A review of tool wear estimation using theoretical analysis and numerical simulation technologies

Bin Li *

Department of Mechanical Engineering, Luoyang Institute of Science and Technology, Wangcheng Avenue 90, Luoyang 471023, Henan Province, PR China

A R T I C L E   I N F O

Article history:
Received 21 February 2012
Accepted 7 May 2012

Keywords:
Simulation
Tool wear
Metal cutting

A B S T R A C T

A thorough understanding of the material removal process in cutting is essential in selecting the tool material and in design, and also in assuring consistent dimensional accuracy and surface integrity of the finished product. Tool wear in cutting process is produced by the contact and relative sliding between the cutting tool and the workpiece, and between the cutting tool and the chip under the extreme conditions of cutting area. This paper presents the information on development of study on theoretical analysis and numerical simulation of tool wear in all over the world. Hidden Markov models (HMMs) were introduced to estimate tool wear in cutting, which are strongly influenced by the cutting temperature, contact stresses, and relative strains at the interface. Finite element method (FEM) is a powerful tool to predict cutting process variables, which are difficult to obtain with experimental methods. The objective of this work focuses on the new development in predicting the tool wear evolution and tool life in orthogonal cutting with FEM simulations.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Nowadays, metal cutting is a significant industry in economically developed countries, though small in comparison to the customer industries it serves. The automobile, railway, shipbuilding, aircraft manufacture, home appliance, consumer electronics and construction industries, all these have large machine shops with many thousands of employees engaged in machining. It is estimated that 15% of the value of all mechanical components manufactured worldwide is derived from machining operations [1–3].

A thorough understanding of the material removal process in metal cutting is essential in selecting the tool material and in design, and also in assuring consistent dimensional accuracy and surface integrity of the finished product. Friction of metal cutting influences the cutting power, machining quality, tool life and machining cost. When tool wear reaches a certain value, increasing cutting force, vibration and cutting temperature cause deteriorated surface integrity and dimension error greater than tolerance. The life of the cutting tool comes to an end. Then the cutting tool must be replaced or ground and the cutting process is interrupted. The cost and time for tool replacement and adjusting machine tool increase the cost and decrease the productivity. Hence friction in metal cutting relates to the economics of machining and prediction of friction is of great significance for the optimization of cutting process [4–6].

Although various theories have been introduced hitherto to explain the wear mechanism, the complicity of the processes in the cutting zone hampers formulation of a sound theory of cutting tool wear. The nature of tool wear in metal cutting, unfortunately, is not yet clear enough in spite of numerous investigations carried out over the last 50 years. Friction of metal cutting is a result of complicated physical, chemical, and thermo-mechanical phenomena. Recently, the prediction of friction of metal cutting is performed by calculating tool life according to experimental and empirical tool life equations. Although the equation gives the simple relationship between tool life and a certain cutting parameters, it gives only the information about tool life. For the researcher and tool manufacturer, tool wear progress and tool wear profile are also areas of concern. Tool life equation gives no information about the wear mechanism. But capability of predicting contributions of various wear mechanism is very helpful for the design of cutting tool material and geometry. In addition, such tool life equations are valid under very limited cutting conditions. For example, when tool geometry is changed, new equation must be established by making experiment [7,8].

Metal cutting is a highly nonlinear and coupled thermo-mechanical process, where the coupling is introduced through localized heating and temperature rise in the workpiece, which is caused by the rapid plastic flow in the workpiece and by the friction along the tool–chip interface. With the emergency of more and more powerful computer and the development of numerical technique, numerical methods such as finite element method (FEM) [9–11], finite difference method (FDM) and artificial Intelligence (AI) are widely used in machining industry. Among them, FEM has become a powerful tool in the simulation of cutting process. Various variables in the cutting process such as cutting
force, cutting temperature, strain, strain rate, stress, etc. can be predicted by performing chip formation and heat transfer analysis in metal cutting, including those very difficult to detect by experimental method. Therefore friction prediction method may be developed by integrating FEM simulation of cutting process with wear model [12–15]. The new challenge is the implementation of effective models in a finite element environment in order to perform a reliable prediction of tool wear. In fact, the introduction of the finite element codes and their computational improvements that make this tool able to approach machining process simulation represents a break point with the past and opens a new scenario [16–18].

This paper introduces the experimental evidence for what is the nature of the friction contact between chip and tool during chip formation and the historical development of friction models. The objective of this work focuses on some basic aspects of friction modeling in cutting and in particular on the influence of friction modeling on the effectiveness of the numerical simulation of the process.

2. Tool wear mechanisms and models

Tool wear has a large influence on the economics of the machining operations. Prediction of tool wear is complex because of the complexity of machining system. Tool wear in cutting process is produced by the contact and relative sliding between the cutting tool and the workpiece and between the cutting tool and the chip under the extreme conditions of cutting area; temperature at the cutting edge can exceed 1000 °C. Thus, knowledge of tool wear mechanisms and capability of predicting tool life are important and necessary in metal cutting. Any element changing contact conditions in cutting area affects tool wear [19–22]. These elements come from the whole machining system comprising workpiece, tool, interface and machine tool.

Workpiece material and its physical properties determine the cutting force and energy for the applied cutting conditions. The optimal performance of a cutting tool requires a right combination of the above tool parameters and cutting conditions [23]. Interface involves the interface conditions. Increasingly new technologies of interface, such as the minimum liquid lubrication, have been developed to reduce the cost of coolant. The dynamic characteristic of the machine tool, affected by the machine tool structure and all the components taking part in the cutting process, plays an important role for a successful cutting. Instable cutting processes with large vibrations result in a fluctuating overload on the cutting tool and often lead to premature failure of the cutting edge by tool chipping and excessive tool wear [24,25].

Type of tool wear contains crater wear on rake face and flank wear on flank face, which is shown in Fig. 1. Crater wear (Fig. 1) produces a wear crater of the tool rake face. The depth of the crater KT is selected as the tool life criterion for carbide tools. The other two parameters, namely, the crater width KB and the crater center distance KM are important if the tool undergoes re-sharpening [26].

Flank wear (Fig. 1) results in the formation of a flank wear land. For the purpose of wear measurement, the major cutting edge is considered to be divided into the following three zones: (a) Zone C is the curved part of the cutting edge at the tool corner; (b) Zone N is the quarter of the worn cutting edge of length b farthest from the tool corner; and (c) Zone B is the remaining straight part of the cutting edge between Zones C and N. The maximum VBb max and the average VBb average width of the flank wear are measured in Zone B, the notch wear VBn is measured in Zone N, and the tool corner wear VBk is measured in Zone C. As such, the following criteria for carbide tools are normally recommended: (a) the average width of the flank wear land VBb = 0.3 mm, if the flank wear land is considered to be regularly worn in Zone B; (b) the maximum width of the flank wear land VBb max = 0.6 mm, if the flank wear land is not considered to be regularly worn in Zone B. Besides, surface roughness for finish turning and the length of the wear notch VN = 1 mm can be used. However, these geometrical characteristics of tool wear are subjective and insufficient. First, they do not account for the tool geometry (the flank angle, the rake angle, the cutting edge angle, etc.), so they are not suitable to compare wear parameters of cutting tools having different geometries. Second, they do not account for the cutting regime and thus do not reflect the real amount of the work material removed by the tool during the tool operating time, which is defined as the time needed to achieve the chosen tool life criterion [26,27].

In order to find out suitable way to slow down the wear process, many works are carried out to analyze the wear mechanism in metal cutting [28–30]. It is found that tool wear is not formed by a unique tool wear mechanism but a combination of several tool wear mechanisms. Tool wear mechanisms in metal cutting include abrasive wear, adhesive wear, solution wear, diffusion wear, oxidation wear, etc., illustrated in Fig. 2 [31].

2.1. Abrasive wear

Tool material is removed away by the mechanical action of hard particles in the contact interface passing over the tool face. These hard particles may be hard constituents in the work material,
2.2. Adhesive wear

Adhesive wear is caused by the formation and fracture of welded asperity junctions between the cutting tool and the workpiece.

2.3. Diffusion wear

Diffusion wear takes place when atoms move from the tool material to the workpiece material because of the concentration difference. The rate of diffusion increases exponentially with the increase of temperature.

2.4. Oxidation wear

A slight oxidation of tool face is helpful to reduce the tool wear. It reduces adhesion, diffusion and current by isolating the tool and the workpiece. But at high temperature soft oxide layers, are formed rapidly, and then taken away by the chip and the workpiece. According to the temperature distribution on the tool face, it is assumed that crater wear is mainly caused by abrasive wear, diffusion wear and oxidation wear, but flank wear mainly dominated by abrasive wear due to hard second phase in the workpiece material [32].

As the tool wear in cutting operations involves complex wear mechanisms, researchers have attempted to directly correlate the results of tool life to the applied machining parameters (cutting speed, feed, etc.). Many models are developed to describe tool wear in quantity. They can be categorized into two types: tool life models and tool wear rate models, which are shown in Table 1 [33–36].

Tool life models give the relationship between tool life and cutting parameters or variables. For example, Taylor’s tool life equation reveals the exponential relationship between tool life and cutting speed [37], see Table 1. The constants \( n, C, A \) and \( B \) are defined by doing a lot of experiments with cutting speed changing and fitting the experimental data with the equation. It is very convenient to predict tool life by using this equation. In various sizes of cutting database, Taylor’s tool life equation and its extension versions under different cutting conditions appear most frequently. Tool life equations are suitable to very limited range of cutting conditions. As the new machining technologies, e.g. high-speed cutting or dry cutting, are getting spread in manufacturing industry, the existing tool life equations need to be updated with new constants and a lot of experimental work has to be done. In addition, except that tool life can be predicted by these equations, it is difficult to get further information about the tool wear progress, tool wear profile or tool wear mechanisms that are sometimes important for tool designers [38–40].

Tool wear rate models are derived from one or several wear mechanisms. They provide the information about wear growth rate due to some wear mechanisms. In these modes, the wear growth rate, i.e. the rate of volume loss at the tool face (rake or flank) per unit contact area per unit time, is related to several cutting process variables that have to be decided by experiment or using some methods [41,42]. Takeyama [34] derived a fundamental wear rate equation by considering abrasive wear, which is proportional to cutting distance, and diffusive wear. Mathew [35] analyzed the tool wear of carbide tools when machining carbon steels and results have shown that the Takeyama’s diffusion equation can be used to effectively relate the tool wear rate to the average contact temperature of the tool. At cutting temperature higher than 800 °C, the first abrasive term \( G(V, f) \) in Table 1, can be neglected. Molinari [43] proposed a new diffusion wear model by considering the contact temperature to be the main parameter controlling the rate of diffusion in the normal direction to the tool–chip interface. Usui [44] in the tool wear study for carbide tools derived a wear rate model based on the equation of adhesive wear, which involves temperature, normal stress, and sliding velocity at the contact surface. Usui’s model is derived from Shaw’s equation of adhesive wear [45,46]. Except the constants \( A \) and \( B \), Usui’s equation includes three variables: sliding velocity between the chip and the cutting tool, tool temperature and normal pressure on tool face. These variables can be predicted by FEM simulation of cutting process or combining analytical method and FDM. Therefore Usui’s equation is very practical for the implementation of tool wear estimation by using FEM or by using the combination of FDM and analytical method [47].

## Table 1

<table>
<thead>
<tr>
<th>Empirical tool life models</th>
<th>Tool wear rate models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taylor’s basic equation:</td>
<td>Takeyama’s wear model:</td>
</tr>
<tr>
<td>( L^n = C_1 (n, C_1 = \text{constants}) )</td>
<td>(considering abrasive wear and diffusive wear):</td>
</tr>
<tr>
<td>Taylor’s extended equation:</td>
<td>( G = C(V, f) + D \exp(-\frac{E}{RT}) )</td>
</tr>
<tr>
<td>( L = \frac{1}{p, q, r, C_1} )</td>
<td>( G, D = \text{constants} )</td>
</tr>
<tr>
<td>Taylor’s extended equation:</td>
<td>Usui’s wear model: (considering adhesive wear):</td>
</tr>
<tr>
<td>( V = \frac{F}{p, q, r, C_2} )</td>
<td>( B = B_0 (V, \exp(-\frac{E}{RT})) )</td>
</tr>
<tr>
<td>Taylor’s basic equation:</td>
<td>(considering abrasive wear):</td>
</tr>
<tr>
<td>( T^n = C_3 (n, C_4 = \text{constants}) )</td>
<td>( A, B = \text{constants} )</td>
</tr>
</tbody>
</table>

\( L = \text{tool life}; \sigma_0 = \text{normal stress}; T = \text{cutting temperature}; V = \text{cutting speed}; f = \text{feed rate}; E = \text{activation energy}; V_s = \text{sliding velocity}; d = \text{depth of cut}; R = \text{universal gas constant}; BHN = \text{workpiece hardness}; dW/dt = \text{wear rate (volume loss per unit contact area/unit time)}. \)
3. Theoretical analysis procedure of tool wear

Tool wear is a result of physical and chemical interactions between the cutting tool and workpiece that remove small parts of material from the cutting tool [48,49]. Even though abrasive and adhesive wear are the most commonly encountered mechanisms, the wear in cutting tools cannot be described completely by these mechanisms alone. Tool wear is a highly complicated process associated with many variables such as contact stress, nature and composition of the workpiece and cutting tool, temperature on the cutting edge, and cutting conditions [50].

3.1. Prediction theory of gray model

The gray model GM(1,1) is a time series forecasting model. It has three basic operations: accumulated generation, inverse accumulated generation, and gray modeling. The gray forecasting model uses the operations of accumulated to construct differential equations. Intrinsically speaking, it has the characteristics of requiring less data. The gray model GM(1,1), i.e., a single variable first-order gray model, is summarized as follows [51]:

For an initial time sequence,

\[ X^{(0)}(0) = \left\{ x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(i), \ldots, x^{(0)}(n) \right\} \tag{1} \]

where \( x^{(0)}(i) \) is the time series data at time \( i \), and \( n \) must be equal to or larger than 4.

On the basis of the initial sequence \( X^{(0)}(0) \), a new sequence \( X^{(1)}(0) \) is set up through the accumulated generating operation in order to provide the middle message of building a model and to weaken the variation tendency,

\[ X^{(1)}(0) = \left\{ x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(i), \ldots, x^{(1)}(n) \right\} \tag{2} \]

where

\[ x^{(1)}_{(k)} = \sum_{i=1}^{k} x^{(0)}(i), \quad k = 1, 2, \ldots, n \tag{3} \]

Mean sequence is

\[ z^{(1)}_{(k)} = 0.5 x^{(1)}_{(k)} + 0.5 x^{(1)}_{(k-1)}(k = 2, 3, \ldots, n) \tag{4} \]

and it also means

\[ z^{(1)}(k) = \left( z^{(1)}(2), z^{(1)}(3), \ldots, z^{(1)}(n) \right) \tag{5} \]

So the differential equations of this gray model can be expressed as

\[ x^{(0)}(k) + az^{(1)}(k) = b \quad (k = 2, 3, \ldots, n) \tag{6} \]

In the differential equation, \( x^{(0)}(k) \) is named as gray derivative, \( a \) is named as evolution system, \( z^{(1)}(k) \) is named as evolution system background values, \( b \) is named as gray action, if the numbers \( k = 2, 3, \ldots, n \) are introduced in the equation, the equation can be expressed as

\[
\begin{align*}
    x^{(0)}(2) + az^{(1)}(2) &= b \\
    x^{(0)}(3) + az^{(1)}(3) &= b \\
    \vdots \\
    x^{(0)}(n) + az^{(1)}(n) &= b \\
\end{align*}
\tag{7}
\]

and from Eq. (7), it is easy to get

\[
\begin{bmatrix}
    x^{(0)}(2) \\
    x^{(0)}(3) \\
    \vdots \\
    x^{(0)}(n) \\
\end{bmatrix} =
\begin{bmatrix}
    -z^{(1)}(2) & 1 \\
    -z^{(1)}(3) & 1 \\
    \vdots & \vdots \\
    -z^{(1)}(n) & 1 \\
\end{bmatrix} \times \begin{bmatrix} a \\ b \end{bmatrix} \tag{8}
\]

where \( a \) and \( b \) are the coefficients to be identified. Let

\[ Y_n = \begin{bmatrix} x^{(0)}(2), x^{(0)}(3), \ldots, x^{(0)}(n) \end{bmatrix}^T \tag{9} \]

and

\[ B = \begin{bmatrix} -z^{(1)}(2) & 1 \\
    -z^{(1)}(3) & 1 \\
    \vdots & \vdots \\
    -z^{(1)}(n) & 1 \end{bmatrix} \tag{10} \]

Also take

\[ Z^{(1)}(k+1) = \frac{1}{2} \left( x^{(1)}(k) + x^{(1)}(k+1) \right) k = 1, 2, \ldots, (n-1) \tag{11} \]

and

\[ A = [a, b]^T \tag{12} \]

where \( Y_n \) and \( B \) are the constant vector and the accumulated matrix respectively. Apply ordinary least-square method to Eq. (8) on the basis of Eqs. (9)–(12), coefficient \( A \) becomes

\[ A = \left( B^T B \right)^{-1} B^T Y_n \tag{13} \]

Substituting \( A \) in Eq. (7) with Eq. (13), the approximate equation becomes the following

\[ \hat{x}^{(1)}(k+1) = \left( x^{(0)}(1) - b/a \right) e^{-ak} + b/a \tag{14} \]

where \( \hat{x}^{(1)}(k+1) \) is the predicted value of \( \hat{x}^{(1)}(k+1) \) at time \( k+1 \). After the completion of an inverse accumulated generating operation on Eq. (14), \( x^{(0)}(k+1) \), the predicted value of \( x^{(0)}(k+1) \) at time \( k+1 \) becomes available and therefore,

\[ \hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \quad k = 0, 1, 2, 3, \ldots, n \tag{15} \]

3.2. Prediction theory of Markov chain

Hidden Markov models (HMMs) were introduced by Baum and colleagues at the end of the 1960s. Then, Baker and Jelinek implemented this theory for speech processing applications in the 1970s. Currently, HMM is a state-of-the-art technique for speech recognition because of its elegant mathematical structure and the availability of computer implementation of these models. Recently, HMM applications have been spreading steadily to other engineering fields that include communication systems, target tracking, machine tool monitoring, fault detection and diagnosis, robotics and character recognition [52–58].

If the observations are continuous signals (or vectors), a continuous probability distribution function in the form of a finite mixture is assigned to each state instead of a set of discrete probabilities [59]. In this case, the parameters of the continuous probability
distribution function are often approximated by a weighted sum of $M$ Gaussian distributions $\eta$.

$$b_j(O_i) = \frac{M}{m=1} \sum_{i=1}^{M} \ell \left( \mu_{jm}, U_{jm}, O_i \right), \quad 1 \leq j \leq N \tag{16}$$

where $\ell$ is a weighting (mixture) coefficient, $\mu_{jm}$ is the mean vector, $U_{jm}$ is the covariance matrix and $M$ is the number of mixture components, a typical observation sequence is denoted as $O = \{O_1, O_2, ..., O_T\}$. $\eta(\mu, U, O)$ is a multivariate Gaussian probability density function with mean vector $\mu$ and covariance matrix $U$, that is

$$\eta(\mu, U, O) = \frac{1}{\sqrt{(2\pi)^n |U|}} \exp \left( -\frac{1}{2} (O - \mu)^T U^{-1} (O - \mu) \right) \tag{17}$$

where $n$ is the dimensionality of $O$.

Consider the random process $\{x_n, n \in T\}$ and set of discrete conditions $I = \{i_0, i_1, i_2, ..., i_l\}$. If for any integer $n \in I$, the conditional probability as

$$p = \{x_{n+1} = i_{n+1}|x_0 = i_0, x_1 = i_1, ..., x_n = i_n\} = p = \{x_{n+1} = i_{n+1}|x_n = i_n\} \tag{18}$$

where the random process $\{x_n, n \in T\}$ can be called Markov chain, and the probability

$$p^{(k)}_{ij} = p(x_{m+k} = j|x_m = i), \quad (i, j \in I) \tag{19}$$

expresses the system probability at $m + k$ moment for state $j$, and the known condition is $m$ moment for state $i$, order the $P^{(k)}_I$, and we can get the following matrix:

$$P^{(k)} = \begin{bmatrix} p^{(k)}_{11} & \cdots & p^{(k)}_{1m} \\ \vdots & \ddots & \vdots \\ p^{(k)}_{n1} & \cdots & p^{(k)}_{nn} \end{bmatrix} \tag{20}$$

This matrix is called $k$ step transition probability matrix of Markov chain, and

$$\sum_{j=1}^{n} p^{(k)}_{ij} = 1 \tag{21}$$

Hidden Markov models are an extension of Markov chains. A Markov chain is a random process of discrete-valued variables involving a number of states linked by a number of possible transitions. At different times, the system is in one of these states; each transition between the states has an associated probability, and each state has an associated observation output for tool wear estimation.

### 3.3. Prediction of tool wears with state division

The classification methods of state division are relative value method, relative error method, standardization of residual deviation method and parallel curve method [60]. Based on the prediction data of $\hat{x}^{(k)}(k)$ curves in the gray model GM(1,1), and bar areas paralleled with prediction data curves are divided, every bar areas constitute the corresponding state interval $Q_e$ which can be expressed as:

$$Q_i = [Q_{1i}, Q_{2i}], \quad i = 1, 2, ..., n$$

$$Q_{1i} = \hat{x}^{(0)}(k) + C_i$$

$$Q_{2i} = \hat{x}^{(0)}(k) + D_i \tag{22}$$

where $C_i$ and $D_i$ are the constant vector based on data of tool wears, and $n$ is the number of state interval.

Suppose the component of state transition probability matrix $P^{(k)}_j$ is the $P^{(k)}_{ij}$, and

$$P^{(k)}_{ij} = M_{ij}(k)/M_i, \quad i, j = 1, 2, ..., n \tag{23}$$

expresses the system probability for state $j$, and the known condition is $m$ moment for state $i$, order the $P^{(k)}_I$, which we can get in Eq. (20), $M_i$ is the sample number of primary data falling into the state $i$ according to certain probability; $M_{ij}(k)$ is the sample number of primary data from system moment $Q_k$ to moment $Q_l$ by $k$ step.

When we confirm the transfer moment $Q_k$, also the change interval $[Q_{1k}, Q_{2k}]$ could be ascertained, let the prediction data $Y(k)$ to the halfway point as follows:

$$Y(k) = \left( Q_{1k} + Q_{2k} \right) / 2 = \hat{x}^{(0)}(k) + (\hat{A}_1 + \hat{B}_1)/2 \tag{24}$$

### 4. Simulation procedure of tool wear

It is postulated in this study that the growth of tool wear can be evaluated at discrete points in time, although in reality it is a continuous process. The proposed procedure for predicting tool wear is shown in Fig. 3 [32]. In every calculation cycle, chip formation and heat transfer analysis jobs are submitted to analyze the steady-state cutting process and obtain the cutting process variable values necessary for the calculation of wear rate at steady state. Nodal wear rate is calculated by using the tool wear mathematical model. Based on the calculated nodal wear rate, a suitable cutting time increment is searched by program according to a user-specified VB increment value. Then the nodal displacement due to wear in the cutting time increment is calculated at every tool face node, and the tool geometry is updated according to the calculated nodal displacement. If the produced flank wear VB is smaller than the user-defined tool reshape criterion VB_max, a second tool wear calculation cycle starts with the updated tool geometry [61,62].

![Fig. 3. Procedure for predicting tool wear.](image-url)
4.1. Chip formation analysis

The chip formation model was created using the implicit FEM software, which uses an updated Lagrangian formulation. This means that the material is attached to the mesh, with periodic remeshing to avoid severe element distortion [63]. A very fine mesh was used near the tool tip in the deformation zone of the workpiece and the edge of the tool as shown illustratively in Fig. 4. Automatic end adaptive remeshing was used to take into account the severe distortion of mesh elements near the tool tip.

It is well accepted that the influence of strain, strain rate and temperature on work flow stress must be included in successful simulations of machining: many empirical equations have been developed to describe this. Johnson–Cook constitutive equation was used to describe the material's performance in the cutting process. The full expression for flow stress $\sigma$, including strain path effects, could be calculated as follow [64]:

$$\sigma = (B\varepsilon^n) \left[ 1 + C \ln \left( \frac{\dot{\varepsilon}}{1000} \right) \right] \left( \frac{\theta - \theta_m}{\theta_r - \theta_m} \right) + a \exp \left( -0.00005(\theta - 700)^2 \right)$$

(25)

where $\dot{\varepsilon}$, $\varepsilon$ reflect the strain rate and strain, $\theta$ is the temperature; $\theta_m$ is the temperature of melting point; $\theta_r$ is the room temperature, and the value is 20 °C. The coefficients $B$, $C$, $n$ and $a$ are material constants.

The shear failure model is based on the value of the equivalent plastic strain at element integration points [65]. When the equivalent plastic strain reaches the strain at failure $\dot{\varepsilon}_{pl}^f$, then the damage parameter $w$ exceeds 1, and material failure takes place. If at all the integration point material failure takes place, the element is removed from the mesh. The damage parameter, $w$, is defined as

$$w = \sum \left( \frac{\Delta \varepsilon_{pl}^f}{\varepsilon_{pl}^f} \right)$$

(26)

where $\Delta \varepsilon_{pl}^f$ is an increment of the equivalent plastic strain. The summation is performed over all increments in the analysis.

Normally, equivalent plastic strain at failure, $\dot{\varepsilon}_{pl}^f$, is obtained by using experimental methods [66]. By employing the continuous chip formation analyzing methods, it is possible to determine strain at failure without making any experiment. Observing the movement of material points on the chip underside and the machined workpiece surface in steady-state chip formation process, a separation area of the workpiece material could be found.

The velocity of material points at workpiece nodes on the chip underside and the machined surface is shown in Fig. 5, the separation area is between Node 1 and Node 3. The material above the separation area moves upwards into the chip and the material below the separation area moves downwards to join in the machined surface.

According to the sliding velocities of the workpiece material points along tool–chip interface, a more exact position of the max principal stress can be displayed with point tracking technology in Fig. 6. It can be assumed that material failure is taking place in this area. By varying cutting parameters or tool geometry, the dependency of strain at failure on temperature, strain rate, pressure, etc. also can be studied.

4.2. Heat transfer analysis

To reach a thermally steady state for the cutting tool, the simulation needs to be run for a much longer cutting period on the order of a half second. This will considerably increase the computational time and data storage capacity, making cutting simulations expensive. Therefore, an approximate method was used to obtain the steady state tool temperatures, which involves the pure heat transfer analysis for the tool only. In the cutting phase the cutting tool is heated by the heat flux acted on the tool–chip and tool–workpiece interface. The total heat flux is composed of frictional heat flux $q^f$ and conductive heat flux $q^c$. Frictional heat flux is created due to the sliding friction between the workpiece material and the tool face [67]. The amount of frictional heat flux into the cutting tool is calculated by Eq. (27) [68].

$$q^f = (1 - f) \eta \tau v_s$$

(27)

where $\tau$ is the frictional stress; $v_s$ is the sliding velocity; $\eta$ specifies the fraction of mechanical energy converted into thermal energy; and $f$ gives the fraction of the generated heat flowing into the workpiece.

Therefore frictional heat flux is influenced by chip form, sliding condition and contact with the tool face.

Conductive heat flux is caused by the temperature difference of tool–chip and tool–workpiece at the interface. It is governed by Eq. (28).

$$q^c = k(\theta_h - \theta_s)$$

(28)

where $q^c$ is the conductive heat flux crossing the interface from point A on the workpiece to point B on the cutting tool; $k$ is the gap conductance; and $\theta$ is the nodal temperature on the surface.

Therefore conductive heat flux is temperature dependent. Both heat flux components are varying from node to node and the basic nodal heat flux data can be obtained from the chip formation analysis.

The calculated heat flux on the tool surface that represents the interface heating from the chip/workpiece objects was applied as customized thermal boundary conditions on the contact elements. For
this purpose, a user subroutine was developed [69,70]. Fig. 7 plots the distribution of temperature on the chip and tool face at steady state in different cutting speeds by the finite element analysis. The cutting tool is aluminum based ceramic tool, and the workpiece is AISI 1045 steel, and the simulation is with the following parameters: cutting speed \( v = 160 – 320 \) m/min, cutting depth \( d_p = 0.2 \) mm, and feed rates \( f = 0.1 \) mm/r.

As the chip flow and tool temperature solutions were obtained in two separate phases, the process variables for the chip/workpiece and the tool had to be combined by an external program to display the results in their entirety in the post-processor [71].

4.3. Nodal wear rate calculation

Nodal wear rate varies with the cutting time. In the cutting phase, tool wear takes place under the contact of the tool with the workpiece. In cooling phase, nodal wear rate is equal to zero and no wear produced. Nodal average wear rate is calculated by Eq. (29) [32].

\[
\bar{w}_{ij} = \frac{\int_{t_k}^{t_{k+1}} w(t) \, dt}{Z}
\]

where \( \bar{w}_{ij} \) is nodal average wear rate; \( w(t) \) is the nodal wear rate; \( Z \) is the time span of one milling cycle; \( i \) is the nodal label; and \( j \) is the milling cycle number.

It is very difficult to get the function of nodal wear rate. But nodal wear rate values at some discrete time points can be obtained by sampling cutting process variables during chip formation and heat transfer analysis and then calculating the individual nodal wear rate values. Based on these nodal wear rate values, an approximate nodal average wear rate can be calculated by following Eq. (30).

\[
\bar{w}_{ij} = \frac{\sum_{k=1}^{n} (w_{ij,k} + w_{ij,k+1}) (t_{k+1} - t_k)}{Z}
\]

where \( n \) means that the entire milling cycle is divided into \( n - 1 \) small portions by \( n \) evenly spaced time points; \( k \) is the time point number; and nodal wear rate is calculated at every time point. In the real calculation, sampling of cutting process variables and the calculation of nodal wear rate are not performed in the entire milling cycle because no wear takes place in the cooling phase.

4.4. Tool wear prediction on rake and flank face

The tool wear on the rake face and tool tip normally follows a crater shape in the cross section. Therefore, tool rake adjustments by moving individual surface nodes in the normal direction can be used to account for varying wear depths along the tool rake face. The wear depth for each node is described by its nodal wear rate alone, as the wear rate (volume loss per unit area per unit time) is equivalent to the wear depth under the two-dimensional condition. Therefore, for the given cutting time increment \( \Delta t_k = t_{k+1} - t_k \), the nodal displacement caused by tool wear between two simulation cycles can be approximated by Eq. (31) [24].

\[
\Delta d_{ik} = \bar{w}_{ik} \cdot \Delta t_k, \quad i = 1, \ldots, N
\]
where \( w_{i,k} \) is the wear rate for the node \( i \) at time \( t_k \) and is assumed to be constant throughout the time period \( \Delta t_k \). \( N \) is the total number of the contact nodes. \( \Delta d_{i,j} \) the nodal displacement for the node \( i \) corresponding to \( \Delta t_k \). The total wear depth for the node \( i \) is thus equal to the vector sum of all incremental nodal displacements associated with the node \( i \) throughout the total cutting time.

Individual nodal movement method can also be applied to the tool flank face, if the obtained nodal wear rates on the flank wear land are uniform so that a realistic flat flank wear land can be generated. A relationship between the flank wear rate \( w \) and the flank wear width \( V_{FB} \) was derived from their geometrical definition. For an infinitesimal time increment \( dt \), the increase in \( V_{FB} \) can be expressed by Eq. (32) [25].

\[
\Delta V_{FB} = dl \tan \alpha + \frac{dl}{\tan \gamma} \approx \frac{dl}{\tan \gamma} = \frac{w dt}{\tan \gamma}
\]

where the tool rake angles \( \alpha \) and \( \gamma \) are the tool relief angles; and \( dl \) is the increase in flank wear depth \( (= w dt) \).

Fig. 8 reports the tool wear prediction on rake and flank faces with the tool model parameters \( KT \) (crater depth), and \( KM \) (crater position) and flank wear width \( V_{FB} \). The cutting tool is aluminum based ceramic tool, and the workpiece is AISI 1045 steel. It is seen from the figure that the location of the maximum wear rate is on the tool rake face and is nearly coincident with that of the maximum cutting temperature. In addition, a region of very low wear rates is observed close to the tool radius on the rake face side.

5. Conclusion

This paper presents the information on the development of study on theoretical analysis and numerical simulation of tool wear. Hidden Markov models were introduced to estimate tool wear in cutting, which are strongly influenced by the cutting temperature, contact stresses, and relative strains at the interface. Finite element method is a powerful tool to predict cutting process variables, which are difficult to obtain with experimental methods. The objective of this work focuses on the new development in predicting the tool wear evolution and tool life in orthogonal cutting with FEM simulations. The following conclusions were obtained:

1. A Markov chain is a random process of discrete-valued variables involving a number of states linked by a number of possible transitions. At different times, the system is in one of these states; each transition between the states has an associated probability, and each state has an associated observation output for tool wear estimation.

2. The simulations using a cutting tool with constantly updated rake face and flank face geometries have shown that it is possible to predict the evolution of tool wear at any given cutting time from FEM simulations by using the methodology proposed in this study.

3. As the chip flow and tool temperature solutions were obtained in two separate phases, the process variables for the chip/workpiece and the tool had to be combined by an external program to display the results in their entirety in the post-processor.

Acknowledgments

The work described in this paper is supported by the National Natural Science Foundation of China (No. 51105188), the Natural Science Foundation of Education Department of Henan Province (No. 12B460016), and the Development Program for Science and Technology of Luoyang. Thanks to the government support of Henan Province and Luoyang City.

References
