

Imaging Method: An Efficient Algorithm for Moving Target Tracking by UWB Radar

Dusan Kocur, Jan Gamec, Maria Švecová, Maria Gamcová, Jana Rovňáková

Dept. of Electronics and Multimedia Communications
Technical University of Košice
Park Komenského 13, 041 20 Košice, Slovak Republic
dusan.kocur@tuke.sk, jan.gamec@tuke.sk, maria.svecova@tuke.sk,
maria.gamcova@tuke.sk, jana.rovnakova@tuke.sk

Abstract: This paper presents the imaging method applied for moving target tracking by a multistatic ultra-wideband radar system. The imaging method consists of such signal processing phases as raw radar data pre-processing, background subtraction, the fusion of the data obtained by the particular receiving channels of a radar, detection, localization and the tracking itself. In this paper, the particular phases of the imaging method are described. Firstly, the theoretical base of the particular phases is denoted and then an overview of the signal processing methods which can be applied within corresponding phase is given. Here, we will also outline that a signal cross-talk, exponential averaging, 2D double-stage (N,k) detector, a target gravity center estimation and a low-complex nonlinear two-stage tracking filter can be used within the mentioned radar signal processing phases as convenient signal processing methods. The good performance of the imaging method presented in this contribution is illustrated by processing signals obtained by the ultra-wideband radar for a scenario represented by through wall tracking of a single moving target.

1 Introduction

The word radar is an abbreviation for Radio Detection And Ranging. In general, radar systems use modulated waveforms and antennas to transmit electromagnetic energy into a specific volume in space to search for targets. Objects (targets) within a search volume will reflect portions of this energy (radar returns or echoes) back to the radar. Then, these echoes are processed by the radar receiver to extract target information such as range, velocity, angular position, and other target identifying characteristics [9]. If the fractional bandwidth of the signals emitted by the radar is greater than 0.20 or if these signals occupy 0.5 GHz or more of the spectrum, the radar is referred to as the ultra-wideband radar (UWB

radar). Example waveforms of that kind include impulse (video pulse), coded impulse trains (e.g. M-sequence), stepped frequency, pulse compression, random noise, and other signal formats that have high effective bandwidths [22]. UWB technology in radar allows for a very high accuracy ranging, rigidity to multi-path propagation and external EMI [28]. With regard to these properties, UWB radars have become very popular for military, rescue, automotive, and medical applications as well as for material characterization in recent years.

For the purpose of moving target detection and tracking, several UWB radars have been developed with promising results for through wall tracking of moving people during security operations, for through wall imaging during fire, for through rubble localization of trapped people following an emergency (e.g. an earthquake or explosion) and for through snow detection of trapped people after an avalanche. For these applications, UWB radar sensor networks operating in monostatic, bistatic, and multistatic modes can be used. Monostatic radars are systems in which the transmitter and receiver are collocated. In contrast, in the case of bistatic radars, the transmitter and receiver are not collocated. In multistatic radars a single transmitter is monitored by multiple, dispersed receivers [26].

UWB radar sensor output, referred to as raw radar data (signals), can be interpreted as a set of impulse responses from the surroundings, through which the electromagnetic waves emitted by the transmitting antenna of the radar were propagated. Then, the issue of moving target tracking by a multistatic radar system is comprised of the estimation of a target trajectory based on the processing of raw radar data obtained from all receiving antennas included in the UWB radar sensor network.

There are two basic approaches to raw radar signal processing which can be applied for that purpose. The former approach was originally introduced in [17] for through wall tracking of a moving target by using an M-sequence UWB radar equipped with one transmitting and two receiving antennas. Here, target coordinates as the function of time are evaluated by using an estimation of time of arrival corresponding to a target to be tracked and electromagnetic wave propagation velocity along the line transmitting antenna-target-receiving antenna.

The later approach is based on radar imaging techniques, when the target locations are not calculated analytically, but rather targets are seen as radar blobs in gradually generated radar images [4]. For the radar image generation, different modifications of a back projection algorithm can be used [4], [5], [19]. With regards to the fundamental concept of the method – the radar image generation based on raw radar data – the method is sometimes referred to as the imaging method. In the imaging method, moving target tracking, i.e. determining target coordinates as the continuous function of time, is a complex process that includes the following phases-tasks of radar signal processing [7]: raw radar data pre-processing, background subtraction, fusion of the data obtained by all receiving channels of the radar, detection, localization and tracking itself.

In this paper, the imaging method for moving target tracking by multistatic UWB radar systems will be described. In the next section, a real through wall scenario of single target tracking will be presented. As the radar device considered for the scenario, the M-sequences UWB radar equipped with one transmitting and two receiving antennas will be used [1], [18]. Section 3 is the core of our contribution. In this Section, the particular phases of imaging method will be introduced. In order to provide a brief but comprehensive characterization of the particular phases, the significance of the particular phases will be introduced firstly and then a review of signal processing methods frequently used for the phase task solution will be presented. The outputs of each phase of the imaging method will also be illustrated by the results of signal processing obtained in the scenario described in Section 2. Conclusions will be drawn and final remarks made in Section 4.

2 Basic Scenario of Moving Target Localization

The basic scenario of the target tracking by a multistatic UWB radar system used for the illustration of the imaging method is outlined in the Figs. 1-3. The scenario is represented by moving target positioning through two light concrete walls. The thickness of the first and the second walls were 50 cm and 40 cm, respectively. The walls were arranged parallel way at distance 2.25 m. A person walked along the perimeter of Room 2 from Pos. 14 through Pos. 15, Pos. 21 and Pos. 20 back to Pos. 14 (Fig. 1).



Figure 1

Measurement scenario



Figure 2
M-sequence UWB radar system

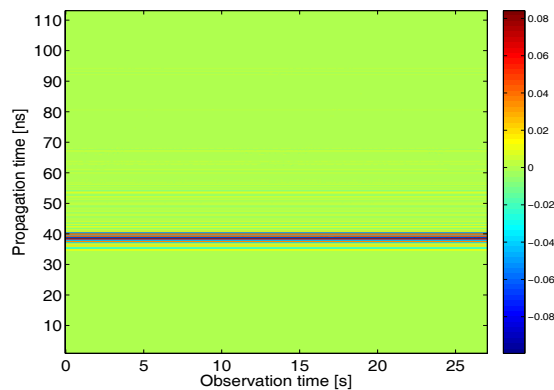
The raw radar data analyzed in this scenario were acquired by means of a multistatic M-sequence UWB radar (Fig. 2) with one transmitting (T_x) and two receiving channels (R_{x1} , R_{x2}) [1], [18]. The system clock frequency for the radar device is about 4.5 GHz, which results in an operational bandwidth of about DC-2.25 GHz. The M-sequence order emitted by the radar is 9, i.e. the impulse response covers 511 samples regularly spread over 114 ns. This corresponds to an observation window of 114 ns leading to an unambiguous range of about 16m and a radar resolution of about 3,3 cm. 256 hardware averages of environment impulse responses are always computed within the radar head FPGA to provide a reasonable data throughput and to improve the SNR by 24 dB. The additional software averaging can be provided by the basic software of the radar device. In our measurement, the radar system was set in such a way as to provide approximately 10 impulse responses per second. The total power transmitted by the radar was about 1 mW. The radar was equipped with three double-ridged horn antennas placed in a line (Fig. 1). Here, one transmitting antenna was located in the middle between two receiving antennas.



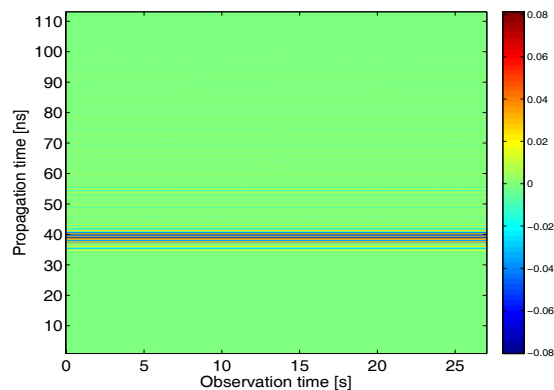
Figure 3
Measurement room interior

The raw radar data obtained by the measurement according to above described scenario are aligned to each other to create a 2D picture called a radargram, where the vertical axis is the time propagation (t) of the impulse response and the horizontal axis is the observation time (τ). The radargrams obtained for the scenario described in this Section are given in the Fig. 4.

The mentioned impulse responses (i.e. the impulse responses of the surroundings, through which the electromagnetic waves emitted by the transmitting antenna of the radar are propagated) are given by a correlation analysis of digital signals obtained by the analogue-to-digital conversion of the voltage at the terminals of the receiving antenna of the radar. With regard to the physical significance of the discussed impulse responses and taking into account the complex process of their creation, the impulse responses and the other quantities obtained based on their processing are considered to be dimensionless physical quantities.



(a)



(b)

Figure 4

Basic scenario: radargrams: (a) Receiving channel Rx1; (b) Receiving channel Rx2

3 Imaging Method: Basic Phases and Performance Illustration

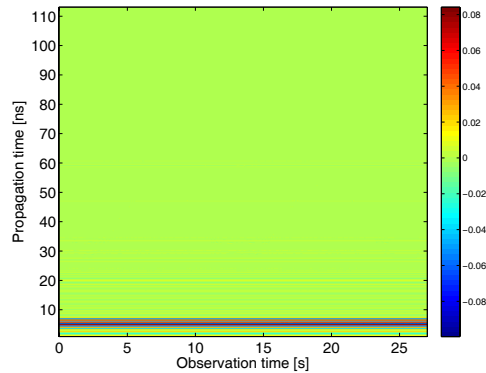
In the case of UWB radar signal processing by imaging method applied for moving target tracking by a multistatic UWB radar, target positioning is a complex process that includes such signal processing phases as raw radar data pre-processing, background subtraction, fusion of the data obtained from the particular receiving channels (antennas) of the radar, detection, localization and tracking [4], [7], [17]. In the next parts of this Section, the significance of particular phases will be outlined and a list of the methods most frequently used for these phases will be given. The outputs of the particular phases will be illustrated by the results of processing of the signals obtained in the basic scenario.

3.1 Raw Radar Data Pre-Processing

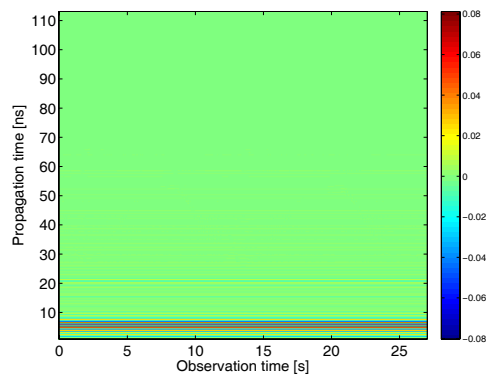
The intention of the raw radar data pre-processing phase is to remove or at least to decrease the influence of the radar systems itself on the raw radar data. In our contribution, we will focus on the problem of time-zero setting.

In the case of the M-sequence UWB radar, its transmitting antenna transmits M-sequences periodically. The exact time instant at which the transmitting antenna starts emitting the first elementary impulse of the M-sequence (so-called chip) is referred to as the time-zero. This depends e.g. on the cable lengths between the transmitting / receiving antennas and the transmitting/receiving amplifiers of the radar, the total group delays of the radar device's electronic systems, etc.; this especially depends on the chip position at which the M-sequence generator started to generate the first M-sequence. This position is randomly changed after every power supply reconnection. To find the time-zero means to rotate all the received impulse responses in such a way that their first chips correspond to the spatial position of the transmitting antenna. There are several techniques for finding the number of the chips needed for such rotating of the impulse responses. The method most often used is to utilize signal cross-talk [29]. The significance of the time-zero setting comes from the fact that the targets cannot be localized correctly without the correct time-zero setting.

The examples of the radargrams obtained by the measurement according to the basic scenario with the correct time-zero setting utilizing the signal cross-talk method are given for the first and second receiving channel in Figs. 5(a) and 5(b), respectively.



(a)



(b)

Figure 5

Pre-processed radargrams: (a) Receiving channel Rx1; (b) Receiving channel Rx2

3.2 Background Subtraction

It can be observed from the radargrams with the correct zero-time setting that it is impossible to identify any targets in these radargrams. The reason is the fact that the components of the impulse responses due to the target are much smaller than those of the reflections from the front wall and the cross-talk between the transmitting and receiving antennas or from those of other large or metal static objects. In order to be able to detect, localize and track a target, the ratio of signal scattered by the target to noise has to be increased. For that purpose, background subtraction methods can be used. They help to reject especially the stationary and correlated clutter, such as antenna coupling, impedance mismatch response and ambient static clutter, and they allow the response of a moving object to be detected.

Let us denote the signal scattered from the target and received by the n -th receiving antenna (Rx_n) as $s^n(t, \tau)$ and all other waves and noises received by Rx_n are denoted jointly as background $b^n(t, \tau)$. Let us assume also that there is no jamming at the radar performance, and the radar system can be described as linear one. Then, the raw radar data obtained from Rx_n can be simply modelled by the following expression:

$$h^n(t, \tau) = s^n(t, \tau) + b^n(t, \tau). \quad (1)$$

As is indicated by the name, the background subtraction methods are based on the idea of subtracting the background (clutter) estimation from the pre-processed raw radar data.

Then, the result of the background subtraction phase can be expressed as

$$\begin{aligned} h_b^n(t, \tau) &= h^n(t, \tau) - \hat{b}^n(t, \tau) = \\ &= s^n(t, \tau) + [b^n(t, \tau) - \hat{b}^n(t, \tau)] \end{aligned} \quad (2)$$

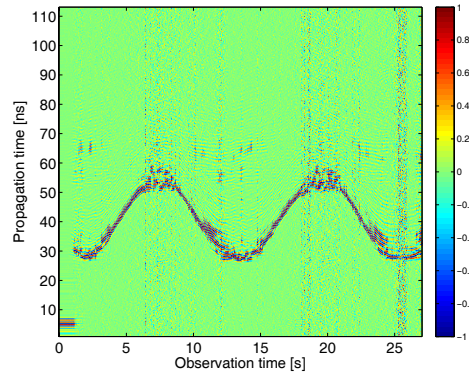
where $h_b^n(t, \tau)$ represents a set of the impulse responses with subtracted background and

$$\hat{b}^n(t, \tau) = \left[h^n(t, \tau) \right]_{\tau_1}^{\tau_2} \quad (3)$$

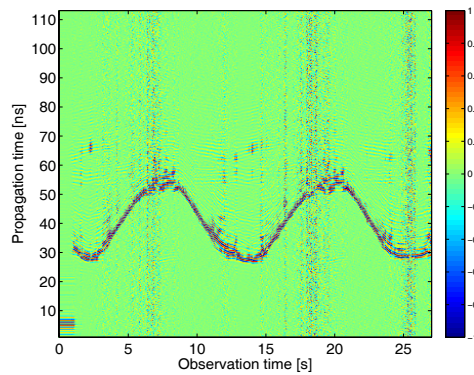
is the background estimation obtained by $h^n(t, \tau)$ processing over the interval $\tau \in \llbracket \tau_1, \tau_2 \rrbracket$.

In the above outlined scenario, it can very easily be seen that $s^n(t, \tau)$ for $t = \text{const.}$ represents a non-stationary component of $h^n(t, \tau)$. On the other hand, $b^n(t, \tau)$ for $t = \text{const.}$ represents a stationary and correlated component of $h^n(t, \tau)$. Therefore, the methods based on the estimation of the stationary and correlated components of $h^n(t, \tau)$ can be applied for the background estimation.

It has been shown in the literature that methods such as basic averaging (mean, median) [13], exponential averaging [30], adaptive exponential averaging [30], adaptive estimation of Gaussian background [27], Gaussian mixture method [20], moving target detection by FIR filtering [11], moving target detection by IIR filtering [12], prediction [24], principal component analysis [23], etc. can be used for the background subtraction. These methods differ in relation to assumptions concerning the properties of the clutter as well as to their computational complexity and suitability for online signal processing.



(a)



(b)

Figure 6

Radargram with subtracted background: (a) Receiving channel Rx1; (b) Receiving channel Rx2

Because of the simplicity of the scenario discussed in this contribution, a noticeable result can be achieved by using the simple exponential averaging method [30], where the background estimation is given by

$$\hat{b}^n(t, \tau) = \alpha \hat{b}^n(t, \tau - 1) + (1 - \alpha) h^n(t, \tau) \quad (4)$$

where $\alpha \in (0, 1)$ is a constant exponential weighing factor controlling the effective length of the window over which the mean value and the background of $h^n(t, \tau)$ are estimated. The results of the background subtraction by using the exponential averaging method applied for raw radar data given in Fig. 5 are presented in Fig. 6. In this figure, high-level signal components representing the signal scattered by a moving target can be observed. In spite of that fact, there are still a number of impulse responses where it is difficult or even impossible to identify signal components due to electromagnetic wave reflection by a moving target.

3.3 Data Fusion

The intention of the fusion of the data obtained from the particular receiving channels of the radar is to create a radar image $I(x, y, \tau)$ expressing the total level of the signal scattered by the scanned area at the coordinates (x, y) for the observation time instant τ . To transform the impulse response with subtracted background (impulse response $h_b^n(t, \tau)$) into a radar image, different modifications of a back projection algorithm can be used [3], [4], [19].

Let us assume that the coordinates of the transmitting antenna Tx and the coordinates of the receiving antenna Rx_n ($n=1, 2, \dots, K$) are known and they are given by $Tx(x_{Tx}, y_{Tx})$ and $Rx_n(x_R^n, y_R^n)$ ($n=1, 2, \dots, K$). Then for a homogeneous scanned area, the back projected signal $I(x_i, y_i, \tau)$ at the pixel $P_i(x_i, y_i)$ of the room image plane is given by [4]:

$$I(x_i, y_i, \tau) = \sum_{n=1}^K I_n(x_i, y_i, \tau) = \sum_{n=1}^K |h_b^n(t_i(n), \tau)| \quad (5)$$

where

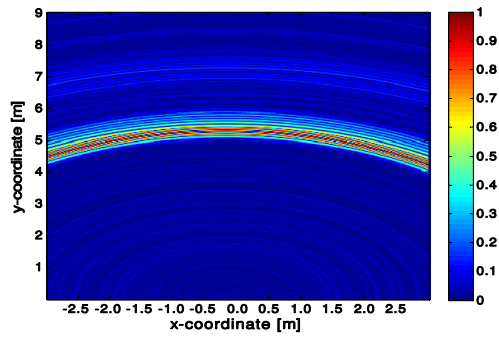
$$t_i(n) = (\|Tx P_i\| + \|P_i Rx_n\|) / v \quad (6)$$

$$\|Tx P_i\| = \sqrt{(x_i - x_{Tx})^2 + (y_i - y_{Tx})^2} \quad (7)$$

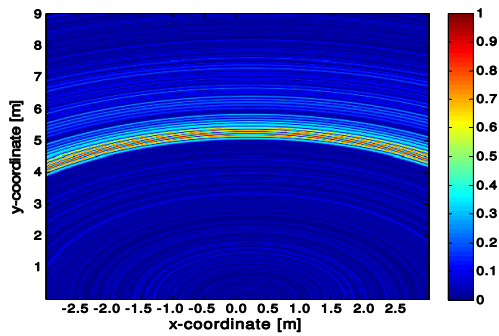
$$\|P_i Rx_n\| = \sqrt{(x_i - x_R^n)^2 + (y_i - y_R^n)^2} \quad (8)$$

In these expressions, $I_n(x_i, y_i, \tau)$ is the radar image corresponding to the impulse response $h_b^n(t, \tau)$, v is the velocity of the propagation of the electromagnetic wave emitted by the radar, and $t_i(n)$ is the total time for the transmitted signal to travel from Tx to the pixel $P_i(x_i, y_i)$ and then travel back from the pixel $P_i(x_i, y_i)$ to Rx_n . In expressions (6)-(8), the symbol $\|XY\|$ is set for the Euclidean distance between the points X and Y .

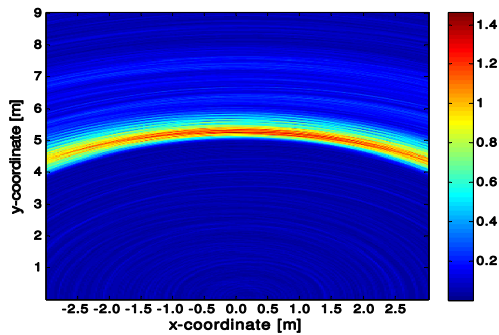
The result of this procedure is given by a sequence of 2D radar images $I(x, y, \tau)$ providing both the range and direction of a potential target motion. The radar images obtained by the described back projection algorithm for the scenario outlined in Section 2 and one chosen observation time instant τ are given in Fig. 7. In this figure, the radar images $I_1(x, y, \tau)$, $I_2(x, y, \tau)$ and $I(x, y, \tau)$ corresponding to the receiving channel $Rx1$, $Rx2$ as well as the final radar image obtained by the addition of the particular radar images are given, respectively.



(a)



(b)



(c)

Figure 7

The data fusion: (a) Radar image $I_1(x, y, \tau)$ corresponding to the receiving channel $Rx1$; (b) Radar image $I_2(x, y, \tau)$ corresponding to the receiving channel $Rx2$; (c) Radar image $I(x, y, \tau)$ created by the data fusion from both receiving channels

3.4 Detection

Detection is the next step in the radar signal processing which comes after the data fusion obtained from the receiving antennas of the radar. It represents a class of methods that determine whether a target is absent or present in the examined radar signals.

The solution of the target detection task is based on statistical decision theory [10], [22]. The detection methods analyze the radar image $I(x, y, \tau)$ obtained in the data fusion phase for the propagation time instant τ and reach a decision whether a signal scattered from a target $s(x, y, \tau)$ is absent (hypothesis H_0) or is present (hypothesis H_1) in $I(x, y, \tau)$. These hypotheses can be mathematically described as follows:

$$\begin{aligned} H_0 : I(x, y, \tau) &= n_{BS}(x, y, \tau) \\ H_1 : I(x, y, \tau) &= s(x, y, \tau) + n_{BS}(x, y, \tau) \end{aligned} \quad (9)$$

where $n_{BS}(x, y, \tau)$ represents the residual noise included in $I(x, y, \tau)$. A detector discriminates between the hypotheses H_0 and H_1 based on the comparison of testing statistics $X(x, y, \tau)$ and threshold $\gamma(x, y, \tau)$. Then, the output of the detector is a binary image $I_d(x, y, \tau)$ given by

$$I_d(x, y, \tau) = \begin{cases} 0 & \text{if } X(x, y, \tau) \leq \gamma(x, y, \tau) \\ 1 & \text{if } X(x, y, \tau) > \gamma(x, y, \tau) \end{cases} \quad (10)$$

The detailed structure of a detector depends on the selected strategy and optimization criteria of the detection [10], [14], [22]. Then the selection of the detection strategies and the optimization criteria results in a testing statistic specification and threshold estimation methods.

The most important groups of the detectors applied for radar signal processing are represented by sets of optimum or sub-optimum detectors. Optimum detectors can be obtained as a result of the solution of an optimization task usually formulated by means of probabilities or likelihood functions describing the detection process. Here, the Bayes criterion, the maximum likelihood criterion or the Neymann-Pearson criterion is often used as the bases for the detector design. However, the structure of the optimum detector be extremely complex. Therefore, sub-optimum detectors are also very often applied [22]. For the purpose of target detection by using UWB radars, detectors with fixed threshold, (N,k) detectors [25], IPCP detectors [22] and constant false alarm rate detectors (CFAR) [2] have been proposed.

For the purpose of the target detection at the imaging method, the 2D double-stage (N,k) detector can be used with advantage. This detector is a modification of the

(N, k) detector originally proposed in [25] and the 2D single-stage (N, k) detector. Following [25], the output of the single-stage 2D (N, k) detector for $N = 2l + 1$ can be described by (10) where:

$$X(x, y, \tau) = \sum_{i=1}^k I_{(N,k)}^2(i, \tau) \quad (11)$$

In this expression, $I_{(N,k)}(i, \tau)$ for $i = 1, 2, \dots, k$ represents the k -maximum values of $I^2(v, w, \tau)$ from its $N^2 = (2l + 1)^2$ samples if

$$v \in \{x - l, x - l + 1, \dots, x, \dots, x + l - 1, x - l\} \quad (12)$$

$$w \in \{y - l, y - l + 1, \dots, y, \dots, y + l - 1, y - l\} \quad (13)$$

In the case of (N, k) detectors it is assumed that $\gamma(x, y, \tau) = \text{const}$. Let us denote the input-output relations of the 2D single-stage (N, k) detector described by (10)-(13) as

$$I(x, y, \tau) \xrightarrow{(N,k)} I_d^{(N,k)}(x, y, \tau) \quad (14)$$

Then, the 2D double-stage (N, k) detector is given by the following expressions:

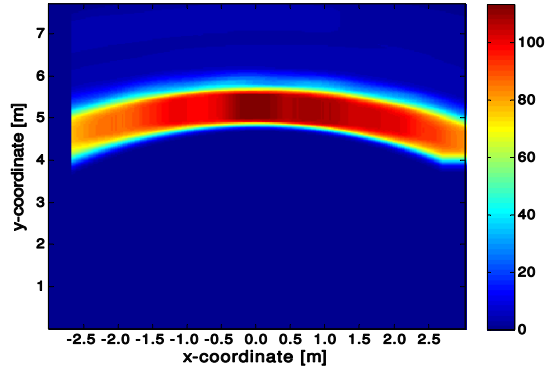
$$I(x, y, \tau) \xrightarrow{(N,k)} I_{d,1}^{(N,k)}(x, y, \tau) \quad (15)$$

$$I_X(x, y, \tau) = I_{d,1}^{(N,k)}(x, y, \tau) X(x, y, \tau) \quad (16)$$

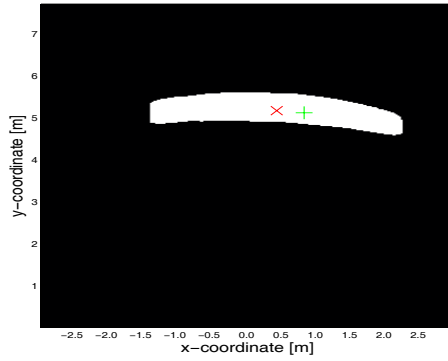
$$I_X(x, y, \tau) \xrightarrow{(N,k)} I_{d,2}^{(N,k)}(x, y, \tau) \quad (17)$$

$$I_d(x, y, \tau) = I_{d,2}^{(N,k)}(x, y, \tau) \quad (18)$$

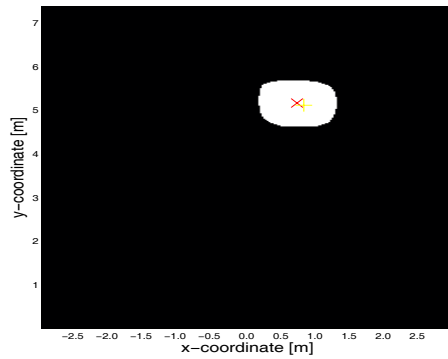
where $X(x, y, \tau)$ is given by (11). A detailed description and explanation of the 2D double-stage (N, k) detector given by (10)-(17) is beyond the scope of this article and they are not given here. Instead, the performance of the 2D double-stage (N, k) is illustrated for the basic scenario considered in this article. With this intention, the testing statistics for the second stage of the 2D double-stage (N, k) detector, the output of the first stage and the second stage of the detector denoted as $I_{d,1}^{(N,k)}(x, y, \tau)$ and $I_{d,2}^{(N,k)}(x, y, \tau)$ are presented in Fig. 8. These results have been obtained by the processing of the radar image $I(x, y, \tau)$ given in Fig. 7 (c).



(a)



(b)



(c)

Figure 8

The 2D double-stage (N, k) detector performance illustration: (a) Testing statistics of the second stage of the detector; (b) The output of the first stage of the detector; (c) The output of the second stage of the detector. The symbols “x” and “+” represent the true and estimated positions of the target gravity center, respectively

3.5 Localization

The aim of the localization phase is to determine the target coordinates in the defined coordinate systems. The target positions estimated in the consecutive time instants create a target trajectory.

If a target is represented by only one non-zero pixel of the detector output $I_d(x, y, \tau)$, then the target is referred to as a simple target. However, in the case of the scenario analyzed in this contribution, the radar range resolution (3,3 cm) is considerably finer than the physical dimensions of the target to be detected (a moving person). This results in the detector output being expressed not by only one non-zero pixel of $I_d(x, y, \tau)$; rather, the detector output is given by a complex binary image. The set of non-zero samples of $I_d(x, y, \tau)$ represents multiple-reflections of the electromagnetic waves scattered by a target or false alarms. The multiple-reflections due to the target are concentrated around the true target position at the detector outputs. In this case, the target is the distributed target. In the part of $I_d(x, y, \tau)$ where the target should be detected not only non-zero but also by zero samples of $I_d(x, y, \tau)$ can be observed. This effect can be explained by a complex target radar cross-section due to the fact that the radar resolution is much higher than that of the target size and taking into account the different shape and properties of the target surface. The set of false alarms is due to especially weak signal processing under very strong clutter presence.

Because the detector output for a distributed target is very complex, the task of distributed target localization is more complicated than for a simple target. For that purpose, an effective algorithm has been proposed in [17] for UWB radar signal processing by using time of arrival estimation. The basic idea of distributed target localization introduced in [17] consists in the substitution of the distributed target with a proper simple target. This basic idea can be applied in a modified form also for UWB radar signal processing by the imaging method. In this case, the distributed target is substituted by the simple target located in the center of gravity of the distributed target. The coordinates of the target gravity center $[x_T(\tau), y_T(\tau)]$ for observation time instant τ can be evaluated by

$$x_T(\tau) = \frac{1}{\sum_i \sum_j I(i, j, \tau)} \sum_i \sum_j i I_d(i, j, \tau)$$

$$y_T(\tau) = \frac{1}{\sum_i \sum_j I(i, j, \tau)} \sum_i \sum_j j I_d(i, j, \tau)$$
(19)

where the summation is made through all pixels belonging to the target to be tracked. Then the target coordinates as the output of the target localization phase are given by $[x_T(\tau), y_T(\tau)]$.

The influence of the particular detection stages upon the target positioning accuracy, if the target is substituted with its gravity center, is illustrated in Fig. 8 (b) and (c). In these figures, the symbols “x” and “+” represent the true and estimated positions of the target gravity center, respectively. The comparison of the target positions given in these figures indicates that the 2D double-stage (N,k) detector has the potential to overcome the 2D single-stage (N,k) detector performance.

The true trajectory and the trajectory estimated by the localization method of the target moving according to basic scenario outlined in Section 2 are given in Fig. 9. Here, the distributed target position has been substituted by the simple target position by use of the center of the gravity of the distributed target according to (19) applied to the output of the 2D double-stage (N,k) detector. It can be observed from Fig. 9 that the estimation of the target position based on the localization phase is very noisy, deformed and shifted along the y -axis. The detailed analysis of the noise present at the target trajectory estimate has shown that this effect is due to imperfect radar system performance under the hard conditions and the quality of the signal processing methods applied for raw radar data processing. In order to smooth out the target trajectory estimate, target tracking algorithms can be applied with advantage.

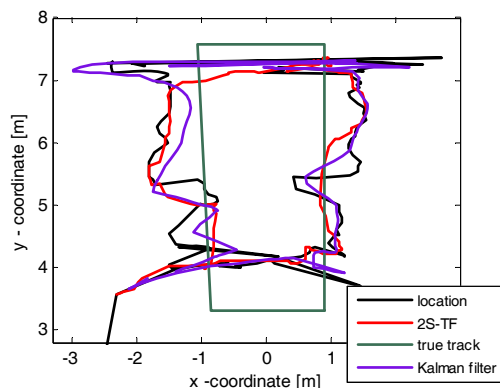


Figure 9

Target localization and tracking

The deformation and shifting of the target trajectory estimate observed in Fig. 9 are especially due to the wall effect [16]. In the case of the described imaging method we have assumed for simplicity that the environment through which the electromagnetic waves emitted by the radar are radiated is solely air. This is not true for through wall localization because the wall is a medium with different permittivity and permeability than that of air and hence, the electromagnetic wave propagation velocity in the air and wall are different. In addition to the mentioned quantities, wall thickness also has a strong influence on the target location

precision. Therefore, this so-called wall effect displaces the target outside of its true position, if through wall localization is based on the above mentioned simplified assumption. With regard to that fact, the precision of through wall localization can be improved if the quantities such as the permittivity, permeability and thickness of the wall are included in the target positioning procedure. For that purpose, the wall effect compensation method proposed in [4] can be used with advantage. Because, in our scenario, the wall permittivity and permeability were not known, the wall effect compensation was not considered here.

3.6 Tracking

The target position estimate can be improved by target tracking. Target tracking provides a new estimation of the target location based on its previous positions. Most of the tracking systems utilize a number of basic and advanced modifications of Kalman filters, such as for example linear, nonlinear and extended Kalman filters [15], [8], and particle filters [5]. In addition to Kalman filter theory, further methods of tracking are available. They are usually based on smoothing the target trajectory obtained through target localization methods. Here, the linear least-square method is also widely used (e.g. [6]). Another approach to target tracking is represented by joint target localization and tracking. Here, the Taylor series based tracking algorithm [21] can be given as the example.

The detailed analyses of the target coordinates obtained in the localization phase during through wall tracking of a moving target by the UWB radar have shown that the target coordinate estimation error does not possess the nature of additive Gaussian or impulsive noise. As the consequence, the “traditional” tracking or smoothing algorithms (e.g. averaging, Wiener filters, linear Kalman filters, median filters, etc.) will have reduced efficiency. A possible solution to this problem could be through, for example, the application of extended Kalman filters or particle filters. However, these alternatives are characterized by high computational complexity. Thus, a possible solution of that problem is the application of the low-complex nonlinear two-stage tracking filter (2S-TF) not requiring Gaussian noise assumption [7].

The performance of the linear Kalman filter and the 2S-TF is illustrated in Fig. 9 for the scenario outlined in Section 2. The tracking filters were applied for the improvement of the estimation of the target trajectory obtained in the localization phase also given in Fig. 9. It can be observed from this figure that the 2S-TF provides a much better estimation of the true trajectory of the target than the linear Kalman filter. The estimated trajectory of the target is still deformed and shifted, but it is much smoother and more similar to its true trajectory in comparison with that obtained by the linear Kalman filtering. It is very important to note that the better performance of the 2S-TF was achieved with its much less computational complexity in comparison with that of linear Kalman filter.

Conclusions

In this contribution, the imaging method for moving target tracking by the multistatic UWB radar systems was described. Firstly, we outlined the theoretical base of the particular phases of the imaging method and then we presented an overview of signal processing methods which can be applied in the corresponding phases. The imaging method, including its particular phases, was illustrated by a real UWB radar signal processing. The obtained results expressed by the visual comparison of the true trajectory of the target and the target trajectory estimations (Fig. 9) were revealed that the described imaging method can provide good results for through wall tracking of a single moving target by a multistatic UWB radar system equipped with one transmitting and two receiving antennas. The illustrated performance of the described imaging method was also confirmed by a number of similar results obtained for different scenarios (more precisely, for different kinds of the walls) of through wall tracking of a single moving target.

As follows from the method applied for the data fusion from the particular receiving antennas (expression (5)), the proposed approach of radar signal processing for the purpose of through wall tracking of a moving target can in principle be extended to scenarios employing a general multistatic UWB radar system (one transmitting and $n \geq 2$ receiving antennas) or to radar systems using $m \geq 2$ transmitting and $n \geq 2$ receiving antennas, usually referred to as MIMO radars. In the case of these modified scenarios and under the condition of multiple target tracking, the presented fundamental algorithm of data fusion must be supplemented by suitable methods of radar data association and advanced approaches to target localization. In order to improve the target positioning accuracy, the wall effect, the effect of an incidental dispersion and the effect of multiple reflections must be taken into account. The solution to these challenging problems will be the object of our follow-up research.

Acknowledgement

This work was supported by the Slovak Research and Development Agency under contract No. LPP-0080-09 (50%). This work is also the result of the project implementation Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF (50%).

References

- [1] Crabbe S. et al.: Ultra Wideband Radar for Through Wall Detection from the RADIOTECT Project, Fraunhofer Symposium, Future Security, 3rd Security Research Conference Karlsruhe, Karlsruhe, Germany, 2008
- [2] Dutta P., Arora A. K., Bibyk S. B.: Towards Radar-enabled Sensor Networks, The 5th International Conference on Information Processing in Sensor Networks (IPSN 2006) Special track on Platform Tools and Design Methods for Network Embedded Sensors, Nashville, Tennessee, USA, pp. 467-474, 2006

- [3] Engin E., Ciftcioglu B., Ozcan M., Tekin I.: High Resolution Ultrawideband Wall Penetrating Radar, *Microwave and Optical Technology Letters*, Vol. 49, No. 2, pp. 320-325, 2006
- [4] Gauthier S., Hung E., Chamma W.: Surveillance Through Concrete Walls, *Technical Memorandum, DRDC Ottawa TM 2003-233, Canada*, pp. 1-66, Dec. 2003
- [5] Grewal M. S., Andrews A. P.: *Kalman Filtering: Theory and Practice*, Prentice Hall, 2003
- [6] Hellebrandt M., Mathar R., Scheibenbogen M.: Estimating Position and Velocity of Mobiles in a Cellular Radionetwork, *IEEE Transactions on Vehicular Technology*, Vol. 46, No. 1, pp. 65-71, 1997
- [7] Kocur D., Gamec J., Švecová M., Gamcová M., Rovňáková J.: A New Efficient Tracking Algorithm for Through Wall Tracking of Moving Target by Using UWB Radar, *The 54th Internationales Wissenschaftliches Kolloquium, Ilmenau University of Technology, Ilmenau, Germany*, 2009
- [8] Krokavec, D., Filasová, A.: *Diskrétné systémy*. Elfa, Košice, 2008
- [9] Mafhaza B. R., Elsherbeni A. Z.: *MATLAB Simulation for Radar System Design*, Chapman & Hall, CRC Press LLC, 2004
- [10] Minkler G., Minkler J.: *CFAR: The Principles of Automatic Radar Detection in Clutter*, Magellan Book Company, 1990
- [11] Nag S., Barnes M.: Moving Target Detection Filter for an Ultra-Wideband Radar, *IEEE Radar Conference '03*, pp. 147-153, May 2003
- [12] Nag S., Fluhler H., Barnes H.: Preliminary Interferometric Images of Moving Targets Obtained Using a Time-modulated Ultra-Wide Band Through-Wall Penetration Radar, *IEEE Radar Conference '01*, pp. 64-69, 2001
- [13] Piccardi M.: Background Subtraction Techniques: a Review, *Proceedings of IEEE SMC International Conference on Systems, Man and Cybernetics, The Hague, The Netherlands*, Vol. 4, pp. 3099-3104, Oct. 2004
- [14] Poor H. V.: *An Introduction to Signal Detection and Estimation*, Springer Verlag, 1994
- [15] Ristic B., Arulampalam S., Gordon N.: *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Artech House, 2004
- [16] Rovňáková J., Kocur D.: Compensation of Wall Effect for Through Wall Tracking of Moving Targets, *Radioengineering*, Vol. 18, No. 2, pp. 189-195, 2009
- [17] Rovňáková J., Švecová M., Kocur D., Nguyen T. T., Sachs J.: Signal Processing for Through Wall Moving Target Tracking by M-sequence UWB Radar, *The 18th International Conference Radioelektronika, Prague, Czech Republic*, pp. 65-68, April 24-25, 2008

- [18] Sachs J. et al.: Detection and Tracking of Moving or Trapped People Hidden by Obstacles using Ultra-Wideband Pseudo-Noise Radar, The 5th European Radar Conference (EuRAD), Amsterdam, Netherlands, pp. 408-411, Oct. 2008
- [19] Sachs J., Zetik R., Peyerl P., Friedrich J.: Autonomous Orientation by Ultra Wideband Sounding, International Conference on Electromagnetics in Advanced Applications, Torino, Italy, 2005
- [20] Stauffer C., Grimson W.: Learning Patterns of Activity Using Real-Time Tracking, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 8, pp. 747-757, Aug. 2000
- [21] Švecová, M., Kocur, D.: Taylor Series-based Tracking Algorithm for Through Wall Tracking of a Moving Persons. Acta Polytechnica Hungarica, Vol. 7, No. 1, 2010, pp. 5-21
- [22] Taylor J. D.: Ultra-Wideband Radar Technology, CRC Press, 2001
- [23] Tipping M. E., Bishop C. M.: Mixtures of Probabilistic Principal Component Analysers, Neural Computation, Vol. 11, No. 2, pp. 443-482, Feb. 15, 1999
- [24] Toyama K., Krumm J., Brumitt B., Meyers B.: Wallflower: Principles and Practice of Background Maintenance, 7th IEEE International Conference on Computer Vision, Vol. 1, pp. 255-261, 1999
- [25] Van der Spek G. A.: Detection of a Distributed Target, IEEE Trans. on Aerospace and Electronic Systems, Vol. 7, No. 5, pp. 922-931, 1971
- [26] Withington P., Fluhler H., Nag S.: Enhancing Homeland Security with Advanced UWB Sensors, IEEE Microwave Magazine, Vol. 4, No. 3, pp. 51-58, Sept. 2003
- [27] Wren C. et al.: Pfinder: Real-time Tracking of the Human Body, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, pp. 780-785, July 1997
- [28] Yarovoy A.: Ultra-Wideband Radars for High-Resolution Imaging and Target Classification, European Radar Conference, Munich, Germany, Oct. 2007
- [29] Yelf R.: Where is True Time Zero? The 10th International Conference on Ground Penetrating Radar, Vol. 1, pp. 279-282, 2004
- [30] Zetik R. et al.: Detection and Localization of Persons Behind Obstacles Using M-Sequence Through-the-Wall Radar, Proceedings of SPIE, Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense, Vol. 6201, May 2006

Non-Classical Problems of Irreversible Deformation in Terms of the Synthetic Theory

Andrew Rusinko

Óbuda University

Népszínház u. 8, H-1081 Budapest, Hungary

E-mail: ruszinko.endre@bgk.uni-obuda.hu

Abstract: The paper is concerned with the generalization of the synthetic theory to the modeling of both plastic and creep deformation. Non-classical problems such as creep delay, the Bauschinger negative effect and reverse creep have been analytically described; the calculated results show satisfactory agreement with experiments. These problems cannot be modeled in terms of classical creep/plasticity theories. The main peculiarity of the generalized synthetic theory consists in the fact that the macro-deformation is highly associated with processes occurring on the micro-level of material.

Keywords: plastic deformation; primary creep; steady-state creep; the Bauschinger negative effect; creep delay; reverse creep

1 Introduction

The work presented herein regards the generalization of the synthetic theory of plastic deformation [18] to the modeling of not only plastic, but also creep (both primary and steady-state) deformation. This theory, which is concerned with small strains of work-hardening metals, incorporates (synthesizes) the Budiansky slip concept and the plastic flow theory developed by Sanders.

The key points of the generalized synthetic theory are:

I It is of both mathematical and physical nature. As a mathematical (formal) theory, the synthetic theory is in full agreement with the basic laws and principles of plasticity, such as Drucker's postulate, the law of the deviator proportionality, the isotropy postulate, etc. [18]. As a physical model, the synthetic theory allows for real processes occurring at the micro-level of material during loading, and the macro-behavior of material is fully governed by these processes. Therefore, the synthetic theory is a two-level theory.

II Independently of the type of deformation (creep or plastic) to be modeled, a single notion, irreversible (permanent) deformation is introduced, i.e. the

deformation is not split into “instantaneous” plastic and viscous parts [17]. The manifestation of the plastic or viscous component and their interrelations depend on the concrete loading/temperature-regime. The correctness to use the notion of irreversible deformation follows from the similarity of the mechanism of time-dependent and plastic deformation. Indeed, this mechanism is slips of the parts of crystal grains relative to each other. These slips are induced mainly by the motions of dislocations which, in turn, are induced/accompanied by other micro-structural imperfections (defects) of the crystalline lattice (vacancies, interstitial atoms, etc.). Undoubtedly, the driving forces and concrete configurations of the defects are different under different conditions. Nevertheless, despite the variety of processes occurring in a body subjected to different loading regimes, numerous experiments systematically record the arising of dislocation gliding for any type of inelastic straining. Other facts justifying the similarity of the nature of plastic and time-dependent deformation are **(i)** hydrostatic stress does not affect creep deformation; **(ii)** the axes of principal stress and creep strain rate coincide; **(iii)** no volume change occurs during creep [3]. These observations are the same as those for plastic deformation [6, 7].

III Following the tendency of unified approaches to the determination of irreversible deformation [4, 5], the system of constitutive equations that governs the whole spectrum of inelastic deformation has been worked out. In terms of generalized synthetic theory, the universality of this system is based on:

(i) a single equation provides the relation between a) micro-irreversible deformation, b) defects of crystalline structure inducing this deformation and c) time. Further, the procedure of the transition from micro- to macro-level is also uniformed: the sum of irreversible micro-strains determines the magnitude of macro-strain.

ii) the hardening rule is set in such a way that the transformation of loading surface obeys a unique rule. In addition, the kinetics of the loading surface transformation is not set a priori but is fully determined by the loading regime.

The objectives of this papers are to demonstrate how, by utilizing the uniformed method, the generalized synthetic theory is capable of embracing both plastic and creep deformation. In addition, some non-classical problems such as creep delay, the Bauschinger negative effect and reverse creep [12] are considered. The investigation of reverse creep is of great importance due to the fact that this phenomenon contradicts the hypothesis of creep potential [3, 13]. The advantages of synthetic theories above classical theories of creep and plasticity are considered.

2 Fundamentals of the Synthetic Theory of Plastic Deformation

The synthetic theory is based on the Budiansky slip concept [2] and the plastic flow theory developed by Sanders [19]. Below, the basic principles of the synthetic theory [18] are briefly reviewed.

A) The establishment of strain-stress relationships takes place in the Ilyushin stress deviatoric space, \mathbf{S}^5 , [8]. A load is presented by stress-deviator vector, $\bar{\mathbf{S}}$, whose components are defined as

$$\begin{aligned} S_1 &= \sqrt{3/2}S_{xx}, \quad S_2 = S_{xx}/\sqrt{2} + \sqrt{2}S_{yy}, \quad S_3 = \sqrt{2}S_{xz}, \\ S_4 &= \sqrt{2}S_{xy}, \quad S_5 = \sqrt{2}S_{yz}, \end{aligned} \quad (2.1)$$

where S_{ij} ($i, j = x, y, z$) are the stress-deviator tensor components; $|\bar{\mathbf{S}}| = 3\sqrt{2}J_2$, where J_2 is the second invariant of stress deviator tensor [7]. Further throughout we will consider the cases when $\bar{\mathbf{S}} \in \mathbf{S}^3$ ($S_4 = S_5 = 0$).

B) **Yield criterion and yield surface.** One of the key points consists in the construction of planes tangential to the yield surface in \mathbf{S}^5 instead of the yield surface itself. The inner-envelope of tangent planes constitutes the yield surface. By making use of this method, a new yield criterion is introduced, which coincides with neither the Tresca nor the von-Mises yield criterion in \mathbf{S}^5 . At the same time, the new criterion is reduced to the von-Mises yield criterion in \mathbf{S}^3 meaning that the trace of the five-dimensional yield surface takes the form of a sphere in \mathbf{S}^3 ($S_4 = S_5 = 0$):

$$S_1^2 + S_2^2 + S_3^2 = 2\tau_s^2, \quad (2.2)$$

where τ_s is the yield limit of a material in pure shear.

C) **Loading surface.** Following Sanders [19], the stress deviator vector shifts planes tangential to the yield surface on its endpoint during loading. The movements of the planes located on the endpoint of stress deviator vector are translational, i.e. without a change in their orientations. Those planes which are not on the endpoint of the stress deviator vector remain unmovable. Despite the fact the $\bar{\mathbf{S}} \in \mathbf{S}^3$, the displacements of planes tangential to the **five-dimensional yield surface** must be considered. On the other hand, the positions of these planes can be set by their traces in \mathbf{S}^3 . As a result, any plane in \mathbf{S}^3 (either tangential to the sphere (2.2) or locating beyond this sphere) is the trace of the plane tangential to the five-dimensional yield surface [18].

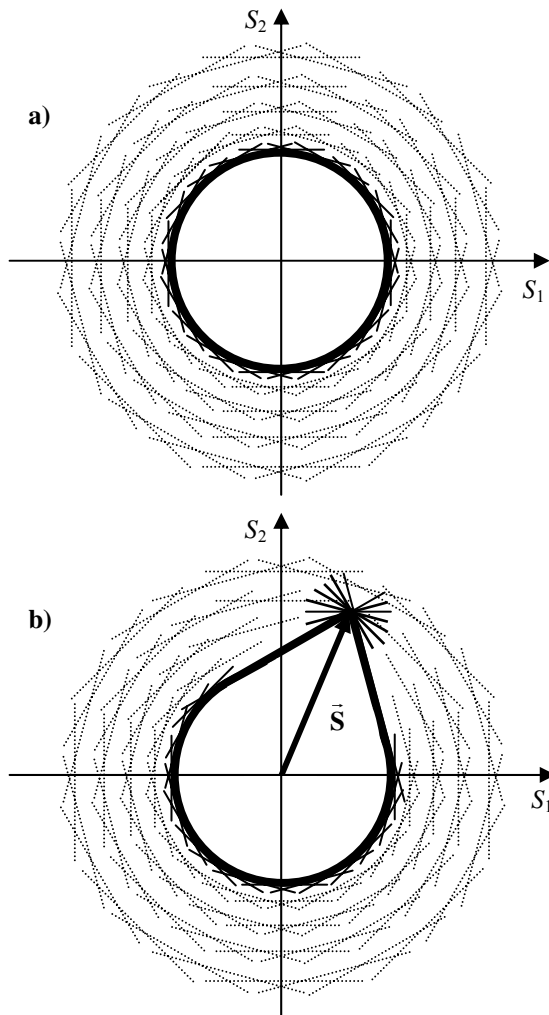


Figure 1

Yield and loading surface in terms of synthetic theory

The loading surface constructed as the inner-envelope of the tangential planes takes the shape fully determined by the current positions of the planes. Therefore, the behavior of the loading surface is not prescribed a priori, but is fully determined by the hodograph of the stress deviator vector.

For simplicity, let S_1 - S_2 coordinate-plane play the role of \mathbf{S}^3 . Then, Fig. 1a illustrates the yield surface (2.2) (circle) in the virgin state of the material. The planes (lines) tangential both to the five-dimensional yield surface and to its trace in \mathbf{S}^3 are shown as solid lines. The lines filling up the S_1 - S_3 plane beyond the

circle (the traces of the planes tangential to only the five-dimensional yield surface) are shown as dotted lines.

Fig. 1b shows loading surface due to the action of vector $\vec{S} \in \mathbf{S}^3$ which shifts a set of planes. It is easy to see that the corner point arises on the loading surface at the endpoint of \vec{S} (loading point). This fact is of great importance for the description of the peculiarities of plastic straining at non-smooth (orthogonal) loading trajectories [18] where any theory with regular loading surface has proved to be unsuitable.

The condition that the tangent plane is located on the end-point of the stress deviator vector can be expressed as

$$H_N = \vec{S} \cdot \vec{N}, \quad (2.3)$$

where H_N is the distance between the origin of coordinates and the tangent plane in \mathbf{S}^5 ; \vec{N} is the unit vector normal to the tangent plane, which defines the orientation of the plane. If the plane is not reached by \vec{S} , $H_N > \vec{S} \cdot \vec{N}$. The distance to plane in \mathbf{S}^5 can be expressed through that to its trace in \mathbf{S}^3 , h_m , as

$$H_N = h_m \cos \lambda, \quad (2.4)$$

where index m indicates the unit vector, \vec{m} , normal to the tangent plane in \mathbf{S}^3 :

$$\vec{m}(\cos \alpha \cos \beta, \sin \alpha \cos \beta, \sin \beta), \quad (2.5)$$

In expression (2.4), λ is the angle between the vectors \vec{m} and \vec{N} . The angles α and β are shown in Fig. 2.

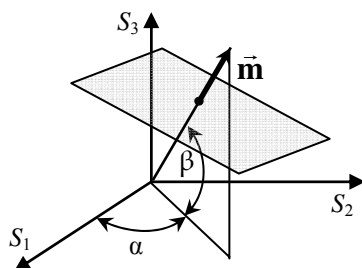


Figure 2
Orientation of normal vector \vec{m}

In addition, the \vec{N} and \vec{m} vector components are related to each other as [18]

$$N_k = m_k \cos \lambda, \quad k = 1, 2, 3$$

$$N_1 = \cos \alpha \cos \beta \cos \lambda, \quad N_2 = \sin \alpha \cos \beta \cos \lambda, \quad N_3 = \sin \beta \cos \lambda, \quad (2.6)$$

Therefore, expressions (2.3) and (2.6) give that

$$H_N = \vec{\mathbf{S}} \cdot \vec{\mathbf{m}} \cos \lambda = (S_1 m_1 + S_2 m_2 + S_3 m_3) \cos \lambda, \quad \vec{\mathbf{S}} \in \mathbf{S}^3. \quad (2.7)$$

As follows from formulae (2.4) and (2.6), if $\lambda = 0$, then $H_N = h_m$ and $N_k = m_k$ ($k = 1, 2, 3$). This holds true only for the planes which are tangential both to the five-dimensional yield surface and to its trace in \mathbf{S}^3 . It is these planes with $\lambda = 0$ that govern the transformation of the loading surface in \mathbf{S}^3 [18].

D) Plastic strain vector components. Similarly to the *Batdorf-Budiansky slip concept*, the synthetic theory is of a two-level nature. Each tangent plane represents an appropriate slip system at a point in a body (microlevel, see Fig. 3), and the plane motion symbolizes an elementary process of plastic deformation within this slip system.

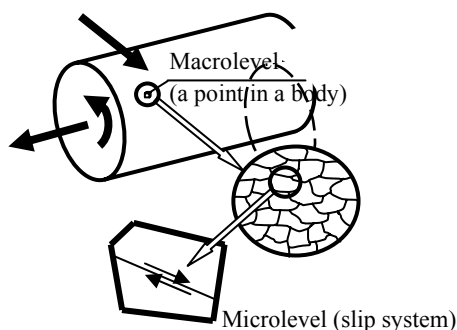


Figure 3

Two levels of the determination of deformation

To define an average, continuous measure of plastic slip within one slip system, we introduce a scalar magnitude, plastic strain intensity (φ_N), is proposed as

$$r\varphi_N = H_N - \sqrt{2}\tau_S = \vec{\mathbf{S}} \cdot \vec{\mathbf{N}} - \sqrt{2}\tau_S. \quad (2.8)$$

Formula (2.8) holds true for the planes displaced by the stress deviator vector, i.e. if $H_N = \vec{\mathbf{S}} \cdot \vec{\mathbf{N}}$. If $H_N > \vec{\mathbf{S}} \cdot \vec{\mathbf{N}}$, φ_N is set to be zero. An incremental plastic strain-vector, $d\vec{\mathbf{e}}^S$, (micro plastic deformation on the lower(micro)-level) is assumed to be in the direction of the outer normal to the plane and determined as

$$d\vec{\mathbf{e}}^S = \varphi_N \vec{\mathbf{N}} dV. \quad (2.9)$$

In expression (2.9), dV is an elementary volume constituted of the elementary set of planes in \mathbf{S}^3 that covered an elementary distance due to an infinitesimal increase in the stress vector [1]:

$$dV = \cos \beta d\alpha d\beta d\lambda. \quad (2.10)$$

The total (macro) plastic strain-vector at a point in a body, $\bar{\mathbf{e}}^S$, is determined as the sum (three-folded integral) of the micro plastic strains ‘produced’ by movable planes:

$$\bar{\mathbf{e}}^S = \int_V \varphi_N \bar{\mathbf{N}} dV \quad \text{or} \quad \dot{\bar{\mathbf{e}}}^S = \int_V \dot{\varphi}_N \bar{\mathbf{N}} dV \quad (2.11)$$

The strain vector components relate to the strain-deviator tensor components e_{ij} ($i, j = x, y, z$) as [8]

$$\begin{aligned} e_1 &= \sqrt{3/2} e_{xx}, & e_2 &= e_{xx}/\sqrt{2} + \sqrt{2} e_{yy}, & e_3 &= \sqrt{2} e_{xz}, \\ e_4 &= \sqrt{2} e_{xy}, & e_5 &= \sqrt{2} e_{yz}. \end{aligned} \quad (2.12)$$

By using equations (2.6) and (2.10), equation (2.11) becomes

$$\begin{aligned} e_k^S &= \iiint_{\alpha \beta \lambda} \varphi_N m_k \cos \lambda \cos \beta d\alpha d\beta d\lambda \quad \text{or} \\ \dot{e}_k^S &= \iiint_{\alpha \beta \lambda} \dot{\varphi}_N m_k \cos \lambda \cos \beta d\alpha d\beta d\lambda, \quad k = 1, 2, 3 \end{aligned} \quad (2.13)$$

The integration in (2.13) must be taken over planes shifted by the stress deviator vector.

3 The Generalization of the Synthetic Theory

To extend the boundaries of the applicability of the synthetic theory, the following is proposed.

D) To reflect the well-known fact that the defects of the crystal structure of metals are the *carriers* of irreversible deformation, a new notion, defect intensity (ψ_N), is introduced. ψ_N represents an average continuous measure of the defects (dislocations, vacancies, etc.) generated by irreversible deformation within one slip system.

II) To model the influence of loading rate upon irreversible straining, a new function of time and loading rate, the so called integral of non-homogeneity (I_N), is introduced. By considering the physical nature of irreversible deforming, the formula for I_N will be strictly derived in 3.1.2.

III) Instead of (2.8), the defect intensity is related to H_N and I_N :

$$\psi_N = H_N - I_N - \sqrt{2}\tau_P = \tilde{\mathbf{S}} \cdot \tilde{\mathbf{N}} - I_N - \sqrt{2}\tau_P, \quad (3.1)$$

where τ_P is the creep limit of material in pure shear. In terms of the generalized synthetic theory, the yield and creep limits are related to each other by equation derived further (see 4.1)). The establishment of a relationship between ψ_N and H_N is fully logical, because the distance H_N characterizes the degree of work-hardening. Indeed, the greater the plane distance, the greater a stress deviator vector is needed to reach the plane, i.e. to induce irreversible strain.

IV) To establish a relationship between irreversible deformation, defects and time (t), the following equation is proposed

$$d\psi_N = r d\varphi_N - K\psi_N dt, \quad (3.2)$$

where, r is the model constants and K is a function of homological temperature, Θ , and $|\tilde{\mathbf{S}}|$ (see 5). The units of quantities in (3.2) are $[\psi_N] = \text{Pa}$, $[\varphi_N] = 1$, $[r] = \text{Pa}$ and $[K] = \text{sec}^{-1}$.

In what follows, the parameter of non-homogeneity and the detailed analysis of the proposed generalizations are considered.

3.1 The Integral of Non-Homogeneity

3.1.1 Local Micro-Stresses and the Physics of Primary Creep

As is well known, plastic deformation is accompanied by the formation of dislocation pile-ups, tangles of dislocations, unmovable jogs, grains boundaries, etc (the nucleation of dislocations is also observed at elastic deformation). These defect-formations, being of strongly local character, raise an uneven stress/strain distribution through the microstructure of metal that, in turn, leads to considerable distortions of the crystal lattice where the strain energy is mainly stored.

The considerable non-homogeneity and concentration of micro-strains/stresses of the second and third kind were observed in experiments performed on specimens of pure copper, iron and titanium [9]. The experiments show that both stresses and strains are distributed non-homogeneously within grains (under both elastic and plastic loading). In addition, if the strain is greater than its average value through

the grain, then the stress inducing this strain is smaller than average stress and vice versa. At the same time, the total over- and under-loading is equal to zero.

The non-homogeneous stress distribution makes the metal structure more unstable than in an annealed state. Once favorable conditions arise (for example, if the stress stops increasing), the relaxation of crystal lattice distortions is observed. It is the difference between the local and average stresses that is the driving force for the relaxation that occurs mainly due to spontaneous slips in grains induced by the movements of dislocation. Indeed, under thermal fluctuations, locked and tangled dislocations and the obstructions in their way themselves become progressively movable, thereby promoting the development of deformation. Therefore, the time dependent relaxation of the crystal lattice distortions governs the progress of the primary creep deformation.

The local stresses arising around the lattice distortions we will call local microstresses. These stresses display the following properties: 1) they, being directly correlated with dislocation density, make the material stronger; 2) the greater the loading rate, the greater the local stresses; 3) they are unstable: as soon as favorable conditions arise, they decrease with time. It must be noted that the local microstress relaxation is also observed during slow loading.

Therefore, on the one hand, the local microstresses cause the “rate-hardening” of the material during active loading but, on the other hand, they can relax resulting in the softening of the material. Time-dependent macro-deformation is the result of the concurring processes of the hardening and softening.

3.1.2 The Integral of Non-Homogeneity as the Mathematical Measure of Local Stresses

To establish a relation between the microstress non-homogeneity and elastic strain energy, consider an elementary volume of body (treated as point) consisting of a large number of microparticles. Let $\bar{\sigma}_{kq}^0$ denote the average stress deviator tensor components (macrostress) acting at the given point. The microstress non-homogeneity can be expressed through the stress deviator tensor components acting in each microparticle, $\bar{\sigma}_{kq}$, as

$$\bar{\sigma}_{kq} = \bar{\sigma}_{kq}^0 + \bar{\sigma}_{kq}', \quad (3.3)$$

where $\bar{\sigma}_{kq}'$ are random quantities expressing the over/under-loading in each particle. We set the reaction of $\bar{\sigma}_{kq}'$ on the change in the average stress as

$$d\bar{\sigma}_{kq}' = C_{ijkq} d\bar{\sigma}_{ij}^0, \quad (3.4)$$

where C_{ijkl} are random numbers that vary from particle to particle, which are assumed to be independent from $\bar{\sigma}_{ij}^0$. Let us suppose that all random numbers C_{ijkl} have an identical distribution function, F , and are independent of each other. Since $d\bar{\sigma}_{ij}^0$ are macroscopic (average) stress components, the mathematical expectation of parameters C_{ijkl} is

$$\int_{-\infty}^{\infty} C_{ijkl} F(C_{ijkl}) dC_{ijkl} = 0 \quad \Xi \quad (3.5)$$

Formula (3.5) means that the total over/under-loading with respect to the average stress is equal to zero. In addition,

$$\int_{-\infty}^{\infty} F(C_{ijkl}) dC_{ijkl} = 1 \quad \Xi, \quad (3.6)$$

As was pointed out earlier, the local stresses are unstable and can relax with time. The equation governing their time-dependent behavior is proposed as

$$d\bar{\sigma}_{ij}' = C_{ijpq} d\bar{\sigma}_{pq}^0 - p\bar{\sigma}_{ij}' dt. \quad (3.7)$$

The first item on the right side in the above formula characterizes the rise of $\bar{\sigma}_{ij}'$ given by (3.4); term $-p\bar{\sigma}_{ij}' dt$ gives the time-dependent decrease of microstresses, which is taken to be proportional to $\bar{\sigma}_{ij}'$. The solution of the differential equation (3.7) for $\bar{\sigma}_{ij}'$ is

$$\bar{\sigma}_{ij}' = C_{ijkq} I_{kq}(t), \quad I_{kq}(t) = \int_0^t \frac{d\bar{\sigma}_{kq}^0}{ds} \exp(-p(t-s)) ds. \quad (3.8)$$

Now, expression (3.3) becomes

$$\bar{\sigma}_{ij} = \bar{\sigma}_{ij}^0 + C_{ijkq} I_{kq}(t). \quad (3.9)$$

As is well known, elastic strain energy can be expressed as

$$U = \frac{1}{12G} \left[(\bar{\sigma}_{xz} - \bar{\sigma}_{yy})^2 + (\bar{\sigma}_{yy} - \bar{\sigma}_{zz})^2 + (\bar{\sigma}_{zz} - \bar{\sigma}_{xx})^2 + 6(\tau_{xy}^2 + \tau_{yz}^2 + \tau_{zx}^2) \right], \quad (3.10)$$

where G is the elastic shear modulus. By substituting stresses $\bar{\sigma}_{ij}$ from (3.9) into (3.10), we obtain

$$\begin{aligned}
U = & \frac{1}{12G} \left\{ \left(\bar{\sigma}_{xx}^0 + C_{xxkq} I_{kq} - \bar{\sigma}_{yy}^0 - C_{yykq} I_{kq} \right)^2 + \left(\bar{\sigma}_{yy}^0 + C_{yykq} I_{kq} - \bar{\sigma}_{zz}^0 - C_{zzkq} I_{kq} \right)^2 + \right. \\
& + \left. \left(\bar{\sigma}_{zz}^0 + C_{zzkq} I_{kq} - \bar{\sigma}_{xx}^0 - C_{xxkq} I_{kq} \right)^2 + \right. \\
& + \left. 6 \left[\left(\tau_{xy}^0 + C_{xykq} I_{kq} \right)^2 + \left(\tau_{yz}^0 + C_{yzkq} I_{kq} \right)^2 + \left(\tau_{xz}^0 + C_{xzkq} I_{kq} \right)^2 \right] \right\} \quad (3.11)
\end{aligned}$$

The mean value of U is determined by the following relation

$$\langle U \rangle = \underbrace{\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty}}_{36} U F(C_{xxx}) \dots F(C_{xzx}) dC_{xxx} \dots dC_{xzx} \quad (3.12)$$

$\langle U \rangle$ can be decomposed in two parts:

$$\langle U \rangle = J_1 + J_2, \quad (3.13)$$

$$J_1 = \underbrace{\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty}}_{36} U_0 F(C_{xxx}) \dots F(C_{xzx}) dC_{xxx} \dots dC_{xzx} = U_0, \quad (3.14)$$

$$\begin{aligned}
J_2 = & \frac{1}{12G} \underbrace{\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty}}_{36} \left(C_{xxx}^2 I_{xx}^2 + 2C_{xxx} I_{xx} \bar{\sigma}_{xx}^0 + \dots \right) F(C_{xxx}) \dots F(C_{xzx}) dC_{xxx} \dots dC_{xzx} = \\
= & \frac{1}{12G} I_{xx}^2 \underbrace{\int_{-\infty}^{\infty} C_{xxx}^2 F(C_{xxx}) dC_{xxx}}_{35} \cdot \underbrace{\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} F(C_{xxy}) \dots F(C_{xzx}) dC_{xxy} \dots dC_{xzx}}_{35} + \\
& + \frac{1}{6G} I_{xx} \bar{\sigma}_{xx}^0 \underbrace{\int_{-\infty}^{\infty} C_{xxx} F(C_{xxx}) dC_{xxx}}_{35} \cdot \underbrace{\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} F(C_{xxy}) \dots F(C_{xzx}) dC_{xxy} \dots dC_{xzx}}_{35} + \dots
\end{aligned} \quad (3.15)$$

where U_0 is the strain energy for the case of homogeneous stress distribution determined by formula (3.10) at $\bar{\sigma}_{ij} = \bar{\sigma}_{ij}^0$. In arriving at the result (3.14), expression (3.6) has been taken into account. In order to evaluate integral J_2 , it is enough to investigate its first two terms. Indeed, formula (3.5) implies that all the integrals in (3.15) containing C_{ijkl} are equal to zero. The integrals containing C_{ijkl}^2 give the variance of random numbers C_{ijkl} , B_1 :

$$\int_{-\infty}^{\infty} C_{ijkl}^2 F(C_{ijkl}) dC_{ijkl} = B_1 \quad \Sigma \quad (3.16)$$

As a consequence,

$$J_2 = \frac{2B_1}{G} (I_{xx}^2 + I_{yy}^2 + I_{zz}^2 + 2I_{xy}^2 + 2I_{yz}^2 + 2I_{zx}^2).$$

Finally, expression (3.13) is

$$\langle U \rangle = U_0 + \frac{2B_1}{G} (I_{xx}^2 + I_{yy}^2 + I_{zz}^2 + 2I_{xy}^2 + 2I_{yz}^2 + 2I_{zx}^2). \quad (3.17)$$

By subtracting from the right-hand side in (3.17) the expression $2B_1/3G(I_{xx} + I_{yy} + I_{zz})^2$, which is equal to zero due to $\bar{\sigma}_x^0 + \bar{\sigma}_y^0 + \bar{\sigma}_z^0 = 0$, we obtain

$$\langle U \rangle = U_0 + \frac{2B_1}{3G} [(I_{xx} - I_{yy})^2 + (I_{yy} - I_{zz})^2 + (I_{zz} - I_{xx})^2 + 6(I_{xy}^2 + I_{yz}^2 + I_{zx}^2)]. \quad (3.18)$$

Substituting I_{ij} from (3.8) into (3.18) and converting the variables σ_{ij} to the stress vector components S_n by formula (2.1), the expression for the mathematical expectation of elastic strain energy is obtained as

$$\langle U \rangle = U_0 + \frac{2B_1}{3G} \sum_{n=1}^5 \left[\int_0^t \frac{dS_n}{ds} \exp(-p(t-s)) ds \right]^2 \quad (3.19)$$

The value of $\langle U \rangle$ is seen to consist of two parts; the term U_0 corresponds to homogeneous stress distribution and the second term characterizes the time-dependent deviation of stresses from their average value. If a body is ideally homogeneous, the distribution functions of random numbers C_{ijkl} degenerate in the Dirac delta-function and, according to (3.16), we obtain $B_1 = 0$. As seen from formula (3.19), $\langle U \rangle$ depends not only on the rate of stress vector components \dot{S}_n at a given instant, but on its values for the all history of loading as well. For the case $\dot{S}_n = const$,

$$\langle U \rangle = U_0 + \frac{2B_1}{3G} \dot{S}_n \dot{S}_n \left[\int_0^t \exp(-p(t-s)) ds \right]^2. \quad (3.20)$$

Since $\dot{S}_n \dot{S}_n = \dot{S}^2$ (S denotes the length of stress vector),

$$\langle U \rangle = U_0 + \frac{2B_1}{3G} \left[\int_0^t \frac{dS}{ds} \exp(-p(t-s)) ds \right]^2. \quad (3.21)$$

In the case that the stress deviator vector has only one non-zero component, expressions (3.19) and (3.21) are identical. We take the square root in the right-hand side in relation (3.21) to be the scalar measure of micro-non-homogeneity:

$$I = B \int_0^t \frac{dS}{ds} \exp(-p(t-s)) ds, \quad B = \sqrt{\frac{2B_1}{3G}} = \text{const}. \quad (3.22)$$

We will term I as the integral (parameter) of non-homogeneity. To work with the integral of non-homogeneity on the microlevel of material, we replace S in (3.22) by scalar product $\vec{S} \cdot \vec{N}$. This replacement reflects the fact that the driving force of plastic flow within a slip system is not the whole macro-stress vector \vec{S} but only its projection $\vec{S} \cdot \vec{N}$ (resolved stress). Thus, finally, the characteristic of local micro-stresses has the form

$$I_N = B \int_0^t \frac{d\vec{S}}{ds} \cdot \vec{N} \exp(-p(t-s)) ds. \quad (3.23)$$

In contrast to (3.22), the adopted integral (3.23) depends on angles α , β , and λ thereby allowing for the orientation of tangent planes in the Ilyushin subspace \mathbf{S}^3 .

Let us analyze the integral of non-homogeneity for the loading regime shown in Fig. 4 ($\vec{v} = d\vec{S}/dt = \text{const}$). On the first portion of the loading, formula (3.23) gives

$$I_N(t) = B(\vec{v} \cdot \vec{N}) \int_0^t \exp(-p(t-s)) ds = \frac{B(\vec{v} \cdot \vec{N})}{p} [1 - \exp(-pt)], \quad t \in [0, t_1]. \quad (3.24)$$

As seen from (3.24), $I_N(t)$ grows from the very beginning of loading. If we take the loading rate $v = S/t$ to be infinitely large, we can approximate the function $\exp(-pS/v)$ in (3.24) by the Taylor series, which results in the following relation

$$I_N = B(\vec{S} \cdot \vec{N}) \quad \text{as } v \rightarrow \infty. \quad (3.25)$$

For the range $t > t_1$ when $\vec{v} = 0$, let us split the range of integration in formula (3.23) into two parts, $[0, t_1]$ and $[t_1, t]$:

$$I_N(t) = B(\vec{v} \cdot \vec{N}) \int_0^{t_1} \exp(-p(t-s)) ds = \frac{B(\vec{v} \cdot \vec{N})}{p} [\exp(pt_1) - 1] \exp(-pt), \quad t \geq t_1. \quad (3.26)$$

From (3.24) and (3.26) the following properties of the integral of non-homogeneity can be indicated: (i) during loading, it grows proportionally to the loading rate; (ii) it decreases under constant loading. Therefore, the time-dependent behavior of the integral of non-homogeneity correlates with that of local microstresses.

The condition $I_N = 0$ symbolizes the end of transformations occurring in the crystal lattice under primer creep and transition to the steady-state stage of creep.

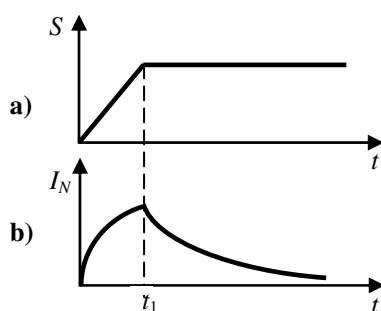


Figure 4
 I_N - t diagram

Intermediate discussion. The sum of the two quantities in equation (3.1), $\psi_N + I_N = \vec{S} \cdot \vec{N} - \sqrt{2}\tau_P$, characterizes the straining state of the material and determines the stress to induce irreversible deformation. The parameters ψ_N and I_N have a common trait; they can relax in time (see (3.2), and (3.26)). On the other hand, there is an essential difference between these quantities: ψ_N expresses the number of defects that produce irreversible deformation, whereas I_N characterizes the loading-rate-dependent development of these defects. The integral I_N behaves in a different way depending on loading regime: a) under loading, I_N symbolizes the load-rate strengthening of material; b) under constant stress, I_N drops expressing the lattice distortion relaxation that results in time-dependent, progressive deformation. The behavior of ψ_N and I_N is governed by different equations; I_N depends on loading-rate-history, formula (3.23), whereas ψ_N is related to irreversible deformation by (3.2).

3.2 System of Constitutive Equations

Formulae (3.1), (3.23), (3.2) and (2.13) constitute the base of the generalized synthetic theory:

$\psi_N = H_N - I_N - \sqrt{2}\tau_p,$	(A)
$I_N = B \int_0^t \frac{d\bar{S}}{ds} \cdot \bar{N} \exp(-p(t-s)) ds,$	(B)
$d\psi_N = rd\varphi_N - K\psi_N dt,$	(C)
$e_k^i = \iiint_{\alpha \beta \lambda} \varphi_N m_k \cos \lambda \cos \beta d\alpha d\beta d\lambda \quad \text{or}$ $\dot{e}_k^i = \iiint_{\alpha \beta \lambda} \dot{\varphi}_N m_k \cos \lambda \cos \beta d\alpha d\beta d\lambda, \quad k = 1,2,3$	(D)

The procedure of the calculation of irreversible strain vector components (e_k^i) is the following, (i) at a given stress deviator vector and loading rate, the defect intensity is determined by (A) and (B), (ii) the strain intensity can be found by (C) and, finally, (iii) formula (D) gives the values of strain(rate) vector components.

Expression (C) is one of the most important in terms of the generalized synthetic theory. It reflects the well-known fact that the defect intensity $d\psi_N$ grows with the increase in deformation ($rd\varphi_N$) and simultaneously decreases (relaxes) with time ($-K\psi_N dt$). Owing to (C), one does not need to split a deformation into its “instantaneous” (plastic) and viscous parts; both of them develop simultaneously. The degree of this development depends on concrete loading- and temperature-regimes. That is why, further throughout, we will use a single notion, irreversible deformation, by which we mean the deformation progressing with time (independently of whether we consider very short-termed loadings at plastic deformations or loadings lasting several hours or days as in creep tests).

The (A)-(D) system governs all types of irreversible deformation for any state of stresses and loading regimes.

Regard must be paid to the integration limits in formula (D). When founding the boundary values of angles α , β and λ , one must follow a single rule – only tangent planes which are on the endpoint of the stress deviator tensor produce irreversible strains. Since the plane distances are related to ψ_N , the limits of integration in (D) are determined from the conditions $\psi_N = 0$,

$$0 \leq \lambda \leq \lambda_1, \quad \cos \lambda_1(\alpha, \beta) = \frac{\sqrt{2}\tau_P}{(\bar{\mathbf{S}} \cdot \bar{\mathbf{m}}) - I_N}. \quad (3.27)$$

The condition $\lambda_1 = 0$ gives the equation for the boundary values of angles α and β :

$$\bar{\mathbf{S}} \cdot \bar{\mathbf{m}} - I_N = \sqrt{2}\tau_P. \quad (3.28)$$

4 Irreversible Deformation in Terms of the Generalized Synthetic Theory

4.1 Creep-Yield Limit Relation

Consider the case of arbitrary stress state, and assume that the loading rate is infinitely small so that the parameter of non-homogeneity tends to zero. If an irreversible deformation does not occur, $\psi_N = 0$, formula (A) gives that the tangential planes in \mathbf{S}^3 ($\lambda = 0$) are equidistant from the origin of coordinates:

$$h_m(\alpha, \beta) = H_N(\alpha, \beta, \lambda = 0) = \sqrt{2}\tau_P. \quad (4.1)$$

The above formula implies that the creep surface (creep locus in \mathbf{S}^3 setting the condition for the onset of first plastic flow at infinitesimal loading rate), being constructed as the inner-envelope of tangential planes, takes the form of the sphere of radius $\sqrt{2}\tau_P$:

$$S_1^2 + S_2^2 + S_3^2 = S_P^2, \quad S_P = \sqrt{2}\tau_P. \quad (4.2)$$

For the case of pure shear, expression (2.1) gives $S_3 = \sqrt{2}\tau_{xz}$ and $S_1 = S_2 = 0$ meaning that the stress vector $\bar{\mathbf{S}}(0, 0, S_3)$ acts along S_3 -axis. Let τ_P denote the value of shear stress when vector $\bar{\mathbf{S}}(0, 0, \sqrt{2}\tau_P)$ reaches the first tangential plane on the sphere (4.2). Since this plane is perpendicular to S_3 -axis ($\beta = \pi/2$ and $\lambda = 0$), formulae (2.5) and (2.7) give $H_N = \sqrt{2}\tau_P$. Therefore, τ_P expresses the creep limit of metal in pure shear.

Now, our goal is to establish the relation between τ_S and τ_P . It is worth starting with the case of pure shear. Let the loading be of constant rate, $v = \dot{S}_3 = \dot{S} = const$, $S = |\bar{\mathbf{S}}|$. Then expression (3.23) becomes

$$I_N = Bv \sin \beta \cos \lambda \int_0^t \exp(-p(t-s)) ds. \quad (4.3)$$

Until the stress vector reaches the tangential planes, $\psi_N = 0$, formula (A) takes the form

$$H_N = I_N + S_P. \quad (4.4)$$

By integrating in (4.3) and inserting the result of the integration in (4.4), we obtain

$$H_N = \frac{Bv}{p} [1 - \exp(-pt)] \sin \beta \cos \lambda + S_P, \quad t \in [0, t_1] \text{ in Fig. 4a.} \quad (4.5)$$

As seen from (4.5), the plane distances grow due to the increase in I_N . This means that formula (4.5) describes the movements of planes in the direction away from the origin of coordinate. Since these movements are not caused by the “pushing” action of stress deviator vector, they do not cause irreversible strain. The inner envelope of the planes with distances from (4.5) is shown in Fig. 5a (only tangent planes with $\lambda = 0$ are shown). As seen, the action of the integral of non-homogeneity does not result in the formation of a corner point.

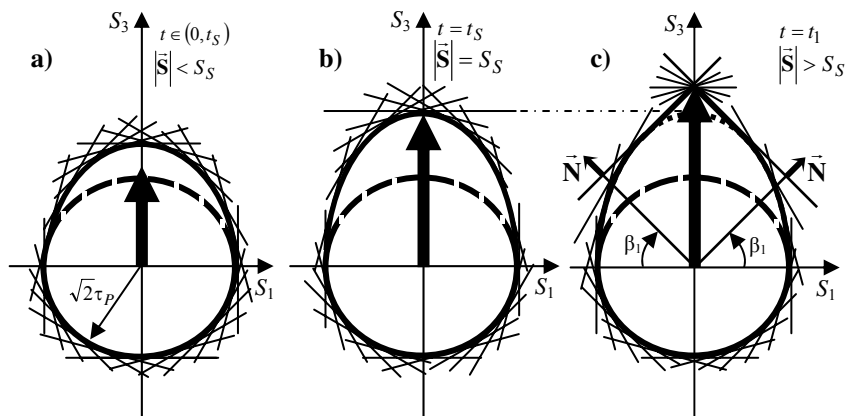


Figure 5

The transformation of yield (a and b) and loading (c) surface

Let S_S denote the length of the stress deviator vector which at time t_S ($t_S \in [0, t_1]$) reaches the first plane ($\beta = \pi/2$ and $\lambda = 0$), i.e. the plastic flow starts developing (Fig. 5b). For this plane, formula (2.7) gives $H_N = S_S$. The replacement of H_N by S_S in (4.5) leads to the equation for S_S :

$$S_S = \frac{Bv}{p} (1 - \exp(-pt_S)) + S_P, \quad S_S = vt_S \quad (4.6)$$

The plot of S_S as the function of v constructed on the base of (4.6) is shown in Fig. 6. As follows from Eq. (4.6), curve $S_S = S_S(S_P)$ has a horizontal asymptote, which is at a distance of $S_P/(1-B)$ from the abscissa that corresponds to the case of an infinitely large loading rate.

For $S > S_S$, the stress deviator vector translates some set of plane (Fig. 5c), and angle β_1 gives the boundary planes on the endpoint of \vec{S} .

Further, let us find the yield limit for an arbitrary proportional loading with a constant loading rate. Now, expression (3.23) is

$$I_N = B \cos \lambda \times \int_0^t \left(\frac{dS_1}{ds} \cos \alpha \cos \beta + \frac{dS_2}{ds} \sin \alpha \cos \beta + \frac{dS_3}{ds} \sin \beta \right) \exp[-p(t-s)] ds \quad (4.7)$$

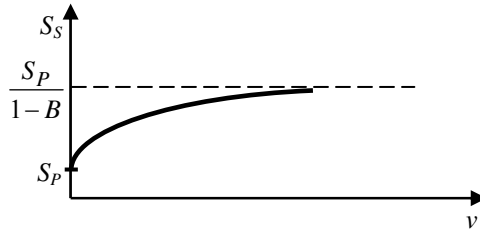


Figure 6
Yield limit vs loading rate plot

In the direction of the action of the stress deviator vector whose orientation is given by angles α_0 and β_0 , relations $\cos \alpha_0 = S_1 \cdot (S_1^2 + S_2^2)^{-1/2}$ and $\sin \beta_0 = S_3 \cdot S^{-1}$ hold true and formula (4.7) yields the form

$$I_{N_0} = B \cos \lambda \int_0^t \frac{S_i}{S} \frac{dS_i}{ds} \exp[-p(t-s)] ds \quad (4.8)$$

Since \dot{S} can be expressed as $\dot{S} = \frac{dS}{dt} = \frac{S_i}{S} \frac{dS_i}{dt}$, Eq. (4.8) gives

$$I_{N_0} = B \cos \lambda \int_0^t \frac{dS}{ds} \exp[-p(t-s)] ds \quad (4.9)$$

The integral I_{N_0} is identical to that from (4.3) at $\beta = \pi/2$ meaning that formula (4.6) is applicable to the determination of yield limit via the creep limit for an arbitrary state of stress.

Summarizing, formula (4.6) is of great importance due to the fact that it allows working with only one material constant, creep limit. In contrast to classical theories of plastic/creep deformation that use separately yield limit or creep limit depending on the problem to be solved, the generalized synthetic model is constructed in such a way that the creep limit plays the role of the material constant, while the yield limit is a function of loading rate.

4.2 The Modeling of Irreversible Deformation

Consider the case of proportional loading when the loading trajectory is a straight line in \mathbf{S}^3 . Further, let the plot of $S(t)$ have the form as in Fig. 4. Since the synthetic theory provides the fulfillment of the law of the deviator proportionality [14-16, 18], the formulae obtained for the case of, e.g., pure shear are fully applicable (up to constants) to arbitrary rectilinear loading path in \mathbf{S}^3 .

For the case of pure shear, expressions (A), (2.5) and (2.7) give the defect intensity as

$$\psi_N = [S_3 - I] \sin \beta \cos \lambda - S_P = \left(\frac{\Omega}{a} - 1 \right) S_P, \quad S_3 > S_S, \quad (4.10)$$

$$\Omega = m_3 \cos \lambda = \sin \beta \cos \lambda, \quad (4.11)$$

$$a = \frac{S_P}{S_3 - I}. \quad (4.12)$$

In formulae (4.10) and (4.12)

$$I = B \int_0^t \frac{dS_3}{ds} \exp[-p(t-s)] ds, \quad (4.13)$$

$$I|_{t=t_1} \equiv I_1 = \frac{Bv}{p} [1 - \exp(-pt_1)], \quad v = \dot{S}_3 = \dot{S} = const, \quad (4.13a)$$

$$I|_{t>t_1} \equiv I_2 = \frac{Bv}{p} [\exp(pt_1) - 1] \exp(-pt). \quad (4.13b)$$

According to (3.27) and (3.28), the defect intensity in expression (4.10) is positive for

$$0 \leq \alpha \leq 2\pi, \quad \beta_1 \leq \beta \leq \pi/2, \quad 0 \leq \lambda \leq \lambda_1, \quad \cos \lambda_1 = \frac{\sin \beta_1}{\sin \beta}, \quad \sin \beta_1 = a \quad (4.14)$$

The loading surface at $t = t_1$ is shown in Fig. 5c or 7a.

The defect intensity increment is expressed from (4.10) as

$$d\psi_N = (dS_3 - dI)\Omega. \quad (4.15)$$

Beyond the angles-diapason given by (4.14), we have $\psi_N = d\psi_N = 0$ and $\varphi_N = d\varphi_N = 0$.

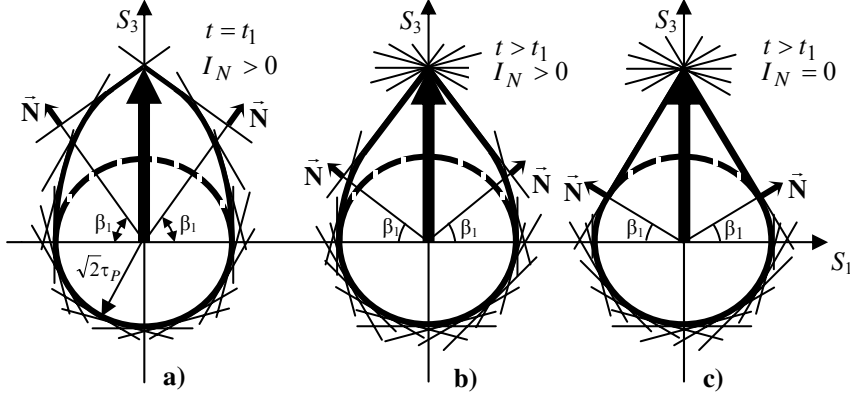


Figure 7

Kinetics of loading surface at creep

Consider the transformation of the loading surface for $t > t_1$ when $\dot{S}_3 = 0$. For the tangent planes that are beyond the diapason (4.14), formula (A) at $\psi_N = 0$ gives

$$H_N = S_P + I_2 \sin \beta \cos \lambda. \quad (4.16)$$

Because of the descending character of I_2 , we infer that H_N decreases for $t > t_1$ meaning that planes that are not on the endpoint of the stress deviator vector at $t = t_1$ start to move towards the origin of the coordinates. These movements result in the greater number of planes becoming located at the endpoint of the stress deviator vector. This situation is illustrated by Fig. 7b from which it is seen that the boundary angle β_1 determined by (4.14) and (4.13b) decreases with time. As integral I_2 tends to zero, formula (4.16) gives $H_N = S_P$ meaning that the planes stop moving and the boundary angle β_1 takes its minimal value (Fig. 7c).

Further, formula (B) gives the strain intensity as

$$rd\varphi_N = d\psi_N + K\psi_N dt = (dS_3 - dI)\Omega + KS_P \left(\frac{\Omega}{a} - 1 \right) dt. \quad (4.17)$$

Finally, formulae (C) gives the increment in irreversible-strain-vector-component, Δe_3^i , as

$$\Delta e_3^i = \frac{1}{2r} \int_0^{2\pi} d\alpha \int_{\beta_1}^{\pi/2} \sin 2\beta d\beta \int_0^{\lambda_1} \Delta\varphi_N \cos \lambda d\lambda, \quad (4.18)$$

where $\Delta\varphi_N$ is given by (4.17). In formula (4.18), the symbol Δ stands for the time-dependent increment of Δe_3^i and $\Delta\varphi_N$. By integrating over α , β and λ in (4.18), we obtain

$$\Delta e_3^i = a_0 [\Delta\Phi(a) + K\Phi(a)\Delta t], \quad (4.19)$$

where

$$a_0 = \frac{\sqrt{2}\pi\tau_p}{3r} = \text{const}, \quad \Phi(a) = \frac{\arccos a}{a} - 2\sqrt{1-a^2} + a^2 \ln \frac{1+\sqrt{1-a^2}}{a}, \quad \Phi(1) = 0. \quad (4.20)$$

The analysis of (4.20) shows that the function Φ is in inverse proportion with its argument a , Fig. 8.

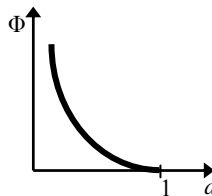


Figure 8
 $\Phi(a)$ function

By integrating over time in (4.19), we obtain the formula for the irreversible strain component in pure shear as:

$$e_3^i = a_0 \left[\Phi(a) + \int_{t_S}^t K\Phi(a) dt \right]. \quad (4.21)$$

To evaluate the integral (4.21), one needs to know the function $K(S_3(t), \Theta)$. This question will be considered in detail in 5.

Following the law of deviator proportionality [14], formula (4.21) can be rewritten for the case of an arbitrary stress state as

$$e_k^i = a_0 \left[\Phi(a) + \int_{t_S}^t K\Phi(a) dt \right] \frac{S_k}{S} \quad k = 1, 2, 3 \quad (4.22)$$

where, instead of (4.12),

$$a = \frac{S_P}{S - I}, \quad (4.23)$$

and I is calculated by (4.13a) and (4.13b) where $v = \dot{S} = const$.

Formula (4.22) is of a general character; it is applicable to the modeling of any type of deformation, both plastic and unsteady/steady-state creep. At $t = t_1$, we obtain the plastic strain vector components; at $t > t_1$ we get the total, plastic and creep, strain components.

4.3 The Analysis of the System of Constitutive Equations. Partial Cases

1) Consider the case of steady-state creep when $dS = 0$ and $I_N = 0$. It is clear that expression (4.15) gives $d\psi_N = 0$, i.e. the defect intensity (density) does not change during the steady state creep, reflecting the well-known fact that the steady-state creep deformation develops under the equilibrium between the processes of hardening and softening. Therefore, formula (C) gives the constant strain intensity rate:

$$r\dot{\phi}_N = K\psi_N = const, \quad K(S, \Theta) = const. \quad (4.24)$$

Another consequence from conditions $dS = 0$ and $I_N = 0$ is $a = S_P/S = const$ and $\Phi(a) = const$ (see (4.20) and (4.23)). According to (4.22), the steady-state creep strain(rate) components, e_k^P , can be written as

$$e_k^P = a_0\Phi(a)\frac{S_k}{S} + a_0K\Phi(a)\frac{S_k}{S}t \quad \text{or} \quad \dot{e}_k^P = a_0K\Phi(a)\frac{S_k}{S} = const \quad (4.25)$$

where $a_0\Phi(a) \cdot (S_k/S)$ is the value of strain at the end of primary creep. Formula (A) shows that the plane distances do not change with time, meaning that the steady state creep deformation is “produced” by the set of motionless planes which are located on the endpoint of the stress deviator stress (Fig. 7c). Since function K appears in the formula for steady-state creep rate, we can infer that it takes very small values, and the manifestation of the second term in (4.22) becomes material only under long-termed loadings. At the same time, it is important to emphasize that the role of the time integral in (4.22) grows with the increase in the duration of loading especially at elevated temperatures.

2) On the basis of the above, we can neglect the second term in (4.22) or the term $K\psi_N dt$ in formula (C) when plastic or unsteady state creep strains are investigated. This is absolutely justifiable due to the fact that the second term in (4.22) is comparable with the term $a_0\Phi(a)$ only under very long-termed loading (at least several tens of hours). Therefore, formula (C) at $K = 0$ takes the form

$$rd\varphi_N = d\psi_N \quad (4.26)$$

and expression (4.22) gives

$$e_k^i = a_0 \Phi(a) \frac{S_k}{S}, \quad (4.27)$$

where e_k^i is the total, plastic + primary creep, strain components. The function $\Phi(a)$ is not to be thought of as being of a time-independent nature because its argument a contains the integral of non-homogeneity, which regulates the time-dependent development of deformation.

Formula (4.26) reflects another well-known fact that the increase in plastic deformation causes that in the defects of crystal lattice leading to the work-hardening of material and, consequently, the plastic deformation progress requires an increase in acting stress.

The plastic deformation (e_k^S) at the end of active loading ($t = t_1$) is calculated by (4.27) in which

$$a \equiv a_1 = \frac{S_P}{S - I_1}. \quad (4.28)$$

The deformation produced for $t > t_1$, when $dS/dt = 0$, is calculated by (4.27) in which

$$a \equiv a_2 = \frac{S_P}{S - I_2}. \quad (4.29)$$

The analysis of expressions (4.27) and (4.20) shows that the increase in plastic deformation during active loading is modeled by the growth of S in (4.28) (since the condition $S > S_S$ is hold, the growth in S prevails over that in I_1). Further, the progress of deformation for $t \geq t_1$ is regulated by the decrease of I_2 in (4.29). The condition $I_2 \rightarrow 0$ symbolizes the end of primary creep.

Pure primary creep strain components (without the initial plastic strain components e_k^S), e_k^C , are calculated as the difference between the total and plastic strain components:

$$e_k^C = a_0 [\Phi(a_2) - \Phi(a_1)] \frac{S_k}{S}. \quad (4.30)$$

It is easy to see that

$$e_k^S + e_k^C = a_0 \Phi\left(\frac{S_P}{S}\right) \frac{S_k}{S} = \text{const as } I_2 \rightarrow 0. \quad (4.31)$$

As follows from (4.31), the synthetic theory states that a material possesses some reserve of irreversible deformation which can be manifested in the form of plastic or creep deformation dependent on the loading rate in active loading. Formulae (4.28), (4.29) and (4.27) lead to the following relations between the magnitudes of plastic and primary creep: the growth in loading rate, on the one hand, leads to a decrease in plastic deformation, but, on the other hand, causes the increase in the magnitude of primary creep strain. Another conclusion is that slow loading does not result in following primary creep at all due to $I = 0$ at slow loading.

It is worthwhile to emphasize that only the function Φ is applicable to the calculation of any type of deformation.

3) Consider the case when a complete or partial unloading follows the loading which has produced some irreversible deformation. It is clear that $d\varphi_N = 0$ in unloading and formula (C) becomes

$$d\psi_N = -K\psi_N dt . \quad (4.32)$$

The solution of the differential equation above is

$$\psi_N = \psi_{N_0} \exp(-Kt), \quad (4.33)$$

where ψ_{N_0} is the defect intensity accumulated during the initial irreversible straining. Expression (4.32) describes the process of defects relaxation.

4.4 Loading Criterion

While the limits of integration in formula (D) for proportional loading can be determined relatively simply, this is not the case for arbitrary (curvilinear) loading paths. To express analytically the integration limits for curvilinear loading paths is a very difficult task and, consequently, computer assisted methods must be applied. Nevertheless, a general criterion for the development of irreversible straining must be formulated. Let a current stress vector \vec{S} have produced some irreversible strain, i.e. some set of tangent planes are on its endpoint. For these planes, formulae (A) and (2.3) give

$$S_P + \psi_N + I_N = \vec{S} \cdot \vec{N} . \quad (4.34)$$

If the vector \vec{S} acquires increment $d\vec{S}$, for planes that are on the endpoint of vector $\vec{S} + d\vec{S}$ we have

$$S_P + \psi_N + d\psi_N + I_N + dI_N = \vec{S} \cdot \vec{N} + d\vec{S} \cdot \vec{N} . \quad (4.35)$$

Therefore, formulae (4.34) and (4.35) give that

$$d\psi_N = d\vec{S} \cdot \vec{N} - dI_N . \quad (4.36)$$

We propose the following criterion: the planes that produce irreversible strains due to a given vector \vec{S} continue to do this due to the vector $\vec{S} + d\vec{S}$ if for these planes $d\psi_N \geq 0$:

$$d\vec{S} \cdot \vec{N} - dI_N \geq 0. \quad (4.37)$$

Eq. (4.37) must be applied to the determination of integration limits in formula (D) for arbitrary loading paths.

As seen from the series of Rusinko's works [14-16], the synthetic theory demonstrates good agreement with experimental data.

5 Steady-State Creep. Function K

The steady-state creep rate is governed by expression (4.25), which for the case of uniaxial tension is

$$\dot{\epsilon}_1^P = Ka_0\Phi(a), \quad a = \sigma_P/\sigma_x. \quad (5.1)$$

The function K is proposed as the product of two functions:

$$K = K_1(\Theta)K_2(\sigma). \quad (5.2)$$

Let us split the range of homological temperature into three diapasons: $0 < \Theta \leq \Theta_1$ (low temperature), $\Theta_1 \leq \Theta \leq \Theta_2$ (elevated temperature), and $\Theta_2 \leq \Theta \leq \Theta_3$ (high temperature). The values of Θ_i are $\Theta_1 \approx 0,25$, $\Theta_2 \approx 0,5$ and $\Theta_3 \approx 0,7$ for pure metals and $\Theta_1 \approx 0,3$, $\Theta_2 \approx 0,55$ and $\Theta_3 \approx 0,75$ for alloys. The range $\Theta_3 \leq \Theta \leq 1$ is not considered below.

Within the range $0 < \Theta < \Theta_1$, the temperature is not enough for the thermal activation of dislocation motion, so steady creep does not occur, $K_1(\Theta) = 0$. For $\Theta_2 \leq \Theta \leq \Theta_3$ the dislocation climb is a dominating mechanism of creep. Since the rate of dislocation climb is controlled by the intensity of the diffusional processes, it is natural to assume that the function $K_1(\Theta)$ is proportional to the quantity of migrating (activated) atoms that regulate the vacancy-motion intensity. The relative quantity of activated atoms is

$$\int_{U_0}^{\infty} p dU = \exp\left(-\frac{U_0}{RT}\right), \quad (5.3)$$

where U_0 is the atom-migration-activation energy. In arriving at the result (5.3) we have utilized the Maxwell-Boltzmann energy distribution

$p = \frac{1}{RT} \exp\left(-\frac{U}{RT}\right)$. Thus, (denoting through \tilde{T} the melting point of metal):

$$K_1(\Theta) = \exp\left(-\frac{U_0}{RT}\right) = \exp\left(-\frac{U_0}{R\tilde{T}\Theta}\right). \quad (5.4)$$

Within the range of elevated temperatures, $\Theta_1 \leq \Theta \leq \Theta_2$, various processes govern the steady-state creep, none dominating over another. Therefore, the establishing of function $K_1(\Theta)$ from physical reasoning can lead to unreliable results. As a result, we propose the linear form of function $K_1(\Theta)$ so that $K_1(\Theta_1) = 0$ and $K_1(\Theta_2)$ takes the form of expression (5.4) (Fig. 9):

$$K_1(\Theta) = \frac{\Theta - \Theta_1}{\Theta_2 - \Theta_1} \exp\left(-\frac{U_0}{R\Theta_2\tilde{T}}\right) \text{ for } \Theta_1 < \Theta < \Theta_2. \quad (5.5)$$

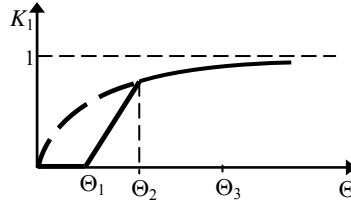


Figure 9
 $K_1(\Theta)$ function

Function K_2 can be found by making use of the empirical formula for steady-state creep rate in uniaxial tension:

$$\dot{\epsilon}_x^P = C f(\Theta) \sigma_x^k, \quad \dot{\epsilon}_x^P = 0 \text{ at } \sigma_x < \sigma_P. \quad (5.6)$$

It is clear that expressions (5.1) and (5.6) give $\dot{\epsilon}_x^P = 0$ as $\sigma_x < \sigma_P$.

Further, as follows from (4.20), function Φ behaves as $\pi \sigma_x / (2\sigma_P)$ for $\sigma_P / \sigma_x \rightarrow 0$ and the strain-rate in (5.1) (together with (2.12)) can be written as

$$\dot{\epsilon}_x^P = \frac{\pi^2}{9r} K_1(\Theta) K_2(\sigma_x) \sigma_x. \quad (5.7)$$

Taking function f in (5.6) to be in the form of K_1 and equating the right-hand sides of formulae (5.6) and (5.7) to each other, we obtain

$$K_2(\sigma_x) = \frac{9Cr}{\pi^2} \sigma_x^{k-1}. \quad (5.8)$$

6 Creep Delay

Consider the case when a loading regime is as in Fig. 10a.

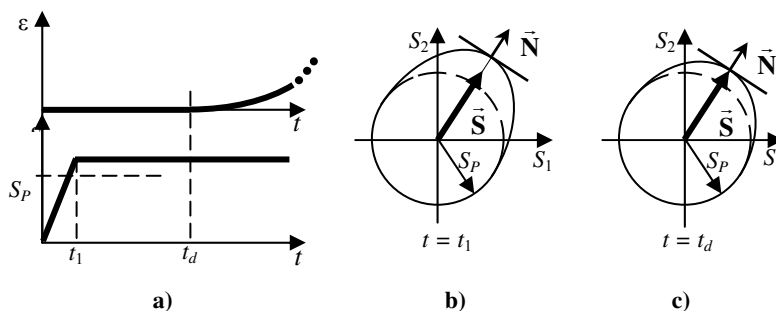


Figure 10
Creep delay

Let the modulus of stress deviator vector at $t = t_1$ ($S > S_p$) not reach any tangential plane (Fig. 10b), i.e. there is no plastic deformation ($\psi_N = 0$ and $\varphi_N = 0$ for all the planes). According to formula (3.26), once $dS/dt = 0$, the integral of non-homogeneity starts to decrease and formula (A) becomes

$$H_N = S_p + \frac{vB}{p} [\exp(pt_1) - 1] \exp(-pt), \quad t > t_1. \quad (6.1)$$

The formula above means that the tangent planes move back. The instant that a first plane touches the endpoint of the stress vector, $t = t_d$, symbolizes the start of creep deformation (Fig. 10c). The period of time when the creep deformation is absent, $[0, t_d]$, is referred to as creep delay [10]. It is clear that the first plane to be on the endpoint of the stress deviator vector is perpendicular to it, $H = \vec{S} \cdot \vec{N} = S$. Replacing H_N in (6.1) by S gives the following equation for t_d :

$$S_p + \frac{vB}{p} [\exp(pt_1) - 1] \exp(-pt_d) = S \quad (6.2)$$

As seen from (6.2), the duration of creep delay grows with the loading rate and is absent at slow loading when $I_N = 0$.

6.1 Steady-State Creep as a Function of Initial Deformation

According to formulae (5.1), (5.4), (5.5) and (5.8), the steady-state creep rate is a single-valued function of acting stress and temperature independently of whether the stress magnitude exceeds the yield limit of the material or not. This is in full agreement with numerous experiments. In this context, the investigations of Namestnikov, V. and Chvostunkov, A. [11] carried out as long ago as in the 1960s are of great importance. They deal with the comparison of the strain-hardening creep theory to experimental results for the cases when a creep deformation develops from an elastic or plastic state. First, the case when the creep diagram starts with some initial plastic deformation is considered. The model constants are determined so that the calculated steady-creep rates best fit experimental ones. As it has turned out, if we use these constants for the case when the creep deformation develops from an elastic state, the calculated results show considerable discrepancy with experiments.

Let us investigate, in terms of the generalized synthetic theory, whether the presence or absence of the initial plastic deformation affects the steady-state creep rate. Consider the case when at a given temperature one and the same stress deviator vector produces or not plastic deformation at $t = t_1$ (Figs. 11a and 11b) in the loading regime shown in Fig. 4a. Such a situation can be obtained at different loading rates. Following the techniques of the construction of loading surface discussed in 4.1. and 4.2, it is easy to see that the loading surfaces for the steady-state creep state are identical in both cases (Figs. 11c and 11d). This simple illustration shows that the generalized synthetic theory provides experimentally confirmed results that a secondary creep rate is a single-valued function of stress and temperature.

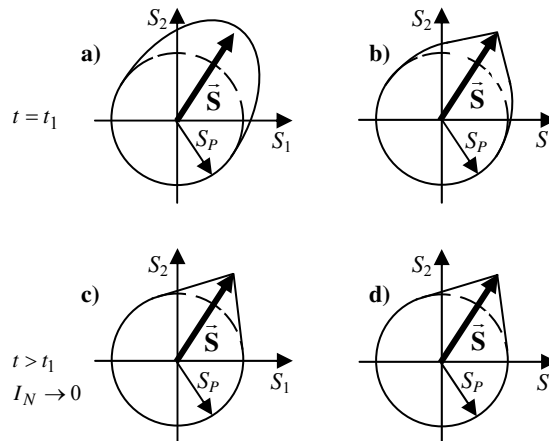


Figure 11

Identical steady-state creep surfaces with different initial straining states

Therefore, in contrast to classical creep theories, the generalized synthetic theory leads to correct results independently of whether the creep deformation develops from a plastic or elastic state (after some creep-delay-period).

7 The Bauschinger Negative Effect

The Bauschinger effect – an initial plastic deformation of one sign reduces the resistance of the material with respect to a subsequent plastic deformation of the opposite sign – is explained by the fact that the repulsive forces, which arise within dislocations conglomeration generated in initial loading, hasten the onset of the plastic deformation of opposite sign and, consequently, the smaller stresses are needed to induce plastic deformation. Starting from a certain initial-plastic-strain, the so-called Bauschinger negative effect is observed [7] when the compressive plastic deformation starts to develop at a positive magnitude of acting stress, Fig. 12a.

In order to allow for the Bauschinger effect, one must replace the equation proposed in [18] by the following:

$$\Psi_{-N} = -\Psi_N, \quad (7.1)$$

where an index $-N$ symbolizes the plane whose outward-pointing normal vector $-\vec{N}$ is in the opposite direction to the vector \vec{N} .

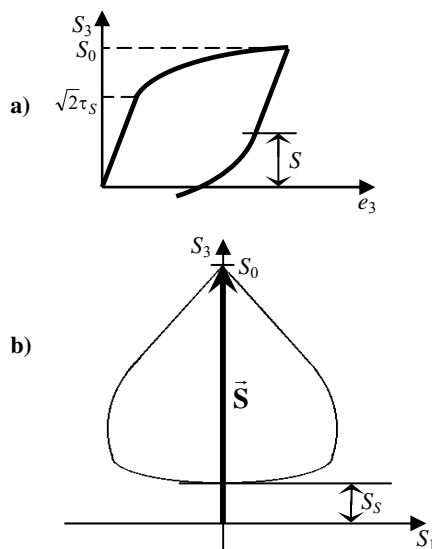


Figure 12
Bauschinger negative effect

The parameter of non-homogeneity for tangent planes with normal vectors $-\vec{\mathbf{N}}$, I_{-N} , is

$$I_{-N} = B \int_0^t \frac{d\vec{\mathbf{S}}}{ds} \cdot (-\vec{\mathbf{N}}) \exp[-p(t-s)] ds = -B \int_0^t \frac{d\vec{\mathbf{S}}}{ds} \cdot \vec{\mathbf{N}} \exp[-p(t-s)] ds = -I_N \quad (7.2)$$

It is clear that a rate-hardening occurring due to a plastic loading does not exert influence upon that in the following loading of opposite sign. To rephrase this in terms of synthetic theory, we say that if I_N is positive, then I_{-N} is set to be zero and vice versa.

The distance to the planes with vectors $-\vec{\mathbf{N}}$, on the basis of expressions (A), (7.1) and (7.2), can be written as

$$H_{-N} = S_P + \psi_{-N} + I_{-N} = S_P - \psi_N - I_N. \quad (E)$$

Formula (E) symbolizes that an initial plastic deformation of one sign reduces the resistance of material with respect to a subsequent plastic deformation of the opposite sign. Indeed, distance H_{-N} decreases with the growth in defect intensity ψ_N and integral I_N due to the plastic straining in directions $\vec{\mathbf{N}}$. The decrease in H_{-N} means that tangential planes with normals $-\vec{\mathbf{N}}$ near the origin of coordinates. If the initial loading is of such a magnitude that $\psi_N + I_N > S_P$, the distance to the planes calculated by formula (E) becomes negative, meaning that the planes with normals $-\vec{\mathbf{N}}$ have gone over the origin of coordinate, i.e. the Bauschinger negative effect is manifested. Fig. 12b shows the loading surface whose lower part is constructed on the basis of formula (E) for the case of the Bauschinger negative effect in pure shear.

Formula (E), which governs the relation between the hardening/softening processes occurring in opposite directions, must be included into the system of constitutive equations (A)-(D).

8 Reverse Creep

Consider the time-dependent deformation of a specimen of aluminum alloy PA4 (chemical composition 0.7-1.2% Mg, 0.6-1.0% Mn, 0.7-1.2% Si, 0.5% Fe, the rest Al) under the stepwise uniaxial loading shown in Fig. 13 [12]. The $\varepsilon-t$ curve consists of the following portions:

(1-2) unsteady creep under constant tensile stress σ_1 at room temperature for $t \in [0, t_c]$ (portion 0-1 is the initial plastic deformation due to σ_1);

(2-3) the drop of stress by the amount of $\Delta\sigma$ that results in the plastic **compressive!** strain $\Delta\varepsilon^S$; this is the manifestation of the Bauschinger negative effect;

(3-4) **compressive reverse creep!** Despite the stretching stress acts, the specimen undergoes compressive creep deformation for $t \in [t_c, t_c + t_r]$;

(4-5) the creep deformation does not progresses for the range $t \in [t_c + t_r, t_c + t_r + t_d]$;

beyond point 5, for $t > t_c + t_r + t_d$, the creep in the direction of acting stress is resumed.

As stated in [12], the reverse creep is observed only if the Bauschinger negative effect takes place.

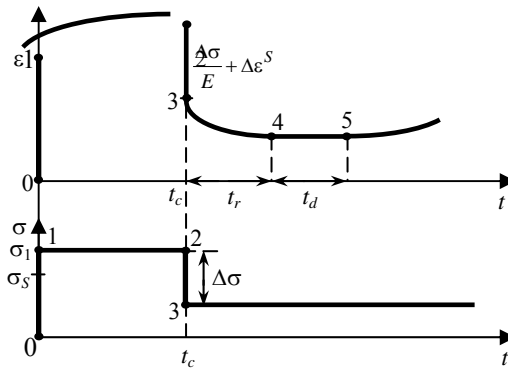


Figure 13

Creep diagram under stepwise loading

Let the magnitude of time-periods t_r and t_d be the main task of this Section.

For the case of uniaxial tension, according to expression (2.1), the stress deviator tensor components are $S_1 = \sqrt{2}\sigma_1/\sqrt{3}$ and $S_2 = S_3 = 0$, i.e. vector $\vec{S}(S_1, 0, 0)$ is co-directed with S_1 -axis. To determine t_r and t_d , it is quite sufficient to study the displacements of two planes with normals \vec{N} and $-\vec{N}$, which are tangential to sphere (4.2) in the virgin state ($\lambda = 0$) and perpendicular to the vector \vec{S} . These planes will be denoted by **I** and **I'**, respectively. The perpendicularity of planes **I** and **I'** to S_1 -axis implies that the orientations of their normals \vec{N} and $-\vec{N}$ are set by angles $\alpha = 0$ and $\beta = 0$, and $\alpha = \pi$ and $\beta = 0$, respectively.

Formulae (2.5) and (2.7) give that

$$H_N = \vec{S} \cdot \vec{m} \cos \lambda = S_1 m_1 \cos \lambda = S_1 \Omega, \quad \Omega = m_1 \cos \lambda = S_1 |\cos \alpha| \cos \beta \cos \lambda \quad (8.1)$$

It is clear that $\Omega = 1$ for planes **I** and **I'** and, further throughout, all formulae will be written at $\Omega = 1$.

By making use of the basic formulae of the generalized synthetic theory, let us study the positions of planes **I** and **I'** on each portion of loading (the indexes in further formulae follow the marks in Fig. 13). Fig. 14 illustrates the positions of planes **I** and **I'**. In Fig. 14 a, b, d and e, together with planes **I** and **I'**, the set of planes shifted by the stress deviator vector are also shown. Since we restrict ourselves only to the determination of t_r and t_d , the number of these planes is immaterial and thus they are shown purely schematically.

Portion 0-1. The plane **I** is displaced by the stress deviator vector $\vec{S}(S_1, 0, 0)$ (Fig. 14a). Assuming the loading rate on portion **0-1** to be infinitely large, formulae (3.25) and (A), together with (8.1), give that

$$I_{N1} = BS_1 > 0, \quad H_{N1} = S_1, \quad \psi_{N1} = S_1(1 - B) - S_P. \quad (8.2)$$

The distance to the plane **I'** can be calculated by formula (E) as

$$H_{-N1} = S_P - \psi_{N1} = 2S_P - S_1(1 - B). \quad (8.3)$$

In arriving at the result (8.3) we have taken into account that $I_{-N1} = 0$. Dependent on the values of S_P , S_1 and B , the distance H_{-N1} can take both positive and negative values.

Portion 1-2. The distance to plane **I** remains unchangeable for $0 \leq t \leq t_c$ due to the fact that it continues to be on the endpoint of the stress deviator vector. Expressions (3.26) and (A) are

$$I_{N(1-2)} = BS_1 \exp(-pt), \quad (8.4)$$

$$\psi_{N(1-2)} = S_1(1 - B \exp(-pt)) - S_P, \quad \psi_{N2} = \psi_{N(1-2)} \Big|_{t=t_c}. \quad (8.5)$$

The distance to the plane **I'** is governed by formula (E) at $I_{-N(1-2)} = 0$ and formula (8.4):

$$H_{-N(1-2)} = S_P - \psi_{N(1-2)} = 2S_P - S_1(1 - B \exp(-pt)),$$

$$H_{-N2} = H_{-N(1-2)} \Big|_{t=t_c}. \quad (8.6)$$

To ensure the arising of the Bauschinger negative effect at the unloading in portion **2-3**, we require that $S_1 \gg S_P$ so that the magnitude of distance H_{-N2} is negative, i.e. plane **I'** has gone over the origin of coordinates. The positions of planes **I** and **I'** at $t = t_c$ are shown in Fig. 14b.

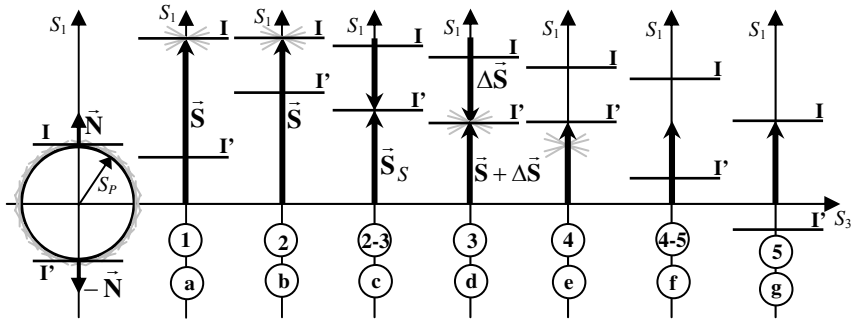


Figure 14
Positions of planes I and I'

Portion 2-3 (stress drop). The stress increment vector $\Delta\vec{S}$ shifts a set of planes with normals $-\vec{N}$ (plane \mathbf{I}' is also among them) thereby producing the compressive plastic-strain-increment $\Delta\epsilon^S$. Since $\Delta\epsilon^S$ is produced under the action of positive stress, $\vec{S} + \Delta\vec{S}$, it is clear that the Bauschinger negative effect occurs. Let us find a yield limit, S_S , in the partial unloading. To do this, we need to determine the distance to the plane \mathbf{I}' when the vector $\Delta\vec{S}$ reaches this plane. At $\Delta S = S_1 - S_S$, the compressive plastic strain starts to develop (Fig. 14c).

Before we turn to the determination of S_S , special attention must be paid to the integral of non-homogeneity I_{-N} . Since the reverse creep is of a compressive nature, its development can be modeled only by means of planes with normals $-\vec{N}$. Therefore, we require that I_{-N3} be positive. This means that the reverse creep will develop if the compressive (negative) plastic deformation has occurred, i.e. the material needs to obtain some strain energy which can be released in the form of time-dependent deformation. To meet condition $I_{-N3} > 0$, we require that the integral of non-homogeneity I_{N3} be negative. I_{N3} , due to the stress drop of ΔS , is $I_{N3} = B[S_1 \exp(-pt_c) - \Delta S]$. The inequality $I_{N3} < 0$ holds true if to require that

$$\Delta S > S_1 \exp(-pt_c). \quad (8.7)$$

As a result,

$$I_{-N3} = -I_{N3} = B[\Delta S - S_1 \exp(-pt_c)] > 0. \quad (8.8)$$

The yield limit S_S is equal to the magnitude of the stress vector when it reaches the plane \mathbf{I}' , which is at the following distance from the origin of coordinate (Fig. 14c):

$$H_{-N} = \vec{S} \cdot (-\vec{N}) = -S_S \quad (8.9)$$

Since the planes with normals \vec{N} are not on the stress deviator vector $\vec{S} + \Delta\vec{S}$, there is no increment in the defect intensity, $\Psi_{N3} = \Psi_{N2}$. Therefore, expressions (8.5) and (A) give at $\Delta S = S_1 - S_S$ that

$$H_{-N3} = S_P + \Psi_{-N3} + I_{-N3} = 2S_P - S_1(1 - B) - BS_S. \quad (8.10)$$

By letting H_{-N3} in (8.10) be equal to $-S_S$, the equation for the yield limit in the partial unloading is:

$$S_S = S_1 - \frac{2S_P}{1 - B}. \quad (8.11)$$

Summarizing, the occurrence of $\Delta\varepsilon^S$ is possible if the magnitude of S_S is positive and the stress $S_1 - \Delta S$ is less than S_S (Fig. 14d). These conditions, in the view of (8.11), can be met if

$$S_1 > \frac{2S_P}{1 - B} \quad \text{and} \quad \Delta S > \frac{2S_P}{1 - B}. \quad (8.12)$$

As $\Delta S \leq S_1$, the fulfillment of the second inequality in (8.12) provides the fulfillment of the first one.

Portion 3-4 (reverse creep). The integral of non-homogeneity for $t > t_c$ can be obtained if we multiply I_{-N3} from (8.8) by $\exp[-p(t - t_c)]$:

$$I_{-N(3-4)} = B[\Delta S - S_1 \exp(-pt_c)] \exp[-p(t - t_c)]. \quad (8.13)$$

Since plane **I** is on the endpoint of the vector $\vec{S} + \Delta\vec{S}$, one can write that

$$H_{-N(3-4)} = S_P + \Psi_{-N(3-4)} + I_{-N(3-4)} = -(S_1 - \Delta S), \quad (8.14)$$

$$\Psi_{-N(3-4)} = -(S_1 - \Delta S) - I_{-N(3-4)} - S_P, \quad (8.15)$$

$$\dot{\Psi}_{-N(3-4)} = -\dot{I}_{-N(3-4)} = pI_{-N(3-4)}. \quad (8.16)$$

Reverse creep strain rate intensity, $\dot{\phi}_{-N(3-4)}$, can be found from formula (C):

$$r\dot{\phi}_{-N(3-4)} = \dot{\Psi}_{-N(3-4)} + K\Psi_{-N(3-4)} = pI_{-N(3-4)} + K\Psi_{-N(3-4)}. \quad (8.17)$$

Now, formulae (8.13) and (8.15)-(8.17) give that

$$r\dot{\phi}_{-N(3-4)} = B(p - K)[\Delta S \exp(pt_c) - S_1] \exp(-pt) - K[(S_1 - \Delta S) + S_P]. \quad (8.18)$$

As seen from (8.14), tangent planes with negative normals start to move towards the origin of the coordinate leaving the endpoint of the stress deviator vector. The decrease in the number of planes on the endpoint of the stress deviator vector

leads to a decrease in the $r\dot{\varphi}_{-N(3-4)}$ from (8.18), i.e. the decrease in the reverse creep rate is modeled. As long as the right-hand side in (8.18) is positive, tangent planes with negative normals produce irreversible (creep) strain at constant $S_1 - \Delta S$. By letting $\dot{\varphi}_{-N(3-4)} = 0$, we express the fact that the last plane (**I** plane) has left the endpoint of stress deviator vector (Fig. 14e), and we can calculate the duration of reverse creep t_r as

$$t_r = \frac{1}{p} \ln \frac{B(p-K)(\Delta S - S_1 \exp(-pt_c))}{K(S_1 - \Delta S + S_P)}. \quad (8.19)$$

$t_r > 0$ if the numerator is greater than the denominator in (8.19):

$$[B(p-K) + K]\Delta S > [B(p-K)\exp(-pt_c) + K]S_1 + KS_P. \quad (8.20)$$

Therefore, the formulae derived are valid if the magnitude of partial unloading ΔS satisfies expressions (8.8), (8.12) and (8.20).

As seen from (8.19), the reverse creep time t_r grows with ΔS if we hold S_1 and t_c fixed; this is true for the whole range of ΔS , from $S_1 - S_S$ to S_1 (complete unloading). Another result is that the reverse creep time t_r grows with the initial creep duration t_c at fixed values of S_1 and ΔS . Furthermore, the function $t_r(t_c)$ is bounded above by horizontal asymptote

$$\max t_r = \frac{1}{p} \ln \frac{B(p-K)\Delta S}{K(S_1 - \Delta S + S_P)} \quad \text{as } t_c \rightarrow \infty. \quad (8.21)$$

These results agree with experimental data.

Portion 4-5 (creep delay). The defect intensity for plane **I** at $t = t_c + t_r$ is determined by formulae (7.1), (8.10) and (8.14) at $t = t_c + t_r$:

$$\psi_{N4} = -\psi_{-N4} = \frac{P}{p-K}(S_1 + S_P - \Delta S). \quad (8.22)$$

Since tangent planes with neither positive or negative normals are not on the endpoint of $\vec{S} + \Delta\vec{S}$ (Fig. 14f), irreversible straining does not occur for $t > t_c + t_r$, $d\varphi_N = 0$. Therefore, we arrive at defects relaxation equation (4.33) which, together with initial condition (8.22), takes the form

$$\psi_{N(4-5)} = \frac{P}{p-K}(S_1 + S_P - \Delta S)\exp(-K(t - t_c - t_r)). \quad (8.23)$$

As $I_{N(4-5)} = 0$, the distance to plane **I** is

$$H_{N(4-5)} = S_P + \frac{P}{p-K} (S_1 + S_P - \Delta S) \exp(-K(t - t_c - t_r)). \quad (8.24)$$

This means that plane **I** moves in the direction towards the origin of the coordinates. The instant of time when the plane is on the endpoint of vector $\vec{S} + \Delta\vec{S}$ (Fig. 14g) can be found from (8.24) by letting

$$H_{N(4-5)} = S_1 - \Delta S. \quad (8.25)$$

As a result, we obtain the duration of creep delay t_d as

$$t_d = \frac{1}{K} \ln \frac{p(S_1 + S_P - \Delta S)}{(p-K)(S_1 - S_P - \Delta S)}. \quad (8.26)$$

For $t > t_c + t_r + t_d$ the creep of positive sign develops in time due to the fact that the planes come back to the endpoint of vector $\vec{S} + \Delta\vec{S}$. It is worth noting that the formula for t_d holds true if

$$S_1 - \Delta S > S_P. \quad (8.27)$$

This inequality expresses an obvious condition for the occurring of the positive creep deformation that the acting stress $S_1 - \Delta S$ must exceed the creep limit S_P .

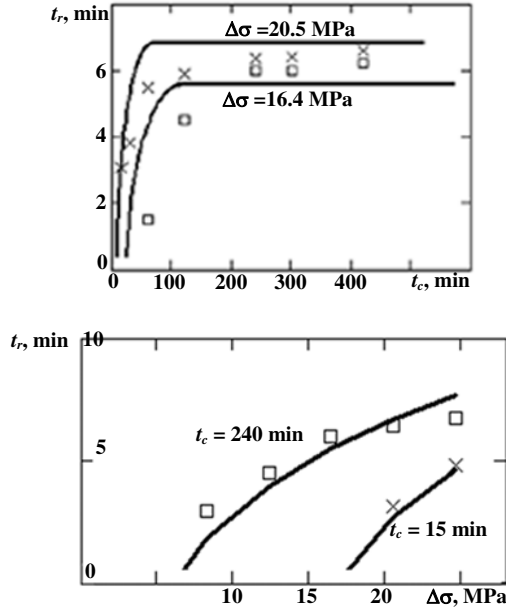


Figure 15

Experimental and calculated $t_r - t_c$ and $t_r - \Delta\sigma$ curves

In Fig. 15, $t_r \sim t_c$ and $t_r \sim \Delta\sigma$ curves are plotted on the basis of (8.19) (symbols \square and \times indicate experimental points), for the following values of acting stress, creep limit, and the model constants, $\sigma_1 = 227 \text{ MPa}$, $\sigma_p = 10 \text{ MPa}$, $B = 0.05$, $K = 2.5 \cdot 10^{-4}$, $p = 0.2 \text{ min}^{-1}$. The comparison between the calculated result and experimental data shows satisfactory agreement.

Discussion

The phenomenon of reverse creep is of great importance relative to the theories based on the hypothesis of creep potential. According to the concept of potential, a creep rate is a single-valued function of the state of stress and stress values, meaning that a loading prehistory does not affect the creep rate. On the other hand, the reverse creep strongly depends on loading regime. Furthermore, the sign of the reverse creep is opposite to that of the acting stress. Therefore, the generalized theory provides broader possibilities than classical creep theories.

Conclusions

The generalized synthetic theory of irreversible deformation is capable of modeling a very wide circle of problems ranging from plastic and steady/unsteady-state creep deformation to non-classical problems of irreversible deformation such as creep delay, the Bauschinger negative effect and reverse creep. This capability results from (i) the uniformed approach to the modeling of irreversible deformation and (ii) the intimate connection between macro-deformation and the processes occurring on the macro-level of material.

Acknowledgement

The author expresses thanks to Prof. K. Rusinko (Budapest University of Technology and Economics, Hungary) for many useful conversations on the topics presented in this article.

References

- [1] Andrusik, J., Rusinko, K.: Plastic Strain of Work-Hardening Materials under Loading in Three-Dimensional Subspace of Five-Dimensional Stress-Deviator Space (in Russian). *Izv. RAN (Russian Academy of Sciences), Mekh. Tverd. Tela* **2**: pp. 92-101, 1993
- [2] Batdorf, S. B., Budiansky, B.: A Mathematical Theory of Plasticity Based on the Concept of Slip. *NACA, Technical Note*, 1949
- [3] Betten, J.: *Creep Mechanics*, Springer, Berlin, 2005
- [4] Chaboche, J. L.: "Unified Cyclic Viscoplastic Constitutive Equations: Development, Capabilities and Thermodynamic Frame Work", *Unified Constitutive Laws of Plastic Deformation*, Kraus and Kraus zeds, Academic Press, 1996

- [5] Chaboche, J. L., Lesne, P. M., Maire, J. F.: Thermodynamic Formulation of Constitutive Equations and Applications to the Viscoplasticity and Viscoelasticity of Metals and Polymers, *Int. J Solid Structures*, 1997
- [6] Chakrabarty, J.: *Applied Plasticity*, Springer-Verlag, New York/Berlin/Heidelberg, 2000
- [7] Chen, W. F., Han, D. J.: *Plasticity for Structural Engineers*, New York, 1988
- [8] Ilyushin, A. A.: *Plasticity*, Moscow, 1963
- [9] Kuksa, L., Lebedev, A., Koval'chuk, B.: Laws of Distribution of Microscopic Strains in Two-Phase Polycrystalline Alloys under Simple and Complex Loading, *Strength of Materials*, **18**: pp. 1-5, 1986
- [10] McLean, D.: *Mechanical Properties of Metals and Alloys*, John Wiley, New York and London, 1977
- [11] Namestnikov, V., Chvostunkov, A. 1960. Creep of Duralumin under Constant and Alternating Loading (in Russian), *Prikl. Mekh. i Tekhn. Fiz.* **4**
- [12] Osipyuk, V.: Explanation and Analytical Description of Delayed Creep, *International Applied Mechanics*, **27**: pp. 374-378, 1991
- [13] Rabotnov, Y.: *Creep Problems in Structural Members*, North-Holland, Amsterdam/London, 1969
- [14] Rusynko, A.: Creep with Temperature Hardening, *J. Materials Science* **33**: pp. 813-817, 1997
- [15] Rusynko, A.: Mathematical Description of Ultrasonic Softening of Metals within the Framework of Synthetic Theory, *J. Materials Science* **37**: pp. 671-676, 2001
- [16] Rusinko, A.: Analytic Dependence of the Rate of Stationary Creep of Metals on the Level of Plastic Prestrain, *J. Strength of Metals* **34**: pp. 381-389, 2002
- [17] Rusinko, A.: Bases and Advances of the Synthetic Theory of Irreversible Deformation, *XXII International Congress of Theoretical and Applied Mechanics (ICTAM) 25-29 August 2008, Adelaide, Australia*, 2008
- [18] Rusinko, A., Rusinko, K.: Synthetic Theory of Irreversible Deformation in the Context of Fundamental Bases of Plasticity, *Int. J. Mech. Mater.* **41**: pp. 106-120, 2009
- [19] Sanders, I.: Plastic Stress-Strain Relations Based on Linear Loading function. *Proc. 2nd US Nat. Congr. Appl. Mech.*, pp. 455-460, 1954

Performance of a Magnetic Fluid-based Short Bearing

R. M. Patel¹, G. M. Deheri², Pragna A. Vadher³

¹ Department of Mathematics, Gujarat Arts and Science College
380 006 Ahmedabad, Gujarat State, India, patel@rediffmail.com

² Department of Mathematics, Sardar Patel University
388 120 Vallabh Vidyanagar, Gujarat State, India, deheri@rediffmail.com

³ Department of Physics, Government Science College
382 016 Gandhinagar, Gujarat State, India, pragnavadher@rediffmail.com

Abstract: An effort has been made to study and analyze the performance of a magnetic fluid based infinitely short hydrodynamic slider bearing. The Reynolds' equation is solved with appropriate boundary conditions. The expressions for various performance characteristics such as pressure, load carrying capacity and friction are obtained. Results are presented graphically. It is clearly seen that the load carrying capacity increases considerably due to the magnetic fluid lubricant. Further, the film thickness ratio increases the load carrying capacity. It is found that the load carrying capacity increases as the ratio of the length to outlet film thickness increases while it decreases with respect to the increasing values of the ratio of the width to the outlet film thickness. In addition, it is investigated that the magnetic fluid lubricant unalters the friction. Lastly, this article makes it clear that the negative effect induced by the ratio of the width to the outlet film thickness can be neutralized up to a considerably large extent by the combined positive effect of the magnetization parameter, the film thickness ratio and the ratio of the length to outlet film thickness. This study provides ample scope for extending the life period of the bearing system.

Keywords: short bearing; magnetic fluid; Reynolds' equation; pressure; load carrying capacity

1 Introduction

The classical analysis of the hydrodynamic lubrication of slider bearings was presented by Pinkus and Sternlicht [1]. Exact solutions of Reynolds' equation for slider bearings with various simple film geometries were discussed in a number of

books and research papers (Cameron [2], Archibald [3], Lord Rayleigh [4], Charnes and Saibel [5], Basu, Sengupta and Ahuja [6], Majumdar [7], Hamrock [8], Gross, Matsch, Castelli, Eshel, Vohr and Wildmann [9]). Prakash and Vij [10] analyzed the hydrodynamic lubrication of a plane slider bearing taking different geometries into consideration. Bagci and Singh [11] dealt with the optimal designs for hydrodynamic lubrication of finite slider bearings considering the effect of one dimensional film shape. Osterle and Saibel [12] studied the effect of bearing deformation in slider bearing lubrication. Here it was concluded that the performance was mostly adversely affected in the sense that the load carrying capacity decreased. Patel and Gupta [13] considered the effect of slip velocity on hydrodynamic lubrication of a porous slider bearing and proved that the load carrying capacity decreased due to the velocity slip. Abramovitz [14] investigated the performance of a pivoted slider bearing with convex pad surfaces. Here it was concluded that the load carrying capacity was more for a convex pad than for a flat one and that such a bearing might be centrally pivoted. Maday [15] extended the analysis of Rayleigh [4] to verify that the optimum slider contains only one step. The design of the optimum one dimensional slider bearing in terms of load carrying capacity was also investigated by McAllister, Rohde and McAllister [16]. Morgan and Cameron [17] were the first to present an approximate analytical solution for the performance characteristics of a porous metal bearing. Later on, Rouleau [18] obtained an exact solution of the above problem.

All these above studies considered conventional lubricants. Agrawal [19] dealt with the configuration of Prakash and Vij [10] with a magnetic fluid lubricant and found its performance better than the one with conventional lubricant. Bhat and Deheri [20] modified and extended the analysis of Agrawal [19] by considering a magnetic fluid based porous composite slider bearing with its slider consisting of an inclined pad and a flat pad. Here it was established that magnetic fluid increased the load carrying capacity, unaltered the friction, decreased the coefficient of friction and shifted the centre of pressure towards the inlet. Prajapati [21] investigated the performance of a magnetic fluid based porous inclined slider bearing with velocity slip and concluded that the magnetic fluid lubricant minimized the negative effect of the velocity slip. Also, Prajapati [22] analyzed the hydrodynamic lubrication of an inclined porous slider bearing with variable porous matrix thickness. The analysis presented was modified and extended by Deheri, Patel and Patel [23] to study the effect of transverse surface roughness on the above configuration in the presence of a magnetic fluid lubricant.

Here an attempt has been made to study and analyze the performance of a magnetic fluid based short bearing wherein, the magnitude associated with the magnetic field is represented by a trigonometric function.

2 Analysis

The configuration of the bearing which is infinitely short in Z – direction is shown below.

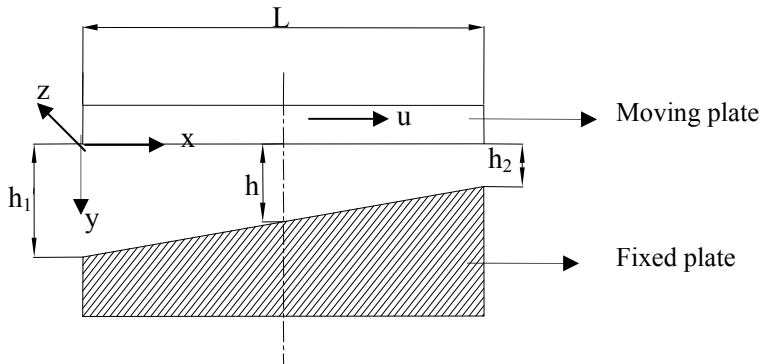


Figure 1

Configuration of the bearing system

The slider moves with the uniform velocity u in the X – direction. The length of the bearing is L and the breadth B is in Z – direction, where $B \ll L$. The dimension B being very small the pressure gradient $\partial p / \partial z$ is much larger than the pressure gradient $\partial p / \partial x$ and hence the latter can be neglected. The lubricant film is considered to be isoviscous and incompressible and the flow is laminar. The magnetic fluid is a suspension of solid magnetic particles of diameter approximately 3-15 nanometer stabilized by a surfactant in a liquid carrier. By applying an external magnetic field these fluids can be confined, positioned, shaped and controlled as desired. The magnetic field is oblique to the stator as in Agrawal [19]. The magnetic field is

$$\bar{H} = (H(z)\cos\phi, 0, H(z)\sin\phi); \phi = \phi(x, z) \quad (1)$$

and the inclination angle is determined from

$$\cot\phi \frac{\partial\phi}{\partial z} + \frac{\partial\phi}{\partial x} = - \frac{1}{H} \frac{dH}{dz} \quad (2)$$

Following the discussions conducted in Verma [25], Prajapati [22] and Bhat and Deheri [26] we consider the magnitude expressed in the form of

$$H^2 = kB^2 \cos\left(\frac{\pi z}{B}\right) \quad (3)$$

where k is chosen to suit the dimensions of both sides and the strength of the magnetic field. Under usual assumptions of magnetohydrodynamic lubrication the

governing Reynolds' equation (Agrawal [19], Prajapati [22], Bhat and Deheri [27]) turns out to be

$$\frac{d^2}{dz^2} \left(p - \frac{\mu_0 \bar{\mu} H^2}{2} \right) = \frac{6\mu u}{h_2^3 \left\{ 1 + m \left(1 - \frac{x}{L} \right) \right\}^3} \frac{dh}{dx} \quad (4)$$

where μ_0 is the magnetic susceptibility, $\bar{\mu}$ is free space permeability, μ is lubricant viscosity and m is the aspect ratio. The associated boundary conditions are

$$p = 0 \text{ at } z = \pm(B/2)$$

and

$$\frac{dp}{dz} = 0 \text{ at } z = 0 \quad (5)$$

Integrating equation (1) with the boundary conditions (2) one obtains the pressure distribution as

$$p = \frac{\mu_0 \bar{\mu} k B^2 \cos\left(\frac{\pi z}{B}\right)}{2} + \frac{3\mu u h_2}{L h_2^3 \left\{ 1 + m \left(1 - \frac{x}{L} \right) \right\}^3} \left(\frac{B^2}{4} - z^2 \right) \quad (6)$$

Introducing the dimensionless quantities

$$\begin{aligned} m &= \frac{h_1 - h_2}{h_2} & X &= \frac{x}{L} & P &= \frac{h_2^3 p}{\mu u B^2} \\ \mu^* &= \frac{h_2^3 k \mu_0 \bar{\mu}}{\mu u} & Y &= \frac{y}{h} & Z &= \frac{z}{B} \end{aligned}$$

one gets the pressure distribution in dimensionless form as

$$P = \frac{\mu^* \cos(\pi Z)}{2} + \frac{3m h_2}{L \{1 + m(1 - X)\}^3} \left(\frac{1}{4} - Z^2 \right) \quad (7)$$

The non-dimensional load carrying capacity is given by

$$W = \frac{h_2^3 w \pi}{\mu u B^4} = \pi \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_0^1 P(X, Z) dX dZ \quad (8)$$

Therefore, the dimensionless load carrying capacity of the bearing is given by

$$W = \frac{\mu^* L h_2}{h_2 B} + \frac{\pi h_2}{4 B} \left[1 - \frac{h_2^2}{h_1^2} \right] \quad (9)$$

The frictional force \bar{F} per unit width of the lower plane of the moving plate is obtained as

$$\bar{F} = \int_{-1/2}^{1/2} \bar{\tau} dZ \quad (10)$$

where

$$\bar{\tau} = \left(\frac{h_2}{\mu u} \right) \tau \text{ is non-dimensional shearing stress} \quad (11)$$

while

$$\tau = \frac{dp}{dz} \left(y - \frac{h}{2} \right) + \frac{\mu u}{h} \quad (12)$$

On simplifications this yields,

$$\bar{\tau} = \frac{dP}{dZ} \frac{B}{h_2} \{1 + m(1 - X)\} \left(Y - \frac{1}{2} \right) + \frac{1}{\{1 + m(1 - X)\}} \quad (13)$$

At $Y = 0$ (at moving plate), we find that

$$\bar{\tau} = \frac{B\mu^* \pi}{4h_2} \{1 + m(1 - X)\} \sin(\pi Z) + \frac{3mh_2 ZB}{Lh_2 \{1 + m(1 - X)\}^2} + \frac{1}{\{1 + m(1 - X)\}} \quad (14)$$

Thus, in non-dimensional form the frictional force assumes the form

$$F_0 = \frac{1}{\{1 + m(1 - X)\}} \quad (15)$$

Further, at $Y = 1$ (at fixed plate), one obtains that

$$\bar{\tau} = -\frac{B\mu^* \pi}{4h_2} \{1 + m(1 - X)\} \sin(\pi Z) - \frac{3mh_2 ZB}{Lh_2 \{1 + m(1 - X)\}^2} + \frac{1}{\{1 + m(1 - X)\}} \quad (16)$$

Lastly, in dimensionless form the frictional force comes out to be

$$F_1 = \frac{1}{\{1 + m(1 - X)\}} \quad (17)$$

3 Results and Discussion

Equations (7) and (9) present the variation of pressure distribution and load carrying capacity while the frictional force is determined from Equation (10). A comparison with the conventional lubricants indicates that the non-dimensional pressure increases by

$$\frac{\mu^* \cos(\pi Z)}{2}$$

while the load carrying capacity enhances by

$$\frac{\mu^* \left(\frac{L}{h_2} \right)}{\left(\frac{B}{h_2} \right)}.$$

Furthermore, a closed scrutiny of the results presented here and the results of the investigation carried out by Patel [24] reveals that the load carrying capacity is approximately four times more here. Besides, the magnetic fluid lubricant unalters the friction which is a clear message from Equations (15) and (17).

In Figures 2-4 we have the analytical non-dimensional pressure distribution with respect to Z for different values of aspect ratio, the ratio of length to outlet film thickness and X respectively. These figures suggest that the pressure is more in the case of ratio of the length to outlet film thickness in the sense that the non-dimensional pressure is more with respect to the ratio of length to outlet film thickness as compared to that of m and X . Besides, it is observed that the pressure increases significantly with respect to the aspect ratio up to the value 0.75 and then it increases marginally.

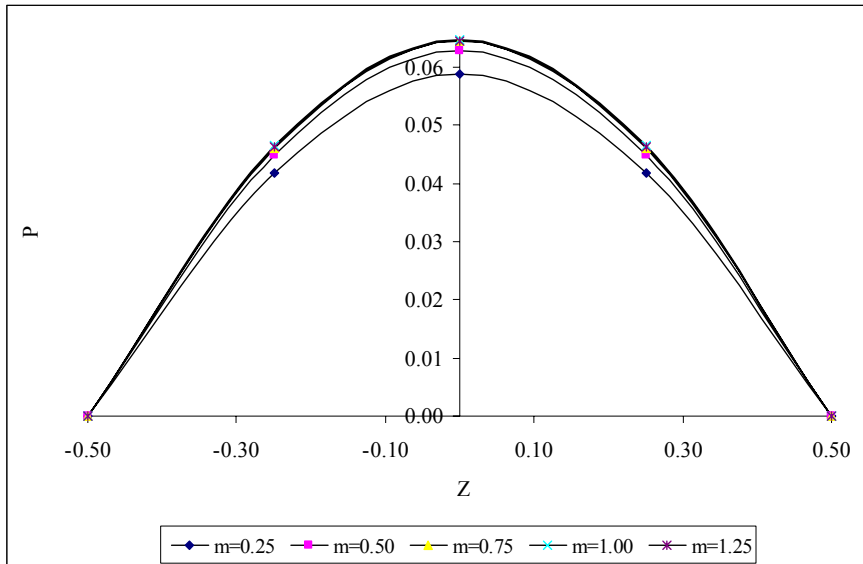


Figure 2

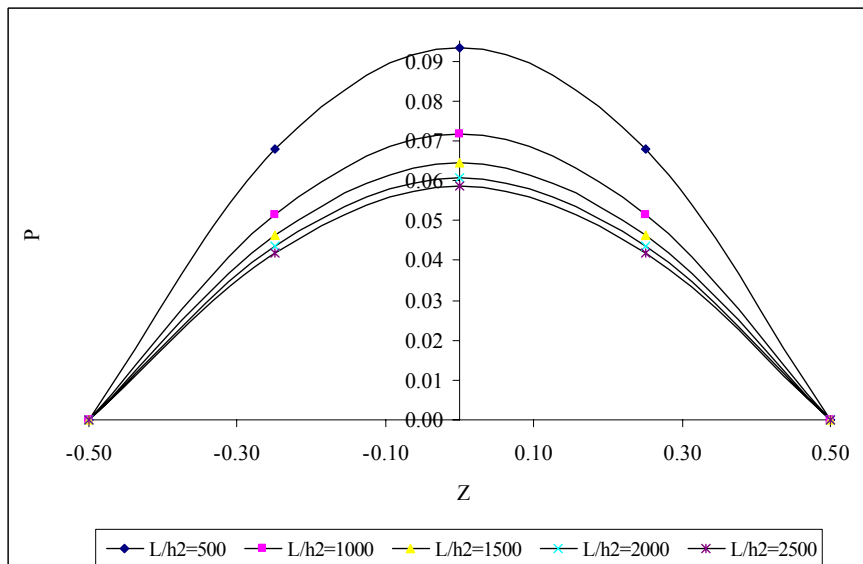
Variation of pressure with respect to Z and m for $\mu^* = 0.001$ 

Figure 3

Variation of pressure with respect to Z and L/h_2 for $\mu^* = 0.001$

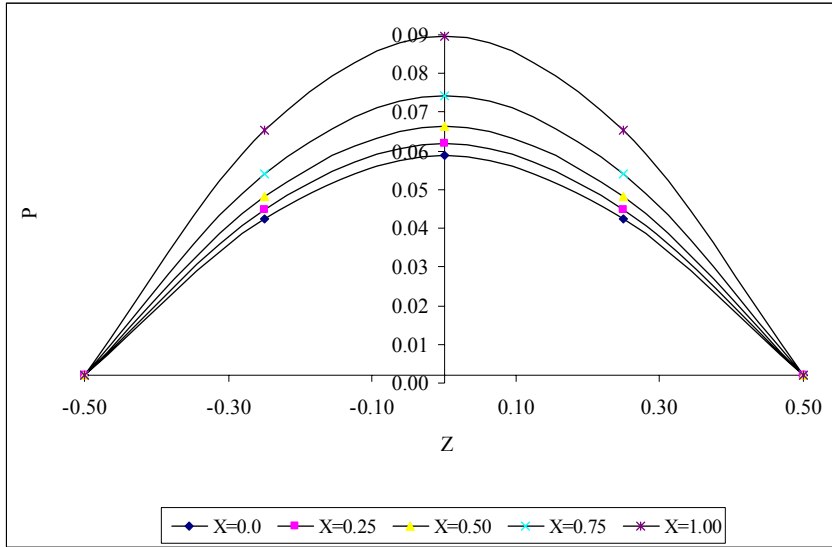


Figure 4
Variation of pressure with respect to Z and X for $\mu^* = 0.001$

In Figures 5-7 the magnitude of the pressure at a given Z coordinate is depicted in function of μ^* for various values of m, L/h_2 and X. These figures reveal that the effect of the ratio of length to outlet film thickness, X and m are negligible with respect to the magnetization parameter.

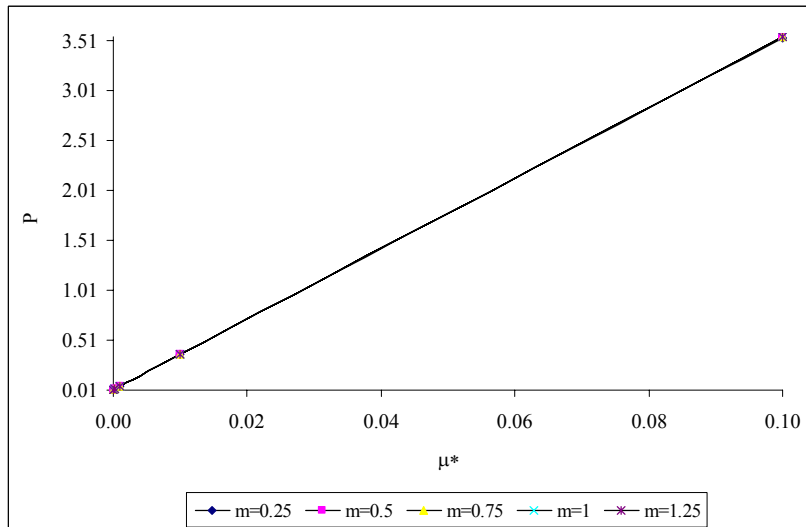


Figure 5
Variation of pressure with respect to μ^* and m for Z = 0.25

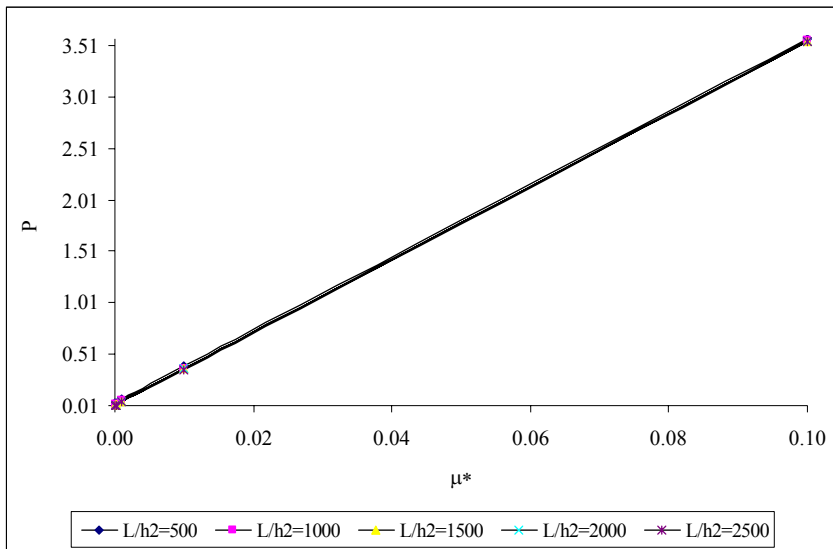


Figure 6

Variation of pressure with respect to μ^* and L/h_2 for $Z = 0.25$

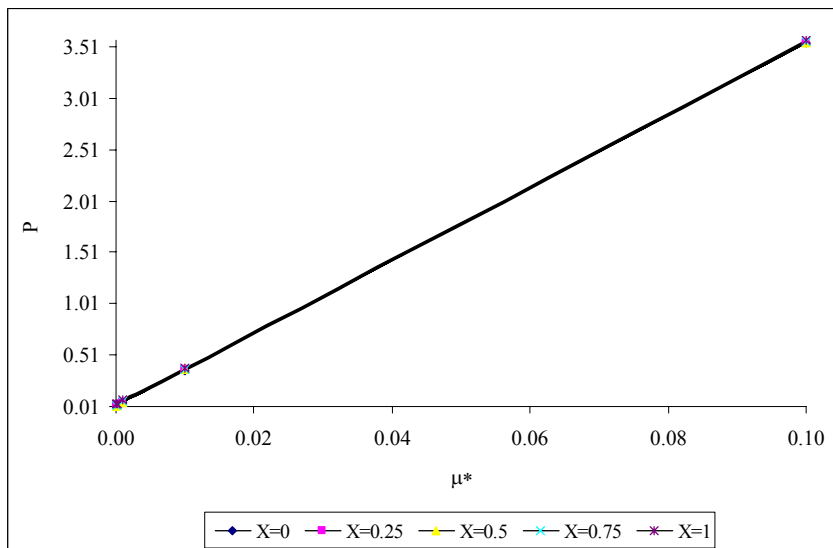


Figure 7

Variation of pressure with respect to μ^* and X for $Z = 0.25$

The variation of load carrying capacity with respect to the magnetization parameter for various values of the ratio of the length to outlet film thickness, the ratio of width to outlet film thickness and the film thickness ratio is presented in

Figures 8-10 respectively. It is clearly seen that the load increases significantly with respect to the magnetization parameter thereby telling that the magnetism induces a positive effect. However, the effect of the film thickness ratio with respect to the magnetization parameter is negligible as indicated by Figure 10.

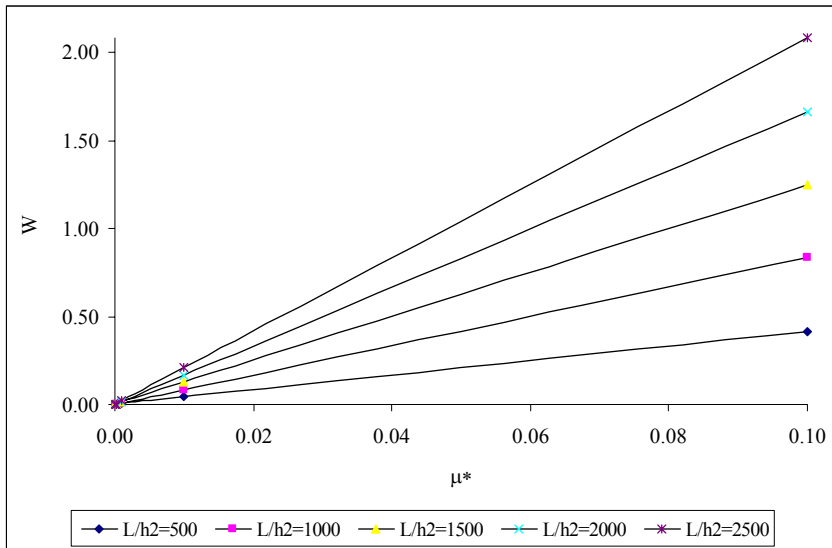


Figure 8
Variation of load carrying capacity with respect to μ^* and L/h_2

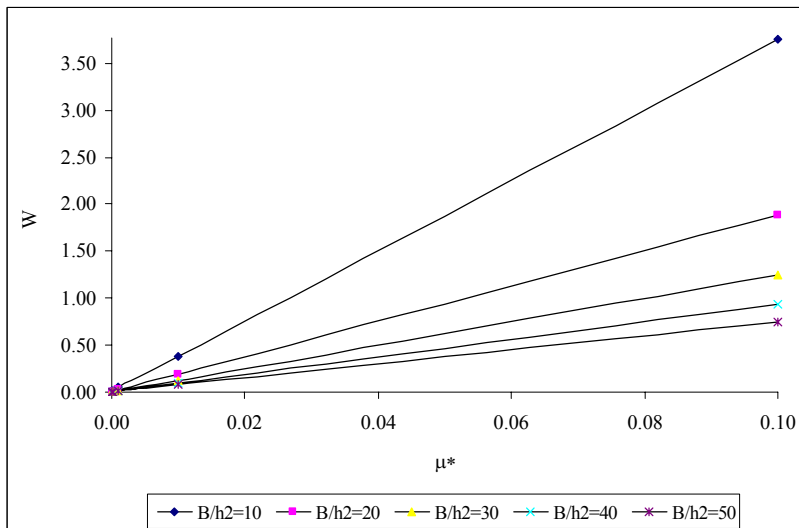


Figure 9
Variation of load carrying capacity with respect to μ^* and B/h_2

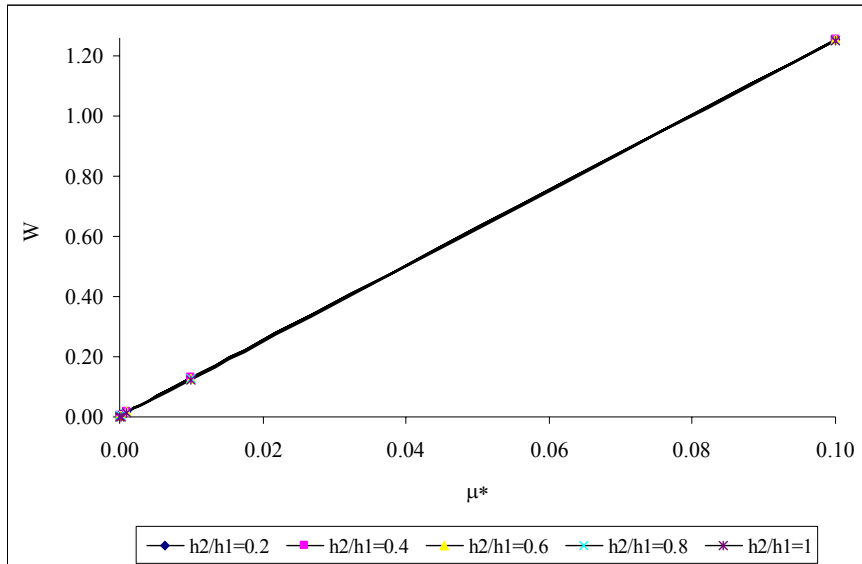


Figure 10

Variation of load carrying capacity with respect to μ^* and h_2/h_1

Figures 11 and 12 describe the profile for load carrying capacity with respect to the ratio of the length to outlet film thickness for several values of the ratio of width to outlet film thickness and the film thickness ratio. These two figures tell that the load carrying capacity increases considerably due to the ratio L/h_2 . Further, it is observed that the ratio of width to outlet film thickness and the film thickness ratio h_2/h_1 decrease the load carrying capacity. In addition, the rate of increase in load carrying capacity with respect to L/h_2 remains uniform in the case of the film thickness ratio while it increases in the case of the ratio of width to outlet film thickness. Thus, the aspect ratio increases the load carrying capacity considerably.

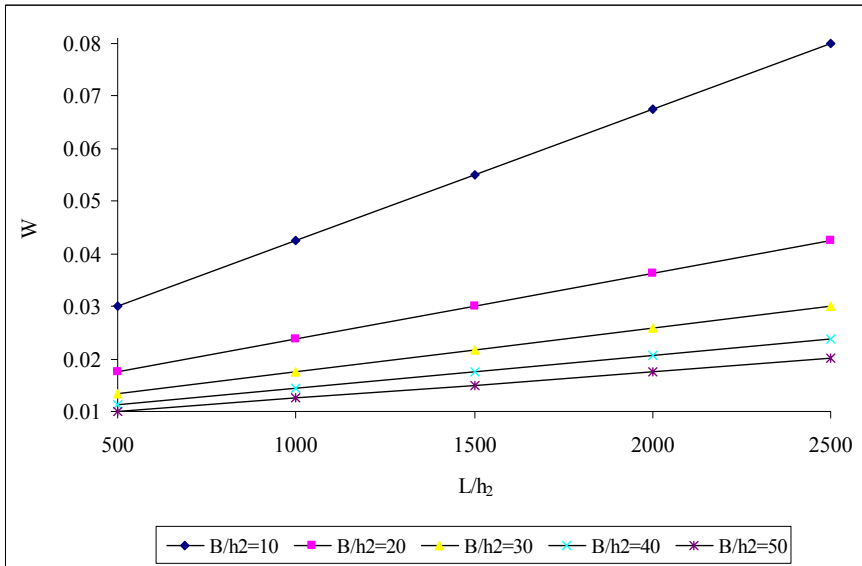


Figure 11
Variation of load carrying capacity with respect to L/h_2 and B/h_2

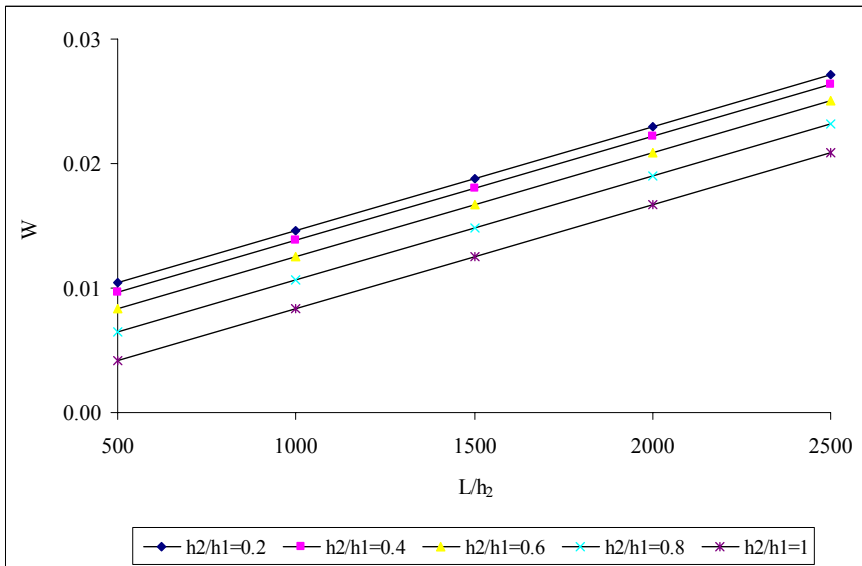


Figure 12
Variation of load carrying capacity with respect to L/h_2 and h_2/h_1

Lastly, Figure 13 presents the variation of load carrying capacity with respect to the ratio of width to outlet film thickness for various values of film thickness ratio h_2/h_1 . It is clearly visible that the combined effect of these two parameters namely B/h_2 and h_2/h_1 , is considerably adverse. The decrease in the load carrying capacity due to B/h_2 is more at the beginning while the reverse is true with respect to h_2/h_1 .

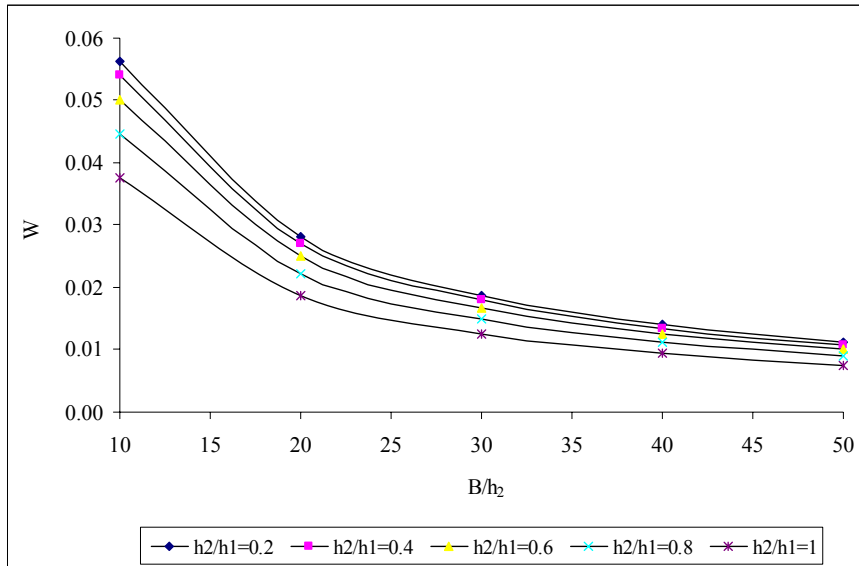


Figure 13

Variation of load carrying capacity with respect to B/h_2 and h_2/h_1

This article reveals that the negative effect induced by B/h_2 can be neutralized almost completely by the combined positive effect of magnetization parameter, L/h_2 and the aspect ratio.

Conclusion

A comparison of this present article with those of the studies carried out by Patel [24] and Bhat and Deheri [27] makes it clear that there is at least a four-time increase in the load carrying capacity here. The importance of this work lies in the fact that it offers an additional degree of freedom from a magnitude point of view in light of the investigations conducted by Prajapati [21] and Verma [25]. Moreover, the present article may open a new route towards the optimal performance of the bearing system. In addition, the results presented here tend to suggest that there exists enough scope for extending the life period of the bearing system by controlling lubricant loss.

Acknowledgements

The minute observations, fruitful suggestions and critical comments of the referee are gratefully appreciated. We acknowledge with thanks the positive attitude

adopted by the referee during the course of reviewing. Two of the authors, R. M. Patel and G. M. Deheri thank UGC for the funding of U. G. C. major research project (U. G. C. F. No. 32-143/2006 (SR) – “Magnetic fluid based rough bearings”) under which this study has been carried out.

References

- [1] Pinkus, O., Sternlicht, B., Theory of Hydrodynamic Lubrication, McGraw-Hill, New York (1961)
- [2] Cameron, A., The Principles of Lubrication, 1966, Longmans, London
- [3] Archibald, F. R., A Simple Hydrodynamic Thrust Bearing, ASME, 72, 1950, p. 393
- [4] Lord Rayleigh, Notes on the Theory of Lubrication, Phil. Mag., 35, 1918, pp. 1-12
- [5] Charnes, A., Saibel, E., On the Solution of the Reynolds' Equation for Slider Bearing Lubrication, Part 1, ASME, 74, 1952, p. 867
- [6] Basu, S. K., Sengupta, S. N., Ahuja, B. B., Fundamentals of Tribology, Prentice-Hall of India Private Limited, New Delhi, 2005
- [7] Majumdar, B. C., Introduction to Tribology of Bearings, S. Chand and Company Limited, New Delhi, 2008
- [8] Hamrock, B. J., Fundamentals of Fluid Film Lubrication, McGraw-Hill, Inc. New York, 1994
- [9] Gross, W. A., Matsch Lee, A., Castelli, V., Eshel, A., Vohr, J. H., Wildmann, M., Fluid Film Lubrication, 1980, A Wiley-Interscience Publication, John Wiley and Sons, New York
- [10] Prakash, J., Vij, S. K., Hydrodynamic Lubrication of Porous Slider, Journal of Mechanical Engineering and Science, 15, 1973, pp. 232-234
- [11] Bagci, C., Singh, A. P., Hydrodynamic Lubrication of Finite Slider Bearing: Effect of One Dimensional Film Shape and their Computer-aided Optimum Designs, Journal of Lubrication Technology, ASLE Transaction, 105, 1983, pp. 48-66
- [12] Osterle, F., Saibel, E., The Effect of Bearing Deformation in Slider Bearing Lubrication, Journal of Lubrication Technology, ASLE Transactions, 1, 1958, pp. 213-216
- [13] Patel, K. C., Gupta, J. L., Hydrodynamic Lubrication of a Porous Slider Bearing with Slip Velocity, Wear, 85, 1983, pp. 309-317
- [14] Abramovitz, S., Theory for a Slider Bearing with a Convex Pad Surface; Side Flow Neglected, Journal of Franklin Inst., 259, 1955, p. 221
- [15] Maday, C. J., A Bounded Variable Approach to the Optimum Slider Bearing, ASME Journal of Lubrication Technology, 99, 1947, p. 180

- [16] McAllister, G. T., Rohde, S. M., McAllister, M. W., A Note on the Optimum Design of Slider Bearing, ASME Journal of Lubrication Technology, 102(1), 1980, p. 117
- [17] Morgan, V. T., Cameron, A., Mechanism of Lubrication in Porous Metal Bearings, Proc. of the Conf. on Lubrication and Wear. Inst. Mech. Engrs., London, Paper 89, 1957, pp. 151-157
- [18] Rouleau, W. T., Hydrodynamic Lubrication of Narrow Press-fitted Porous Metal Bearings, ASME Journal of Lubrication Technology, 85, 1963, p. 123
- [19] Agrawal, V. K., Magnetic Fluid-based Porous Inclined Slider Bearing, Wear, 107, 1986, pp. 133-139
- [20] Bhat, M. V., Deheri, G. M., Porous Composite Slider Bearing Lubricated with Magnetic Fluid, Japanese Journal of Applied Physics, 30, 1991, pp. 2513-2514
- [21] Prajapati, B. L., Magnetic Fluid-based Porous Inclined Slider Bearing with Velocity Slip, Prajna, 1994, pp. 73-78
- [22] Prajapati, B. L., On Certain Theoretical Studies in Hydrodynamics and Electro Magnetohydrodynamic Lubrication, Dissertation, S. P. University, Vallabh Vidhyanagar, 1995
- [23] Deheri, G. M., Patel H. C., Patel, R. M., Behavior of Magnetic Fluid-based Squeeze Film between Porous Circular Plates with Porous Matrix of Variable Thickness, International Journal of Fluid Mechanics Research, 34(6), 2007, pp. 506-514
- [24] Patel, N. S., Analysis of Magnetic Fluid-based Hydrodynamic Slider Bearing, Thesis M. Tech., S. V. N. I. T., Surat, 2007
- [25] Verma, P. D. S., Magnetic Fluid-based Squeeze Films, Int. J. Eng. Sci., 24(3), 1986, pp. 395-401
- [26] Bhat, M. V., Deheri, G. M., Squeeze Film Behavior in Porous Annular Discs Lubricated with Magnetic Fluid, Wear, 151, 1991, pp. 123-128
- [27] Bhat, M. V., Deheri, G. M., Porous Slider Bearings with Squeeze Film Formed by a Magnetic Fluid, Pure and Applied Mathematika Sciences, 39 (1-2), 1995, pp. 39-43

Nomenclature:

h	Fluid film thickness at any point (mm)
h_1	Maximum film thickness (mm)
h_2	Minimum film thickness (mm)
H^2	Magnetic field
k	Suitably chosen constant associated with the magnetic field
L	Length of the bearing (mm)

B	Breadth of the bearing (mm)
m	Aspect ratio
p	Lubricant pressure (N/mm ²)
P	Dimensionless pressure
u	Uniform velocity in X direction
w	Load carrying capacity (N)
W	Non-dimensional load carrying capacity
ϕ	Inclination angle of the magnetic field
μ	Lubricant viscosity (N.s/mm ²)
μ_0	magnetic susceptibility
$\bar{\mu}$	Free space permeability
μ^*	Dimensionless magnetization parameter
τ	Shear stress (N/mm ²)
$\bar{\tau}$	Dimensionless shear stress
F	Frictional force (N)
\bar{F}	Dimensionless frictional force
F_0	Non-dimensional frictional force (at moving plate)
F_1	Non-dimensional frictional force (at fixed plate)
\bar{H}	Magnetic field

A Copula-based Approach to the Analysis of the Returns of Exchange Rates to EUR of the Visegrád Countries¹

Magda Komorníková¹, Jozef Komorník²

¹ Faculty of Civil Engineering, Slovak University of Technology, Radlinského 11, 813 68 Bratislava, Slovakia, magda@math.sk

² Faculty of Management, Comenius University, Odbojárov 10, P.O.BOX 95, 820 05 Bratislava, Slovakia, jozef.komornik@fm.uniba.sk

Abstract: The currencies of the Visegrád countries (Poland, the Czech Republic, Hungary, and Slovakia) have been considered by the international financial community as a basket of currencies which are closely related, especially in times of their depreciations. On July 1, 2008 the official terminal exchange rate SKK/EUR was fixed. During the following 8 months, the remaining three currencies (PLN, CZK, HUF) changed their long-term behaviour to one of strong parallel depreciation. On the other hand, in the first selected long-term period (January 4, 1999 – June 30, 2008), a relatively mixed development of HUF seemed to exhibit a rather low degree of interdependence with CZK (that had been appreciating very intensively). The values of the Kendall's correlation coefficient calculated for all 3 remaining couples of returns substantially rose in the second period (indicating that similarities between the returns of these exchange rates are stronger in the times of crises). We have performed modeling and fitting of the dependencies of the above mentioned couples of returns of currencies in both the mentioned time periods by several classes of bivariate copulas, as well as by (optimized) convex combinations of their elements.

Keywords: bivariate copula; return of exchange rates; Kendall's tau; convex combinations of copulas; goodness of fit (GOF) test

1 Introduction

The aim of this paper is to further extend our earlier studies of the relations between the returns of couples of exchange rates of the Visegrád countries to EUR ([9], [10]). We have again extended the considered time span until the end of July

¹ The preliminary version of this contribution was presented at international summer school AGOP 2009 in Palma de Mallorca.

2009. We have also deepened the analytical tools of our former copula approach analyses, inspired by several preceding papers dealing with exchange rates modeling ([7, 8, 12]).

The currencies of the Visegrád countries (PLN, CZK, HUF, SKK) were considered by the international financial community as a basket of currencies that were closely related especially in turbulent times. Consequently, several common features in their behavior were expected, and were often also observed.

On July 1, 2008 the official terminal exchange rate SKK/EUR was fixed. Although this country officially entered the EUR zone only 6 months after, that exchange rate was essentially frozen in the meantime.

During the following 13 months, the remaining three currencies (PLN, CZK, HUF) changed their long-term behavior to a strong parallel depreciation until March 2009, when they started to appreciate again. This change was an obvious consequence of the extremely severe crisis of the global financial system that started in the middle of 2008 and which has slightly reversed since March 2009. Let us specify that for daily values of EUR, in the considered currencies the corresponding returns are defined by $R_t = (P_t - P_{t-1})/P_{t-1}$ where P_t is the exchange rate in time t . The time series of daily values of EUR in the considered currencies and the corresponding returns are presented in the Figures 1a-1d.

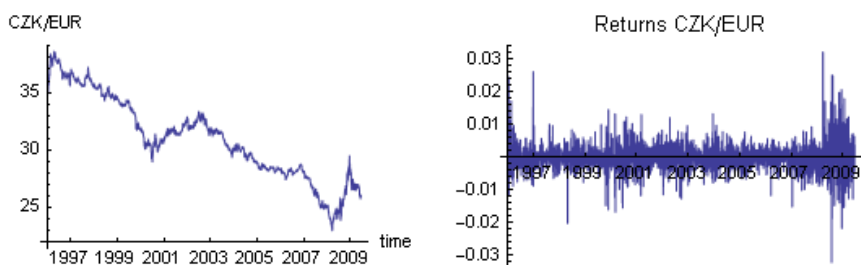


Figure 1a

Exchange rates of the Czech Crowns to EUR and the corresponding returns

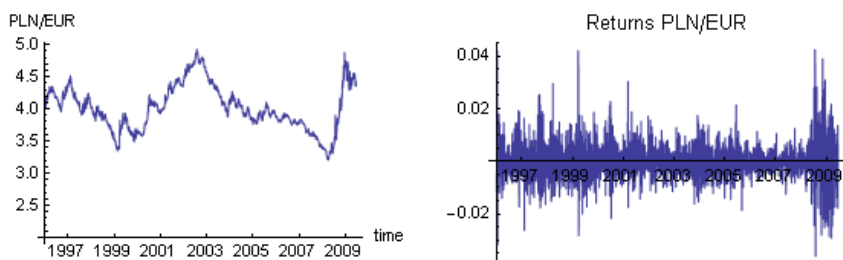


Figure 1b

Exchange rates of the Polish Zloty to EUR and the corresponding returns

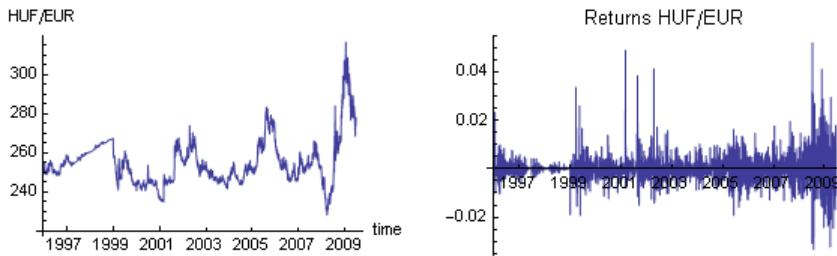


Figure 1c

Exchange rates of the Hungary Forint to EUR and the corresponding returns

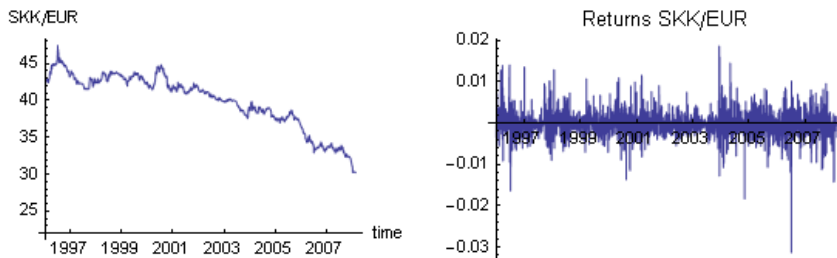


Figure 1d

Exchange rates of the Slovak Crowns to EUR and the corresponding returns

The values of the empirical versions of Kendall's correlation coefficients (cf. e.g. [1, 4, 5]) calculated for all considered couples of returns (presented in Table 1 for the first period January 4, 1999 – June 30, 2008, the second crisis period July 1, 2008 – July 31, 2009 and for the merged period January 4, 1999 – July 31, 2009) are in accordance with the previous qualitative reasoning. Their respective values for all three remaining couples of returns of exchange rates rose substantially in the second period (indicating that similarities between returns of exchange rates of the couples of these currencies are stronger in times of crises).

Table 1

The values of the empirical Kendall's coefficients τ for returns

Couple	4.1.1999 – 30.6.2008	4.1.1999 – 31.7.2009	1.7.2008 - 31.7.2009
(SKK/EUR, CZK/EUR)	0,231	x	x
(SKK/EUR, PLN/EUR)	0,214	x	x
(SKK/EUR, HUF/EUR)	0,240	x	x
(CZK/EUR, HUF/EUR)	0,167	0,209	0,443
(CZK/EUR, PLN/EUR)	0,217	0,246	0,423
(PLN/EUR, HUF/EUR)	0,319	0,345	0,509

We subsequently performed modeling and fitting of the dependencies of the above mentioned couples of returns of currencies separately for 2 periods, before and after July 1, 2008, as well as for the whole considered time period (January 4, 1999 – July 31, 2009) by several classes of bivariate copulas as well as by convex combinations of their elements. Based on our previous modeling experiments we utilized 3 well known 1–parametric classes of Archimedean copulas (Gumbel, Clayton, Frank) and the 2–parametric Joe BB1 copula.

2 Theoretical Basis

Recall that the most important applications of 2–dimensional copulas are related to a well known and very convenient alternative for expressing the joint distribution function F of a vector of continuous random variables (X, Y) in the form

$$F(x, y) = C(F_X(x), F_Y(y)), \quad (1)$$

where F_X, F_Y are the marginal distribution functions. Note that the copula $C : [0, 1]^2 \rightarrow [0, 1]$ is unique whenever X and Y are continuous random variables (see e.g. [11]).

2.1 Tail Dependencies between Random Variables

For a given copula $C(x, y)$, the upper and lower tail dependencies can be defined with reference to how much probability is in regions near $(1, 1)$ (upper-right-quadrant tail) and $(0, 0)$ (lower-left-quadrant tail). Let (X, Y) be a vector of continuous random variables with marginal distribution functions F_X, F_Y . The coefficient λ_U of *upper tail dependence* of (X, Y) is (see e.g. [2])

$$\lambda_U = \lim_{u \uparrow 1} P\{Y > F_Y^{-1}(u) | X > F_X^{-1}(u)\} = \lim_{u \uparrow 1} \frac{1 - 2u + C(u, u)}{1 - u} \quad (2)$$

provided that the limit $\lambda_U \in [0, 1]$ exists. If $\lambda_U > 0$, X and Y are said to be asymptotically dependent in the upper tail; X and Y are said to be asymptotically independent in the upper tail if $\lambda_U = 0$.

The coefficient λ_L of *lower tail dependence* of (X, Y) is

$$\lambda_L = \lim_{u \downarrow 0} P\{Y < F_Y^{-1}(u) | X < F_X^{-1}(u)\} = \lim_{u \downarrow 0} \frac{C(u, u)}{u} \quad (3)$$

provided that the limit $\lambda_L \in [0, 1]$ exists. If $\lambda_L > 0$, X and Y are said to be asymptotically dependent in the lower tail; X and Y are said to be asymptotically independent in the lower tail if $\lambda_L = 0$.

2.2 Some Classes of Bivariate Copulas

Table 2 presents a summary of basic facts (presented e.g. in [2, 11, 14]) that are related to the families of classes copulas that we utilize in our analyses.

It is well known ([14]) that the Gumbel class is a limiting case of the Joe BB1 class for $a \rightarrow 0$, while its special case for $b = 1$ is the Clayton class.

Table 2
Characteristics for some Archimedean copulas

Family of copulas	Parameters	Bivariate copula $C(u,v)$	λ_L	λ_U
Gumbel	$b \geq 1$	$\exp\left\{-\left[(-\ln(u))^b + (-\ln(v))^b\right]^{1/b}\right\}$	0	$2 - 2^{1/b}$
Clayton	$a > 0$	$(u^{-a} + v^{-a} - 1)^{-1/a}$	$2^{-1/a}$	0
Frank	$\theta \in \mathfrak{R}$	$-\frac{1}{\theta} \ln\left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{(e^{-\theta} - 1)}\right)$	0	0
Joe BB1	$b \geq 1, a > 0$	$\left\{1 + \left[(u^{-a} - 1)^b + (v^{-a} - 1)^b\right]^{1/b}\right\}^{-1/a}$	$2^{-1/ab}$	$2 - 2^{1/b}$

We can observe that the coefficients λ_L and λ_U can attain values in the whole interval $(0, 1)$ for Joe BB1 copulas, while the same holds for λ_L for strict Clayton copulas and for λ_U in case of Gumbel copulas. Both λ_U and λ_L are equal to 0 for Frank copulas, while $\lambda_L = 0$ for Gumbel copulas and $\lambda_U = 0$ for Clayton copulas. It is also well known (see [2]) that $\lambda_L = \lambda_U = 0$ for so-called normal copulas. More detailed analyses (accompanied by graphical illustrations) related to the tail dependence coefficients can be found in [14]. These coefficients (called also parameters) are there treated as the limit values of the left and right tail concentration functions

$$L(u) = P(V < u \mid U < u) = P(U < u \mid V < u)$$

and

$$R(u) = P(V > u \mid U > u) = P(U > u \mid V > u)$$

with $U = F_X(x)$, $V = F_Y(y)$, (that yields $P(U < u) = P(V < u) = u$).

For the Joe BB1 class, it is shown in [14] that the values of the theoretical Kendall correlation coefficient $\tau = 1 - \frac{2}{b(a+2)}$ determine a growing system (in τ) of decreasing dependencies between λ_L and λ_U (which can attain maximum values

of λ_U slightly greater than τ). Consequently, Gumbel copulas have λ_U greater than any Joe BB1 copulas with the same value of τ . Similarly, Clayton copulas have λ_L greater than any Joe BB1 copulas with the same value of τ .

2.3 Convex Combinations of Copulas

A very useful tool for fitting the investigated copulas of time series has been obtained in the classes of convex combinations of copulas $C_{\theta_1}(u, v)$ and $C_{\theta_2}(u, v)$ with the weight coefficients α and $(1 - \alpha)$ that have the form

$$C_{\theta_1, \theta_2, \alpha}(u, v) = \alpha C_{\theta_1}(u, v) + (1 - \alpha) C_{\theta_2}(u, v)$$

for $\alpha \in [0, 1]$. It is obvious that the relations

$$\lambda_L = \alpha \lambda_{1,L} + (1 - \alpha) \lambda_{2,L}, \quad \lambda_U = \alpha \lambda_{1,U} + (1 - \alpha) \lambda_{2,U}$$

hold for the coefficients of lower and upper tail dependencies of the considered original and resulting copulas.

2.4 Fitting of Copulas

In practical fitting of the data we utilized the *maximum pseudolikelihood method* (MPLE) of parameter estimation (with initial parameters estimate received by the minimalization of the mean square distance to the empirical copula C_n presented e.g. in [5]). It requires that the copula $C_\theta(u, v)$ is absolutely continuous with

density $c_\theta(u, v) = \frac{\partial^2}{\partial u \partial v} C_\theta(u, v)$. This method (described e.g. in [5]) involves

maximizing a rank-based log-likelihood of the form

$$L(\theta) = \sum_{i=1}^n \ln \left(c_\theta \left(\frac{R_i}{n+1}; \frac{S_i}{n+1} \right) \right) \quad (4)$$

where n is the sample size and θ is vector of parameters in the model. Note that arguments $\frac{R_i}{n+1}, \frac{S_i}{n+1}$ equal to corresponding values of empirical marginal distributional functions of random variables X and Y .

2.5 Goodness of Fit (GOF) Test

We followed the approach of [13] and [15] for goodness of fit test measuring the size of misspecification in the form of the statistics with asymptotical distribution of the type $\chi_{p(p+1)/2}^2$ where p is the number of the estimated parameters. We use a

simplified version of this statistics suggested for practical purposes in [15]. As a compensation for this simplification we only reject the tested models if the corresponding P-value $< 0, 01$.

To compare goodness of fit of the models from several classes of copulas, we apply the Takeuchi criterion TIC ([6]) that is a robustified version of the famous Akaike criterion.

3 Review of Results

For each of considered periods (4.1.1999 – 31.7.2009, 4.1.1999 – 30.6.2008, 1.7.2008 – 31.7.2009), each couple of considered returns of exchange rates and each class of the considered Archimedean copulas (as well as for all convex combinations of their couples) we perform the following sequence of procedures:

- 1 least squares initial estimates of the model parameters θ (by minimizing the L_2 distance $d(C_\theta, C_n)$ from the empirical copula),
- 2 calculation of MPLE estimates of the model parameters θ and TIC ,
- 3 goodness of fit tests (rejecting the models with P-value $< 0, 01$).
- 4 Finally, we choose among the considered classes of copulas with non-rejected models according to the minimalization of the TIC criterion. Subsequently, we calculate lower and upper tail dependencies λ_L and λ_U (using their relations to the model parameters, where we enter the MPLE estimates of those parameters).

3.1 Models for the First Period (4. 1. 1999 – 30. 6. 2008)

a) Archimedean Copulas

Among 4 considered Archimedean copulas only the Gumbel class provided models for all 6 considered couples that had not been subsequently rejected by the GOF test described above. The Clayton class provided such models for the first, fourth, fifth and sixth couples, while the Frank class did it for the last three couples.

The values of the TIC criterion were minimized for the Gumbel class models for the first four couples and for the Frank class models for the remaining two couples.

Note that no models in the 2-parametric Joe BB1 class passed the GOF tests.

Tables 3(a) and 4(a) present the MPLE estimates $\hat{\theta}$ of parameters for optimal copulas for all 6 couples of currencies, P-values corresponding to goodness of fit test statistics χ^2 as well as the minimizing values of *TIC* and the respective values of the L_2 distances to empirical copulas (which may be reduced in comparison with local minima found in the original least squares error approximation). Finally we also present the values of the coefficients of tail dependencies λ_L and λ_U . Note that the values of the coefficients $\hat{\theta}$ are close to each other and also the distances $d(C_\theta, C_n)$ from the corresponding empirical copulas are not dramatically different.

b) Convex Combinations of Copulas

The optimal models for all couples of currencies with the corresponding results of model parameters, P-values, *TICs*, L_2 -distances, λ_L and λ_U for optimal models are presented in Tables 3(b) and 4(b).

Table 3
Results for the pairs of returns of exchange rates including SKK/EUR

a) Archimedean copulas class

Couple	(SKK/EUR, CZK/EUR)	(SKK/EUR, PLN/EUR)	(SKK/EUR, HUF/EUR)
Copula's type	Gumbel	Gumbel	Gumbel
θ	1,279	1,248	1,294
P-value	0,470	0,166	0,155
<i>TIC</i>	-500,47	-271,09	-265,26
$d(C_\theta, C_n)$	0,420	0,403	0,347
λ_L	0,000	0,000	0,000
λ_U	0,281	0,258	0,291

b) The optimal convex combinations of Archimedean copulas

Couple	(SKK/EUR, CZK/EUR)	(SKK/EUR, PLN/EUR)	(SKK/EUR, HUF/EUR)
Copula's type	Frank+Joe BB1	Clayton+Gumbel	Gumbel+Joe BB1
α	0,140	0,124	0,864
θ_1	6,369	1,595	1,208
$\theta_2 = (b_2; a_2)$	(1,189; 0,066)	(1,215; x)	(2,210; 0,296)
P-value	0,118	0,123	0,344
<i>TIC</i>	-529,82	-279,76	-288,89
$d(C_\theta, C_n)$	0,262	0,305	0,315
λ_L	0,0004	0,080	0,047
λ_U	0,180	0,202	0,280

Interestingly, models including the Joe BB1 copulas also passed the GOF tests. This enables us to model simultaneously non-zero lower and upper tail dependencies (which is also possible for convex combinations of Gumbel and Clayton classes). Furthermore, we can observe that for most couples the values of λ_U of optimal models are substantially larger than those of λ_L (this first period was dominated by the appreciation of the considered currencies, mainly SKK and CZK). The only exception is the couple (CZK/EUR, HUF/EUR) where the model in the combination of classes Clayton – Frank had a lower value of *TIC* than one in the Gumbel – Joe BB combination (with $\lambda_U > \lambda_L > 0$), which also passed the GOF test.

Table 4

Results for the returns of exchange rates for the remaining couples of exchange rates for the first period
4.1.1999 - 30.6.2008

a) Archimedean copulas class

Couple	(CZK/EUR, PLN/EUR)	(CZK/EUR, HUF/EUR)	(HUF/EUR, PLN/EUR)
Copula's type	Gumbel	Frank	Frank
θ	1,245	1,554	3,117
P-value	0,067	0,033	0,111
<i>TIC</i>	-262,98	-152,03	-500,06
$d(C_\theta, C_n)$	0,545	0,309	0,418
λ_L	0,000	0,000	0,000
λ_U	0,255	0,000	0,502

b) The optimal convex combinations of Archimedean copulas

Couple	(CZK/EUR, PLN/EUR)	(CZK/EUR, HUF/EUR)	(HUF/EUR, PLN/EUR)
Copula's type	Frank+Gumbel	Clayton+Frank	Frank+Joe BB1
α	0,666	0,503	0,658
θ_1	1,568	0,001	3,101
$\theta_2 = (b_2; a_2)$	(1,459; x)	(3,407; x)	(1,363; 0,138)
P-value	0,151	0,100	0,202
<i>TIC</i>	-279,44	-154,88	-587,87
$d(C_\theta, C_n)$	0,269	0,284	0,427
λ_L	0,000	0,000	0,009
λ_U	0,131	0,000	0,115

3.2 Models for the Second Period (1. 7. 2008 - 31. 7. 2009)

The results for the second period are presented in Table 5. For all 3 considered pairs of exchange rates, optimal models in all three 1-parametric Archimedean copulas classes passed the GOF tests. The optimal models in the Joe BB1 class again did not pass the GOF tests for either of the 3 couples of exchange rates. The minimal values for the TIC criterion were attained for the optimal model in the Gumbel class for the first couple and in the Frank class for remaining 2 pairs.

This time we have $\lambda_L > \lambda_U$ for the last 2 pairs for exchange rates and the dominance of λ_U over λ_L is also dramatically reduced for the first pair. This dramatic change (in comparison to the corresponding models for the first period) can be related to the fact that all 3 considered currencies strongly depreciated in the second period.

Note that the value of $d(C_\theta, C_n)$ increased dramatically in comparison to the corresponding value for the first period.

Table 5

Results for the returns of exchange rates for the remaining couples of exchange rates for the second period 1.7.2008 - 31.7.2009

a) Archimedean copulas class

Couple	(CZK/EUR, PLN/EUR)	(CZK/EUR, HUF/EUR)	(HUF/EUR, PLN/EUR)
Copula's type	Gumbel	Frank	Frank
θ	1,716	4,831	5,952
P-value	0,078	0,454	0,159
<i>TIC</i>	-126,07	-130,47	-177,49
$d(C_\theta, C_n)$	2,688	2,177	1,770
λ_L	0,000	0,000	0,000
λ_U	0,502	0,000	0,000

b) The optimal convex combinations of Archimedean copulas

Couple	(CZK/EUR, PLN/EUR)	(CZK/EUR, HUF/EUR)	(HUF/EUR, PLN/EUR)
Copula's type	Clayton+Gumbel	Frank+Joe BB1	Frank+Joe BB1
α	0,450	0,196	0,221
θ_1	0,554	9,052	3,639
$\theta_2 = (b_2; a_2)$	(2,632; x)	(1,367; 0,454)	(1,674; 0,665)
P-value	0,016	0,247	0,043
<i>TIC</i>	-139,34	-140,96	-188,53
$d(C_\theta, C_n)$	1,758	1,551	1,286
λ_L	0,129	0,273	0,417
λ_U	0,384	0,263	0,379

3.3 Models for Whole (Merged) Period

Despite the dramatic differences between respective models for the first and the second period, we also calculated models for 3 pairs of currencies that can be analyzed through the whole merged period (4. 1. 1999 – 31. 7. 2009).

The results of computations for the whole period are presented in the Table 6. We can observe that the resulting optimal models have distances to the empirical copulas that are comparable to those of the corresponding models for the dominating first period.

On the other hand, despite the fact that the second period represents less than 10% of the data, the parameters of tail dependencies for the resulting optimal models among convex combinations of Archimedean copulas for the whole time period moved disproportionately closer to those for the corresponding models for the second period.

Table 6

Results for the returns of exchange rates for the remaining couples of exchange rates for the whole (merged) period

a) Archimedean copulas class

Couple	(CZK/EUR, PLN/EUR)	(CZK/EUR, HUF/EUR)	(HUF/EUR, PLN/EUR)
Copula's type	Gumbel	Gumbel	Gumbel
θ	1,319	1,266	1,483
P-value	0,01	0,06	0,17
<i>TIC</i>	-462,33	-356,65	-807,75
$d(C_\theta, C_n)$	0,374	0,349	0,677
λ_L	0,000	0,000	0,000
λ_U	0,309	0,271	0,402

b) The optimal convex combinations of Archimedean copulas

Couple	(CZK/EUR, PLN/EUR)	(CZK/EUR, HUF/EUR)	(HUF/EUR, PLN/EUR)
Copula's type	Gumbel+Joe BB1	Frank+Joe BB1	Frank+Joe BB1
α	0,295	0,115	0,285
θ_1	1,007	-3,089	2,013
$\theta_2 = (b_2; a_2)$	(1,390; 0,236)	(1,251; 0,229)	(1,420; 0,335)
P-value	0,011	0,338	0,207
<i>TIC</i>	-498,06	-399,79	-868,51
$d(C_\theta, C_n)$	0,306	0,280	0,469
λ_L	0,088	0,079	0,166
λ_U	0,251	0,229	0,264

Conclusions

- The character of dependencies between the first and the second period changed dramatically (resembling situations described by the regime switching methodology in the time series theory (which has been presented in detail e.g. in [3]).
- Utilizing models in the form of convex combination of Archimedean copulas helped to improve substantially the quality of fitting of empirical copulas. Also tail dependencies for the second period became more pronounced for these types of models. This is in accordance with the occurrence of frequent simultaneous highly extremal returns (of both orientations) in this period.
- Although the couple (PLN/EUR, HUF/EUR) has the largest values of the Kendalls correlation coefficient, its optimal models reach closest fit only for the second (crisis) period.
- Although the Kendalls correlation coefficients were larger for the second period (for all 3 considered couples), the quality of fit for this period (measured by $d(C_{\theta}, C_n)$) was much worse than for the first period.

Acknowledgement

The research summarized in this paper was supported by the Grants APVV/0012/07 and LPP-0111-09.

References

- [1] Bacigál, T.: Advanced Methods of Time Series Modeling and their Application in Geodesy. STU Bratislava, 2008, ISBN 978-80-227-2815-7
- [2] Embrechts, P., Lindskog, F., McNeil, A.: Modeling Dependence with Copulas and Applications to Risk Management. In: Rachev, S. (Ed.) Handbook of Heavy Tailed Distributions in Finance, Vol. 26 (1) 2001, Elsevier, Chapter 8, pp. 329-384
- [3] Franses, P. H., Dijk, D.: Non-Linear Time Series Models in Empirical Finance. Cambridge University Press, 2000
- [4] Frees, E. W., Valdez, E. A.: Understanding Relationships Using Copulas. North American Actuarial Journal, Vol. 2, 1998, pp. 1-25
- [5] Genest, C., Favre, A. C.: Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. Journal of Hydrologic Engineering, Vol. July/August, 2007, pp. 347-368
- [6] Gronneberg, S., Hjort, N. L.: The Copula Information Criterion. Statistical Research Report, No. 7, Dept. of Math. University of Oslo, 2008, ISSN 0806-3842

-
- [7] Hurd, M., Salmon, M., Schleicher, C.: Using Copulas to Construct Bivariate Foreign Exchange Distributions with an Application to the Sterling Exchange Rate Index. Working Paper no. 334, 2007, The Bank of England's working paper series
- [8] Jaworski, P.: Value at Risk for Foreign Exchange Rates – the Copula Approach. *Acta Physica Polonica B*, Vol. 37 (11), 2006, pp. 3005-3015
- [9] Komorníková, M., Komorník, J.: Using Bivariate Copulas for Analysis of Recent Development of the Returns of Exchange Rates of Visegrád Countries to Eur. Submitted to *AUCO Czech Economic Review*
- [10] Komorník, J., Komorníková, M.: Applications of Regime-Switching Models Based on Aggregation Operators. *Kybernetika*, Vol. 43 (4), 2007, pp. 431-442
- [11] Nelsen, R. B.: *An Introduction to Copulas*, Lecture Notes in Statistics, Vol. 139, Springer-Verlag, New York, 1999
- [12] Patton, A.: Modelling Asymmetric Exchange Rate Dependence. *International Economic Review*, Vol. 47 (2), 2006, pp. 527-556
- [13] Prokhorov, A.: A Goodness-of-Fit Test for Copulas. MPRA Paper No. 9998, 2008, online at <http://mpra.ub.uni-muenchen.de/9998>
- [14] Venter, G. G.: Tails of Copulas. Preprint CAS, 2013d01, 2003
- [15] White, H.: Maximum Likelihood Estimation of Misspecified Models. *Econometrica*, Vol. 50, 1982, pp. 1-26

The Dynamic Study of the Weft Insertion of Air Jet Weaving Machines

Lóránt Szabó, István Patkó, Gabriella Oroszlány

Rejtő Sándor Faculty of Light Industry and Environmental Protection
Engineering, Óbuda University, Hungary
E-mail: szabo.lorant@rkk.uni-obuda.hu, patko@uni-obuda.hu,
oroszlany.gabriella@rkk.uni-obuda.hu

Abstract: The application of the air jet loom is widespread in the textile industry because of its high productivity, convenient controllability, high filling insertion rate, low noise and low vibration levels. Air stream in confusor guides can be classified into two types. A weft yarn ejected with high speed air flow is given the drag force caused by friction between the weft yarn and the air flow. In this article we show the study of the dynamics of the type P air jet weaving machines and the definition of the skin friction coefficient for multifilament weft. We have given a calculation procedure for the dynamic description of the insertion process of weaving machines marked P.

Keywords: air jet weft insertion; confusor air guides; maintained air jet; fiction force; skin friction coefficient; friction force measurement; momentum; multifilament weft yarn

1 Introduction

In air jet weaving machines weft is accelerated and taken through the shed by the flow impedance between the flowing air and the weft. Air jet weaving machines belong to the set of intermittent-operation weaving machines. The energy resulting from air pressure directed from the central air tank to the weaving machine changes into kinetic energy in the nozzle, which accelerates and delivers the weft in the air channels differently shaped by machine types. The air leaving the nozzle mixes with the still air, it disperses, and the speed of the axis of the flow drops quickly as it moves away from the nozzle; therefore, in order to reach bigger reed width, the air speed must be kept up in the line of the weft course [3].

Three different systems have mainly been used on commercial air jet weaving machines [6]:

- single nozzle with confusor guides,
- multiple nozzles with guides,
- multiple (relay) nozzles with tunnel reed.

Svaty (former Czechoslovakia) patented the confusor drop wires to guide the air in 1949, which resulted in a wide spread of air jet weaving machines marked P. In 1979 the Nissan company started to use plastic confusor air guides closing at the top. Since the 1980s weaving machines with tunnel reeds and relay nozzles have been the focus of developments [4].

On type P air jet weaving machines, the local transonics speed of the air exiting the nozzle speeds up the weft to $v = 30 - 40 \text{ ms}^{-1}$. By maintaining the velocity of the air jet along the reed width, high weft insertion speeds may be reached. We discussed the aerohydrodynamic study of air jet implementing the weft insertion in our articles [1] and [7]. We use the results of our aerohydrodynamic studies in this article.

2 The Relationship and Definition of the Force Applying to the Weft Inserted in the Air Stream and the Friction Force

Resistance force is formed on the surface of the body placed into the flowing air due to the effect of the flowing air. This force has two components [5]:

- form drag,
- friction force.

Figure 1 shows the ratio of resistance forces applied to the weft as a result of the relationship between the form and the flow.

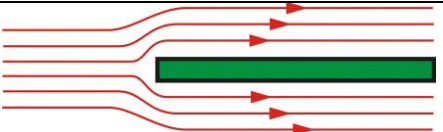
Form and flow	Form drag	Friction force
	<p align="center">approx 0 %</p>	<p align="center">approx 100 %</p>

Figure 1

The ratio of form drag and friction force in the case of weft

The friction force derives from the viscous shearing, which is created between the body surface and the boundary layer of the flowing medium. In general the elemental friction force applying to the elemental surface of the weft may be determined on the basis of the following relationship:

$$dF_f = c_f \cdot \tau \cdot dA \quad (2.1)$$

where:

dF_f : elemental friction force; [N],

c_f : skin friction coefficient; [-],

τ : shear stress derived from the speed of the flowing medium; [Nm^{-2}],

dA : circumfluent surface element; [m^2].

The quantity of the surface element with diameter D and elemental length dx in case of thread:

$$dA = D \cdot \pi \cdot dx \quad (2.2)$$

The shear stress originating between the moving weft and the air jet is given on the base of Bernoulli's equation (Figure 2):

$$\tau = \frac{1}{2} \cdot \rho \cdot (u - v)^2 \quad (2.3)$$

where:

ρ : density of flowing air; [kgm^{-3}],

u : flow speed of air; [ms^{-1}],

v : speed of weft; [ms^{-1}].

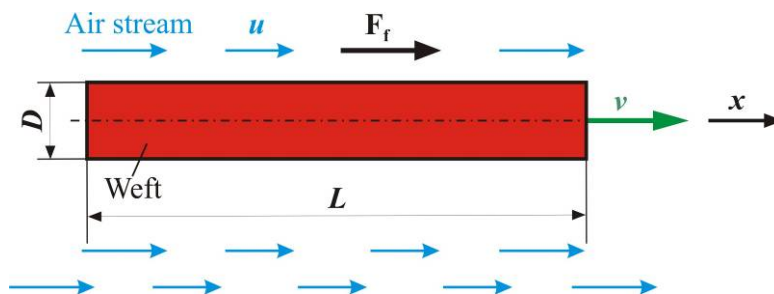


Figure 2

Force acting on a moving weft in air flow [5]

Our studies were made for motionless weft ($v = 0$); therefore we get the following for shear stress:

$$\tau = \frac{1}{2} \cdot \rho \cdot u^2 \quad (2.4)$$

We get the elementary friction force by substituting equations (2.2) and (2.4) into relation (2.1):

$$dF_f = \frac{1}{2} \cdot c_f \cdot \rho \cdot u^2 \cdot D \cdot \pi \cdot dx \quad (2.5)$$

Integrating both sides of equation (2.5) on the length L:

$$F_f = \frac{1}{2} \cdot c_f \cdot \rho \cdot D \cdot \pi \int_0^L u^2 dx \quad (2.6)$$

Considering that the speed of air u is constant on the whole length L:

$$F_f = \frac{1}{2} \cdot c_f \cdot \rho \cdot u^2 \cdot \underbrace{D \cdot \pi \cdot L}_A \quad (2.7)$$

where:

D : diameter of weft; [m],

L : length of weft along the reed width; [m],

A : surface of weft under the impact of the air; $A = D \cdot \pi \cdot L$ [m^2].

With a knowledge of the surface of the weft laid in the air stream of length L as well as the force applying to it, the skin friction coefficient may be defined as follows on the basis of equation (2.7):

$$c_f = \frac{2 \cdot F_f}{\rho \cdot A \cdot u^2} \quad (2.8)$$

If we know the air speed (u), the friction force (F_f) applying to the motionless weft of length dx may be determined from measuring the two forces (Figure 3):

$$F_f = F_n - F_1 \quad (2.9)$$

where:

F_1 : the force measured at measuring point 1; [N],

F_n : the force measured at measuring points $n = 2, 3, 4, \dots, 7, 8$; [N].

The friction force depends on the structure and surface of the weft yarn; these characteristics may be taken into consideration in the skin friction coefficient.

The force applying to a given length of weft may be measured in an air stream of constant speed. For this purpose we have developed the measuring system shown in Figure 3, which consists of a glass tube of inner diameter ($D_{tube} = 7$ mm) con-

nected to the nozzle. We made the speed measurements with a U-tube manometer connected to a Prandtl's tube placed at the end of a glass tube, assuming that in the case of a tube, the speed of the air stream is constant in the axis of the stream.

The weft enters the glass tube through the nozzle needle with tight cross section situated inside the nozzle. It is important that the weft should not get in contact with the side wall of the glass tube during the measurement, which can be checked visually.

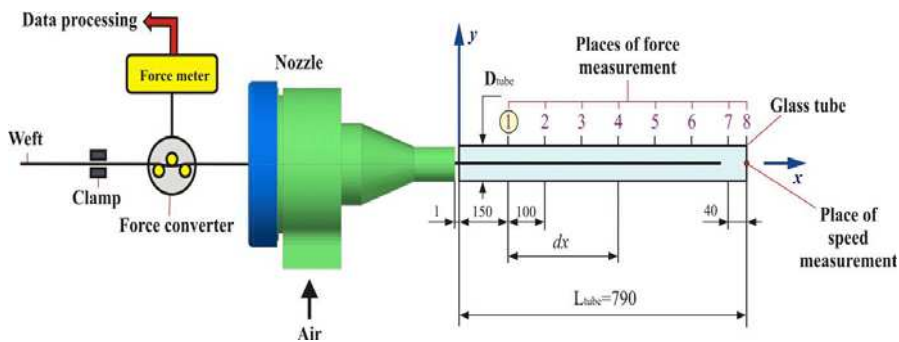


Figure 3

The layout of measuring the skin friction coefficient

The diameter of the weft was determined by a microscope. By averaging multiple measurements, the diameter of the 80 tex multifilament weft was $D = 6.34 \cdot 10^{-4}$ m.

The initial force (F_1) came from the friction force applying to the weft fixed at measuring point 1. By directing the weft in the glass tube to the measuring point, fixed in all cases, there were forces (F_n) corresponding to the measuring points. From equations (2.9) and (2.8) we calculated the different values of skin friction coefficient.

Figure 4 shows the relationship between the skin friction coefficient (c_f) and constant air speeds at $u = 50, 81.6$ and 135.4 ms^{-1} values in the case of different lengths multifilament weft. It can be seen from the figure that the skin friction coefficient is dependent on the air speed but independent of the inserted weft length.

The purpose of our further examinations was to determine the $c_f = f\left(\frac{u}{u_0}\right)$

relation for an 80 tex multifilament weft. The measurements were done in the $u = 30 - 174.3 \text{ ms}^{-1}$ flow speed range and skin friction coefficients were determined here.

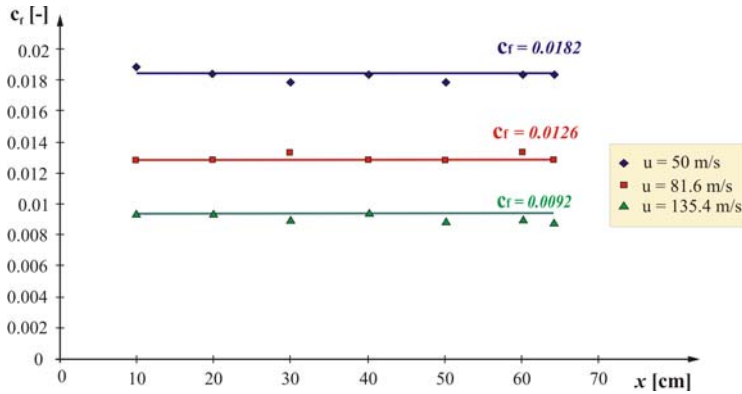


Figure 4

At different air speeds the variation in the skin friction coefficient – in case of constant air speeds – as a function of x

First the initial force F_1 was measured in measuring point 1, then the end of the weft was fixed according to measuring point 8, and force F_8 was measured at the air speed belonging to the previous measuring point (Figure 3). We obtained the friction force applying to the weft of $x = 640$ mm in length:

$$F_f = F_8 - F_1 \quad (2.10)$$

By applying equations (2.10) and (2.8) we obtained the results of Table 1 for skin friction coefficients for a multifilament (80 tex) weft.

Table 1
The values of skin friction coefficient at different speeds

Air speed: u [m/s]	Undimensioned speed: $\left(\frac{u}{u_0}\right)$ [-]	Friction force: F_f [$\cdot 10^{-2}$ N]	Skin friction coef- ficient: c_f [-]
30	0.17	1.5	0.022
39	0.22	2.25	0.019
50	0.28	3.4	0.0118
63	0.36	4.5	0.015
76.3	0.43	5.5	0.0127
81.6	0.47	5.75	0.0113
100	0.57	9	0.011
115.4	0.66	11	0.0108
135.4	0.77	12	0.0086
150	0.86	13	0.0076
$u_0 = 174.3$	1.0	15	0.0074

The u_0 (at $p_t = 3$ bar tank pressure) is the maximum air speed exiting the nozzle. If we use the results from Table 1 and if we un-dimension the air speed by u_0 , we get the function shown in Figure 5, which shows the values of c_f as a function of different nondimensional air speeds (u/u_0).

On the basis of the set of measurement points we used power function approximation to determine function $c_f = f\left(\frac{u}{u_0}\right)$ (Figure 5).

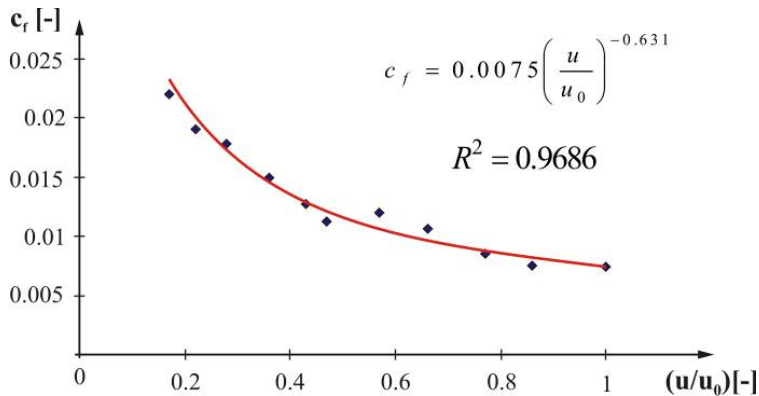


Figure 5

The variation of skin friction coefficient in case of power approximation

From Figure 5 we can read the approximation function describing function

$$c_f = f\left(\frac{u}{u_0}\right):$$

$$c_f = 0.0075 \left(\frac{u}{u_0}\right)^{-0.631} \quad (2.11)$$

3 The Dynamic Examination of Weft Moving in Confusor Guides

Weft insertion by air jet is a complicated movement, which is not a fully controlled technological process. When the motionless weft threaded in the nozzle gets into the air stream, at the moment of starting the insertion, the weft makes an accelerating move, which is created by the friction force deriving from the relation of the weft and the air. This section of weft insertion (when the speed of weft is

smaller than the speed of air stream) is called the acceleration condition of the weft (Figure 6). The acceleration condition may be divided into two parts:

- intense acceleration section, the initial part after the nozzle pipe; the speed of weft increases abruptly,
- weak acceleration section, in which the speed of the weft increases slowly until it is equivalent with the air speed.

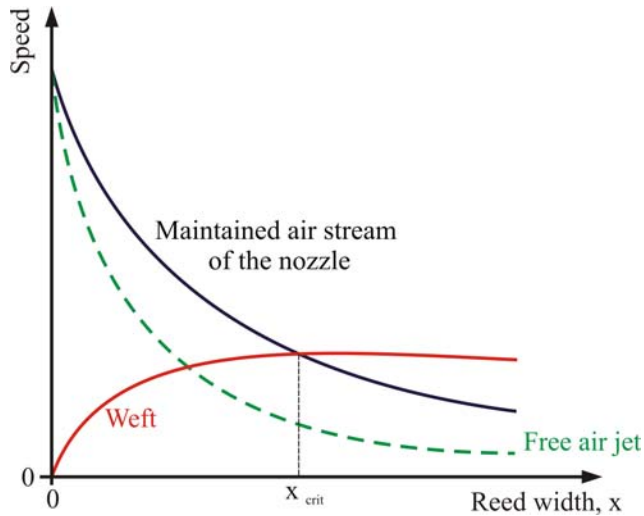


Figure 6

The possibility of modifying the weft speed with air stream from one nozzle maintained by confusor guides

The weft acceleration stoppage is caused by the change in the direction of the friction force, which slows down the weft ($x > x_{crit}$). The weft keeps moving due to its inertia and its speed does not increase any more. Therefore, the ideal situation for air jet machines would be if the reed width were smaller than x_{crit} .

In order to achieve this, different air guide systems are used to maintain high air speed (and thus high weft speed) along the reed width of the machine. In the case of air jet weaving machines, the force moving the weft is exclusively defined by the drag emerging between the air and the thread. The motive force applying in the direction of insertion increases with the air speed and the diameter of the weft. During weft insertion the elemental friction force applying to the elemental weft section may be determined from equation (2.5) and considering that the weft speed is $v \neq 0$, on the basis of the following equation:

$$dF_f = \frac{1}{2} \rho \cdot c_f \cdot D \cdot \pi \cdot (u - v)^2 dx \quad (3.1)$$

The equilibrium of forces applies to the weft inserted in the air stream:

$$\frac{d}{dt} I = F_f - F_S \quad (3.2)$$

where:

I : momentum of the weft; [$kgms^{-1}$],

F_f : friction force; [N],

F_S : force deriving from the friction of the weft and other solid state; [N].

During the course of our further examinations we disregard friction force F_S because the weft removal from the holder and its passage through the guide ring is almost frictionless. Inside the nozzle the relation is generated between the weft and the air stream. Our research only focused on the dynamic study of the weft and air stream exiting the nozzle and moving in different air guide systems. By neglecting friction force (F_S), the created model is not complete, but it is suitable for studying the force which is generated along the axis of the insertion and which moves the weft [2].

In this way we get the differential equation describing the relation between the weft and the air stream inserting the weft on air jet weaving machines marked P:

$$\frac{d}{dt} I = \frac{d}{dt} mv = F_f \quad (3.3)$$

where:

m : the mass of weft yarn in the air stream; [kg],

v : the speed of the weft end at the place of study; [ms^{-1}].

On the laboratory bench shown in Figure 7 we measured the friction forces applying to the 80 tex multifilament weft inserted in a continuous air stream. The size of the slot distance is shown in the figure:

$$R_t = 5 \cdot d_0 = 35 \text{ mm} \quad (3.4)$$

where:

d_0 : inside diameter of the nozzle at the exit; $d_0 = 7 \text{ mm}$.

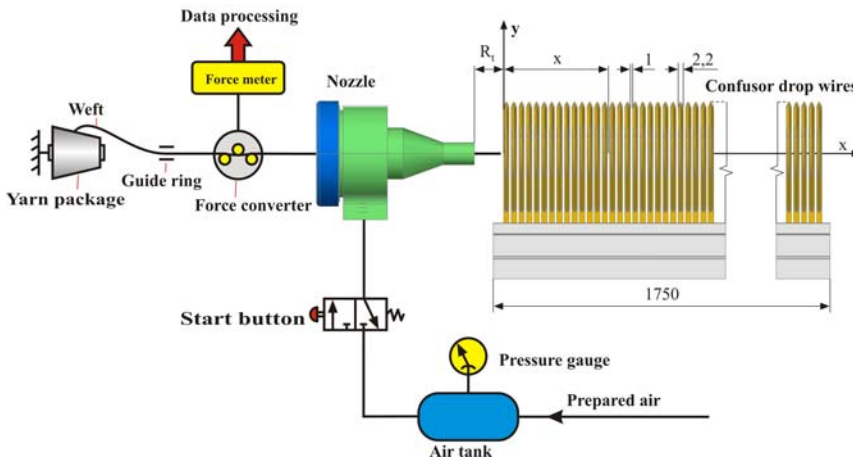


Figure 7

The arrangement of laboratory measurements of forces applying to the weft

Based on our experience it is understood that the friction force applying to the weft depends on:

- the tank pressure used,
- the air guide system,
- the length of the inserted weft in the confuser guides, up to about $x = 170$ cm,
- the yarn structure and diameter of the weft,
- the yarn surface characteristics.

Afterwards, we examined the relation between the theoretically identifiable friction force applying to the weft and the force applying to the motionless weft inserted into the continuous air stream. By substituting weft speed $v = 0$ into equation (3.1) we regain equation (2.5):

In our previous study [1] our aim was to make a function which would define the flow speed at any given point of the reed width for any air guide system, as a function of the initial conditions. In the case of closed plastic drop wires for nondimensional velocity distribution $\left(\frac{u}{u_0}\right)$ we obtained the following functional relationship:

$$\left(\frac{u}{u_0}\right) = \frac{a\left(\frac{x}{r_0}\right)^2 + c\left(\frac{x}{r_0}\right) + b}{\left(\frac{x}{r_0}\right)} = a\left(\frac{x}{r_0}\right) + b\left(\frac{r_0}{x}\right) + c \quad (3.5)$$

where the constants for closed plastic confusor guides are:

$$a = -0.0004 [-],$$

$$b = 5.288 [-],$$

$$c = 0.3243[-],$$

$$r_0 = \frac{d_0}{2} = 3.5 \text{ mm, radius of the applied nozzle at the exit,}$$

$$u_0 = 174.3 \text{ ms}^{-1}, \text{ air speed at the entrance of the confusor guides.}$$

Equation (2.5) undimensioned by values u_0 and r_0 :

$$dF_f = \frac{1}{2} \rho \cdot D \cdot \pi \cdot u_0^2 \cdot r_0 \cdot c_f \cdot \left(\frac{u}{u_0}\right)^2 \cdot d\left(\frac{x}{r_0}\right) \quad (3.6)$$

Furthermore, considering equation (2.11), substituted in place of c_f , we get:

$$\begin{aligned} dF_f &= \frac{1}{2} \rho \cdot D \cdot \pi \cdot u_0^2 \cdot r_0 \cdot \underbrace{0.0075}_{K=0.95 \cdot 10^{-3} \text{ N}} \cdot \left(\frac{u}{u_0}\right)^{-0.631} \cdot \left(\frac{u}{u_0}\right)^2 \cdot d\left(\frac{x}{r_0}\right) \\ &= K \cdot \left(\frac{u}{u_0}\right)^{1.37} d\left(\frac{x}{r_0}\right) \end{aligned} \quad (3.7)$$

and in the case of:

$$\rho = 1.2 \text{ kgm}^{-3},$$

$$D = 6.34 \cdot 10^{-4} \text{ m,}$$

$$u_0 = 174.3 \text{ ms}^{-1},$$

$$r_0 = 3.5 \cdot 10^{-3} \text{ m,}$$

then $K = 0.95 \cdot 10^{-3} \text{ [N]}$. By substituting $z = \left(\frac{x}{r_0}\right)$, and examining the plastic

guides, we can calculate the elemental force applying to the weft yarn with the following equation:

$$dF_f = K \cdot \left(a \cdot z + \frac{b}{z} + c\right)^{1.37} dz \quad (3.8)$$

By integrating both sides of equation (3.8) we get:

$$\int_{z_0}^z dF_f = \int_{z_0}^z K \cdot \left(a \cdot z + \frac{b}{z} + c \right)^{1.37} dz \quad (3.9)$$

that is:

$$F_f(z) - F_f(z_0) = K \cdot \int_{z_0}^z \left(a \cdot z + \frac{b}{z} + c \right)^{1.37} dz \quad (3.10)$$

where:

$F_f(z_0) = F_0$: at $x = 0$ the measured force applying to the weft at the beginning of the air guide system; [N]. On the basis of our measurements demonstrated in Figure 7: $F_0 = 2 \cdot 10^{-2}$ N .

$F_f(z) = F$: in the case of $x > 0$, the drag applying to the weft in the axis of the confusor guides; [N].

By dividing equation (3.10) by value K and by substituting the constants of equation (3.5) into (3.10), we get the undimensioned equation suitable for final integration:

$$F^* = F_0^* + \int_{z_0}^z \left(-0.0004 \cdot z + \frac{5.288}{z} + 0,3243 \right)^{1.37} dz \quad (3.11)$$

where:

$F^* = \frac{F}{K}$: the theoretical undimensioned force applying to the weft in air guide system; [-],

$F_0^* = \frac{F_0}{K} = \frac{2 \cdot 10^{-2} \text{ [N]}}{0.95 \cdot 10^{-3} \text{ [N]}} = 21$ [-]: the undimensioned force measured at the beginning of the air guide system; [-].

We integrated the second member on the right side of equation (3.11) using the program “Maple 13”. Figure 8 demonstrates the graphic solution of equation (3.11).

The figure shows the trend of the measured forces applying to multifilament 80 tex weft placed into continuous air stream in the case of confusor guides compared to the calculated values.

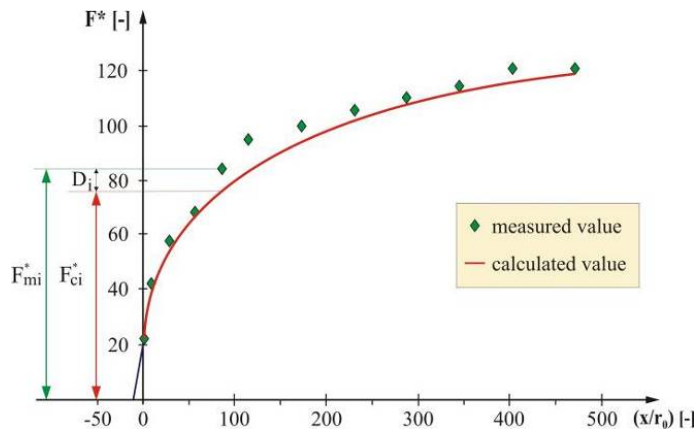


Figure 8

In the case of closed plastic guides, the comparison of the measured and the theoretical values

From the measured and calculated values shown in Figure 8, we determined the deviations in the measuring points (D_i), which can be found in Table 2.

Table 2

The deviation of measured and calculated values as a function of $\left(\frac{x}{r_0}\right)$

Measuring point: i	Place: $\left(\frac{x}{r_0}\right)$ [-]	Measured value: F_{mi}^* [-]	Calculated value: F_{ci}^* [-]	Deviation: D_i [-]	Ratio: R_i [-]
1	14.3	42.1	41.5	0.6	0.985
2	28.6	57.9	54,3	3.6	0.937
3	57.2	68.4	67.5	0.9	0.986
4	85.8	84.2	78.1	6.1	0.927
5	114.4	94.7	84	10.7	0.88
6	171.6	100	94.7	5.3	0.947
7	228.8	105.3	101.5	3.7	0.964
8	286	110.5	108	2.5	0.977
9	343.2	117.1	115.6	1.5	0.987
10	400.4	121	118	3	0.975
11	471.4	121	120	1	0.991

The deviations may be determined from the values in columns 3 and 4 of Table 2:

$$D_i = F_{mi}^* - F_{ci}^* \quad (3.12)$$

where:

D_i : deviation belonging to the measuring point (column 5 in Table 2); [-],

F_{mi}^* : measured value; [-],

F_{ci}^* : calculated value belonging to the measuring point; [-].

We calculate their average from deviation ($n = 11$):

$$\bar{D} = \frac{1}{n} \sum_{i=1}^n D_i \quad (3.13)$$

where:

\bar{D} : the average of the deviations; in our case $\bar{D} = 3.53$ [-].

In order to analyze the deviations between the measured and the calculated values we have introduced the ratio:

$$R_i = \frac{F_{ci}^*}{F_{mi}^*} \quad (3.14)$$

where:

R_i : ratio belonging to the measuring point (column 6 in Table 2); [-].

Similar to equation (3.1) we have determined the average of the ratios. Since the number of our measurements is $n < 30$, it is more practical to use the corrected standard deviation instead of the standard deviation:

$$S^* = \sqrt{\frac{\sum_{i=1}^n (R_i - \bar{R})^2}{n-1}} \quad (3.15)$$

where:

S^* : corrected standard deviation calculated from the ratios belonging to measuring points; [-], which in our case was 0.0342,

\bar{R} : average of ratios; [-].

The value of the corrected standard deviation of the ratios is small. Therefore it may be concluded that the calculated values approximate the measurement results very well. The measured values were bigger at all the measuring points, which may be explained by the fact that the force measuring equipment increased the size of the actual force generated on the weft. The inclusion of the measuring equipment modified the measurement results in a similar manner. This type of error is called systematic error.

Conclusions

Weft insertion through an air stream is a complex and complicated process. The motionless weft threaded in the nozzle gets into the effect of the air stream at the insertion of the weft, as a result of which the weft makes an accelerating move, which is created by the friction force deriving from the contact of the weft and the air. This section of weft insertion is the acceleration section, which may be divided into two parts.

The results of the research:

- We have provided a calculation method for the calculation of air speeds generated in the axis of the different confusor guides: $u = f\left(\frac{x}{r_0}\right) \cdot u_0$ [1].
- For multifilament 80 tex weft we have determined the function $c_f = f\left(\frac{u}{u_0}\right)$ describing the skin friction coefficient of the weft
- We have created a calculation method to determine the force applying to the weft thread – after the insertion – by describing functional relationship $F^* = f\left(\frac{x}{r_0}; \frac{u}{u_0}\right)$.
- Having considered the previous points we have given a calculation procedure for the aerohydrodynamic and dynamic description of the insertion process of weaving machines marked P.

References

- [1] Patkó I., Szabó L.: A szövés és áramlás kapcsolatának vizsgálata légsugaras szövőgépeken. Magyar Textiltechnika LXII. évf. 2009/5, 194-200. o.
- [2] Patkó I.: Lamellák közötti áramlás tulajdonságainak meghatározása. Kandidátusi disszertáció, Budapest, 1994, 74-75. o.
- [3] I. Patkó: The Nozzle's Impact on the Quality of Fabric on the Pneumatic Weaving Machine. Springer, Volume 243, 2009, UK, pp. 583-592
- [4] Szabó R.: Szövőgépek. Műszaki Könyvkiadó, Budapest, 1985, 148. o.
- [5] S. Adanur: Handbook of Weaving. Lancaster, Pennsylvania, 2001, pp. 189-191
- [6] M. Ishida, A. Okajima, Y. Shimada, T. Kurata, F. Hoshiai: Experiments of Flow of Air Jet Loom with Air Guides Part 1: Characteristics of Flow Injected into Air Guides. Journal of the Textile Machinery Society of Japan, 1989, Vol. 36, No. 4, pp. 127-128
- [7] I. Patkó, L. Szabó: The Study of the flow Conditions of Air Jet Weaving Machines. Proceedings of the 10th International Symposium of Hungarian Researchers, November 12-14, 2009, pp. 391-412

Adaptation of Fuzzy Cognitive Maps – a Comparison Study

Ján Vaščák, Ladislav Madarász

Department of Cybernetics and Artificial Intelligence, Faculty of Electrical
Engineering and Informatics, Technical University of Košice
Letná 9, 042 00 Košice, Slovakia
jan.vascak@tuke.sk, ladislav.madarasz@tuke.sk

Abstract: This paper deals with the experimental study and comparison of various adaptation methods for setting-up parameters of fuzzy cognitive maps (FCMs). A survey is given of the best known methods, which are mostly based on unsupervised learning. The authors show better performance using supervised learning, namely least mean square approaches. Experiments were done on a simulation example of autonomous vehicle navigation. The paper is concluded by comparing their efficacy and properties.

Keywords: adaptation methods; fuzzy cognitive learning; Hebbian learning; supervised learning

1 Introduction

A navigation process of a vehicle requires solving a number of problems of different matter, i.e. finding not only a path to a goal but also regarding various constraints such as obstacle avoidance, limitations resulting from cooperation tasks with other vehicles, minimizing fuel consumption, etc. Such a process can be described by a set of conditions, which are at the same time mutually interconnected, creating complex decision chains and recurrences and causing very complicated mutual influences.

Means of artificial intelligence are based mostly on production rules, mainly because of their human-friendly knowledge representation, which will be the case if the rules are mutually independent; i.e. outputs of any rule do not enter antecedents of any other rule and so decision chains and closed loops are not created [10, 17]. Such systems will be named ‘simple’ rule-based systems for the following. However, a complex system is characterized by just such chains and loops. In that case the rule-based knowledge representation loses its main advantage and becomes ‘unreadable’.

Therefore, for overcoming the limits of simple rule bases, fuzzy cognitive maps (FCMs) seem to be very suitable means, and these are able very clearly to represent graphically to a human notions and relations among them as seen in Fig. 1. A simple rule base (a rule set) lacking any chains or loops is the simplest or, in other words, the most degenerated form of an FCM. Hence, all operations and properties of production rules are valid for FCMs, too. Further, FCMs possess other additional properties and abilities (described in the next section) that are also convenient for the analysis and modelling of dynamic systems.

On the other hand, FCMs have the same basic drawbacks as other fuzzy systems: they are not able to self-learn. The design of adaptation approaches is much more difficult because of their complex structure and its variability [16, 18]. Therefore, at least the definition of notions, which are represented by nodes (see Fig. 1), is done manually by an expert and adaptation is limited to setting-up relations, i.e. graph edges. Most adaptation approaches are based on unsupervised learning, mainly Hebbian learning, e.g. [1, 3, 4, 13] but there are also approaches utilizing evolutionary computing [12].

Unsupervised learning is convenient for tasks such as clustering if we need, for instance, to separate data into a few groups. It is significant that elements from a given group have stronger relations among them and usually they react or are activated at the same time, which corresponds to the Hebbian paradigm of learning. However, navigation is a specific kind of activity somewhere between decision and control. It is a set of several processes with different dynamics, and any trial for separation could fail [14]. From this point of view, supervised learning seems to be more convenient. To verify this hypothesis as well, the well-known least mean square (LMS) approach, with some modifications, was used to show its suitability in this area.

In this paper, after introducing some basic notions regarding FCMs and their properties in Section 2, we will concentrate on variations of Hebbian learning and LMS in Section 3 and show properties of these learning approaches on a navigation simulation of a vehicle with their mutual comparison in Section 4. Finally, we will conclude with some remarks regarding the utilization of adaptation methods for navigation.

2 Fuzzy Cognitive Maps

In general a Cognitive Map (CM) is an oriented graph where its nodes represent notions and its edges causal relations (see Fig. 1). Mostly, notions are states or conditions and edges are actions or transfer functions, which transform a state in a node to another one in another node.

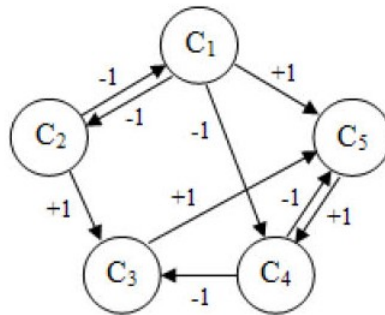


Figure 1

An example of FCM with crisp edges

CM is able to describe complex dynamic systems. It is possible to investigate e.g. limit cycles, collisions, etc. In addition, it is possible to define the strengths of relations, too. Originally they were represented by three values -1 , 0 and 1 . Perhaps the main advantage of CM is its human-friendly knowledge representation in graphical form.

FCM represents an extension of CM and was proposed by Kosko in 1986 [6]. The extension is based on the strength values that are from an interval $[-1; 1]$ as well as the fact that the nodes can be represented by membership functions. Strengths, or rather weights, correspond to rule weights in rule-based systems.

There are several possible formal definitions of FCM, but still the most commonly used one is in form given by Chen [2], which respects the original numerical matrix representation proposed by Kosko, where FCM is defined as a 4-tuple:

$$FCM = (C, E, \alpha, \beta) \quad (1)$$

where:

C – finite set of cognitive units described by their states $C = \{C_1, C_2, \dots, C_n\}$,

E – finite set of oriented connections between nodes $E = \{e_{11}, e_{12}, \dots, e_{nn}\}$,

α – mapping $\alpha \rightarrow [-1; 1]$,

β – mapping $\beta \rightarrow [-1; 1]$.

In other words, C represents nodes, E is for edges, α is a membership function placed in a node and the result is a grade of membership for values entering such a node. Similarly, β has the same meaning as α but for edges. The only difference compared to the definition of a fuzzy set is that the original interval of real values determined for grades of membership $[0; 1]$ was extended to $[-1; 1]$ in order to define as well negative connections, which refer to negations in logic. Similarly, β does the same but it is placed on a connection, i.e. it is its weight represented as a membership function. Further, we will use β only in the form of a singleton (a crisp value).

For computational representation of FCM a transition matrix is used. For the example in Fig. 1 it will look like:

$$E = \begin{bmatrix} 0 & -1 & 0 & -1 & 1 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}. \quad (2)$$

Cognitive units are in each time step in a certain state. Using a transition matrix we can compute their states for next time step and thus repeatedly for further steps. Similarly, as for differential equations we can draw phase portraits. To preserve values in prescribed limits a boundary function L is used as well. So we can compute the states for $t+1$ as follows [6]:

$$C(t+1) = L(C(t) \cdot E). \quad (3)$$

Comparing Fig. 1 and 2 we can see FCMs are an extension of simple fuzzy rule-based systems. Fuzzy rules are totally independent because their consequents do not have any mutual influence, which is possible only in simpler decision cases. Simple rule-based systems do not enable any decision chains or representation of temporal information. From this point of view they are only a very special and restrained case of FCM, which can be depicted as an example in Fig. 2, where a set of m rules with inputs LX_i and outputs LU_i ($i=1, \dots, m$) is figured in the form of an FCM. There is depicted the evaluation process resulting in an accumulated (aggregated) value LU_c being defuzzified into a crisp form LU_c^* , too.

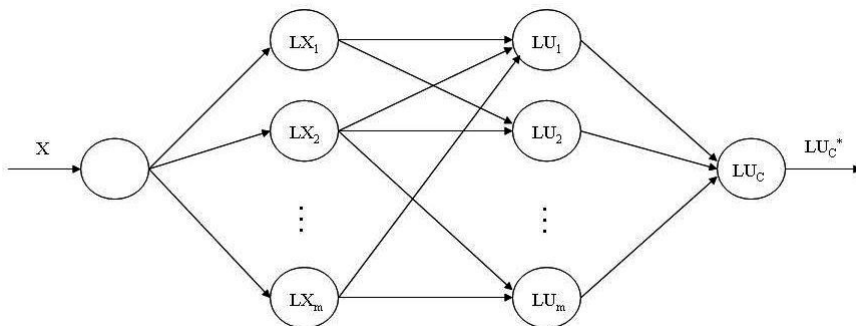


Figure 2

An example of FCM representing a simple rule set

3 Adaptation Approaches to FCM in Navigation

There are two basic approaches for setting-up the parameters of FCM. The first approach is based on expert skills where we need several experts who are able to set-up their own FCM. The weights of edges are then averaged. It seems to be very simple, but firstly we need to have such experts, and secondly the skills of these experts will necessarily be of different quality, which means we should rather do a weighted average, but there does not exist any hint how to determine these weights.

The second approach is based on unsupervised learning, which is well known mainly from neural networks. The creator of FCM, Bart Kosko, proposed using Hebbian learning [7], which is unsupervised. FCM was originally proposed only for causal relations, and looking at (3) we can see the state of a node and its change in next time $t+1$, i.e. $C(t+1)$ depends on $E(t)$ and $C(t)$, so we can write for the change of an element $e_{ij}(t) \in E(t)$ generally:

$$\dot{e}_{ij}(t) = f_{ij}(E(t), C(t)) + g_{ij}(t) \quad (4)$$

where $g_{ij}(t)$ is the so-called *forcing function*. In other words, there should be found a certain mutual correlation among nodes in individual time steps because they are more or less interconnected. So for n nodes we can get

$$\begin{aligned} E(t+n+1) = & q_1 C(t+1)^T \cdot C(t+2) + q_2 C(t+2)^T \cdot C(t+3) + \dots + \\ & q_{n-1} C(t+n-1)^T \cdot C(t+n) + \\ & q_n C(t+n)^T \cdot C(t+1), \end{aligned} \quad (5)$$

which is the sum of correlation matrices with the same dimensions like E and q_i are properly chosen weights.

Further, we can see that FCM, with its topology, resembles a neural network, too. Correlation learning (5) is a form of Hebbian learning and for neural networks it is known in the form:

$$\dot{e}_{ij}(t) = -e_{ij}(t) + C_i(t) \cdot C_j(t) \quad (6)$$

where $C_i(t), C_j(t) \in C(t)$ are directly interconnected nodes – i as output and j as input node. The formula (6) is named *Hebbian correlation learning* (HCL). The principle of Hebbian learning is based on the synchronous (at the same time) activation of both nodes. In such a case, the connection between them will be strengthened. Otherwise, if the nodes do not activate at the same moment, the connection will be weakened. However, in control generally due to various dynamic dependences, it can happen that the nodes will be activated at different time steps although there is a connection between them. Already this fact puts the idea of using Hebbian learning into doubt. Several experiments such as e.g. in [3,

7] as well as our own experience confirm poor results using the original form of Hebbian learning - HCL (6). There are problems with the selection of training patterns. If their activations are small, then probably only a few connections will arise and FCM will be poorly structured. In the opposite case many so-called false connections will remain, which should be removed.

Hebbian learning with damping / decay (HLD) is a modification of HCL using forgetting (decay) factor, which depends on the forgetting parameter α , and it is able to control the learning speed by learning parameter γ :

$$\dot{e}_{ij}(t) = \gamma \cdot C_i(t) \cdot C_j(t) - \alpha \cdot C_i(t) \cdot e_{ij}(t). \quad (7)$$

Another modification is *differential Hebbian learning* (DHL), which takes into consideration changes of node activations, i.e. their derivatives. If the derivatives of $C_i(t)$, $C_j(t)$ are nonzero then

$$\dot{e}_{ij}(t) = -e_{ij}(t) + \dot{C}_i(t) \cdot \dot{C}_j(t). \quad (8)$$

Nonlinear Hebbian learning (NHL) [13] is a more sophisticated method using stopping criteria as well. It uses its own calculations of activation values for nodes $C_j(t)$:

$$C_j(t+1) = f(C_j(t) + \sum_{\substack{i \neq j \\ i=1}}^n C_i(t) \cdot e_{ij}(t)). \quad (9)$$

The weight e_{ij} for the next time step $t+1$ is then calculated as

$$e_{ij}(t+1) = \eta \cdot e_{ij}(t) + \gamma \cdot C_j(t) \cdot (C_i(t) - \text{sgn}(e_{ij}(t))) \cdot e_{ij}(t) \cdot C_j(t) \quad (10)$$

where η is the weight decay learning coefficient and γ is the own learning parameter. However, formula (10) is used only for weights that were initially nonzero, which requires knowledge from an expert, and this method is convenient first of all for final fine-tuning FCM.

There are two stopping criteria for learning: criterion of minimum square error F_1 of nodes, which represent outputs of FCM (denoted as OC) and criterion F_2 indicating whether there are still some significant changes of node activations during learning:

$$F_1 = \sqrt{\sum_{k=1}^p (OC_k - T_k)^2}, \quad (11)$$

$$F_2 = |OC_k(t+1) - OC_k(t)| < \varepsilon. \quad (12)$$

F_1 gives us information about the performance quality of p outputs from FCM (e.g. action variables of a controller in the form of FCM) where T_k is an average value from the interval of allowed values for a given OC and it should be known by experts (e.g. the allowed range of accelerations for a vehicle). The goal is to minimize F_1 under a certain value. F_2 tells us about the stability of the designed FCM, which is characterized by the stabilizing activation values of output nodes. The size of ∂ has been chosen after a series of experiments to be 0,002 [13]. When both criteria are fulfilled then the learning will be stopped.

The modified version, the so-called *improved nonlinear Hebbian learning* (INHL) introduced in [9], is based on the following weight change adaptation:

$$\Delta e_{ij}(t) = \eta \cdot \Delta e_{ij}(t-1) + \gamma \cdot z(t)^2 \cdot (1 - z(t)) \cdot (C_j(t) - e_{ij}(t-1) \cdot C_j(t)) \quad (13)$$

where $z(t) = 1 / (1 + \exp(-C_i(t)))$ and the weight in the next time step $e_{ij}(t+1)$ is calculated as $e_{ij}(t) + \Delta e_{ij}(t)$. There is modified also the criterion F_1 as a sum of OC_k^2 .

Because our training data also contained the desired values of output variables we tried to use a supervised learning method, in this case a well known LMS method defined as

$$e_{ij}(t+1) = e_{ij}(t) + \gamma \cdot (y_d(t) - \sum_{i=1}^n e_{ij}(t) \cdot C_i(t)) \cdot C_i(t) \quad (14)$$

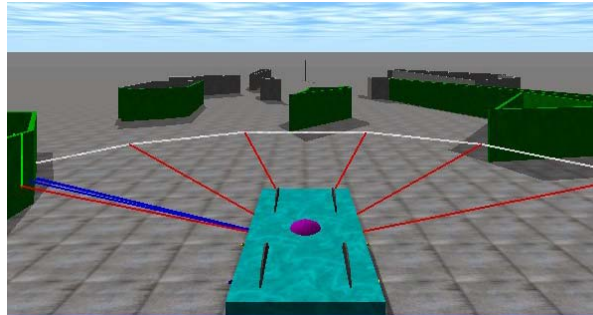
where $\sum_{i=1}^n e_{ij}(t) \cdot C_i(t)$ is the activation value of the input node $C_j(t)$ and $y_d(t)$ is the desired activation value of $C_j(t)$.

4 Modifications and Experimental Comparison

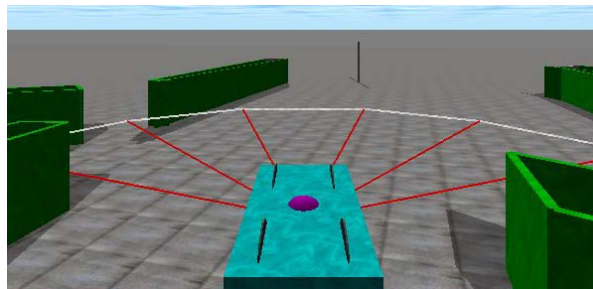
For automatic navigation of a vehicle, a simulator based on ODE library (<http://www.ode.org/>) was proposed [15] as being able to respond to real physical conditions. The main interface consists of an area with various obstacles and a vehicle model that can be controlled manually as well as automatically implementing source code of a given method. The simulator is able to collect data during the movement using 5 sensors that divide the surroundings into the same number of sectors, and from these data it can compute the current position of the vehicle. The goal is depicted as a vertical pole and its position is given in advance, see Fig. 3.

For all methods three basic experiments were performed using different starting positions, as depicted in Fig. 3. For comparison purposes, a manually designed

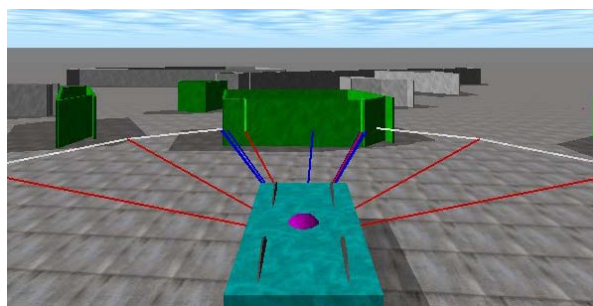
FCM was constructed, which is depicted in Fig. 4. There are in total 16 nodes, where 4 of them are output nodes (*OC*) $C_0 - C_3$ with bold capitals – turn to left (L), go forward (F), turn to right (R), stop (S). Other nodes represent calculated positions of the goal in relation to the vehicle $C_4 - C_7$, with symbols G_L – goal left, G_S – goal straight, G_R – goal right, G_C – goal close, and signals from sensors $C_8 - C_{15}$ where S_i means the number of a sensor, *c* – close and *cc* – critically close. Positive connections are depicted as solid lines and negative connections as dashed ones.



(a)



(b)



(c)

Figure 3

Starting positions of the vehicle simulator

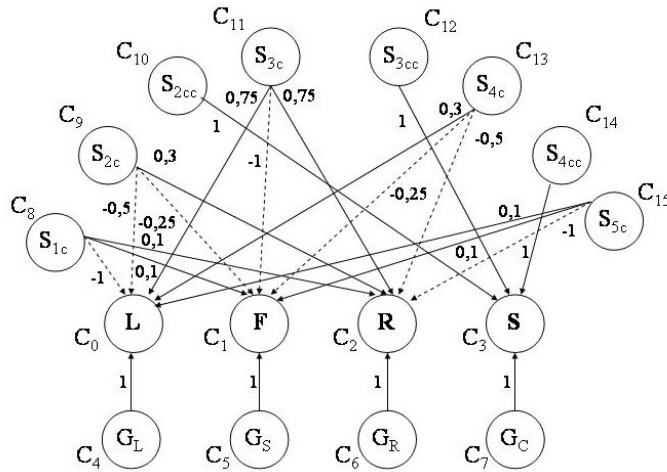


Figure 4

Manually defined reference FCM for the navigation of a simulated vehicle

Concerning the functions, which were implemented in the nodes, they are membership functions (Fig. 5) for evaluating signals from the sensors as *obstacle closeness* – nodes $C_8 - C_{15}$, calculated position of the goal like *goal position* – nodes $C_4 - C_6$ and *goal closeness* – node C_7 and finally, for evaluating action values of output nodes *angle of turning* – nodes $C_0 - C_2$. When the activation value in the node C_3 (stopping) achieves the value $0,95$ then the vehicle will stop.

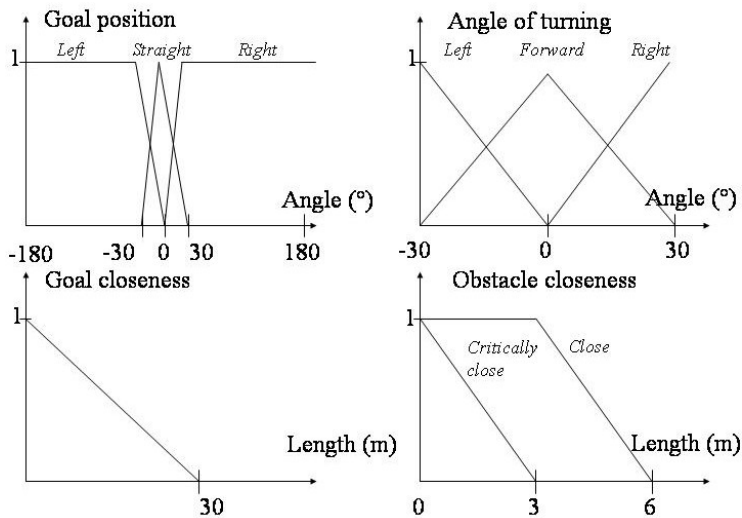


Figure 5

Membership functions of manually designed FCM nodes

In the case of a manual design it is necessary to mention we can consider only suboptimal solutions because we can never state that there is no better solution. Therefore, we suppose after a satisfactory number of cycles in the manual setting-up process that we have potentially the ‘best’ possible solution.

Concerning the evaluation criteria, we chose a subjective classification from 1 – the best to 5 – the worst (unsatisfactory). Generally such aspects were observed and evaluated as the behaviour of the vehicle in the goal area, the quality of obstacle avoidance, the smoothness of the wheel traces (whether the wheels did not turn too chaotically as a beginner), the choice of possible paths and of course the number of collisions with obstacles. The subjectivity of our evaluation is based on these three points: criteria weights, comparison to manually designed FCM as well as the calculation of some criteria. Firstly, the mentioned criteria have different weights of importance, with in this case the most important criteria being the number of collisions and the number of successfully found paths to the goal, which in turn are related to the total number of experiments. These weights were determined manually by a human reflecting his/her opinions. Secondly, resulting behaviour of a vehicle based on the adapted FCM is compared to the behaviour resulting from the manually designed FCM, which again reflects the properties of a given human. Finally, a criterion such as the choice of a possible path is evaluated purely subjectively, where another user may have a different opinion, because we consider uncertain notions such as the complexity of a path.

It seems to be a very ‘non-scientific’ approach but considering some objective criteria such as the shortest trajectory, the minimum number of turnings, the minimum number of collisions with obstacles, the minimum time, etc. there will be many situations when some criteria are evaluated better and some worse, and calculating the total measure of fulfilment is a nondeterministic task dependent on a given application. (This task leads again to weighted averaging, and simply setting-up weights for given criteria is a nondeterministic and very subjective task.) So we can get several solutions with the same or approximate total evaluation value. If we compare subjective and objective evaluation approaches the first approach seems to be more similar to the intuitive human reasoning and this was the most important element for choosing the subjective approach.

As already mentioned above, the task of adaptation methods is only to set-up the weights of connections e_{ij} . The structure (topology) of nodes and their activation (in our case membership) functions are given by an expert in advance. Further, the behaviour description of FCM designed by individual adaptation methods follows.

HCL by (6) showed really very poor results. Only two connections were created and the vehicle stopped very quickly. The reason for this is that in the training data there were many patterns which did not activate any membership function in the nodes, and thus in such a case already created connections were zeroed. Therefore, such patterns were removed before learning started to prevent clearing connections. The number of connections increased considerably but weights were

very small so the result was almost identical with the previous case. Only through using a very large training set and after many learning cycles was this approach able to bring some results, which is unacceptable.

Similarly, HLD by (7) behaved in a manner similar to HCL – without filtering inactive patterns, only a few connections were established; and omitting them, too many connections were established. Either there were strong connections to the stopping node C_3 , which caused premature finish, or too many, and weak connections are in fact false and they should be removed. In any event, such a structure of connections needs manual corrections done by an expert, which is not welcome.

In contrast, concerning Hebbian learning, NHL and INHL represented excellent results [8] after only a few learning cycles. Although NHL needs setting-up of nonzero connections in the matrix E by an expert, it produces very stable solutions without any false or weak connections behaving like noise. There are very small differences between a manually designed matrix E and matrices created by NHL and INHL. As the INHL method also enables the updating of zero initial weights, it tends to create almost a full set of all possible connections but with very small weight values which do not affect the performance of such an FCM. Filtering such connections could be useful because INHL shows slightly smaller stability of learning than NHL, and it needs a few more learning cycles.

Using supervised learning also brought very good results. At first, LMS by (14) was tried on a zero initial weight matrix. The results were only a little worse than those obtained by the manually designed FCM. However, the best results of all were achieved using LMS with unity initial matrix. The smoothness of wheel traces was even better than with an FCM designed by an expert, but it is necessary to do many experiments setting-up the learning parameter to find the best one, which is time consuming work. In Fig. 6 there are depicted for comparison the weight matrices of LMS with unity initialization and the FCM manually designed by an expert. We can see significant differences in matrices although they are functionally similar.

Table I
Successfulness of FCM Learning Methods

Order	Method	Note
1	LMS with unity initialization	1
2	Expert	2
3	Nonlinear Hebbian learning	2
4	Improved nonlinear Hebbian learning	2
5	LMS with zero initialization	3
6	<i>Hebbian learning with damping</i>	5
7	<i>Hebbian correlation learning</i>	5

In the Table I are the mentioned methods ordered according to their degree of success, and total evaluations are assigned to them based on experiments done with the simulation of vehicle navigation in the area with obstacles. HLD and HCL are denoted in italics because they seem to be fully inconvenient for FCM.

Conclusions

Although the proposed experiment does not require the creation of a complex form of FCM containing decision chains or loops, and therefore it was possible to use a simple LMS method, the utilization of supervised learning is a challenge for research in spite of many theoretical difficulties. Unsupervised learning is not a nostrum for solving problems of machinery learning. The Hebbian law is very general as it does not go into the depth of learning principles, and therefore enhanced methods derived from Hebbian learning such as NHL and INHL are perhaps also efficient only for simpler low-order problems such as two-tank systems, etc., as is often presented in literature. Therefore it could also be very useful to focus interest on various interpolation and nonlinear methods already used in conventional rule-based fuzzy systems [5, 11]. There is still one more aspect which should be taken into consideration. Using learning methods we can get indeed two functionally identical systems – FCM but their structure is quite different, as we can see also in Fig. 6. This is a problem because especially FCM should reflect just human representation of knowledge.

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0,1 & 0,1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0,5 & -0,25 & 0,3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0,75 & -1 & 0,75 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0,30 & -0,25 & -0,5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0,1 & 0,1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(a)

$$\begin{bmatrix} 0 & 0 & 0 & 0 & -0,06 & -0,57 & -0,2 & -0,52 & -0,63 & -0,63 & -0,03 & 1 & -0,08 & -0,63 & -0,3 & -0,63 \\ 0 & 0 & 0 & 0 & 0,18 & 0,54 & 0,13 & 0,7 & 0,5 & 0,25 & 0 & 0,45 & -0,07 & 0,25 & 0 & 0,5 \\ 0 & 0 & 0 & 0 & -0,2 & -0,57 & -0,06 & -0,52 & -0,63 & -0,63 & -0,3 & 1 & -0,08 & -0,63 & -0,03 & -0,63 \\ 0 & 0 & 0 & 0 & 0,48 & 0,19 & 0,48 & 0,48 & 0,28 & 0,43 & 1 & 1 & 1 & 0,43 & 1 & 0,28 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(b)

Figure 6

Weight matrices of FCM designed manually (a) and using LMS with unity initialization (b)

Acknowledgement

This work is the result of the project implementation: Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120020) supported by the Research & Development Operational Program funded by the ERDF.

References

- [1] Blanco A., Delgado M., Pegalajar M. C., Fuzzy Automaton Induction Using Neural Network, *International Journal of Approximate Reasoning*, Elsevier, Vol. 27, pp. 1-26, 2001
- [2] Chen S. M., Cognitive-Map-based Decision Analysis Based on NPN Logics, *Fuzzy Sets and Systems*, Elsevier, Vol. 71, No. 2, pp. 155-163, 1995
- [3] Gavalec M., Mls K., Fuzzy Cognitive Maps and Decision Making Support, in: *Proc. of the 21th Int. Conf. Mathematical Methods in Economics*, Prague, pp. 87-93, 2003
- [4] Hueriga A. V., A Balanced Differential Learning Algorithm in Fuzzy Cognitive Maps, in: *Proc. of the Sixteenth International Workshop on Qualitative Reasoning*, Barcelona, Spain, 2002
- [5] Johanyák Zs. Cs., Kovács Sz., A Brief Survey and Comparison on Various Interpolation-based Fuzzy Reasoning Methods, *Acta Polytechnica Hungarica*, Vol. 3, No. 1, pp. 91-105, 2006
- [6] Kosko B., Fuzzy Cognitive Maps, *International Journal of Man-Machine Studies*, Elsevier, Vol. 24, No. 1, pp. 65-75, 1986

- [7] Kosko, B., *Fuzzy Engineering*, Prentice-Hall, 1997
- [8] Kotras M., *Adaptation of Fuzzy Cognitive Maps for Needs of Navigation*, Bc. thesis in Slovak, Technical University in Košice, Slovakia, 2009
- [9] Li S. J., Shen R. M., *Fuzzy Cognitive Map Learning Based on Improved Nonlinear Hebbian Rule*, in: *Proc. of the Third Int. Conference on Machine Learning and Cybernetics*, Shanghai, China, pp. 2301-2306, 2004
- [10] Madarász L., *Intelligent Technologies and their Applications in Complex Systems (in Slovak)*, Technical University in Košice, Slovakia, p. 348
- [11] Oblak S., Škrjanc I., Blažič S., *If Approximating Nonlinear Areas, then Consider Fuzzy Systems*, *IEEE Potentials*, Vol. 25, No. 6, pp. 18-23, 2006
- [12] Papageorgiou E. I., Parsopoulos K. E., Stylios C. D., Groumpos P. P., Vrahatis M. N., *Fuzzy Cognitive Maps Learning Using Particle Swarm Optimization*, *International Journal of Intelligent Information Systems*, Springer, Vol. 25, No. 1, pp. 95-121, 2005
- [13] Papageorgiou E. I., Stylios C. D., Groumpos P. P., *Unsupervised Learning Techniques for Fine-Tuning Fuzzy Cognitive Map Causal Links*, *Int. Journal of Human-Computer Studies*, Elsevier, Vol. 64, pp. 727-743, 2006
- [14] Pozna C., Troester F., Precup R.-E., Tar, J. K., Preitl S., *On the Design of an Obstacle Avoiding Trajectory: Method and Simulation*, *Mathematics and Computers in Simulation*, Elsevier Science, Vol. 79, No. 7, pp. 2211-2226, 2009
- [15] Rutrich M., *Adaptation of Fuzzy Cognitive Maps for Navigation Needs*, MSc. thesis in Slovak, Technical University in Košice, Slovakia, 2009
- [16] Tar, J. K. et al., *Centralized and Decentralized Applications of a Novel Adaptive Control*, *INES 2005*, Budapest, Budapest Tech, 2005, pp. 87-92
- [17] Vaščák J., *Fuzzy Cognitive Maps in Path Planning*, *Acta Technica Jaurinensis*, Vol. 1, No. 3, 2008, pp. 467-479
- [18] Vaščák J., Madarász L., *Automatic Adaptation of Fuzzy Controllers*, *Acta Polytechnica Hungarica*, Budapest, Hungary, Vol. 2, No. 2, 2005, pp. 5-18

Problems of Digital Sustainability

Tamás Szádeczky

Department of Measurement and Automation, Kandó Kálmán Faculty of
Electrical Engineering, Óbuda University
Tavaszmező u. 17, H-1084 Budapest, Hungary
szadeczky.tamas@kvk.uni-obuda.hu

Abstract: The article introduces digital communication by drawing comparisons between the histories of digital and conventional written communication. It also shows the technical and legal bases and the currently reached achievements. In relation to the technical elements, it acquaints the reader with the development and current effects of computer technology, especially cryptography. In connection with the legal basis, the work presents the regulations which have emerged and made possible the legal acceptance of the digital signature and electronic documents in the United States of America, in the European Union and among certain of its member countries, including Hungary. The article reviews the regulations and the developed practices in the fields of e-commerce, electronic invoices, electronic records management and certain e-government functions in Hungary which are necessary for digital communication. The work draws attention to the importance of secure keeping and processing of electronic documents, which is also enforced by the legal environment. The author points to the technical requirements and practical troubles of digital communication, called digital sustainability.

Keywords: electronic archive; digital sustainability; preservation; electronic signature; data security

1 Introduction

We may distinguish three revolutions in the development of written communication [1]. The first revolution was the invention of alphabetical writing carrying phonetic value around 1300 BC., which segregated the text from the content. The second revolution in the 15th Century was the book printing invented by Gutenberg, Johannes Gensfleisch, which made written material widely available. The third revolution – which is called the digital revolution – is currently going on. We cannot accurately define its nature without an historical overview, but we may declare that it has modified all of the essential structures of written communication. At the same time, however, it has kept all the achievements of the previous revolutions. This is the key to digital culture and to the information society.

With the development of digital communication, computers conquered space in the area of applications which traditionally used paper. We can consider records management, accountancy and generally the creation of the public documents, notary documents and simple contracts as those applications. The necessary technical conditions have existed since the 1990s, when the technology of encryption with an asymmetric key was worked out in detail, in addition to computer and network technologies. First this development made creating electronic signature possible, and then came the acceptance of defining the legal consequences. We may solve electronic authentication of documents by electronic signature in the international practice and in Hungarian legal requirements. In the European Union and in the Republic of Hungary there are at present political and legislative endeavours aimed at rapid development in these areas. The problems of this process are referred to as the issue of digital sustainability [2].

The basic question of this research is: What is the difference between conventional and digital communication, especially as regards the long-term storage and usage of electronic documents. The research question was analysed via observation, information gathering and empirical statements based on the personal professional practice of the author.

2 Conditions of the Development of Digital Communication

The technical requirements necessary for the development of digital communication can be divided into three parts: computer technology, networking technology and data security procedures, providing a practical equivalent to the requirements of traditional written communication.

The first Turing-complete digital computer, the Zuse Z3, was constructed by Zuse Konrad in 1941. The creation of the transistor in 1947 revolutionized the world of crude electromechanical computers [3]. The integrated circuit, a mass of transistors fixed onto a single sheet, was developed in 1958. The first personal computer (the IBM PC) appeared in 1981, making it possible for home and office users to have computational performance. This was a huge invention compared to the usage of time slices in computer centres earlier.

The development of computer networks started in 1962, when the Advanced Research Projects Agency (ARPA) set up a research team called the Intergalactic Network, a group of which developed the time-sharing system. This system made possible the sharing of a mainframe computers' services between numerous users on a telex network [4].

In 1969 a research team of ARPA created the first packet switched computer network, the ARPANET. This network connected only a couple of universities and military systems at that time. It was not until the end of the 1970s that the civilian and broad usage of the network was in sight, but already at this time rapid enlargement had started. ARPANET was interlinked with NSFNet during the late 1980s, and the term 'Internet' was used as the name of the new network, which has since become the large and global TCP/IP network.

According to the World Internet Project 2007, 49 percent of Hungarian households (about two million) owned a personal computer at home, and the third of these computers (35%) also had an internet connection [5].

There is another significant requirement – but one which generally receives lesser emphasis – in connection with the development of digital communication: human resources. This means inclination and ability of citizens and clients to use the developments in digital communication and to actively participate in the information society. Because of the limited nature of this paper, the human resources and the economical aspect are beyond its.

In the following parts of the chapter, the development of the data security procedures and the legal-political conditions will be outlined, because of their emphasized significance in the present approach to the topic.

2.1 Data Security Procedures

The development and free public usage of data security procedures ensure a connection between the requirements of written communication and those of digital communication. It is necessary to ensure confidentiality, authenticity and originality in order to achieve the logic safety of electronic data. We accomplish this with mathematical methods, which cryptography (science of encryption) deals with. The encryption of information is practically the same age as writing, because since we started writing messages on paper (or on clay tablets) we have wanted others to be unable to read it.

Modern cryptosystems are wholly based on mathematical encryption methods, made by computers, which were developed at the end of twentieth century. There are two kinds of encryption methods: symmetric and asymmetric. The primary difference between the two is that the same key is used for encryption and decryption in the case of the symmetric one (this method is called single key because of this), while we perform the encryption and decryption processes with two different keys in the case of the asymmetric one (the two key method).

The modern computer-based symmetric encryption was developed in the 1970s. It applies reversible functions on blocks (generally 64-256 bits) by mixing and replacing characters or blocks of the clear text. In this method, it is possible to attain a good security level with the combination and numerous repeating of relatively simple procedures [6]. A popular symmetric cryptosystem of this kind is

the Data Encryption Standard (DES). While it is today insecure because of its short key length (56 bits), it is nevertheless occasionally used. As a result of break competitions on DES¹ we can state that any confidential message encrypted by DES is breakable in minutes with the appropriate hardware. From the principle of symmetric encryption (in civilian use) the cyphertext is breakable in all cases; it depends only on time.² Due to Kerckhoffs' principle, the security of encryption depends only on quality of the applied mathematical algorithm and the length of key, and thus not on the secrecy of the algorithm. Despite this, in individual cases, in the interest of increasing security (DES) or due to reasons of intellectual property (IDEA), the algorithm may be secret. Nowadays the Advanced Encryption Standard (AES) is used more often than anything else. AES, developed in the 2000s, uses 128-256 bits of key length and is hopefully based on an appropriate algorithm. Nobody has yet found a weak point on it, and thus experiments show that AES would be breakable in millions of years with all computers in the world [7].

The other branch of modern cryptography is public or asymmetric key cryptography, developed in mid 1970s. This branch is based on mathematical problems of one way trap door functions: discrete logarithm, discrete square root, and factorization with very large prime factors [8]. The solution to these problems is simple with a certain secret (private key), while without the secret it is more complicated. The application of asymmetric systems requires a larger key (1024-4096 bits) and more time for the coding and decoding processes, but the attained security is equal to that of the symmetric cryptosystems. For asymmetric encryption, it is necessary to generate two keys (a key pair) from a common secret. The common secret is destroyed after their generation. One of the key pairs' members will be the cryptographic private key, which the owner must not let out of his possession under any circumstances. If this occurs after all, it is deemed compromised and it is not allowed to use the key pair any more. If possible, it is necessary to revoke it. The other member of the key pair is the public key, which can be published on the Internet, and it is allowed to be shared on any unsecure communication channels. The usage of the two keys is equivalent, and thus it is possible to encrypt with one key and decrypt with the other key. Thus the application of different directions becomes possible (encryption and electronic signature).

The process of making an electronic signature is the following: you make a fingerprint from the data with a hash function. This function is a trap-door function, which means that the accomplishment of the function is simple in one direction, but in the other direction it is a complicated mathematical problem. This function generates a constant-size (128-512 bits) data set from the discretionary

¹ see <http://www.rsa.com/rsalabs/node.asp?id=2108>

² There is a theoretically unbreakable algorithm, One Time Pad, but its usage is not practical in civil environment, but it worths for military usage.

quantity of data. Changing one single bit in the input should modify at least 50 percent of the output bits (avalanche effect). We call the data set received an output fingerprint, since it characterizes the input data in a unique way. It is impossible to restore the input from the output and it is almost impossible to find two inputs with the same output.³ In practice SHA-1, RIPEMD-160, and Whirlpool algorithms are used. MD5 was suspected of being unsafe for years and its algorithm was broken in December 2008, and since that time its usage has been unsafe [9]. After the execution of this process on the document we have to sign, we encrypt the fingerprint with the cryptographic private key. We receive the electronic signature as a separate file which is independent from the document. We may send the signed document and the signature together to the addressee on a public channel, for example by e-mail. The addressee decrypts the electronic signature with our public key and obtains the fingerprint that we made. During this time he or she makes a fingerprint from the document we have sent with the same hash function and compares them. If they are the same, we can claim with certainty that no changes have occurred in the signed document, and that the signature on the document was made by the pair of our public key. On the other hand, it does not prove that this private key actually belongs to the sender, or that it has not been revoked, and we cannot ascertain the time of the signing. These criteria have to be proved by additional functions and controls. We may solve the problem of authenticity by binding the keys to one person, using two methods: with the web of trust (WoT) method [10], used by Pretty Good Privacy (PGP). According to this method, individuals trusting each other sign each other's keys. If the addressee trusts any of the persons signing the sender's key, or can follow back the signatures to a reliable person, this assures the sender's authenticity. The disadvantage of this method is that vast confidential networks are necessary in which two unknown people may have common acquaintances. The other method is Public Key Infrastructure (PKI). In this method, a third party who is entrusted by all communicating parties attests the reliability of all customers. This occurs with the use of a certificate, which is an electronic data set, and implies the public key. The third party is entrusted by the state or the public, who verifies the owner of the key and the key's relation prior to issuing the certificate (for example with the request of an ID or a verification e-mail). If the Certification Authority (CA), which is the national root, is trusted, other elements of the path are automatically entrusted. The publicly entrusted entity is called Certification Service Provider (CSP).⁴ The Timestamping Authority (TSA), who electronically signs the accurate time that the sender builds into the electronic signature of the document, does the authentication statement of the time of the electronic signature. A request for a timestamp generally happens on-line through the internet. The trustworthiness of the TSA is proven by its certificate issued by a trusted CA. The usage of key pairs, or rather the certificates, can be limited. A key pair can be used for an electronic

³ weakly or strongly collision free security, see [8]

⁴ CSP and CA are often used in same sense

signature, encryption, authentication or certificate issuance (such as CA). If we want to use more functions from among these but the certificate is restricted, we will need more key pairs and certificates.

The outlined technology has been developed to be suitable for the service of trusted digital communication. This fact alone is not sufficient for reaching this aim without considering other aspects.

2.2 Regulations and Practice

After the algorithmic and technical infrastructural realisation of the electronic signature, it was necessary to actualize practical usage in order for legislators to allow paper-based signatures to be replaced with electronic signatures. Utah, in the U.S., in early 1995 was the first state to pass a digital signature act. In Europe, Germany was the first country that accepted electronic signatures in 1996 (Gesetz zur digitalen Signatur), followed by the United Kingdom in 1999 (Building Confidence in Electronic Commerce – A Consolidation Document) and the European Union in 1999 (Directive 1999/93/EC of the European Parliament and of the Council of 13 December 1999 on Community framework for electronic signatures).

The first step in Hungarian legislation was Act XXXV of 2001 on Electronic Signatures [11].⁵ Electronic governmental services started partly on the basis of this rule of law. Electronic tax returns became available from the end of the nineties in several phases, and this improved considerably in 2002 and 2006 [12]. The electronic data-supplier service of the land registry (TakarNet) started in 2003, client gate (detailed later) began in 2005, regulations on electronic records management have existed since 2006, electronic public procurement started in 2007, electronic company registration was launched in 2008, as was the electronic auction of revenue authority and the change of Act on Accounting to simplify the storage of electronic invoices.

The Hungarian electronic signature law – corresponding to Directive 1999/93/EC – distinguishes three levels of electronic signatures independent from the technical background. At the bottom level there is the ‘simple’ electronic signature, which means a name written down into the electronic documents without any safety requirement (for example the signature at the end of an email). The legislator does not bind any special legal consequence to it, but only makes its acceptance subject to free consideration. The second safety level is the advanced electronic signature, which has to meet the requirements for the capability of identifying the signatory, uniquely linked to the signatory, which is created using devices that the signatory can maintain under his sole control; and it is linked to the document to which it relates in such a manner that any change to the data of the document made

⁵ abbreviated as ‘Eat.’

subsequent to the execution of the signature is detectable.⁶ The binding legal consequence is the correspondence to the legal requirements to put into written form excluding several specified fields.⁷ From among the electronic signatures, the safest one that meets the highest requirements is the qualified electronic signature. The qualified electronic signature is an advanced electronic signature, which is created with a secure signature creation device (SSCD) and which comes with a certificate issued with it.⁸ The requirements for qualified certificates are the strictest and these certificates are subject to additional regulation. An electronic document supplied with a qualified electronic signature is a full, conclusive private agreement. Thus, its authenticity and the fact of its belonging to the signer cannot be disputed until proof to the contrary. In the case of the last two levels, an independent audit and authority supervision is compulsory for the operation of Certification Service Providers and the production of certificates. The audit is carried out by independent audit companies (currently two) designated by the minister responsible for informatics and supervised by National Communication Authority. The security and quality of electronic signature and certification services in Hungary is ensured by these controls. Presently four Certificate Service Providers are operating in the Republic of Hungary.

Several non-PKI-based identification systems are used in Hungarian e-government applications, such as a smartcard solution in tax revenue procedures between 2004 and 2006 and 'Client Gate' from 2005. 'Client Gate' is a web-based online authentication system operated by Senior State Secretariat for Informatics. Registration occurs in local administrative offices with an identity-check; later several governmental services are accessible via this portal after SSL username and password authentication.

3 Problems of Digital Sustainability

The long-term preservation of electronic documents, especially of the electronically signed ones, is a complex task. The electronic data and its physical form must be protected from destruction. It is necessary to solve the long-term probative value of the electronic signature by the storage of its certification path, and it is also necessary to ensure access to the application capable of opening the given document.

The long-term preservation of the soundness of electronic data is aggregate combination of physical, logical and operational safety tasks [13]. In all cases, the redundancy of the data storage system and safe long-term storage of data storage devices is necessary.

⁶ Eat. 2. § 15.

⁷ Eat. 3. § (2-3) family law and judicial proceedings, unless the law explicitly allows

⁸ Eat. 2. § 17.

3.1 Excessive Velocity

The Hungarian government is struggling to move citizens' and other clients' paper-based activities to electronic procedures. This means in the long term, in practice, the creation and usage of electronic documents exclusive. This endeavour seems an exaggerated foreshadowing process in certain cases.

Legislators have not provided time for the conversion from paper-based processes to electronic ones. For example, the electronic tax return, electronic records management and the electronic firm registration happened this way. The majority of companies had less than one year to switch to the electronic tax return in the firm registration procedure; some companies had only half a year. There was no possibility for lawyers to register companies without adopting new technologies and knowledge [14]. We can consider this as switching to digital communication from the traditional written communication. We practiced the traditional communication for several thousand years, yet the number of illiterate people is one hundred thousand in Hungary [15]; therefore, a paradigm shift of this scale seems hopeless in several years. Well-developed countries, which have already introduced these steps towards digital communication maintain the opportunity to use paper-based documents. For example, in Austria the electronic firm registration procedure was available in the 80s with computers via the telephone network, a considerable part of public and criminal procedures became electronic in 90s, and a refining of the informatics opportunities for payment became available in this decade. Despite this intense and continuous development, the clients still have the opportunity to use paper-based documents in the above procedures, which he or she can download from the website of the Ministry of Justice.

3.2 Diversity of Formats

The variety of formats and the resulting diversity of processes and differences of applications have caused considerable difficulties until now and probably will cause them later as well. The well-known and more or less widely used electronic document file formats are the plain text file (txt), Microsoft Rich Text Format (RTF), and the Portable Document Format (PDF). Currently in Hungary TXT, RTF 1.7, PDF 1.3 formats fall under interpretational obligation by authorities in electronic administrative procedures [16]. The primary disadvantage of these formats is that these are unstructured; therefore, these can only clumsily be processed in automatic systems. The Microsoft Word document (DOC) is more widely used, but it is a licensed, proprietary format; its precise construction is Microsoft's trade secret. This has made it impossible for any other companies' software to be fully compatibility with it. To correct these generation faults both developer sides (Microsoft and OpenDocument Foundation) made new formats based on XML. Extensible Markup Language (XML) is a general aim descriptor

language, with the goal of the formation of special descriptor languages. XML, which is an improvement on the SGML language standardized by ISO in 1986 [17], became a W3C recommendation in 1998 [18]. The aim of XML is to structure data. It is licence-free, platform-independent and widely supported. An XML document is valid if it is well-formed (suits for the syntax of the XML language) and it matches a defined content rule, which defines the accepted value types and value places. This definition of the rules can be done with Document Type Definition (DTD) or XML Schema Definition (XSD). The OpenDocument format (ODF) developed by OpenDocument Foundation [19] and Office Open XML (OOXML) developed by Microsoft [20] are based on this technological framework. These formats may become the widely used future document formats. The forms possess unique XSD with severe bindings based on this technology and can be automatically processed.

The general and the records management metadata problems are professionally particular, but in administrative informatics there are considerable questions. These are about in what kind of form, value and entity the metadata (data about data) appears. Many initiatives have been initiated to try to resolve this issue: the Dublin Core Metadata Initiative (DCMI), Managing Information Resources for e-Government (MIREG), GovML and PSI Application Profile [21].

A foreign country example for the format problem is that it was not possible to restore the measurement data of the Viking space probe stored on magnetic tape in 1976. It was necessary to type in everything again from the earlier printed documents because the stored format was unknown.

3.3 On-Line Data Security

If we store the electronic data in a working computer system on-line, it is necessary to protect the system physically against disaster losses (fire, water damage, even by the public utility drinking water supply or drainage, earthquake, the destruction of the object due to other reasons); against incidents due to a deficiency of technical requirements (deficiency of power supply, power supply disturbances, the deterioration of climatic circumstances (which may be temperature or humidity problem), informatics network problems)); against electromagnetic disturbances (even in case of intentional damage); and against technical reliability problems (production mistakes, fatigue, other breakdowns). The logic safety covers the reliability of software elements (the operating system, applications), protection against intentional damage (viruses, worms, malicious programs, network attacks, and hacker activity), the security of network protocols, setting rights, identity management and access management. [22] The reliability of software components could be enhanced by defensive programming techniques, with mitigating risks like data leakage vulnerability. [23]

3.4 Offline Data Security

The storage of the digital data can be done on optical or magnetic media if the data does not change, and this solution seems to be more beneficial in terms of expenses. The primary medium of optical data storage is the DVD nowadays. This is not an eternal data storage solution, contrary to public belief. Depending on the quality of the disk we may be confident of the survival of the data up to at most 10 years, but based on the author's own experience, the data can be lost from the disk even after two years. Arising from this, it follows that in the case of optical data storage use, periodic regeneration is inevitable, which in practical terms means copying the discs. The benefit of the optical data storage is the insensitivity to electromagnetic fields, but it is necessary to maintain the suitable temperature, humidity and mechanics at the time of storage. The other widely used data storage is magnetic tape, which has a longer history than optical drives. The capacity of magnetic tape storage still exceeds the capacity of the optical; a cassette tape can have one Terabyte capacity.⁹ Over all of requirements of optical storage, the requirement of protection against electromagnetic interference (EMI) also appears. Regular regeneration of the media is necessary because the magnetic carrier demagnetises over time.

Those who store or archive data must provide the necessary environment for the opening of electronic documents, but users and operators are inclined to forget this requirement, despite its particularly large significance. Presupposing the worst case, imagine we have to open a stored electronic document 100 years later. The word processor application was made by a software-developing firm working in a garage in 1992. The format of the document does not comply with any kind of standards, and the full computer architecture has changed, but legal regulations do not allow the document to be destroyed. In this naturally strongly polarized example we have to store the archived document, the application able to open the document, the operating system capable of running the application and the full hardware configuration capable of running the operating system. The simplification of this procedure can be done by the emulation of a system capable of all of the above or a special realisation of the emulation and the migration called Universal Virtual Computer (UVC) [24]. This problem arose already at the time of the research of the state security documents after the political transformation in Hungary, when the researchers were unable to read important documents stored on magnetic tapes, because the proper reader could not be obtained or reproduced.¹⁰ Of course the interest in keeping this information secret may be playing a part in this issue.

⁹ see <http://www.ibm.com/systems/storage/tape/>

¹⁰ Based on the oral information of Dr. Trócsányi Sára, head of department of Office of Data Protection and Freedom of Information Commissioner, Hungary, Pécs, November 15, 2008.

As a mixed on-line and off-line data security solution for general documents, records and book digital libraries can be used as a part of conventional libraries [25]. This solution has been under research for a decade in the Western countries, but is not well known in Central and Eastern Europe.

3.5 Authenticity of an Electronic Signature

The validity of the electronic signature is provable after a couple of hours or days. The reason of this is that in order to check the validity, we need to know the current certificate revocation list (CRL) of the CA at the time of the signature, in which the certificate service provider publishes the certificates revoked due to being compromised or to something else. This time can significantly be reduced with the use of Online Certificate Status Protocol (OCSP). After this procedure, the authenticity of the electronic signature is continuously provable, if the revocation lists and the full certification path is accessible. The electronic archiving activity receives a role here: at the moment of archiving, the valid authentication data is stored and re-authenticated by the electronic archiving service provider. Thus the authenticity of an electronic document signed by the archiving service provider depends only on the existence of the electronic archiving service provider, but is not influenced by the falling out of a part of the certification path, for example by the terminating of CA activity or the compromise of the CA key. This is the point where the electronic archiving service provider's activity rises above the issue of the simple safe data storage.

Conclusions

It follows from the previously expounded manifold requirements and problems that during storage and processing of electronic documents, an organization faces serious difficulties. The risk stemming from the outlined problems is diverging in different countries. The faster that digital communication develops, the harder it is to find time to solve problems. Therefore the chance of a later escalating of troubles increases.

The rapid development of administrative informatics in Central and Eastern European countries belongs to this category [26]. In the author's opinion, these countries did not analyse well enough the 30 years of experience of Western European states in this field, even though avoiding some mistakes would be possible that way. These undiscovered problems may cause serious data loss in the government sector.

Digital preservation deals with the dangers stemming from the above-mentioned problems and with the protection against them. The hypothesis of an apocalyptic future because of all the above is called the digital dark age [27]. According to this theory, most of the electronic documents made in the 21st Century will disappear with just a few written memories remaining, similar to the Middle Ages. Leading representatives of this theory are the Getty Research Institute researchers [28], and

Kuny, Terry [29]. However, a refutation was also born which states that the experiences until now are only examples of deficiencies in data restoration, not in data loss [30]. As a comprehensive solution to the technical problems, an ISO reference model was made called Open Archival Information System (OAIS) [31].

Against the excessive speed, better policymaking may provide protection in Hungary; for the format problem and partially data security problem, stricter regulation may provide protection. The efficient handler of the electronic signature and the partially data security problems should be the electronic archiving service providers working on the market. Certainly, because the complexity of these problems and the deficiency of the market, many claim there is no electronic archiving service provider working well (in a suitable measure used) in Hungary. On the other hand, the organisations obliged to preserve electronic documents reckon that they are able to do justice to these requirements with their own infrastructure and human resources. Based on the author's own experience the village local governments with ten employees also think that they will be able to satisfy this task on their own, but it does not seem so in the long run.

As a problem statement for further research we could find out the exact specialties of digital sustainability in Central and Eastern Europe and make recommendations on reducing the gap between WEU and CEE states.

References

- [1] Rubin, J. S., 'The Printed Book: Death or Transfiguration', *Journal of the World Book Community*, 1990, Vol. 1, No. 1, pp. 14-20
- [2] Kevin Bradley, *Defining Digital Sustainability*, *Library Trends*, 2007, http://findarticles.com/p/articles/mi_m1387/is_1_56/ai_n21092805/ [2009. 11. 01.]
- [3] Köpeczi, B. (eds.) 1974, *Az embergéptől a gépemberig*, Minerva, Budapest, p. 206
- [4] Kita, C. I., 'J.C.R. Licklider's Vision for the IPTO', *IEEE Annals of the History of Computing*, 2003, No. 3, p. 65
- [5] ITTK, *Hungarian Information Society Report 1998-2008*, ITTK, Budapest, 2008, p. 38
- [6] Horváth, L., Lukács, Gy., Tuzson, T., Vasvári Gy., *Informatikai biztonsági rendszerek*, BMF-E&Y, Budapest, 2001, p. 111
- [7] Virasztó, T., *Titkosítás és adatretjtés. Biztonságos kommunikáció és algoritmikus adatvédelem*, Netacademia, Budapest, 2004, p. 50
- [8] Menezes, A., Oorschot, P. van, Vanstone, S., *Handbook of Applied Cryptography*, CRC Press, 1996, p. 284

-
- [9] Sotirov, A., Stevens, M., Appelbaum, J., Lenstra, A., Molnar, D., Osvik, D. A., Weger, B. de, MD5 Considered Harmful Today. Creating a Rogue CA Certificate, 2008, <http://www.win.tue.nl/hashclash/rogue-ca/> [2009. 12. 03.]
- [10] Abdul-Rahman, A., The PGP Trust Model, EDI-Forum, Langdale, 1998
- [11] Act XXXV of 2001 on Electronic Signatures in Hungary
- [12] Jacsó, T., Az ügyfélkapu és az eBEV használata, Saldo, Budapest, 2006
- [13] Ross, S., Hedstrom, M., 'Preservation Research and Sustainable Digital Libraries', International Journal on Digital Libraries, 2005, Vol. 5, No. 4, pp. 317-324
- [14] Szilágyi, K. B., Az elektronikus cégeljárás gyakorlati kézikönyve, Jogászoknak Kft., Pécs, 2008, p. 4
- [15] UNESCO, UIS Statistics in brief, UNESCO, Paris, 2008
- [16] Decree No. 12/2005. (X. 27.) IHM of the Minister of Informatics and Telecommunications on the technical rules on documents which can be applied in electronic administrative procedure, 1st appendix
- [17] ISO 8879:1986 Information processing, Text and office systems, Standard Generalized Markup Language (SGML)
- [18] World Wide Web Consortium, XML Core Working Group Public Page <http://www.w3.org/XML/> [2009.11.15.]
- [19] ISO/IEC 26300:2006 Open Document Format for Office Applications (OpenDocument) v1.0
- [20] ISO/IEC 29500:2008, Information Technology – Office Open XML formats; ECMA-376 Office Open XML File Formats - 2nd edition (December 2008)
- [21] Bountouri, L., Papatheodorou, C., Soulikias, V., Stratis, M., 'Metadata Interoperability in Public Sector Information', Journal of Information Science, 2007, No. 7, pp. 1-25
- [22] Illési, Zs., 'Számítógép hálózatok krimináltechnikai vizsgálata', Hadmérnök, 2009, Vol. 4, No. 4
- [23] Schindler, F., 'Coping with Security in Programming', Acta Polytechnica Hungarica, 2006, Vol. 3, No. 2, pp. 65-72
- [24] Lorie, R., The UVC: a Method for Preserving Digital Documents - Proof of Concept. IBM Netherlands, Amsterdam, 2002
- [25] Hamilton, V., 'Sustainability for Digital Libraries', Library Review, 2004, Vol. 53, No. 8, pp. 392-395

- [26] Otjacques, B., Hitzelberger, P., Feltz F., 'Interoperability of E-Government Information Systems: Issues of Identification and Data Sharing', Journal of Management Information Systems, 2007, Vol. 23, No. 4, pp. 29-51
- [27] Wikipedia, Digital dark age.
http://en.wikipedia.org/wiki/Digital_Dark_Age [2009.11.20.]
- [28] MacLean, M., Davis, B. H. (Eds.), Time and Bits: Managing Digital Continuity. Getty Publications, Los Angeles, 2000
- [29] Kuny, T., A Digital Dark Ages? Challenges in the Preservation of Electronic Information. IFLA, Copenhagen, 1997
- [30] Harvey, R., So Where's the Black Hole in our Collective Memory? A Provocative Position Paper (PPP), 2008,
http://www.digitalpreservationeurope.eu/publications/position/Ross_Harvey_black_hole_PPP.pdf [2008. 10. 28.]
- [31] Consultative Committee for Space Data Systems, Reference Model for an Open Archival Information System (OAIS) CCSDS Secretariat, Washington, DC, 2002

The Digital Pre-Operative Planning of Total Hip Arthroplasty

**Monika Michalíková¹, Lucia Bednarčíková¹, Martin Petřík¹,
Jozef Živčák¹, Richard Raší²**

¹ Department of Biomedical Engineering, Automation and Measurement
Faculty of Mechanical Engineering
Technical University of Košice
Letná 9, 042 00 Košice, Slovakia
E-mail: monika.michalikova@tuke.sk, lucia.hutnikova@tuke.sk,
martin.petrik@tuke.sk, jozef.zivcak@tuke.sk

² Trauma Surgery Department
L. Pasteur University Hospital of Košice
Rastislavova 43 04001 Košice, Slovakia
E-mail: rasi@fnlp.sk

Abstract: Pre-operative planning is a very important part of hip arthroplasty (especially reimplantation of total hip and hip joint). Conventional pre-operative planning is realized with caliper, protractor, plastic transparent templates and x-ray images. This conventional templating technique is time consuming with many errors and impractical. This paper presents the current applications of computer technology in the field of surgery and pre-operative planning of total hip implantation. At the present time, orthopaedic surgeons use transparent template radiographs as part of pre-operative planning in order to gauge the suitability and correct size of an implant. The newly developed CoXaM software offers a simple solution to the problems by using digital x-ray images and handmade transparent plastic templates. The utilization of developed software has many advantages in the hospital unit (the elimination of storing large inventories of implants, the minimalization of errors from the magnification of templates and x-ray images, etc.). The proposed methodology provides the opportunity for comfortable, user-friendly and dimensionally accurate computer programming surgical operation. The technique is reliable, cost effective and acceptable to patients and radiographers.

Keywords: digital pre-operative planning; software; x-ray image; hip joint replacement; template

1 Introduction

Computer technology has many applications in different fields of industry, health care and medicine. This encompasses paper-based information processing as well as data processing machines (a Hospital information system or Clinical information system) and image digitalization of a large variety of medical diagnostic equipment (e.g. computer images of X-ray, MR, CT). Many of these applications allow the visualization and classification, respectively the identification and the assessment, of the diagnosed objects. The aim of the computer technology in medicine is to achieve the best possible support of patient care, pre-operative surgery planning and administration by electronic data processing.

Radiographs historically have not been standardized according to magnification. Depending upon the size of a patient, a film will either magnify a bone and joint (of large patients with more soft tissue) or minimize (in the case of thin patients). An orthopedic surgeon must estimate the degree of plus or minus magnification in order to select an implant that is the correct size. The surgeon may be helped by the incorporation of a marker that is of a known size. By calculating the difference between the size of the marker displayed on the film and the actual size of the marker, the orthopedic surgeon can identify the degree of magnification/minimization and compensate accordingly when selecting a prosthetic template. [1]

Accurate preoperative planning improves the procedure's precision [2, 3], shortens its duration [2, 4], and reduces the incidence of prosthesis loosening [5, 6] and loss of bone stock [5, 6, 7]. As well, it lowers the risk of periprosthetic fracture, helps restore femoral offset and leg length in hip arthroplasty, facilitates optimization of alignment and ensures the required implants are available while minimizing the costs [8] and complications (e.g., instability) [5, 9, 10, 11, 12, 13, 14, 15, 16, 17].

Kulkarni *et al.* in 2008 devised a method whereby a planar disc placed on the radiographic cassette accounts for the expected magnification. Digital radiography is becoming widespread. Accurate pre-operative templating of digital images of the hip traditionally involves positioning a calibration object onto its centre. This can be difficult and cause embarrassment. [18]

Digital pre-operative planning enables the surgeon to select from a library of templates and electronically overlay them over the image. Therefore, the surgeon can perform the necessary measurements critical to the templating and pre-operative planning process in a digital environment. The pre-operative planning process is fast, precise, and cost-efficient, and it provides a permanent, archived record of the templating process. [19]

William Murzic *et al.* in 2005 presented a study with the aim of evaluating the accuracy of a specific templating software (with an emphasis on femoral component fit) and comparing it to the traditional technique using standard radiographs. [20]

Incorrect preoperative templating of a THA might lead to inappropriate implant size and position [2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 21, 22], and revision of the prosthesis might be needed. Preoperative analog [3, 9, 10, 12, 16, 23, 24, 25, 26, 27], and digital [4, 19, 21, 24, 26, 27, 28, 29, 30, 31, 32] templating methods have been studied. [17]

2 Anatomy and Morphology of a Hip Joint

The hip joint, scientifically referred to as the acetabulofemoral joint (art. coxae), is the joint between the femur and acetabulum of the pelvis, and its primary function is to support the weight of the body in both static (e.g. standing) and dynamic (e.g. walking or running) postures.

The hip joint (See Fig. 1) is a synovial joint formed by the articulation of the rounded head of the femur and the cup-like acetabulum of the pelvis. It is a special type of spheroidal, or ball and socket, joint where the roughly spherical femoral head is largely contained within the acetabulum and has an average radius of curvature of 2.5 cm. [4]

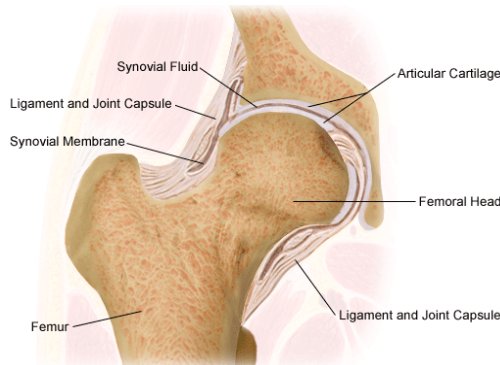


Figure 1

Right hip joint – cross-section view [42]

The hip muscles act on three mutually perpendicular main axes, all of which pass through the center of the femoral head, resulting in three degrees of freedom (See Fig. 2) and three pair of principal directions: Flexion and extension around a transverse axis (left-right); lateral rotation and medial rotation around a longitudinal axis (along the thigh); and abduction and adduction around a sagittal axis (forward-backward); and a combination of these movements (i.e. circumduction, a compound movement in which the leg describes the surface of an irregular cone) [34].

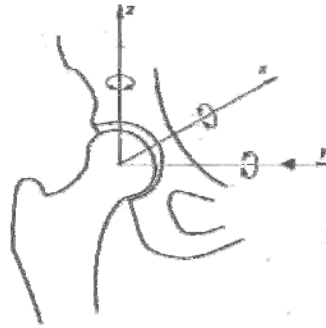


Figure 2
Three degrees of hip joint freedom

The most important morphological specifications (See Fig. 3) which can be measured on an anteroposterior pelvic radiograph are:

- the femoral neck angle (the caput-collum-diaphyseal angle, the CCD angle) – between the longitudinal axes of the femoral neck and shaft, which normally measures approximately 126° in adults,
- the acetabular inclination (the transverse angle of the acetabular inlet plane) – the angle between a line passing from the superior to the inferior acetabular rim and the horizontal plane, which normally measures 40° in adults. [33]

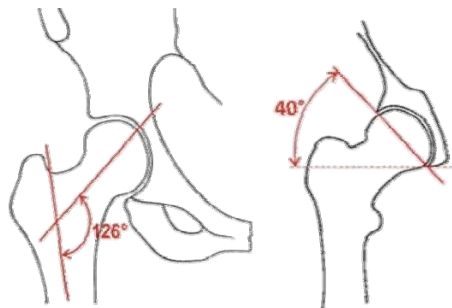


Figure 3
Femoral neck angle and acetabular inclination

The next important morphological specification is the femoral neck anteversion, which is practically unmeasured in AP projection and which can be measured well by CT or MR (3D measurement methods).

A perfect AP radiograph of the femur needs to account for the anteversion of the femoral neck. Patients are required to rotate the leg internally by a mean of 15° . Restricted rotation of the hip in osteoarthritis sometimes makes it difficult to

achieve this position. A study of the radiological dimensions of the femoral canal shows that the AP width of the medullary canal at the isthmus does not change significantly from 20° internal to 40° external rotation. At 20 mm below the lesser trochanter there is no significant change on internal rotation and an apparent increase of 1.1 mm with 20° external rotation. [35, 36]

3 Total Hip Prosthesis

Hip replacement (total hip replacement) is a surgical procedure in which the hip joint is replaced by a prosthetic implant. Replacing the hip joint consists of replacing both the acetabular and the femoral components (See Fig. 4).

Such joint replacement orthopaedic surgery generally is conducted in order to relieve arthritis pain or to fix severe physical joint damage as part of hip fracture treatment. Hip replacement is currently the most successful and reliable orthopaedic operation, with 97% of patients reporting improved outcome.

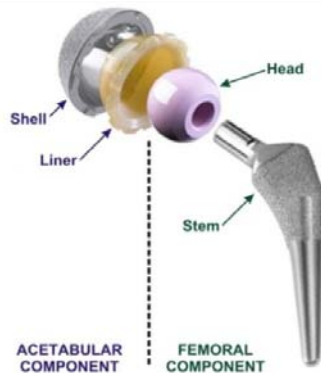


Figure 4

The modular structure of total hip prosthesis

The total hip prosthesis must be anchored securely within the skeleton for correct functioning. A loose sitting total hip prosthesis is painful, and such a loose total hip is also stiff.

There are two methods for securing the fixation of a total hip prosthesis to the skeleton [37]:

- 1 The cemented total hip – the surgeon uses bone cement for fixation of the prosthesis to the skeleton;
- 2 The cementless total hip – the surgeon impacts the total hip directly into the bed prepared in the skeleton;

The construction, form, and rehabilitation after the operation with these two types of prostheses are different.

- 3 The hybrid total hip prosthesis – a cementless cup paired with cemented shaft.

With the increasing utilization of uncemented implants, templating has become more critical. With a higher risk of intra-operative fracture during insertion, it is re-assuring for the surgeon when the pre-operative prediction matches the intra-operative choice of implant. [38]

A tight interference fit is desirable when introducing the femoral component of an uncemented hip replacement. A stem which is too small may not be stable, and attempts to insert one which is too large increases the risk of intraoperative fracture. Such a complication has been reported in 3% to 24% of patients. [23, 35, 39]

Successful surgery requires the precise placement of implants in order that the function of the joint is optimized both biomechanically and biologically. Pre-operative planning is helpful in achieving a successful result in total joint replacement. Pre-operative templating in total hip replacement helps familiarize the surgeon with the bone anatomy prior to surgery, reducing surgical time as well as complications.

This activity takes time and also is subject to mathematical error. Digital pre-op planning allows for an image to be displayed electronically, and with the aid of a known sized marker, automatically calculates the magnification and recalibrates the image so that it is sized at 100% from the perspective of the user. [1]

Typically most reconstructive surgeons have used acetate overlays and radiographs to determine appropriate implant size. Pre-operative planning is realized with caliper, protractor, plastic templates and x-ray images. The measurement is time consuming and can involve multiple errors. Digital images replace radiographs, which can no longer be lost or misplaced in a completely filmless system. X-ray images are viewed on a diagnostic grade monitor, rendering prosthetic overlays useless. [19, 40, 41]

4 Digitalization of the Pre-Operative Planning by CoXaM Software

The “CoXaM” software was developed in Visual Studio 2005 (Microsoft) in the Visual C++ programming language at the Department of Biomedical Engineering, Automation and Measurement at the Faculty of Mechanical Engineering, Technical University of Košice.

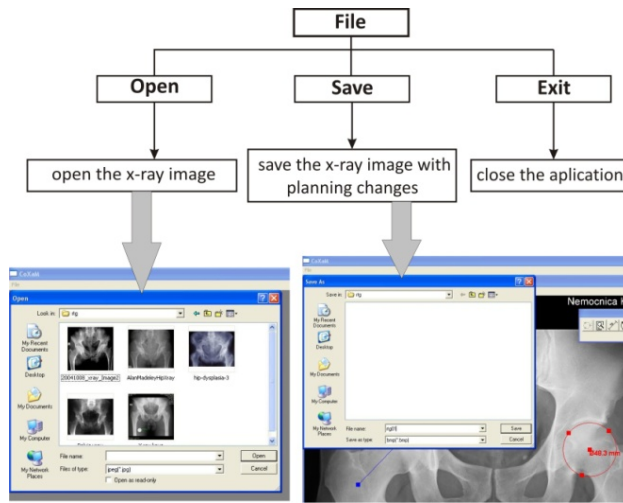


Figure 5
Overview of main menu

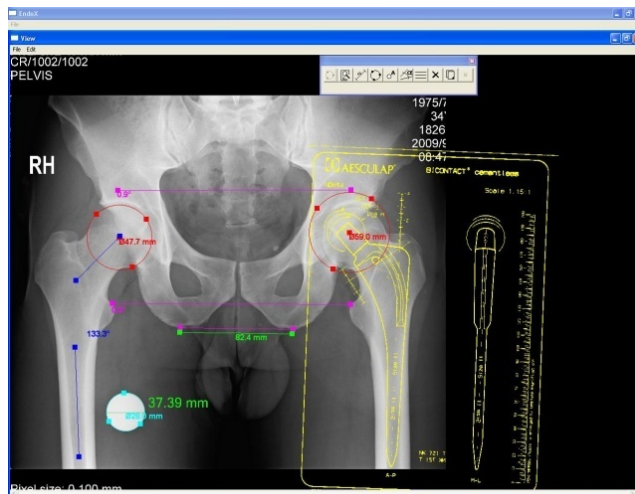


Figure 6
Example of using the “CoXaM” and control panel

The new, sophisticated software “CoXaM” (See Fig. 5) was designed for pre-operative planning and helps to determine on the X-ray image length dimensions, a center of rotation, and angle values (See Fig. 6). These parameters are considered in parallel with guidance lines. The software enables the digitalization of plastic templates from several producers, which will assess the suitability of the type of implant. By using digital templates, the surgeon can employ a sequential

method to determine which size of prosthesis to use and where to place the prosthesis within the bone to ensure optimum functioning of the joint following surgery. The incorporation of the various templates into the software in terms of the “magnification factor” is essential for accurate pre-operative templating and planning.

4.1 Possibilities of CoXaM Software

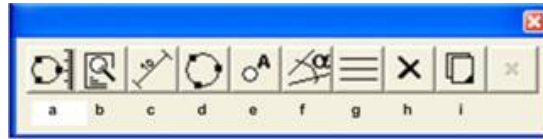


Figure 7

Control panel detail – a) calibration circle, b) center icon, c) measurement of dimension, d) circle, e) text, f) angle, g) examination of three line parallelism, h) removing the planning, i) templates

- Calibration circle (See Fig. 7a) – allows exact conversion of the marking dimensions for a given calibration feature on the X-ray. Determining the three-point plotted circle whose diameter in millimeters the real user enters – in this case 28 mm.
- Center icon (See Fig. 7b) – centers the x-ray image into viewport.
- Measurement of dimension (See Fig. 7c) – calculates the distance between two points. If the calibration is passed the result is in millimeters; otherwise it is displayed in pixels.
- Circle (See Fig. 7d) – from three points the software calculates the circle (center, diameter). If the calibration is passed the result is in millimeters; otherwise it is in pixels. The circles are used for finding the center of the hip joint and defining the dimensions of the femoral head and acetabular component diameter. With help of the circle you can determine the floatable center of rotation before and after surgery.
- Text (See Fig. 7e) – allows the user to enter text. The font used is Arial 12 pts.
- Angle (See Fig. 7f) – The angle between two lines (created from four points). It is not necessary that the two lines have an intersection point.
- Examination of a three-line parallelism (See Fig. 7g) – the L. Spotorno and S. Romagnoli method calculates the parallelism between three lines (created from six points) – the ischial tuberosities flowline (the base line), the superior acetabular rims flowline and the lesser trochanters flowline.
- Clearing the planning (See Fig. 7h) – removes all these tasks and clears the x-ray image.

- Templates (See Fig. 7i) – opens a digital template from the database of scanned templates from total hip prosthesis procedures. This allows calibrating of the templates and inserting it into the x-ray image. The size of the template is equal to the size of x-ray, and it is possible to rotate and move it.

At present, for preoperative planning software is used where it is necessary to define the magnification of a reference object on the x-ray image (See Fig. 8). This magnification will be determined of the real diameter and the diameter of the reference object measured on the x-ray image.

Using CoXaM software it is not necessary to know the x-ray image magnification. The user defines the value of the reference object diameter, thus eliminating the necessity of measurement (See Fig. 8).

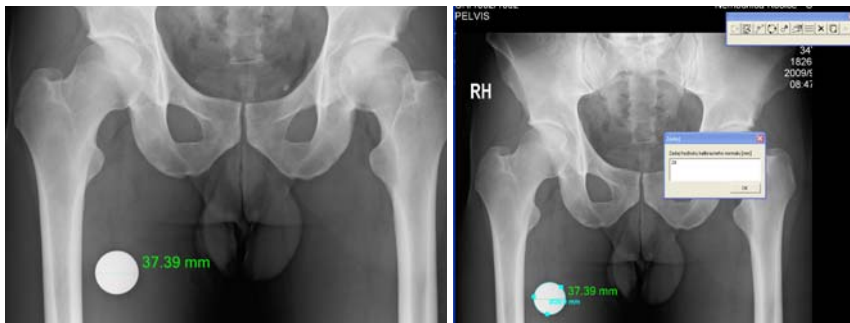


Figure 8

Software used at the present time (diameter evaluation – real diameter of the calibration component is 28.00 mm) and CoXaM software (determining of the reference object diameter)

Traditionally, an orthopedic surgeon places an acetate template enlarged to between 110% and 120% over an X-ray film magnified to between 110% and 130%. [1] An object (a disc or sphere is recommended) of a known size projected onto the film is necessary to determine the magnification.

Hendrikus J. A. Crooijmans MSc et. al. in 2009 designed a templating method using a new way of determining the hip magnification with a linear relationship between magnification of the hip and the reference object on top of the pubis symphysis; the relationship was determined on 50 radiographs. We then compared our method with two other templating methods: an analog method assuming an average hip magnification of 15% and a digital method determining the hip magnification with a one-to-one relationship between the reference object and the hip. [17]

In digital radiograph templating, the template and radiograph can be scaled to obtain identical magnifications. [17]

When properly placed, the magnification of the reference object represented the magnification of the hip (in a one-to-one relationship) and thereby enabled accurate pre-operative templating. The method required the reference object to be properly placed at the same distance from the detector as the center of rotation of the hip. Alternative methods for the correction of magnification, including using a line as a magnification reference [30], using coins placed at various positions [30, 43] as a magnification reference, using software to template digital radiographs [26, 28], or using software to template CT data have previously been described [17, 27].

For pre-operative planning for interventions of hip joint arthroplasty (implantation, reimplantation) plastic templates are commonly used. Each manufacturer offers its own plastic templates for the product of implants (types and sizes) (See Fig. 9).

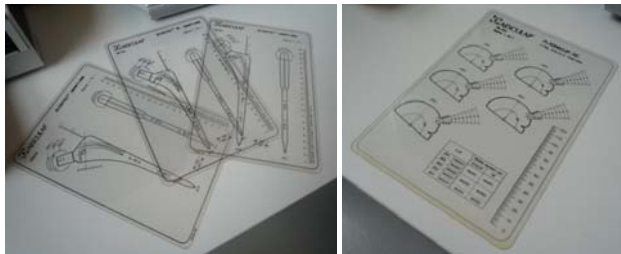


Figure 9
Plastic templates

CoXaM software works with electronic templates that are placed in a folder containing the scanned plastic templates. The user selects the required template from that folder (See Fig. 10).

In the event that the selected template was not calibrated, the next step is to do the calibration. That process is done in the calibration window, in which the abscissa is marked on the scale of the selected templates and the user inserts in the box (See Fig. 11 step 2) value. It is necessary that this value corresponds with the template scale (e.g. the scale of the template is 1,15:1, if the user marks 10 mm abscissa, then the inserted value in the box (See Fig. 11 step 2) is 11,5). After selecting the option “OK” calibration is confirmed. After that, the selected templates are drawn in yellow and applied to the x-ray image. The user can use the tool for template mirroring, in the event that the preoperative planning is for the counteractive hip joint.

It is possible to save the calibration before the confirmation; then for future reference calibration with that template is not necessary (for each template it is necessary to perform calibration when it is first used).

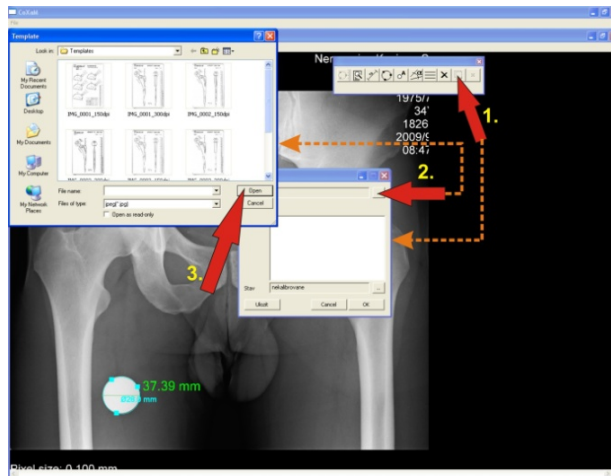


Figure 10

Selection of required template

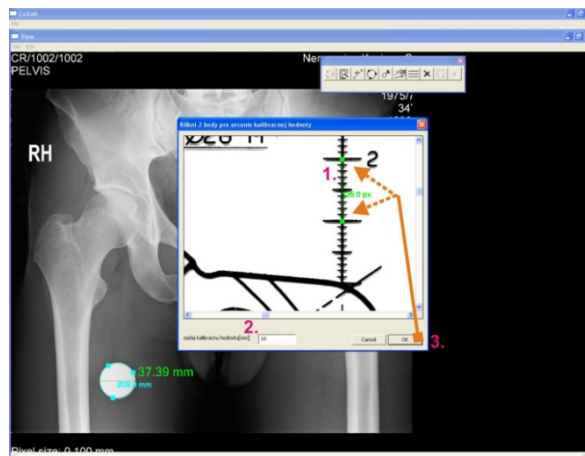


Figure 11

Calibration process of template

The user can to move the template to the required location and rotate it around its center (See Fig. 12).

Selected orthopedic departments used a demonstration version of the CoXaM software. The respondents learnt a basic working knowledge of the software. Afterwards, the orthopedists filled in questionnaires in which they described their opinion of the CoXaM software. Learning how to use CoXaM software takes from 30 to 60 minutes, according to the results from the questionnaires. Preoperative planning for the skilled user takes from 10 to 15 minutes for each case.



Figure 12

Placement of the templates over x-ray image such that optimal fill of both

Conclusions

At present, computer and imaging technologies with electronic outputs are improving slowly but steadily in hospitals. The quality and user comfort of the software equipment often adds value during the hospital surgery planning process.

Hendrikus J. A. Crooijmans MSc *et. al.* in 2009 found a linear relationship between magnification of the reference object at the pubis symphysis and the hip. [17]

CoXaM offers a simple solution to the problem of using digital x-ray images and handmade plastic templates. The problem is solved by the digitalization of templates for use in software. The developed software combines digital x-ray images with digital templates for planning implantation and reimplantation interventions of hip joints.

The new proposed methodology provides the opportunity for comfortable, user-friendly and dimensionally accurate computer programming for surgical operations. The technique is reliable, cost effective and acceptable to patients and radiographers. It can easily be used in any radiography department after a few simple calculations and the manufacture of appropriately-sized discs.

The CoXaM software provides several advantages for orthopaedic surgery. X-ray film is no longer necessary. There are no radiographs to store, lose, or misplace. Over time this results in a cost savings as film and developing supplies are no longer needed. The disadvantages include the initial cost of outfitting the technology.

As digital technology improves and becomes more accessible to the health care industry, digital radiography will be used by an increasing number of hospitals and

orthopedic practices. More practices will become filmless and software programs will be necessary for successful reconstructive planning and templating. Significant clinical studies are planned to statistically confirm the qualitative value of the software and quantitative precision of the output parameters.

Acknowledgement

This research has been supported by the research project 1/0829/08 VEGA - Correlation of Input Parameters Changes and Thermogram Results in Infrared Thermographic Diagnostic.

References

- [1] Bono, J.: Digital Templating in Total Hip Replacement: A New Gold Standard?, *Journal of Bone and Joint Surgery - British Volume*, Vol 86-B, Issue SUPP_IV, 413, 2004
- [2] Blackley H. R., Howell G. E., Rorabeck C. H.: Planning and Management of the Difficult Primary Hip Replacement: Preoperative Planning and Technical Considerations. *Instr. Course Lect.* 2000; 49:3-11
- [3] Knight J. L., Atwater R. D.: Preoperative Planning for Total Hip Arthroplasty: Quantitating its Utility and Precision. *J. Arthroplasty.* 1992; 7:403-409
- [4] Schiffers N., Schkommodau E., Portheine F., Radermacher K., Staudte H. W. [xPlanning and performance of orthopedic surgery with the help of individual templates] (in German). *Orthopaedie.* 2000; 29:636-640
- [5] Haddad F. S., Masri B. A., Garbuz D. S., Duncan C. P.: The Prevention of Periprosthetic Fractures in Total Hip and Knee Arthroplasty. *Orthop Clin North Am.* 1999; 30:191-207
- [6] Müller M. E.: Lessons of 30 Years of Total Hip Arthroplasty. *Clin Orthop Relat Res.* 1992; 274:12-21
- [7] Haddad F. S., Masri B. A., Garbuz D. S., Duncan C. P.: Femoral Bone Loss in Total Hip Arthroplasty: Classification and Preoperative Planning. *Instr. Course Lect.* 2000; 49:83-96
- [8] Kosashvili Y., Shasha N., Olschewski E., Safir O., White L., Gross A., Backstein D.: Digital versus Conventional Templating Techniques in Preoperative Planning for Total Hip Arthroplasty, *Can J. Surg.* 2009 February; 52(1): 6-11
- [9] Cech O., Fassbender M., Kirschner P., Rozkydal Z. [xPreoperative planning and surgical technic in achieving stability and leg length equality in total hip joint arthroplasty] (in Czech). *Acta Chir Orthop Traumatol Cech.* 2002; 69:362-368
- [10] Egli S., Pisan M., Muller M. E.: The Value of Preoperative Planning for Total Hip Arthroplasty. *J. Bone Joint Surg. Br.* 1998; 80:382-390

-
- [11] Goldstein W. M., Gordon A., Branson J. J.: Leg Length Inequality in Total Hip Arthroplasty. *Orthopedics*. 2005; 28(9 suppl):s1037-1040
- [12] Goodman S. B., Huene D. S., Imrie S.: Preoperative Templating for the Equalization of Leg Length in Total Hip Arthroplasty. *Contemp. Orthop*. 1992; 24:703-710
- [13] Lindgren J. U., Rysavy J.: Restoration of Femoral Offset during Hip Replacement: a Radiographic Cadaver Study. *Acta Orthop Scand*. 1992; 63:407-410
- [14] Rubash H. E., Parvataneni H. K.: The Pants too Short, the Leg too Long: Leg Length Inequality after THA. *Orthopedics*. 2007; 30:764-765
- [15] Schmalzried T. P.: Preoperative Templating and Biomechanics in Total Hip Arthroplasty. *Orthopedics*. 2005; 28(8 suppl):s849-s851
- [16] Suh K. T., Cheon S. J., Kim D. W.: Comparison of Preoperative Templating with Postoperative Assessment in Cementless Total Hip Arthroplasty. *Acta Orthop. Scand*. 2004; 75:40-44
- [17] Hendrikus J. A. Crooijmans MSc, Armand M. R. P. Laumen MD, Carola van Pul PhD, Jan B. A. van Mourik MD: A New Digital Preoperative Planning Method for Total Hip Arthroplasties, *Clin Orthop Relat Res* (2009) 467:909-916, DOI 10.1007/s11999-008-0486-y
- [18] Kulkarni A., Partington P., Kelly D., S Muller.: Disc Calibration for Digital Templating in Hip Replacement, *Journal of Bone and Joint Surgery - British Volume*, Vol. 90-B, Issue 12, 1623-1626, doi: 10.1302/0301-620X.90B12.20238
- [19] Bono J., MD: Digital Templating in Total Hip Arthroplasty. *The Journal of Bone and Joint Surgery (American)* 2004; 86:118-122, 2004 The Journal of Bone and Joint Surgery, Inc.
- [20] Murzic J. W. M. D.; Glozman Z. B. S.; Lowe P. R. N.: The Accuracy of Digital (filmless) Templating in Total Hip Replacement. 72nd Annual Meeting of the American Academy of Orthopaedic Surgeons in Washington, DC, February 23-27, 2005
- [21] Wedemeyer C., Quitmann H., Xu J., Heep H., von Knoch M., Saxler G.: Digital Templating in Total Hip Arthroplasty with the Mayo Stem. *Arch Orthop Trauma Surg*. 2007 Nov 10 [xEpub ahead of print]
- [22] Woolson S. T.: Leg Length Equalization during Total Hip Replacement. *Orthopedics*. 1990; 13:17-21
- [23] Carter L. W., Stovall D. O., Young T. R.: Determination of Accuracy of Preoperative Templating of Noncemented Femoral Prostheses. *J. Arthroplasty* 1995; 10:507-13

- [24] González Della Valle A., Comba F., Taveras N., Salvati E. A.: The Utility and Precision of Analogue and Digital Preoperative Planning for Total Hip Arthroplasty. *Int. Orthop.* 2008; 32:289-294
- [25] González Della Valle A., Slullitel G., Piccaluga F., Salvati E. A.: The Precision and Usefulness of Preoperative Planning for Cemented and Hybrid Primary Total Hip Arthroplasty. *J. Arthroplasty.* 2005; 20:51-58
- [26] The B., Diercks R., van Ooijen P., Van Horn J. R.: Comparison of Analog and Digital Preoperative Planning in Total Hip and Knee Arthroplasty. *Acta Orthop.* 2005; 76:78-84
- [27] Viceconti M., Lattanzi R., Antonietti B., Paderni S., Olmi R., Sudanese A., Toni A.: CT-based Surgical Planning Software Improves the Accuracy of Total Hip Replacement Preoperative Planning. *Med Eng Phys.* 2003; 25:371-377
- [28] Davila J. A., Kransdorf M. J., Duffy G. P.: Surgical Planning of Total Hip Arthroplasty: Accuracy of Computer-assisted EndoMap Software in Predicting Component Size. *Skeletal Radiol.* 2006; 35:390-393
- [29] Hananouchi T., Sugano N., Nakamura N., Nishii T., Miki H., Yamamura M., Yoshikawa H.: Preoperative Templating of Femoral Components on Plain X-Rays: Rotational Evaluation with Synthetic X-Rays on ORTHODOC. *Arch Orthop Trauma Surg.* 2007; 127:381-385
- [30] Oddy M., Jones M., Pendegrass C., Pilling J., Wimhurst J.: Assessment of Reproducibility and Accuracy in Templating Hybrid Total Hip Arthroplasty using Digital Radiographs. *J. Bone Joint Surg Br.* 2006; 88:581-585
- [31] Sugano N., Ohzono K., Nishii T., Haraguchi K., Sakai T., Ochi T.: Computed-Tomography-based Computer Preoperative Planning for Total Hip Arthroplasty. *Comput Aided Surg.* 1998; 3:320-324
- [32] Viceconti M., Chiarini A., Testi D., Taddei F., Bordini B., Traina F., Toni A.: New Aspects and Approaches in Pre-Operative Planning of Hip Reconstruction: a Computer Simulation. *Langenbecks Arch Surg.* 2004; 389:400-404
- [33] Schuenke M., Schulte E., Schumacher U., Ross M. L., Lamperti D. E.: *Thieme Atlas of Anatomy (2006)* ISBN-10: 3131420812
- [34] Platzer W.: *Color Atlas of Human Anatomy, Vol. 1, Locomotor System.* 5th revised and enlarged English edition. Stuttgart, New York: Thieme; 2004 ISBN 3-13-533305-1
- [35] Conn, K. S.; Clarke, M. T.; Hallett, J. P.: A Simple Guide to Determine the Magnification of Radiographs and to Improve the Accuracy of Preoperative Templating, *J. Bone Joint Surg [xBr]* 2002; 84-B:269-72. Received 6 July 2001; Accepted after revision 7 August 2001

- [36] Eckrich S. G. J., Noble P. C., Tullos H. S.: Effect of Rotation on the Radiographic Appearance of the Femoral Canal. *J. Arthroplasty* 1994; 9:419-26
- [37] http://www.totaljoints.info/cemented_and_cementless_thr.htm#0
- [38] White, S. P.; Shardlow, D. L.: Effect of Introduction of Digital Radiographic Techniques on Pre-Operative Templating in Orthopaedic Practice, *Annals of The Royal College of Surgeons of England*, Volume 87, Number 1, January 2005, pp. 53-54(2), doi 10.1308/1478708051540
- [39] Schwartz J. T., Mayer J. G., Engh C. A.: Femoral Fracture during Noncemented Total Hip Arthroplasty. *J. Bone Joint Surg [xAm]* 1989; 71-A:1135-42
- [40] http://www.ortho-cad.com/b/Content/Technology_2_2.html
- [41] Michalíková M.: Riešenia tribologických vlastností totálnych náhrad bedrového kĺbu. Dizertačná práca. Košice: Technická univerzita, Strojnícka fakulta, 2009
- [42] http://www.rush.edu/rumc/images/ei_0244.gif
- [43] Wimsey S., Pickard R., Shaw G.: Accurate Scaling of Digital Radiographs of the Pelvis: a Prospective Trial of Two Methods. *J. Bone Joint Surg Br.* 2006; 88:1508-1512
- [44] Živčák J. a kol.: *Biomechanika človeka I.* ManaCon, Prešov, 2007, ISBN 978-80-89040-30-8

Remarks on the Efficiency of Information Systems

András Keszthelyi

Óbuda University, Budapest, Hungary
keszthelyi.andras@kgk.uni-obuda.hu

Abstract: In Hungary there exist two big and well-known scholar information systems (SIS). Both of them have become part of everyday life in the administration of higher education. These SISes have made quite a long journey during their evolution. Even so, they have some annoying disadvantages even today; e.g. they serve too few users at a time too slowly in the case of critical activities, such as during registration for courses or for examinations. There are many circumstances which can result in such poor efficiency. In this paper I try to investigate the role of three-level data modelling in an indirect way. I, with a colleague of mine at Budapest Tech, planned and executed a measurement which shows that (much) better performance can be reached based on a 'good' data model, even in a poorer environment.

Keywords: database efficiency; scholar information system

1 Scholar Information System - SIS

The administration of the scholastic records of students has become a great task for today, one which would need great resources if the administration were done in the historical manner, that is, using a paper-based system. This is due to not only the increasing number of students but also the complexity of the credit system as well.

At this point, it is necessary to clearly see the difference between a 'database' and an 'information system'.

A database is a finite amount of data stored in a suitable manner in order to manage the needed administrative tasks in the shortest time possible. There were no theoretical reasons for demanding that a 'database' should be computer-based, but for historical reasons, we do not use this expression for a paper-based filing cabinet. In a computer-based environment, a 'database' is a finite amount of data stored in a proper *structure* according to the data model. These data represent a) entities, b) the attributes of the entities, and c) the connections between the entities.

An 'information system' is more than the database itself. Of course the heart of an information system is a database. We need some or additional *infrastructure* and technical/administrative *people* to manage the database itself and to serve the administrative tasks or jobs and the users. We need some or more special *rules* in order to work in a correct way and to have the least possible number of errors as far as possible. These all can be called an 'information system'.

In the "old" times, the curriculum was well-defined, which was mandatory for all students in only one possible way. Nowadays, the curriculum is well defined, too, but there are a large number of possible ways to fulfill the requirements. Students themselves can decide the manner and timing of their studies. There is only one main rule, which is a logical one: a student is allowed to register for a course if the prerequisites of that course have been fulfilled; for example, one is not allowed to register for Mathematics II when he or she has not yet succeeded in Mathematics I.

According to the above, it is natural that the students select not only the subjects they want to study in a given semester, but they also personally choose one of the courses of that particular subject as well (if the number of the subscribed students to the given course is less than the maximum number allowed). Course selection can be a complex, iterative process until such time as the students are able to compile suitable timetables suitable for them. The registration for examinations has almost the same attributes as the above-mentioned registration for courses.

Storing the students' results in their different subjects is simpler.

So, managing the administration by hand can scarcely be imagined. Almost everywhere this task is solved by computer-based information systems. Such a system can be called a 'Scholar Information System'.

1.1 Main Questions about a SIS

There are many questions that can be asked in connection with an SIS. These can be grouped in many ways. From the point of view of the end-users, the main questions are the following:

Is the system able to manage a large number of administrative tasks simultaneously? Is it able to serve all or nearly all the students in a given time period who want to or are obliged to do a task in the administrative field. This problem can occur in two typical situations: the first one is the registration for examinations at the end of the semesters; the second one is the registration for courses at the beginning of the semesters.

Is the system realistic? Does it work in accordance with real life? Firstly, does it know and serve the administrative rules of a given institute? Secondly (last but not least), are these rules themselves realistic? Are they in accordance with the rules of logic; are they practical?

Is the system ergonomic? How much time is needed to perform a given task? How many mouse-clicks are needed to perform the most frequent activities?

Can the system handle all the personal and scholastic data of the students in a secure enough way?

In this paper, I examine the first question mentioned above, the load-ability of a database that could be that of a SIS in a typically problematic situation. I planned and executed a quantitative measurement in order to determine how many administrative tasks can be performed almost simultaneously. I have chosen one of the two most critical activities: the registration for examinations. According to our everyday experiences and to a student questionnaire, this task usually forms a bottleneck.

1.2 Existing SISEs

In Hungary there are two well-known scholar information systems that are used in higher education: the ETR (Egységes Tanulmányi Rendszer – the Uniform Scholar System) and Neptun.

We have been using Neptun for nearly a decade at Budapest Tech.

At the beginning there were serious problems even in the functionality, e.g. there was no possibility to administer if a student had a dispensation. The database was not able to cope with the load caused by registration periods, as it should have been expected. After nearly a decade, we are using the third main version of Neptun. Of course it has developed since then, but it has its own weaknesses in the field of load-ability even today. Our nearly ten thousand students are divided into different sets, and each set of students can start their registration on different days even today in order to lower the load on the database.

According to a student questionnaire created by this author in 2008, the most frequent problems observed by the students were: aborted connections, short timeout and slowness.

ETR has the same problems in efficiency, as can be read about even in the Hungarian-language wikipedia (<http://hu.wikipedia.org/wiki/ETR>).

These problems are widely known and these are the problems which make most people very angry in most cases. So it is reasonable to investigate the problem of the efficiency of information systems.

2 Efficiency of Databases

What can we call the efficiency of a database? It is the capability to cope with high loads. This capability is determined by several very different factors: the hardware environment, the software environment (the operating system, the relational database management system, the application programming language and tools, the application programs themselves), and the quality of the data model.

Of course, the influence of the hardware environment is very important. This is the first circumstance which comes into one's mind, but it must be declared that to increase the performance of the hardware in order to have a higher software performance is a 'brute force' method: the more money you have, the higher performance you will get.

There are more sophisticated and, of course, cheaper methods which result in higher software performance.

Let us look at the software environment. The operating system, the relational database management system and the application itself are the most important elements in this field. The first two can only be chosen from a given set based on various ratings of their most important technical and co-operational features. How some technical aspects influence the performance I investigated before and presented some years ago. [9]

In the case of the third element, the application, there are more possibilities to influence the performance. After choosing the programming language and tools, there are two main fields which determine the performance of the developed program(s). These are the quality of the applied algorithms and the quality of program coding. In the case of databases, the 'algorithm' has a more special meaning than generally: the quality of the data model is included as the most important; a necessary but not sufficient circumstance.

2.1 The Quality of Data Models

The main steps in developing an information system are: determining what is wanted as precisely as possible; data modelling (i.e. determining the data structure); and determining the functions to operate on the data structure. In the case of data-intensive systems, the data structure is more important and determines the functionality. [1] (p. 541)

So data modelling is the basis, one which is necessary but of course not sufficient for success. The basis is only a possibility on which a good information system can be constructed. In order to succeed, it is necessary to have three-level data modelling and planning, according to Halassy. [2] (pp. 28-33) These levels are the conceptual, the logical and the physical levels. The names of these levels, and even the "three-level" label, have been widely used, but in most cases without the

appropriate meaning. In the early times of databases, Codd wrote that even the SPARC committee of ANSI used these words without defining them precisely. “The definitions of the three levels supplied by the committee in a report were extremely imprecise, and therefore could be interpreted in numerous ways.” [3] (p. 33)

Unfortunately, the field of data modelling is not, and has never been, in focus. In previous times, at the beginning of relational databases, Peter Chen introduced his entity-relationship model [4], which can be considered the basis of data modelling. There are no books even today which discuss data modelling in a scientific way, except books by Dr. Halassy in the Hungarian language, and there are no books which discuss data modelling in its fullness. This is the second reason why I have investigated database efficiency and the role and importance of data modelling in this field.

As Dr. Halassy states, the conceptual level data model is the one in which are described the entities of reality, their properties and relations, or linkages in natural concepts and corresponding to reality. The logical level is the one where the data structure of the database is planned according to the circumstances and constraints of the technical aspects, accessibility and efficiency. Defining the exact type and size of the data elements, the way they are stored in the storage equipments, and the way they are accessed are described in the physical level plan.

There are general prerequisites of the quality of data models. At the conceptual level, a good data model needs to be understandable, unambiguous, realistic, full and *minimal*. [5] (p. 192) Of these properties, minimality is the one which can be precisely examined by mathematical tools.

Minimality is a very important property because redundancy is dangerous. If a data structure is redundant, the database built upon it needs (much) more storage. If it needs more storage, it will need more time to be handled. These problems can be solved by 'brute force', by quicker storage equipment and processors. The biggest danger of redundancy is the possibility of data errors: redundancy causes certain undesirable characteristics, the so called insertion, update, and deletion anomalies that can lead to the loss of data integrity. We can suppose that at least some of the experienced problems of the two SISes are rooted in model level errors. I am here focusing on efficiency.

I was considering whether the data model of the SIS used by Budapest Tech (Neptun) meets the above requirements. Of course I was not given the model documentation itself because it is a commercial software, so I had to try another, indirect way. I made a data model for such a scholar system, a data model which is considered to be good enough, at least by this author, to examine that one instead of the original one.

The prerequisite is that if I can reach better or at least not worse results in a poorer environment, I can state that the reason for the difference can be identified in the differences of the data models.

My concept was to identify a function which is critical from the aspects of the response time and of the number of concurrent users to be attended. There are two such functions in a scholar system: registration for examinations and registration for courses at the end and at the beginning of the semesters. In these two cases, nearly ten thousand students would like to use the system, and each one of the students needs to register for an average of about four examinations or about twelve courses.

I chose the first process, the registration for examinations. I made the conceptual data model carefully, as well as the logical and physical level plans based upon it. A colleague of mine implemented the plan and developed the part of the application which is needed to do some efficiency measurements. [7]

3 Questions about the Measurement

3.1 What to Measure?

I chose the registration for exams as a critical field to investigate, as was mentioned above. First it was necessary to decide what I wanted to measure in this field. The exact response time of each registration of each student? The number of retries and/or the response times? The number of successful and unsuccessful tries in a certain time-period? Do I need to make an ABC-assay and to rank the responses into three sets, one of them called 'very good', the other called 'acceptable' and the third one called 'poor'? At what values can I mark the boundaries of these sets?

3.2 Measurement Errors and Mistakes

There are numerous random factors which can and, of course, do influence the measuring. Let us look at at least some of them.

Since the measurement is done in a working computer environment, all the other possible activities of the operating system would be taken into account, e.g. saving data as a response to a given query in a local file needs some (a little but significantly greater than zero) time. This is a random error because the moving of the read-write heads of the hard disks and the puffer usage is unpredictable.

The network traffic which is not part of the measurement activities could be eliminated, closing the subnet for the time of the measurement, but even in this case, there are some factors at the ethernet level which could influence the measurement. This is a quasi-random error because it increases if the network traffic increases.

Last but not least, the measurement also influences itself. To measure some computer activities by computers needs one or more, more or less complex programs to run. These programs also need lesser or more resources, while the total amount of resources is a given constant.

Beyond the above-mentioned errors, there are observational and computational errors as well to cope with.

3.3 The Object of the Measurement

To measure the response times to four significant digits (in seconds) would be an interesting measurement task to plan and execute. In such a case, the correct handling of the above-mentioned measurement errors would not be an easy problem to solve.

Luckily it is not important to know the response times precisely. There are two important questions, which are the following: Could the response times be tolerated by the average student or not, even when a large number of students would like to be served? How many of the registration attempts are fulfilled?

In trying to answer these two questions, the influence of the above mentioned measurement errors are negligible. The borderline between the 'tolerable' and 'intolerable' time requisites cannot be defined precisely in a mathematical manner because it is the subjective opinion of the end-users, in this case the students.

Therefore, I decided to measure the average and the maximum response times and the successfulness of the registration attempts. If the response times and the number of the unsuccessful attempts are significantly lower than in the real system, even in a poorer environment, my above statement could be proven: a better three-level data modelling results in a better database application, in better response times, and in fewer unsuccessful attempts. In general: a correct three-level data modelling results in a growth in efficiency.

4 The Measurement

The test environment consisted of PC computers and free software. The database server had an Intel processor of four cores and 8 GB of RAM. It runs Linux as the operating system, httpd server Apache with PHP as an application interface between the users and the database, and MySQL as a relational database management system. Instead of a large number of workstations one simple but strong PC was used, one which had enough RAM to run the needed amount of offline browser **wget**. This circumstance has no effect on the measurement: the number of registrations are the same and each of them goes through the network.

The test database contained 8192 students, about as many as BMF's active students, four examinations for each of them. The number of places was one and a half times bigger than the number of the students' examinations. Registrations themselves were made by a PHP script `index.html` randomly for the test user currently called it via the offline browser of the test client workstation. Each test user registered a date for all examinations.

Normally the selection is done by the students themselves, sitting and thinking in front of their computer. From the point of view of the measurement, there was no difference between a date selection by a human student and a random date selection by a program. The circumstance that this date selection is done at server side increases the load on the server a bit, so the measured results are a bit worse than they would be in reality.

We logged the client system time at the beginning of the connection to the database server and when the response was saved to a local file by the offline browser. The server load was also logged. Test registrations were started almost simultaneously with a two second pause after every one hundred starts. The settings of the offline browser `wget` were: max 4 retries, 30 seconds timeout, 10 to 30 seconds between two retries.

The test environment and application is described more precisely in [7], [8/a].

4.1 The Measured Results

The results were better than I had expected.

All the registrations of all the students were successful.

The total time needed for the 32.768 registrations of the 8.192 students was 3 minutes and 7 seconds, so the average time needed for a test student was 0.0228 second, with a maximum value of 1 (one) second because the offline browser `wget` logs its activities in `hh:mm:ss` format, so fractions of seconds cannot be taken into account.

The maximum value for the server load (1 min load) was 2.85, with an interesting, staircase-of-staircases like diagram as described in [7].

These values are quite good compared to real life experience. Even if the measured values were bigger by a whole order of magnitude (i.e. the 32k registration of the 8k students needed about half an hour) they would be good enough to prove my statement. Therefore, I can state that we have the possibility to develop much more efficient scholar information systems.

Conclusions and Two Open Questions

Summarizing the above, I can state that (much) better results can be achieved based on a 'good' three-level data modelling even in poorer hardware and/or

software circumstances. The quality of the data model influences the quality and the efficiency of the database to such an extent that precise three-level data modelling, according to Dr. Halassy [6] (p. 32), ought to be more important than it is generally considered at present. The quality of the data model is an important variable, if not the most important one. There is no other way to produce good information systems. System development methods and standards alone are not enough for that.

We have been using the two big SISes in higher education for about a decade. So I have two open questions at the end.

The first: Are the faculties of computer sciences in the country of John von Neumann able to, want to and dare to develop a better system? The second: Why has the first question never been asked?

References

- [1] Raffai Mária dr.: Információrendszerek fejlesztése és menedzselése. Novadat Bt., 2003
- [2] Halassy Béla dr.: Adatmodellezés. Nemzeti Tankönyvkiadó Rt., 2002
- [3] Codd Edgar Frank: The Relational Model for Database Management - version 2. Addison-Wesley Publishing Company, 1990
- [4] Chen P.: The Entity-Relationship Model -- Toward a Unified View of Data. In: ACM Transactions on Database Systems (TODS), 1976. március, I. évf. 1. szám
- [5] Halassy Béla dr.: Az adatbázisstervezés alapjai és titkai. IDG Magyarországi Lapkiadó Kft., 1995
- [6] Halassy Béla dr.: Ember - információ - rendszer. IDG Magyarországi Lapkiadó Vállalat, 1996
- [7] Szikora Péter: Measured Performance of an Information System. 7th International Conference on Management, Enterprise and Benchmarking, Budapest, 2009
- [8a] Szikora Péter: The Role of the Tools and Methods of Implementation in Information System Efficiency. 2nd International Conference for Theory and Practice in Education, Budapest, 2009
- [8b] Keszthelyi András: The Role of Data Modeling in Information System Efficiency. 2nd International Conference for Theory and Practice in Education, Budapest, 2009
- [9] Keszthelyi András: Information management in the higher education -- the role and importance of the different technologies. 3rd International Conference on Management, Enterprise and Benchmarking, Budapest, 2005

Six-Phase Matrix Converter Fed Double Star Induction Motor

Bachir Ghalem, Bendiabdellah Azeddine

LDEE laboratory, University of Sciences and the Technology of Oran (USTO)
BP 1505 EL M' naouer Oran, Algeria, bachir@univ-usto.dz

Abstract: two different control strategies applied to a direct AC-AC six-phase matrix converter are investigated in the present paper. The first strategy is derived from the Venturini method, and the second approach is practically an extension of the scalar strategy control. Both strategies were originally applied to the three-phase matrix converter.

The current investigation deals with a comparative performance study of a double star induction motor fed from a six-phase matrix converter using the two above control strategies.

After a theoretical introduction of the six-phase matrix converter, a detailed description of the strategies implementation is presented followed by a discussion of results obtained from simulation results.

Keywords: six-phase; matrix converter; modulation coefficients; double star induction motor

1 Introduction

Over the last 20 years induction machines with a double star have been used in many industrial applications due to their performances in high power fields[4]. The double star induction machine requires a double three-phase supply and has many advantages. Not only does give reduced torque oscillations, but it also requires less powerful electronic components as the current flowing in a six-phase machine is less than that flowing in a three-phase machine. However, as the use of an inverter is necessary when feeding a double star induction machine, this may result in supplementary losses, since such an inverter is a harmonic generator[5].

A six-phase matrix converter can be used as a source for these kind machines, but the main problem is in finding the appropriate control method which produces a six-phase system with low electromagnetic torque oscillations.

2 Theory of the Matrix Converter

A matrix converter is a direct power converter, generating variable amplitude voltage and frequency from a rigid entry. An intermediate DC-link circuit is not necessary as in the classical inverter case. The principle of such a converter is based on a matrix topology connecting each input phase with each output phase by a two-way power switch, allowing the flow of power in both directions and therefore operating in all four quadrants. By using some well suited laws, the matrix converter can reproduce all the existing electronic power conversions (AC / AC, AC / DC, DC / DC and DC / AC).

The basic diagram of a matrix converter can be represented by Figure 1.

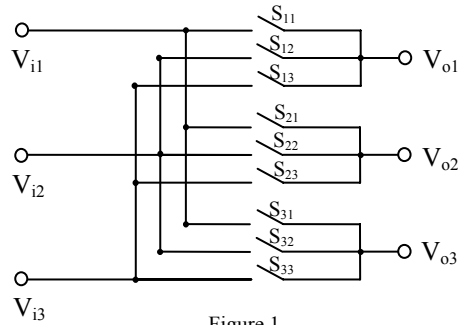


Figure 1

Basic circuit of a matrix converter

The symbol S_{ij} represents the ideal bidirectional switches, where i represents the index of the output voltage and j represents the index of the input voltage.

Let $[V_i]$ be the vector of the input voltages given as:

$$[V_i] = V_{im} \begin{bmatrix} \cos(\omega_i t) \\ \cos(\omega_i t - 2\pi/3) \\ \cos(\omega_i t - 4\pi/3) \end{bmatrix} \quad (1)$$

and $[V_o]$ the vector of desired output voltages.

$$[V_o] = V_{om} \begin{bmatrix} \cos(\omega_o t) \\ \cos(\omega_o t - 2\pi/3) \\ \cos(\omega_o t - 4\pi/3) \end{bmatrix} \quad (2)$$

The problem now consists in finding a matrix $[M]$ known as the modulation matrix, such that

$$[V_o] = [M] \cdot [V_i] \quad (3)$$

The development of the equation (3) gives:

$$\begin{bmatrix} V_{o1} \\ V_{o2} \\ V_{o3} \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \begin{bmatrix} V_{i1} \\ V_{i2} \\ V_{i3} \end{bmatrix} \quad (4)$$

where m_{ij} are the modulation coefficients.

During commutation, the bidirectional switches must function according to the following rules:

- At every instant t , only one switch S_{ij} ($i = 1,2,3$) works in order to avoid short-circuit between the phase.
- At every instant t , at least two switches S_{ij} ($j = 1,2,3$) works to ensure a closed loop load current.
- The switching frequency $f_s = \omega_s / 2\pi$ must have a value twenty times higher than the maximum of f_i, f_o ($f_s \gg 20 \cdot \max(f_i, f_o)$).
- During the period T_s known as the sequential period, which is equal to $1/f_s$. The sum of the time of conduction being used to synthesize the same output phase must be equal to T_s .

Now a time t_{ij} , called the modulation time, can be defined as:

$$t_{ij} = m_{ij} \cdot T_s \quad (5)$$

3 Theory of a Six-Phase Matrix Converter

The basic scheme of a six-phase matrix converter can be illustrated by Figure 2.

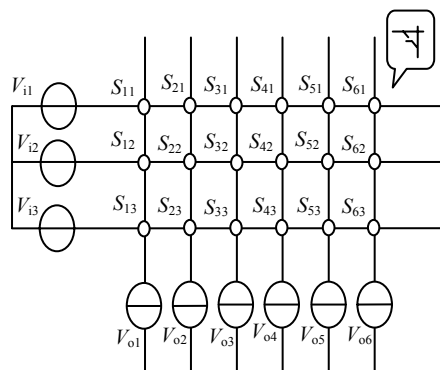


Figure 2

Basic scheme of a six-phase matrix converter

3.1 First Method

This method is derived from that proposed by Venturini [1], [2], [4] and has been used to solve the equation (3), from which a six-phase voltage system at the matrix converter output can be similarly obtained and represented by equation (6).

$$\begin{bmatrix} V_{o1} \\ V_{o2} \\ V_{o3} \\ V_{o4} \\ V_{o5} \\ V_{o6} \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \\ m_{41} & m_{42} & m_{44} \\ m_{51} & m_{52} & m_{53} \\ m_{61} & m_{62} & m_{63} \end{bmatrix} \begin{bmatrix} V_{i1} \\ V_{i2} \\ V_{i3} \end{bmatrix} \quad (6)$$

The calculation of the m_{ij} coefficients in equation (6), will be the same as that for the m_{ij} coefficients for the case of a three phase[1], but expressed as follows:

$$[M] = \begin{bmatrix} m(3,3) & m(4,4) & m(5,5) \\ m(2,4) & m(3,5) & m(4,0) \\ m(1,5) & m(2,0) & m(3,1) \\ m(0,0) & m(1,1) & m(2,2) \\ m(5,1) & m(0,2) & m(1,3) \\ m(4,2) & m(5,3) & m(0,4) \end{bmatrix} \quad (7)$$

where the matrix coefficients are given in a general form as:

$$m(x_1, x_2) = \frac{1}{3} \alpha_1 [1 + 2QZ_1^{-1}(x_1)] + \alpha_2 [1 + 2QZ_1^{-1}(x_2)] \quad (8)$$

$$\text{where } Q = \frac{V_{om}}{V_{im}} \text{ and } Z_\alpha^\beta(y)(t) = \cos[(\alpha\omega_i + \beta\omega_o)t + y\frac{\pi}{3}]$$

$$\text{and } \alpha_1 = \frac{1}{2} \left(1 + \frac{\tan(\theta_i)}{\tan(\theta_o)} \right), \quad \alpha_2 = 1 - \alpha_1$$

with θ_i the phase of input voltages and θ_o the phase of output voltages.

As an example, the average value of the sixth output phase voltage during the k^{th} sequence can be expressed as:

$$V_{o6} = \begin{cases} V_{i1} & 0 \leq t - (k-1)T_s \leq m_{61}^k T_s \\ V_{i2} & m_{61}^k T_s \leq t - (k-1)T_s \leq (m_{61}^k + m_{62}^k) T_s \\ V_{i3} & (m_{61}^k + m_{62}^k) T_s \leq t - (k-1)T_s \leq (m_{61}^k + m_{62}^k + m_{63}^k) T_s \end{cases} \quad (9)$$

3.2 Second Method

The second method is derived from the scalar strategy control [3], [4]. A straightforward approach to generate the active and zero states of matrix switches in fig.1 consists of using the instantaneous voltage ratio of specific input phase voltages. Let us define the following phase voltages present at the input port as:

$$\begin{cases} V_A(t) = V_{im} \cos(\omega_i t) \\ V_B(t) = V_{im} \cos(\omega_i t - 2\frac{\pi}{3}) \\ V_C(t) = V_{im} \cos(\omega_i t - 4\frac{\pi}{3}) \end{cases} \quad (10)$$

At the output port of the converter, the value of any instantaneous output phase voltage may be expressed by equation (11) as follows:

$$V_o(t) = \frac{1}{T_S} (t_K V_K + t_L V_L + t_M V_M) \quad (11)$$

where $t_K + t_L + t_M = T_S$

K-L-M are variable subscripts, which may be assigned the variable A, B or C according to the following rules:

Rule 1: At any instant, the input phase voltage which has a polarity different from both others is assigned to “M”.

Rule 2: The two input phase voltages which share the same polarity, are assigned to K and L, the smallest one of the two, (in absolute value), being “K”. Then t_K and t_L are chosen such that:

$$\frac{t_K}{t_L} = \frac{V_K}{V_L} = \rho_{KL} \quad (12)$$

for the interval where

$$0 \leq \frac{V_K}{V_L} \leq 1 \quad (13)$$

The expressions given by equation (11) and equation (13) are similar to that ones originally proposed by [1]. Equation (12) defines the active time ratio between two out of the three switches, in one commutating leg of the output port (see Fig. 1). This time ratio (t_K/t_L) is proportional to the instantaneous voltage ratio (V_K/V_L) of their associated input phases. The ratio must be established with the smaller instantaneous voltage divided by the larger one, as stated in equation (13).

The converter switching pattern depends only on the scalar comparison of input phase voltages and the instantaneous value (V_o) of the desired output voltage. In

the following, the proper procedure to obtain the respective values of t_K , t_L and t_M during one period T_s of the sequence (or the carrier) frequency f_s is illustrated. For a specific interval where $0 \leq t_K/t_L \leq 1$.

The active times of the three switches associated with the desired output voltage V_o become:

$$t_K = \frac{T_s(V_o - V_M)}{(\rho_{KL}V_K + V_L - (1 + \rho_{KL})V_M)} \quad (14)$$

$$t_M = T_s(1 + \rho_{KL})V_L \quad (15)$$

Using again the current value of ρ_{KL} , equation (14) can be further developed such as:

$$t_L = \frac{T_s(V_o - V_K)V_L}{(V_K^2 + V_L^2 + V_M^2) + (V_K + V_L + V_M)V_M} \quad (16)$$

In a balanced three phase system, the summation of the three instantaneous phase voltage is zero. So the following relationships can be obtained:

$$\frac{t_L}{T_s} = \frac{T_s(V_o - V_K)V_L}{(V_K^2 + V_L^2 + V_M^2)} = \frac{T_s(V_o - V_K)V_L}{1.5V_{im}^2} \quad (17)$$

$$\frac{t_K}{T_s} = \frac{T_s(V_o - V_K)V_K}{1.5V_{im}^2} \quad (18)$$

$$\frac{t_M}{T_s} = 1 - \frac{(t_K - t_L)}{1.5V_{im}^2} \quad (19)$$

The duty cycle of commutators K and L is proportional to the instantaneous value of the corresponding input phase voltage V_K and V_L multiplied by the voltage difference between the desired output voltage V_o and the input phase voltage V_M . It should be noted at this point that the output voltage V_o , (i.e V_A , V_B , V_C), can be any kind of waveform, including DC values.

Solving equations (17, 18 and 19) for a given voltage ratio, $V_{om}/V_{im} = Q \leq 0.5$, will yield a positive value for the times t_K , t_L and t_M as in the case of the Venturini control algorithm.

The desired output voltages with which the coefficients are calculated are:

$$\begin{cases} V_{o1} = V_{om} \cos(\omega_o t) \\ V_{o2} = V_{om} \cos(\omega_o t - \frac{\pi}{3}) \\ V_{o3} = V_{om} \cos(\omega_o t - 2\frac{\pi}{3}) \\ V_{o4} = V_{om} \cos(\omega_o t - 3\frac{\pi}{3}) \\ V_{o5} = V_{om} \cos(\omega_o t - 4\frac{\pi}{3}) \\ V_{o6} = V_{om} \cos(\omega_o t - 5\frac{\pi}{3}) \end{cases} \quad (20)$$

The development of equation (3) for the case of a six-phase matrix converter gives:

$$\begin{bmatrix} V_{o1} \\ V_{o2} \\ V_{o3} \\ V_{o4} \\ V_{o5} \\ V_{o6} \end{bmatrix} = \begin{bmatrix} m_{1K} & m_{1L} & m_{1M} \\ m_{2K} & m_{2L} & m_{2M} \\ m_{3K} & m_{3L} & m_{3M} \\ m_{4K} & m_{4L} & m_{4M} \\ m_{5K} & m_{5L} & m_{5M} \\ m_{6K} & m_{6L} & m_{6M} \end{bmatrix} \begin{bmatrix} V_K \\ V_L \\ V_M \end{bmatrix} \quad (21)$$

The average value of the voltage of the sixth output phase voltage during the K^{th} sequence is given as:

$$V_{o6} = \begin{cases} V_K & 0 \leq t - (k-1)T_s \leq m_{6K}^k T_s \\ V_L & m_{6K}^k T_s \leq t - (k-1)T_s \leq (m_{6K}^k + m_{6L}^k) T_s \\ V_M & (m_{6K}^k + m_{6L}^k) T_s \leq t - (k-1)T_s \leq (m_{6K}^k + m_{6L}^k + m_{6M}^k) T_s \end{cases} \quad (22)$$

4 Double Star Induction Machine Modelling

The mathematical model of the machine is written as a set of state equations, both for the electrical and mechanical parts [6], [8], [9]:

$$[V_{abc,s_1}] = [R_{s_1}][I_{abc,s_1}] + \frac{d[\Phi_{abc,s_1}]}{dt} \quad (23)$$

$$[V_{abc,s_2}] = [R_{s_2}][I_{abc,s_2}] + \frac{d[\Phi_{abc,s_2}]}{dt} \quad (24)$$

$$[V_{abc,r}] = [R_r][I_{abc,r}] + \frac{d[\Phi_{abc,r}]}{dt} \quad (25)$$

$$J \frac{d\Omega}{dt} = T_{em} - T_r - k_f \Omega \quad (26)$$

where J is the moment of inertia of the revolving parts, K_f is the coefficient of viscous friction, arising from the bearings and the air flowing over the motor, and T_{em} is the load torque.

The electrical state variables in the “dq” system are the flux represented by vector $[\Phi]$, while the input variable in the “dq” system are expressed by vector $[V]$.

$$\frac{d\Phi}{dt} = [A][\Phi] + [B][V] \quad (27)$$

$$[\Phi] = \begin{bmatrix} \varphi_{ds_1} \\ \varphi_{ds_2} \\ \varphi_{qs_1} \\ \varphi_{qs_2} \\ \varphi_{dr} \\ \varphi_{qr} \end{bmatrix} ; \quad [V] = \begin{bmatrix} V_{ds_1} \\ V_{ds_2} \\ V_{qs_1} \\ V_{qs_2} \end{bmatrix} \quad (28)$$

The equation of the electromagnetic torque is given

$$T_{em} = p \frac{L_m}{(L_m + L_r)} (\varphi_{dr} (i_{qs_1} + i_{qs_2}) + \varphi_{qr} (i_{ds_1} + i_{ds_2})) \quad (29)$$

The equations of flux are:

$$\Phi_{md} = L_s \left(\frac{\Phi_{ds_1}}{L_{s_1}} + \frac{\Phi_{ds_2}}{L_{s_2}} + \frac{\Phi_{dr}}{L_r} \right) \quad (30)$$

$$\Phi_{mq} = L_s \left(\frac{\Phi_{qs_1}}{L_{s_1}} + \frac{\Phi_{qs_2}}{L_{s_2}} + \frac{\Phi_{qr}}{L_r} \right) \quad (31)$$

Given that the “dq” axes are fixed in the synchronous rotating coordinates system, we have:

$$[A] = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} \end{bmatrix} \quad (32)$$

$$[B] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (33)$$

Where:

$$\begin{aligned}
 a_{11} &= a_{32} = \frac{R_{s1}L_a}{L_{s1}^2} - \frac{R_{s1}}{L_{s1}}; & a_{12} &= a_{34} = \frac{R_{s1}L_a}{L_{s1}L_{s2}}; & a_{56} &= -a_{65} = \omega_{gl} \\
 a_{11} &= a_{32} = -a_{31} = -a_{42} = \omega_s; & a_{14} &= a_{16} = a_{23} = a_{26} = a_{32} = a_{35} = 0; \\
 a_{41} &= a_{45} = a_{52} = a_{54} = a_{61} = a_{62} = 0; & a_{15} &= a_{36} = \frac{R_{s1}L_a}{L_{s1}L_r}; \\
 a_{21} &= a_{42} = \frac{R_{s2}L_a}{L_{s1}L_{s2}}; & a_{22} &= a_{44} = \frac{R_{s2}L_a}{L_{s2}^2} - \frac{R_{s1}}{L_{s1}} & a_{25} &= a_{46} = \frac{R_{s2}L_a}{L_rL_{s2}}; \\
 a_{51} &= a_{62} = \frac{R_rL_a}{L_{s1}L_r}; & a_{52} &= a_{64} = \frac{R_rL_a}{L_{s2}L_r} & a_{55} &= a_{66} = \frac{R_rL_a}{L_r^2} - \frac{R_r}{L_r};
 \end{aligned}$$

5 Simulations and Results Discussion

The simulation was carried out by keeping the supply voltage of the induction motor (i.e the output of the matrix converter) fixed and varying only the frequency f_o in order to be able to compare the motor performance for both strategies presented before. The six-phase matrix converter presented is being simulated for a desired output frequency of $f_o=50$ Hz, with a switching frequency of $f_s = 5$ KHz. Both converters are first feeding a passive R-L load ($R_s=20 \Omega$ and $L_r=0.04$ H) and then a 50 HP, 460 V double star induction motor driving a 100 N.m resistive torque.

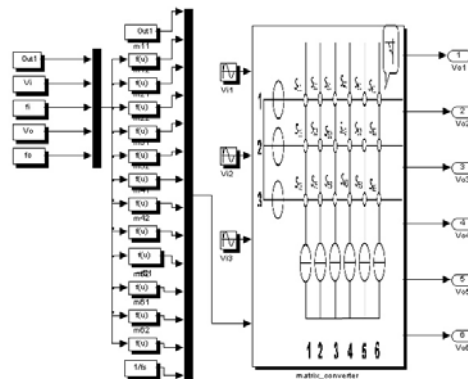


Figure 3

The matrix converter simulink®/Matlab diagram (Venturini method)

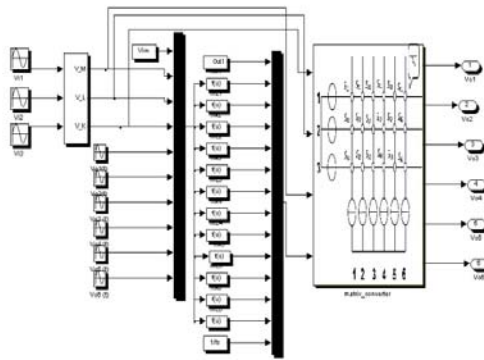


Figure 4

The matrix converter simulink®/Matlab diagram (Scalar strategy control)

5.1 Results Discussion

5.1.1 Results of the First Method ($f_o = 50$ Hz)

Figure 5 shows the matrix converter output voltage and its spectral analysis. The fundamental harmonic lies at the 50 Hz desired frequency and the higher order harmonics are in the neighbourhood of the 5 KHz switching frequency. The current form in a one phase of the R - L load (Figure 6), shows clearly that it approximates a sine wave shape at the desired frequency (50 Hz). It can be noted that the current shape in the case of a six-phase load illustrated in Figure 7 shows that the system is an unbalanced one.

Figure 8 represents a one phase stator current wave form. One can notice the presence of high oscillatory current, resulting from the large electromagnetic torque fluctuations as shows in Figure 10.

The variation of the rotor speed given in Figure 9, shows that its magnitude is not only slightly less than the rotating field speed, but also its shape is an oscillatory one.

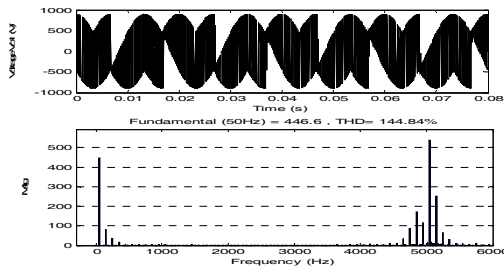


Figure 5

Voltage output and its FFT spectrum

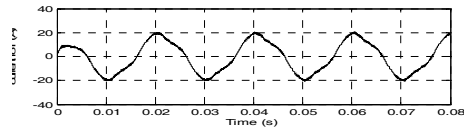


Figure 6

One phase R-L load current wave form

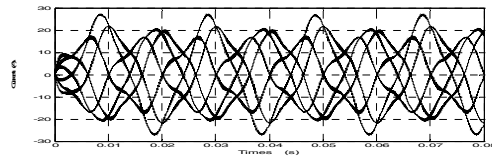


Figure 7

Six-phase R-L load current wave form

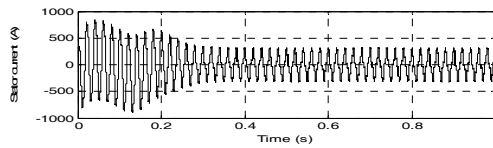


Figure 8

One phase stator current variation

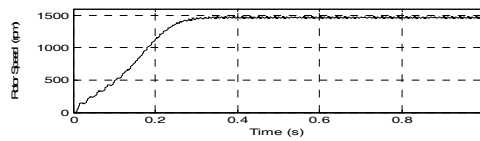


Figure 9

Rotor speed variation

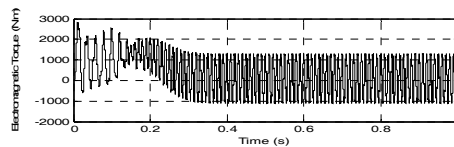


Figure 10

Electromagnetic torque variation

5.1.2 Results of the Second Method ($f_o = 50$ Hz)

Figure 11 represents the matrix converter voltage output and its spectral analysis. It shows a fundamental harmonic lying at the desired frequency (50 Hz) and higher order around the 5 KHz switching frequency.

Figure 12 represents the current form in a phase of the R - L load. This wave form is approaching the sine wave and is at the desired frequency of 50 Hz.

Figure 13 represents the current form in six-phase R-L load. It is visible that it represents a balanced six-phase system.

Figure 14 represents the one phase stator current wave form. In this case, the high oscillations have disappeared leading to a normal operating double star induction machine.

Figure 16 represents the electromagnetic torque wave form. This torque contains very reduced oscillations as that obtained from a balanced voltage system supplying the double star induction machine.

Figure 15 represents the rotor speed variation. This speed is slightly less than the rotating field speed.

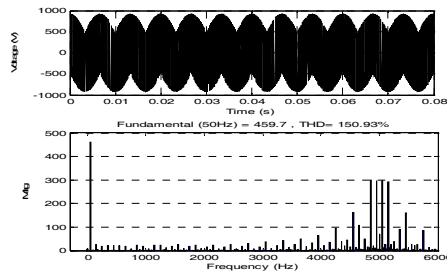


Figure 11

Voltage output and its FFT spectrum

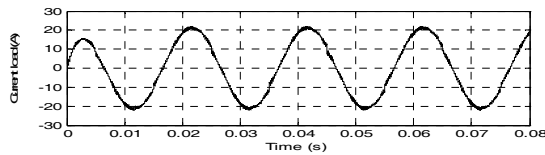


Figure 12

One phase R-L load current wave form

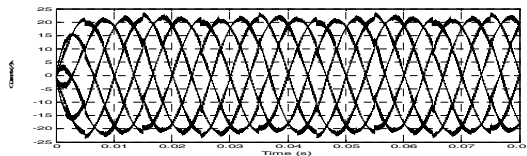


Figure 13

Six-phase R-L load current wave form

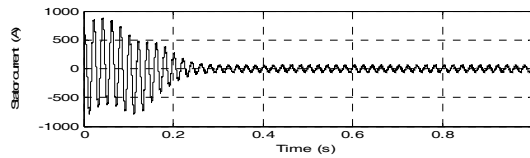


Figure 14

One phase stator current variation

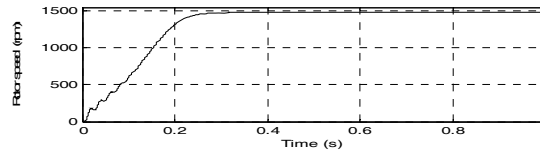


Figure 15

Rotor speed variation

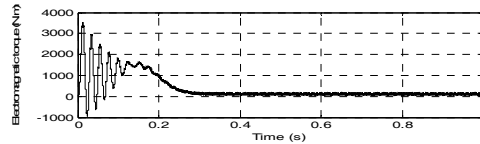


Figure 16

Electromagnetic torque variation

Conclusions

In the present paper, a comparative performance study between two different control strategies applied to a six-phase matrix converter is presented. The first strategy is derived from the Venturini method, and the second one is practically an extension of the scalar strategy control.

This comparative investigation is carried out using first an R-L load and then a double star induction machine. The obtained simulation results have been very useful and helpful in illustrating the merits of each method, though both strategies produce a six-phase voltage system with a load current at the 50 Hz desired frequency.

However, the Venturini method is simple and easy to implement. Unfortunately, the nature of the obtained voltage system is an unbalanced one and needs to be corrected and worked on to produce a balanced one. because an unbalanced voltage system has a direct impact on the motor's performance, giving important electromagnetic torque oscillations and high oscillatory currents in the steady state mode. On the other hand, the scalar strategy is more complex and produces, for the same magnitude, a balanced six-phase voltage system compared to the Venturini method.

References

- [1] Alesina A. and Venturini, M.; "Intrinsic Amplitude Limits and Optimum Design of 9-Switches Direct PWM AC-AC Converters" Power Electronics Specialists Conference, 1988. PESC '88 Record, 19th Annual IEEE, 11-14 Apr. 1988, pp. 1284 -1291 Vol. 2
- [2] G. Roy, G. E. April, "Cycloconverter Operation under a New Scalar Control Algorithm," in *Proc. IEEE PESC '89*, 1989, pp. 368-375
- [3] Bendiabdellah, A., Bachir, G.: "A Comparative Performance Study between a Matrix Converter and a Three-Level Inverter Fed Induction Motor", *Acta Electrotechnica Et Informatica*, No. 2, Vol. 6, 2006
- [4] Ghalem, Bachir; Azeddine Bendiabdallah; "A Comparative Study between Two Control Strategies for Matrix Converter" *Advances in Electrical and Computer Engineering Journal*, Vol. 9, No. 2, 2009, pp. 23-27
- [5] Bachir, G., Bendiabdellah, A.: "Scalar Control for a Matrix Converter" *Acta Electrotechnica Et Informatica*, No. 2, Vol. 9, 2009
- [6] Casadei, D.; Serra, G.; Tani, A.; Zarri, L.; "A Novel Modulation Strategy for Matrix Converters with Reduced Switching Frequency Based
- [7] D. Hadiouche, H. Razik, A. Rezzoug, "Stady and Star Simulation of Space Vector PWM Control of Double-Induction Motors", 2000 IEEE-CIEP, Acapulco, Mexico, pp. 42-47
- [8] J-P. Martin, E. Semail, S. Pierfederici, A. Bouscayrol, F. Meibody-Tabar et B. Davat." Space Vector Control of 5-Phase PMSM Supplied by 5 H-Bridge VSIs". Conference on Modeling and Simulation of Electric Machines, Converters and Systems (ElectrIMACS'02), Montreal (Canada), 2002
- [9] Singh, G. K, Pant, V, Singh, Y. P "Voltage Source Inverter Driven Multi-Phase Induction Machine" *Computers and Electrical Engineering* 29, pp. 813-834, 2003
- [10] Lipo, T. A, "A dq Model of a Six-Phase Induction Machine," *Int. conf. of Electrical Machines, ICEM, Athenes*, pp. 860-867, 1980
- [11] Incze, I. I. Szabo, Cs. Imecs, M. "Voltage-Hertz Strategy for Synchronous Motor with Controlled Exciting Field", *Intelligent Engineering Systems*, 2007, INES 2007, 11th International Conference on, pp. 247-252