul [26]. The study has revealed that the jet impingement angle not only influences the chip flow, but also affects the tool wear. Therefore a compromise has to be found. It was expected that low impingement angles would be most beneficial in terms of chip breaking because the jet would act more directly against the momentum of the chip flow. In terms of the tool life, though, an angle of 20° appeared to be optimal.

Target point

Another important parameter is the jet target point (it is related to the jet impingement angle). Pigott and Colwell [4] suggested placing the jet ¼ to 1/3 of the depth of cut from the nose of the tool. Nandy and Paul [26] report that targeting the jet further away from the tool edge had a positive effect on both flank and crater wear. However, this reduced the efficiency of chip breaking because the jet lost its momentum. Machado et al. [15] has also observed that when the jet could not hit the chip at an optimal spot, chip breaking was inefficient.

Target point becomes especially critical in through-tool cooling (see Figure 1b). Sharma et al. [9] have observed that the distance of the orifice from the cutting edge had to be at least 1.5 times larger than the depth of the cut. At shorter distances the jet was blocked by the extreme pressure exerted by the chip. Moreover, the hole acts as a stress concentrator. Therefore Kovacevic et al. [11] states that the orifice should be located around 1.25 mm from the tool nose to provide sufficient strength and good performance in cutting.

Nozzle diameter

Nozzle diameter in combination with coolant pressure determines jet momentum and its flow rate. The effect of these factors was studied by Dahlman [31]. He has found out that, when machining materials which are known to lead to long chip contact length, high flow rate is necessary to dissipate the big amounts of heat. On the other hand, when chip contact length is expected to be short, heat generation is also expected to be less. Thus it is suggested to use smaller amounts of cutting fluids this way making the process more economic.

Another study on this topic was carried out by Nandy and Paul [26]. Their results show that increasing nozzle diameter from approximately 0.35 to 0.85 mm did not have any significant influence on the average flank wear, and the effect on edge depression was unclear. Maximum flank wear land on the other hand reduced as the diameter was increased. Moreover this resulted in less material adhering to the tools rake face and produced shorter chips.

More support for the idea that bigger orifice can potentially give better results is provided by Mazurkiewicz et al. [25]. They have showed that increasing nozzle diameter from 0.135 to 0.350 mm reduces friction and therefore results in lower cutting forces. The same effect was observed by Kovacevic et al. [11, 12]. In addition bigger orifice resulted in lower surface roughness. These improvements, however, were clearly easing off when nozzle diameter was increased. This suggests that there exist an optimal orifice size.

Coolant pressure

The above parameters describe the system configuration. These are defined in the design phase and would normally remain fixed once the system is in place. Coolant pressure, on the other hand is a control variable and can be adjusted when needed. As discussed in the previous sections, a certain minimum pressure needs to be reached to break the steam barrier and achieve sufficient chip control. On the other hand, we have also discussed that too high rate of cooling or too short tool-chip contact length can reduce tool life. These observations suggest that there exists an optimal coolant pressure.

The existence of the optimal coolant pressure was spotted already in the original study carried out by Pigott and Colwell [4]. They have observed that moving away from the optimum in either direction resulted in shorter tool life. Pigott and Colwell [4] suggested that at high pressure tool surface was swept too fast. As a result the rate of boiling action, which they believed was the reason for improved cooling, was reduced. This phenomenon was further investigated by Nagpal and Sharma [28]. They have found out that all desirable conditions, such as minimum chip curl radius, cutting forces and friction coefficient as well as maximum shear angle occurred simultaneously at a certain coolant pressure.

A different observation has been made by Mazurkiewicz et al. [25], who have experimented with 70 – 280 MPa pressure. At this range both the measured cutting forces and the calculated friction coefficient decreased continuously as coolant pressure was increased. Similarly Wertheim et al. [10] have observed a nearly linear improvement in the life of grooving tools as coolant pressure was increased. Kovacevic et al. [11, 12] have showed that higher pressure resulted in lower cutting forces and better surface roughness in milling. Nevertheless in this case the relative improvement was diminishing. This suggests that an optimal pressure exists and exceeding it is at least not beneficial.

Another interesting result is reported by Nandy and Paul [26] and Nandy et al. [27]. They have observed that higher pressure led to more crater wear. This could be explained by the jet impingement provoked erosion discussed earlier. The proof for this hypothesis can be found in the SEM micrographs provided by Öjmertz and Oskarson [24], where severe wear at the jet targeting point can be seen. We can expect that with the increase in the pressure the eroding capacity of the jet would also grow. This once again supports the hypothesis that an optimal coolant pressure exists.

3 HIGH PRESSURE COOLING WHEN MACHINING WITH CERAMIC TOOLS

Despite the impressive results discussed in the previous section the number of reported studies on the application of high pressure cooling when machining with ceramic tools is scarce. Ezugwu et al. [7] experimented with turning of Inconel 901 with SiC-whiskers reinforced ceramic inserts. They have observed that cooling at a pressure of 14 MPa improved chip breaking. However, it generally led to reduced tool lives as compared to conventional flushing. The reason for this poor performance was said to be a combination of notch and nose wear.

Analogous results were achieved by Öjmertz and Oskarson [24], who have tested rough turning of Inconel 718 with SiC-whiskers reinforced ceramic tools. They have observed that high pressure cooling led to better chip control, reduced tendency to built-up edge formation and therefore better surface quality as compared to dry machining. However, a clear tendency towards increasing depth of cut notch wear, which is said to be the main mechanism determining the life of SiC-whiskers reinforced ceramics, was observed as coolant pressure was increased from 80 to 360 MPa. The scatter of the data points was also bigger at higher pressures. In addition to that, cutting force was increased.

A similar work was performed by Ezugwu and Bonney [14], who have applied coolant at a pressure of 11 – 20 MPa in rough turning of Inconel 718 with SiC-whiskers reinforced ceramic tools. Despite the lower pressure they have also observed that jet cooling caused severe notching and therefore shorter tool life as compared to conventional cooling. Cutting force was also slightly increased, while chip breaking was improved.

Slightly more promising results were achieved when Ezugwu et al. [32] repeated the study under finishing conditions. Besides the improved chip breaking, reduction in cutting force and generally increased tool life was observed at coolant pressures of 11 and 15 MPa. At 20 MPa, thought, tool life dropped significantly due to accelerated notch wear. Tool life was also shorter at 11 MPa and when the speed was increased to 300 m/min.

The above examples show that in general high pressure cooling increases notch wear therefore leading to shorter lifetime of SiC-whiskers reinforced ceramic tools. On the other hand Ezugwu et al. [32] have demonstrated that correct selection of process parameters and coolant pressure can have an opposite effect. Moreover, all studies report about improved chip breaking, which is an advantage, especially in automated manufacturing. Finally, to the best of our knowledge there are no published papers about the attempts to apply high pressure cooling when machining with SiAlON-based ceramics. Therefore, a short experimental study with this type of tools was performed and is presented next.

4 EXPERIMENTAL STUDY

4.1 Experimental set-up

Experiments were carried out with Hessapp DV80 vertical lathe (see Figure 2) equipped with a high pressure pump capable of up to 40 MPa. A custom made nozzle with $\phi 1.0$ mm orifice was fitted into a copper tube, which in turn was epoxy-bonded to a standard tool holder C5-CRSCR-35060-12V. The resulting jet targeting angle was approximately 45°, and the distance from the nozzle to the target point was 23 mm.

The cutting tool was a round SiAlON-based ceramic insert RCGX 12 07 00 E. According to manufacturer's specification, the grade used in our experiments was developed specially for machining of heat resistant super alloys.

Workpiece material was a heat resistant nickel based alloy Inconel 718 used in aerospace industry. This material is characterized as difficult to machine. Its high hardness (36 HRC in case of our test piece) and ability to retain strength at high temperatures result in high cutting forces. Combined with low thermal conductivity (11.1 – 11.4 W/mK at room temperature [33]) this leads to very high temperature (up to 1300°C) being generated in the cutting zone. Therefore machining of Inconel 718 is perfectly suited for high pressure cooling.

4.2 Experimental procedure

For the purpose of comparison two sets of experiments were conducted. In the baseline tests machining was performed under flood cooling. In the second phase high pressure jet was directed at the tool-chip interface and used together with conventional flushing. In both cases after each cut, which on average took 30 seconds to complete, the insert was removed from the tool holder, and the wear was measured with a Mitutoyo toolmaker's microscope. The procedure was continued until one of the tool life criteria defined in Section 4.3 was reached. Further experimental details are given in Table 1.

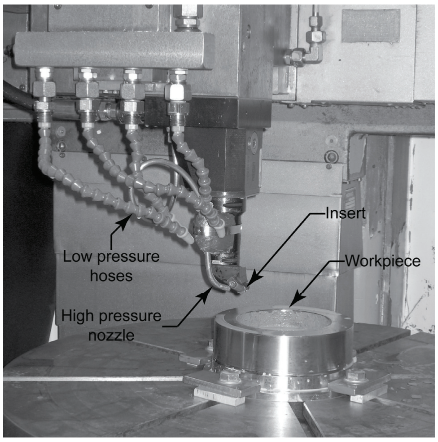


Figure 2: Experimental set-up.

4.3 Tool life criteria

According to ISO 3685 [34], the most common life measures for tools of ceramics are the average and the maximum widths of the flank wear land. Depth of cut notch wear, which was mentioned in the previous sections, is said to depend on the accuracy of repeated depth settings and must therefore be excluded from the flank wear measurements. Another issue with ceramic tools is edge chipping. According to ISO 3685, to a certain extent this type of wear is taken into account by the maximum width of the flank wear land, which for the latter is the recommended measure when edge chipping is expected. Finally a possibility of a catastrophic failure must be had be taken into account when machining hard materials with brittle tools. Therefore, the following tool life criteria were defined for the tests:

- The maximum width of the flank wear land $VB\,\mathrm{max.} = 0.6\,\mathrm{mm}$
- Catastrophic failure

4.4 Results

Tool wear

Six ceramic inserts were tested in this study – three under conventional cooling and three under conventional cooling combined with high pressure cooling. In both cases the wear pattern was similar. At the start of the process the depth of cut marks, shown in Figure 3a, were clearly visible, while flank wear, seen on the left of the same figure, was rather small. Later on

Parameter	Experiment 1	Experiment 2
Coolant pressure		
High pressure cooling	N/A	20.0 MPa
Conventional cooling	0.7 MPa	0.7 MPa
Cutting speed, v_c	150.0 m/min	150.0 m/min
Feed, f_n	0.2 mm/rev	0.2 mm/rev
Depth of cut, a_p	1.0 mm	1.0 mm

Table 1: Experimental conditions.

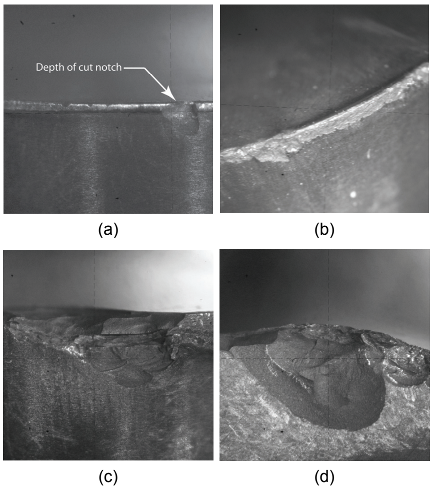


Figure 3: Tool wear: (a) notch wear, (b) flank wear, (c) edge chipping – flank face, (d) edge chipping – rake face.

notch wear stabilized, while flank wear, shown in Figure 3b, grew progressively. Eventually the edge of all but one tool failed as big pieces of the material were chipping off and destroying both the flank face and the rake faces. This can be clearly seen in Figure 3c and Figure 3d, which show the two faces of the same tool.

Tool life

Since no catastrophic failures were observed during this short series of experiments, tool life was determined as the time when the maximum width of the flank wear land reached the 0.6 mm limit. This is illustrated graphically in Figure 4 for conventional and Figure 5 for high pressure assisted cooling. Sudden increase (to over 3 mm) in the width of the flank wear land seen in some figures corresponds to the edge chipping. Note that only the part of the wear curves that was necessary to determine the tool life is shown. Therefore the moment of edge failure is not seen for some of the tools.

From the results shown in the figures on the right it appears that edge failure occurs earlier when high pressure cooling is applied, resulting in shorter tool life. Moreover, the wear measurements seem to be more scattered making the tool life prediction more complicated. Thus, this preliminary study suggests that application of high pressure cooling is detrimental to the life of SiAlON ceramic tools.

Chip formation

Despite the reduction in tool life one clear advantage of high pressure cooling was observed. As can be seen from Figure 6 long continuous chips were produced when conventional flood cooling was applied alone, while short segmented chips were formed when using high pressure assisted cooling. Continuous chips can snarl around the tool and can scratch the workpiece. They also tend to form big lumps (see Figure 7), which can get entangled in the chip disposal equipment and cause unplanned stoppages. For these reasons, continuous chips are undesirable, especially in automated production, where no operator might be available to remove the chips.

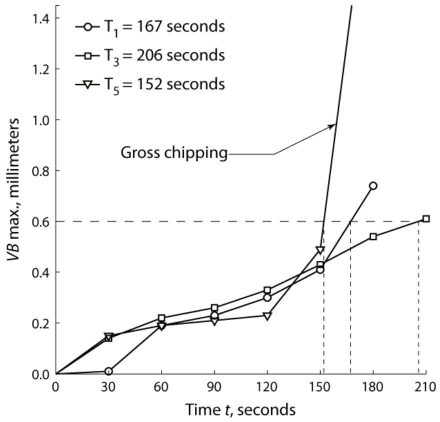


Figure 4: Flank wear – conventional cooling.

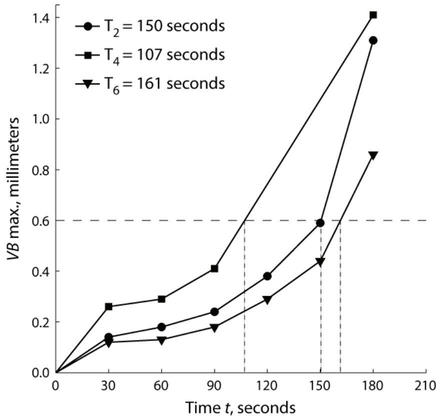
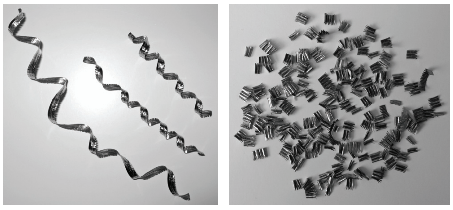


Figure 5: Flank wear – high pressure cooling.



Conventional Conventional + high pressure

Figure 6: Effect of cooling on chip formation.

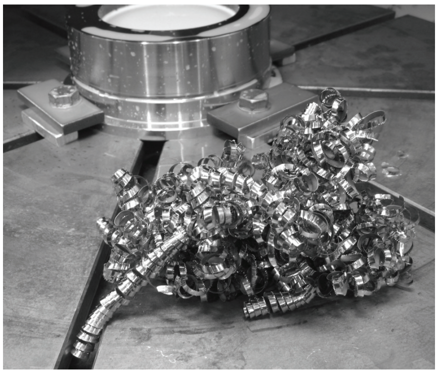


Figure 7: Tangling of continuous chips.

5 DISCUSSION

The reduction in tool life observed in our experiments is inline with the results reported in other studies on application of high pressure cooling when machining with ceramic inserts. The main reason for shorter tool life in these studies was accelerated depth of cut notch wear. This was not the case in our experiments. Nevertheless it is likely that the phenomena leading to shorter tool life are similar.

Ceramics are known to be sensitive to thermal shocks. Therefore the temperature during the cutting process should be as even as possible. However, Öjmertz and Oskarson [24] suggest that in high pressure cooling the chip contact length is varying, causing the amount of heat generated by the friction to change as well. Moreover, oscillation in chip contact length means that the access for the fresh coolant to the cutting edge is opened and closed in a cyclic manner, which further amplifies temperature variation. This hypothesis sounds quite likely. At the start of the growth cycle the chip is too short to be seriously affected by the high pressure cooling. As it gets longer the jet acts on a larger surface area and creates a bigger bending moment. This makes the chip curl up, thus reducing its contact length. Eventually the chip breaks down and the cycle repeats. Under such conditions crack propagation can be initiated and can eventually lead to tool failure. Ezugwu and Bonney [14] suggest that this process can be further accelerated by the jet, which is powerful enough to penetrate into the existing and newly created cracks and make them propagate.

In the previous sections we have also discussed that the reduction in chip contact length can shift the highest temperature and stress zone closer to the cutting edge. This is supported by our results, where edge failure was eventually taking place. We should also note here that the jet impingement angle of 45° was probably too high. This might have created big normal force acting on the chip and this way might have contributed to the increase in stresses on the cutting edge.

Another possible reason for the increase in stresses acting on the tool is work hardening. Öjmertz and Oskarson [24] suggest that at high pressure the cooling might be so intense that hardness of the workpiece would increase, which would make it more difficult to cut. If this is true, however, the problem can be easily resolved by increasing the cutting speed. In that case the benefit would be two-fold. This would raise the temperature to an optimal level. In addition metal removal rate and productivity would go up.

Cyclic processes discussed above mean that the machining process becomes less stable and less predictable. This could explain the increased scatter in wear measurements observed in this study as well as by Öjmertz and Oskarson [24].

6 CONCLUSIONS

The results of this study suggest that application of high pressure cooling has a negative effect on the life of SiAlON-based ceramic cutting tools. Possible reasons for that are thermal cycling, concentration of high stresses and shifting of highest temperature zone closer to the cutting edge as well as workpiece hardening. However, it is likely that reducing the jet angle and increasing the cutting speed could solve some of these problems and yield better outcomes. These adjustments were not possible to make at this stage because of the experimental design of the high pressure system and speed limitations imposed by the machine tool and the size of the workpiece. Therefore a bigger scale study with optimized high pressure system and cutting parameters needs to be performed before making the final conclusion.

ACKNOWLEDGMENT

The authors thank The Research Council of Norway and Volvo Aero Norge AS for supporting this work through the AVROPROS project.

REFERENCES

[1] Narutaki, N., Yamane, Y., Hayashi, K., Kitagawa, T., and Uehara, K., 1993, High-Speed Machining of Inconel 718 with Ceramic Tools, CIRP Annals - Manufacturing Technology, 42: 103-106.

[2] Kitagawa, T., Kubo, A., and Maekawa, K., 1997, Temperature and Wear of Cutting Tools in High-Speed Machining of Inconel 718 and Ti-6Al-6V-2Sn, Wear, 202: 142-148.

[3] Vigneau, J., Boulanger, J. J., and Le Maître, F., 1982, Behaviour of Ceramic Tools During the Machining of Nickel Base Alloys, CIRP Annals - Manufacturing Technology, 31: 35-39.

[4] Pigott, R. J. S., and Colwell, A. T., 1952, Hi-Jet System for Increasing Tool Life, Society of Automotive Engineers - Transactions, 6: 547-564.

[5] Boothroyd, G., and Knight, W. A., 2006, Fundamentals of Machining and Machine Tools, Taylor & Francis, Boca Raton.

[6] Kalpakjian, S., and Schmid, S. R., 2006, Manufacturing Engineering and Technology, Prentice Hall, Singapore.

[7] Ezugwu, E. O., Machado, A. R., Pashby, I. R., and Wallbank, J., 1991, The Effect of High-Pressure Coolant Supply When Machining a Heat-Resistant Nickel-Based Superalloy, Lubrication Engineering, 47: 751 - 757.

[8] Vosough, M., and Svenningsson, I., 2005, Influence of High Pressure Water Jet-Assisted Machining on Surface Residual Stresses on the Work-Piece of Ti-6Al-4V Alloy, in Third International Conference on Experimental Mechanics and Third Conference of the Asian Committee on Experimental Mechanics, Singapore, SPIE - The International Society for Optical Engineering.

[9] Sharma, C. S., Rice, W. B., and Salmon, R., 1971, Some Effects of Injecting Cutting Fluids Directly into the Chip-Tool Interface, Journal of Engineering for Industry, Transactions of the ASME, 93: 441 - 444.

[10] Wertheim, R., Rotberg, J., and Ber, A., 1992, Influence of High-Pressure Flushing through the Rake Face of the Cutting Tool, CIRP Annals - Manufacturing Technology, 41: 101-106.

[11] Kovacevic, R., Cherukuthota, C., and Mohan, R., 1995, Improving Milling Performance with High Pressure Waterjet Assisted Cooling / Lubrication, Journal of Engineering for Industry, 117: 331-339.

[12] Kovacevic, R., Cherukuthota, C., and Mazurkiewicz, M., 1995, High Pressure Waterjet Cooling/Lubrication to Improve Machining Efficiency in Milling, International Journal of Machine Tools and Manufacture, 35: 1459-1473.

[13] Lopez De Lacalle, L. N., Perez-Bilbatua, J., Sanchez, J. A., Llorente, J. I., Gutierrez, A., and Alboniga, J., 2000, Using High Pressure Coolant in the Drilling and Turning of Low Machinability Alloys, International Journal of Advanced Manufacturing Technology, 16: 85-91.

[14] Ezugwu, E. O., and Bonney, J., 2003, Effect of High-Pressure Coolant Supplies When Machining Nickel-Base, Inconel 718, Alloy with Ceramic Tools, Tribology Transactions, 46: 580 - 584.

[15] Machado, A. R., Wallbank, J., Pashby, I. R., and Ezugwu, E. O., 1998, Tool Performance and Chip Control When Machining Ti6Al4V and Inconel 901 Using High Pressure Coolant Supply, Machining Science and Technology, 2: 1 - 12.

[16] Ezugwu, E. O., and Bonney, J., 2004, Effect of High-Pressure Coolant Supply When Machining Nickel-Base, Inconel 718, Alloy with Coated Carbide Tools, Journal of Materials Processing Technology, 153-154: 1045-1050.

[17] Ezugwu, E. O., and Bonney, J., 2005, Finish Machining of Nickel-Base Inconel 718 Alloy with Coated Carbide Tool under Conventional and High-Pressure Coolant Supplies, Tribology Transactions, 48: 76 - 81.

[18] Ezugwu, E. O., Da Silva, R. B., Bonney, J., and Machado, Ã. R., 2005, Evaluation of the Performance of CBN Tools When Turning Ti-6Al-4V Alloy with High Pressure Coolant Supplies, International Journal of Machine Tools and Manufacture, 45: 1009-1014.

[19] Ezugwu, E. O., Bonney, J., Da Silva, R. B., Machado, A. R., and Ugwoha, E., 2009, High Productivity Rough Turning of Ti-6Al-

4V Alloy, with Flood and High-Pressure Cooling, Tribology Transactions, 52: 395 - 400.

[20] Kaminski, J., and Alvelid, B., 2000, Temperature Reduction in the Cutting Zone in Water-Jet Assisted Turning, Journal of Materials Processing Technology, 106: 68-73.

[21] Sørby, K., Bjørnstad, M. S., and Tønnessen, K., 2006, The Effect of High-Pressure Cutting Fluid on Chip Breaking and Tool Life in Turning of Nickel-Base Alloys, In CIRP 2nd International Conference on High Performance Cutting, Vancouver.

[22] Sørby, K., Tonnessen, K., Torjusen, J. E., and Andersen, L., 2007, The Effect of High-Pressure Coolant on Surface Integrity in Turning Ti6Al4V, in Sixth International Conference on High Speed Machining.

[23] Sharman, A. R. C., Hughes, J. I., and Ridgway, K., 2008, Surface Integrity and Tool Life When Turning Inconel 718 Using Ultra-High Pressure and Flood Coolant Systems, Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 222: 653 - 664.

[24] Öjmertz, K. M. C., and Oskarson, H. B., 1999, Wear on SiC-Whiskers Reinforced Ceramic Inserts When Cutting Inconel with Waterjet Assistance, Tribology Transactions, 42: 471 - 478.

[25] Mazurkiewicz, M., Kubala, Z., and Chow, J., 1989, Metal Machining with High-Pressure Water-Jet Cooling Assistance - a New Possibility, Journal of Engineering for Industry, Transactions of the ASME, 111: 7 - 12.

[26] Nandy, A. K., and Paul, S., 2008, Effect of Coolant Pressure, Nozzle Diameter, Impingement Angle and Spot Distance in High Pressure Cooling with Neat Oil in Turning Ti-6Al-4V, Machining Science and Technology, 12: 445 - 473.

[27] Nandy, A. K., Gowrishankar, M. C., and Paul, S., 2009, Some Studies on High-Pressure Cooling in Turning of Ti-6Al-4V, International Journal of Machine Tools and Manufacture, 49: 182-198.

[28] Nagpal, B. K., and Sharma, C. S., 1973, Cutting Fluids Performance - Part 1 - Optimization of Pressure For "Hi-Jet" Method of Cutting Fluid Application, Journal of Engineering for Industry, Transactions of the ASME, 95: 881 - 889.

[29] Sadik, M. I., and Lindström, B., 1995, The Effect of Restricted Contact Length on Tool Performance, Journal of Materials Processing Technology, 48: 275-282.

[30] Senthil Kumar, A., Rahman, M., and Ng, S. L., 2002, Effect of High-Pressure Coolant on Machining Performance, International Journal of Advanced Manufacturing Technology, 20: 83-91.

[31] Dahlman, P., 2002, A Comparison of Temperature Reduction in High-Pressure Jet-Assisted Turning Using High Pressure Versus High Flowrate, Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 216: 467-473.

[32] Ezugwu, E. O., Bonney, J., Fadare, D. A., and Sales, W. F., 2005, Machining of Nickel-Base, Inconel 718, Alloy with Ceramic Tools under Finishing Conditions with Various Coolant Supply Pressures, Journal of Materials Processing Technology, 162-163: 609-614.

[33] Special Metals Corporation, 2007, Inconel Alloy 718, Special Metals Corporation.

[34] ISO 3685, 1993, Tool-Life Testing with Single-Point Turning Tools, International Organization for Standardization.

Report 1

Zydrunas Vagnorius

Estimation of machine tool availability in T-case cell

This document is one of the internal reports written during the project. It is based on a real case study and contains some confidential information. Therefore some parts of the text have been changed or removed.

ESTIMATION OF MACHINE TOOL AVAILABILITY IN T-CASE CELL

Intermediate report – 30.01.2009,
Zydrunas Vagnorius, NTNU

1 INTRODUCTION

Throughout AVROPROS project meetings it was emphasized that availability of machine tools was crucial for the productivity and smooth operation of the T-case cell. Current availability estimates, however, are based on qualified guesses of experienced manufacturing and maintenance engineers. Therefore a more precise evaluation of machine tool availability would be useful and could serve as a baseline for the improvement efforts.

The first study on the availability of the machine tools in T-case cell was based on maintenance bills and feedback reports describing the repair actions. The results of the analysis were presented on 21.05.2008 and showed that equipment availability ranged from 87% to over 99%. The analysis, however, also revealed that the maintenance bills and feedback reports were not sufficient data sources, which is due to the fact that they were built for other specific purposes. Therefore it was concluded that the first estimates did not show the real availability of the machine tools and that additional data were need to improve them.

One possible additional data source was said to be the new e-mail-based failure reporting system, the implementation of which was underway at that time. The primary purpose of this system, originally referred to as KTT VANMA, was to simplify and speed up the process of reporting of the problems to the maintenance company, this way reducing the response time. Moreover the system could show the real-time status of the machine tools and was therefore well suited for availability analysis.

2 DATA COLLECTION

The following data were collected for the analysis:

- KTT VANMA messages for the period from 20.05.2008 to 07.11.2008.
- Maintenance bills. These bills were issued during the period from 01.01.2006 to 16.10.2008 but only the data from 20.05.2008 to 16.10.2008 were used for analysis.
- Feedback reports. Reports covered a period from 01.01.2006 to 16.10.2008 but only the data from 20.05.2008 to 16.10.2008 were used for analysis.
- Planned operating time per week per machine tool.

Figure 1 illustrates the relationship between the data in KTT VANMA and the machine state. In the ideal case KTT VANMA should contain at least three messages for each failure. The first message is sent at time (*b*) and informs about the failure, either critical, when the machine is down, or non-critical, when the machine is in production, but is experiencing some problem. In addition this message shows the time (*a*), when the failure first occurred. The second message informs about the start of repairs at time (*c*). The last message notifies that the repairs have been completed (or at least a temporary solution has been found) at time (*d*) and the machine is ready for production.

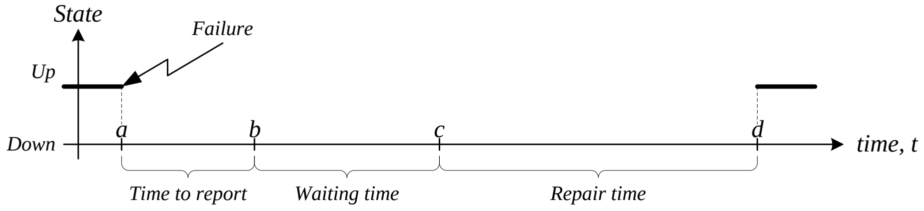


Figure 1: KTT VANMA messages.

As shown in Figure 1, the three messages in KTT VANMA would provide sufficient information to determine the downtime and its components for a particular failure. Considering all failures during a certain time period the availability of the machine would be evaluated. However, the analysis showed that some of the problems were either not reported through KTT VANMA or the information about them was incomplete. Moreover, the system was not used to report the preventive repairs (only two messages related to preventive repairs were found). Thus it was concluded that KTT VANMA messages alone were not sufficient to estimate the availability. For this reason maintenance bills and feedback reports were collected in addition.

Maintenance bills show a complete list of activities on each machine and give an approximate amount of man-hours used per each maintenance order. Feedback reports complement the maintenance bills by describing the problem and repairs in more detail. However, there are a few issues with these reports too:

- The repair hours in maintenance bills are rounded up to the closest half hour, which for the repair time and the downtime are overestimated.
- Maintenance bills do not provide information about parallel repairs, in which case the machine downtime is less than the sum of repair hours. While some parallel repairs are mentioned in the feedback reports, there is evidence that they are more frequent than the reports show. This problem would also lead to downtime being overestimated.
- Neither maintenance bills nor the feedback reports give any information about the waiting time. Since according to the contract between the case and the maintenance companies this time can be up to four hours, ignoring it would lead to downtime being underestimated.

- Neither maintenance bills nor the feedback reports show the time it took before the problem was reported to the maintenance company, which leads to underestimated downtime.

Due to the above problems maintenance bills and the feedback reports give a confusing picture about the downtime. However, they still provide some useful information and could be used in combination with KTT VANMA messages to estimate the availability of the machine tools.

3 METHOD

In general machine availability is determined by the technical downtime and the organisational non-productive time. The latter also includes preventive maintenance. In this study we have focused on the technical problems. However, since in practice preventive maintenance and emergency repairs are often combined, we have included preventive maintenance in the analysis. Then, technical availability of a machine tool can be determined as

$$Technical\ availability = \frac{Planned\ operating\ time - (\sum_{i=1}^{n} DT_{failure,\,i} + \sum_{j=1}^{m} DT_{PM,\,j})}{Planned\ operating\ time} \tag{1}$$

In the above equation n and m are respectively the total number of failures and preventive repairs on a particular machine tool, $DT_{failure,\,i}$ is the downtime due to i-th failure and $DT_{PM,\,j}$ is the downtime resulting from j-th preventive repair.

Downtime associated with an individual failure can be determined from KTT VANMA data. For a critical problem, which leads to a complete machine failure, downtime is approximately the time between points (*a*) and (*d*) in Figure 1. For a non-critical problem, which results in short stoppages, but is eventually resolved by the operator and does not bring the machine completely down, downtime approximately equals the repair time, i.e. time between points (*c*) and (*d*).

When a failure is not reported via KTT VANMA or is reported incompletely, the missing downtime components can be estimated. In the most general case, downtime can be calculated as

$$DT_{failure} = E(TTR) + E(WT) + RT_{KK} \times \varepsilon \tag{2}$$

In the above equation TTR is the time for the machine operator to report about the problem to YYY and WT is the waiting time, i.e. the time for YYY to start the repairs after the failure message is sent out (see Figure 1). The expected values $E(TTR)$ and $E(WT)$ would be found by studying the failures for which the KTT VANMA data existed. Since it is usually impossible to tell from the maintenance bills or the feedback reports whether a problem was critical or not, $E(TTR)$ and $E(WT)$ would be added for all failures.

RT_{KK} is the repair time in the maintenance bills and ε is an error coefficient to compensate for earlier discussed inaccuracies in the data given in these bills. This coefficient would be calculated by comparing the information about the same failures in the bills and KTT VANMA.

As mentioned earlier, preventive repairs were not reported via KTT VANMA, thus, the resulting non-productive time would have to be estimated. Since preventive maintenance activities are usually planed in advance, the machine is down only during the repairs. Then the non-productive time can be estimated as

$$DT_{PM} = RT_{KK} \times \varepsilon \tag{3}$$

Finally, substituting the non-productive time due to individual preventive repairs found from equation (3) and the downtime caused by the individual failures, calculated from equation (2) or KTT VANMA data, into equation (1) the total downtime would be found.

4 RESULTS

In this section the results of the study will be discussed. In addition we will mention a few issues with the use of KTT VANMA system that were discovered throughout this work.

4.1 Downtime Data in KTT VANMA

To determine the downtime directly, messages about the problem and the completed repairs had to be available for all failures and preventive repairs (as will be discussed later, for calculation of the response and the repair time, message about the start of the repairs was also needed). However, the analysis of KTT VANMA data has showed that records in the system were incomplete.

Incomplete means two things: (i) for some failures (and all but two preventive repairs) there were no records in KTT VANMA system, and (ii) in some cases one or two of the required messages were missing. This problem also meant that straightforward analysis of KTT VANMA records was not possible, because a message about a problem and the following message about the completed repairs sometimes referred to two different physical events.

To avoid such errors KTT VANMA's messages were compared against the entries in maintenance bills. The summary is shown in Table 1. It should be noted that KTT VANMA implementation date was different for different machines. Therefore, the data in the table corresponds to different periods and should not be used for comparison of reliabilities. It should also be noted that for some machines, for example M11, the number of problems reported via KTT VANMA was grater than the number of maintenance orders in the bills. This happened when the problem was reoccurring. Then, there were a few physical failures and corresponding KTT VANMA messages, but the costs were assigned to one maintenance order.

Reoccurring failures complicated the matching of KTT VANMA messages with the records in the bills. For example, when the problem occurred twice during the same day, it was impossible to tell which part of that day's repairs, shown in the bills, were related to the first and which to the second set of KTT VANMA messages. Another problem with matching of the records was the lack of information in failure descriptions. We also believe that for some insignificant problems no maintenance orders were opened, which for information did not appear in the invoices.

The last point to emphasize is that, even when both required messages were available, the data was not always useful for calculation of the downtime. We have noticed that sometimes, machine status

Machine	KTT VANMA installation date*	Number of failures / preventive repairs from KTT VANMA installation to 16.10.2008		Matches	Cases with sufficient data for direct downtime determination
		Maintenance bills**	KTT VANMA		
M01	16.07.2008	14	4	3	0
M02	02.09.2008	5	3	2	1
M03	N/A	N/A	N/A	N/A	N/A
M04	07.09.2008	15	8	8	3
M05	13.09.2008	12	9	9	4
M06	28.08.2008	7	1	0	1
M07	13.08.2008	11	8	5	2
M08	20.08.2008	12	10	8	5
M09	10.09.2008	1	1	0	1
M10	N/A	N/A	N/A	N/A	N/A
M11	13.08.2008	22	27	24	14
M12	22.08.2008	10	6	4	3
M13	05.06.2008	37	20	17	10
M14	25.05.2008	31	14	11	8
M15	20.05.2008	31	27	24	11

* Date when the first message was sent via KTT VANMA.
** Only those orders where physical repairs were carried out.

Table 1: Comparison of KTT VANMA records against the data in maintenance bills.

remained unchanged for rather long time. For example, when repairs were started late in the evening, the message that the machine was back in production was sent only next morning. We suspect that in such cases repairs were completed the same evening, but the production was not restarted and the status of the machine was not changed until the start of the new shift. If so, the data could still be used, but the calculated downtime could then include the time, when the machine would not be running under normal conditions (unless it was scheduled to operate 24 hours a day).

Another example is abstract messages, for example: “Forgot to change the status. Machine was in production since yesterday afternoon”. Such messages had to be ignored as it was not clear when the production was restarted.

The conclusion to be drawn here is that KTT VANMA messages did not cover all maintenance activities on the machines, or information was insufficient to determine the downtime. Therefore the remaining data had to be estimated following the procedure described in Section 3.

4.2 Estimation of Mean Time to Report

When a failure occurs, the operator usually tries to fix it. If the problem is resolved, the operator might only send a message informing the maintenance company that such problem has happened or that it occurs time by time and must be looked at some time later. Otherwise the operator informs that the machine is down and must be repaired as soon as possible.

The study of the time it took before operators informed the maintenance company about the problems (either critical or non-critical) is summarised in Table 2. Note that for this analysis the message about the problem had to be available and the time when the problem first occurred had to be specified. However, in 91 out of the total of 165 messages the latter information was missing. We have assumed that in such cases the problem was reported immediately and assumed the time to report was one minute.

Type of event	Number of data points	Time in minutes		
		Average	Minimum	Maximum
Critical failure	112	34	0	800
Non-critical failure	53	87	1	3242
OVERALL	165*	51	0	3242

* Data for the period from 20.05.2008 to 07.11.2008 was used.

Table 2: Estimation of mean time to report.

The average time to report turned out to be quite long. We have observed that the time to report was particularly long for failures, which occurred at night or during the weekend. Maintenance engineers have confirmed that when a problem happens during the night or during the weekend, when they are not at work, machine operators sometimes wait until morning before calling for help.

From Table 2 it can also be noticed that critical failures were reported much quicker. When analysing the maintenance bills, thought, it is often impossible to tell whether the problem was critical or not. Therefore, when estimating the downtime on the bases of the bills, we would usually have to use the less accurate overall average.

4.3 Estimation of Waiting Time

The waiting time is the time it takes before maintenance engineers start the repairs after the information about the failure is sent out. To study it the message about the problem and the message about the start of the repairs were required. However, the latter message turned out to be the most often missing one, which is probably due to the fact that the main focus for both the case and the maintenance companies is the machine uptime, while the repair time itself is of the secondary importance. Abstract messages, such as “Andreas is coming in ten minutes”, or “Gunnar is coming to the machine, but must finish with M15 first” were also common. We suppose that such messages were sent to let the operator know that maintenance engineers were aware of the problem, but for the analysis most of them were useless.

The above problems reduced the number of data points available for analysis, which is summarised in Table 3. Here again we have distinguished between the critical and non-critical failures. In both cases there were a few very long responses, which corresponded to failures that happened during the weekend. Nevertheless, a clear difference between critical and non-critical failures can be observed. Unfortunately, as mentioned before, when analysing the maintenance bills it is often impossible to tell whether the problem was critical or not. Therefore the overall estimates would have to be used, which means that the response time and the downtime would be overestimated.

Moreover, it should be noted that in four additional cases it was possible to determine the lower limit and in 38 more cases the upper limit (response time cannot be longer than the total downtime) of the response time. Considering these data led to slightly shorter overall average response time. Taking into account that the upper limit already means that the response time is overestimated we can conclude that the real response time is shorter than that shown in Table 3.

Type of event	Number of data points	Time in minutes		
		Average	Minimum	Maximum
Critical failure	52	190	0	1698
Non-critical failure	16	1149	30	6550
OVERALL	68*	416	0	6550

* Data for the period from 20.05.2008 to 07.11.2008 was used.

Table 3: Estimation of mean waiting time.

4.4 Determining Coefficient ε

To determine the repair time error coefficient ε three conditions had to be met: (i) the messages about the start and the completion of the repairs had to be available, (ii) the matching order in the maintenance bill had to be found, and (iii) a correlation between the data in KTT VANMA and the bills had to exist.

As discussed in Section 4.3, the message about the start of the repairs was the most often missing one, and in Section 4.1 we have mentioned the difficulties when matching the records in KTT VANMA system and the maintenance bills. For this reason the data available for this analysis were scarce. In total 27 matches were found and were used to check for a possible correlation.

The results of the analysis are illustrated in Figure 2. It revealed that no clear correlation between the two data sets existed. Unfortunately this meant that the error coefficient ε, which was necessary for the estimation of the downtime, could not be found. Thus, at this point we had to stop the analysis and conclude that the existing data was insufficient to determine the availability of the machine tools.

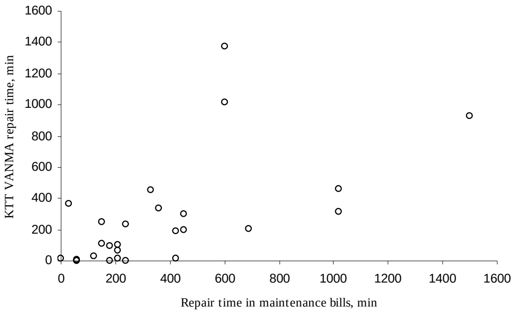


Figure 2: Relationship between repair time in KTT VANMA and maintenance bills.

5 CONCLUSIONS

In this study we have attempted to estimate the availability of the machine tools in the T-case cell. Copies of the messages from KTT VANMA failure reporting system as well as maintenance bills and feedback reports were collected for this work. None of these data sources alone were sufficient to estimate the availability of the machines, thus a model for combining the data from all three sources was developed.

The analysis, however, revealed that in some cases:

- Records in KTT VANMA system were incomplete or missing.
- There were difficulties to mach the records in KTT VANMA system and the maintenance bills due to assignment of cost of reoccurring failures to the same maintenance order, lack of information in failure description and neglecting of insignificant problems.
- After the late repairs the status of the machines was not changed until next morning.
- Abstract and usually useless messages were sent.
- The time when the problem first occurred was not specified.
- Some of the failures occurring late at nigh or during the weekend were not reported until maintenance engineers were back at work.
- There was no clear correlation between the repair time appearing in maintenance bills and the repair time calculated from KTT VANMA data.

Most of the above issues can probably be attributed to the newness of the KTT VANMA system. It can be expected that as people get more used to it the records should become more accurate and the analysis could be improved. So far, however, it must be concluded that on the on the bases of the available data it was not possible to determine all the required parameters and estimate the availability of machine tools in the T-case cell.