Automatic Adaptation of Fuzzy Controllers

Ján Vaščák, Ladislav Madarász

Department of Cybernetics and Artificial Intelligence, Faculty of Electrical
Engineering and Informatics, Technical University of Košice
Letná 9, 041 20 Košice, Slovakia
Jan.Vascak@tuke.sk, Ladislav.Madarasz@tuke.sk

Abstract: The main drawback of 'classical' fuzzy systems is the inability to design and maintain their database. To overcome this disadvantage many types of extensions adding the adaptivity property to those systems were designed. This paper deals with two of them: an improved the so-called self-organizing fuzzy logic controller designed by Procyk and Mamdani as well as a new hybrid adaptation structure, called gradient-incremental adaptive fuzzy controller connecting gradient-descent methods with the first type. Both types of adaptive fuzzy controllers are shown on design of an automatic pilot and control of LEGO robots. The results and comparison to a 'classical' (non-adaptive) fuzzy controller designed by a human operator are also shown here.

Keywords: Fuzzy adaptive controller, Gradient-descent methods, Jacobian, Gradient-incremental adaptation

1 Introduction

Fuzzy logic has found many successful applications, especially in the area of control, but there are some limits of its use that are connected with the inability of the knowledge acquisition and adaptation to changed external conditions or parameters of the controlled system. To overcome this problem there were published lots of papers, e.g. [1, 5, 10, 11], which deal with structures of **Adaptive Fuzzy Controllers** (AFC) using mostly approaches based on many variations of gradient-descent methods, the least square method [8], linear and non-linear regression or the linguistically based rule extraction (e.g. [13, 14]).

Further, we will focus our attention only on 'pure' AFC. The main reason why to deal with this type of AFC is that they are with their nature and calculus the most similar systems to the non-adaptive (classical) FC. The properties of FC are well known, more than in the case of neural networks or genetic algorithms, in general. Fuzzy logic is able to simulate the human vague thinking very efficiently [18] and therefore it seems to be very advantageous only to add the ability of the

knowledge acquisition to 'classical' fuzzy systems and to preserve their properties, too.

In this paper we will describe and analyze the so-called **Self-Organizing Fuzzy Logic Controller** (SOFLC) proposed by Procyk and Mamdani [11], which was modified in many papers, e.g. [7, 8]. Further, its modification and implementation will be shown on two examples: automatic pilot and control of LEGO robots. Results of experiments with these systems are summarized in the concluding part of this paper.

2   Structure of SOFLC

SOFLC (see Fig. 1) belongs to the so-called *performance-adaptive controllers,* which evaluate the control quality by a criterion (or more criteria) like transition time, energy consumption, overshoots, etc. Such a quality measure is called **performance measure** *p(k).*
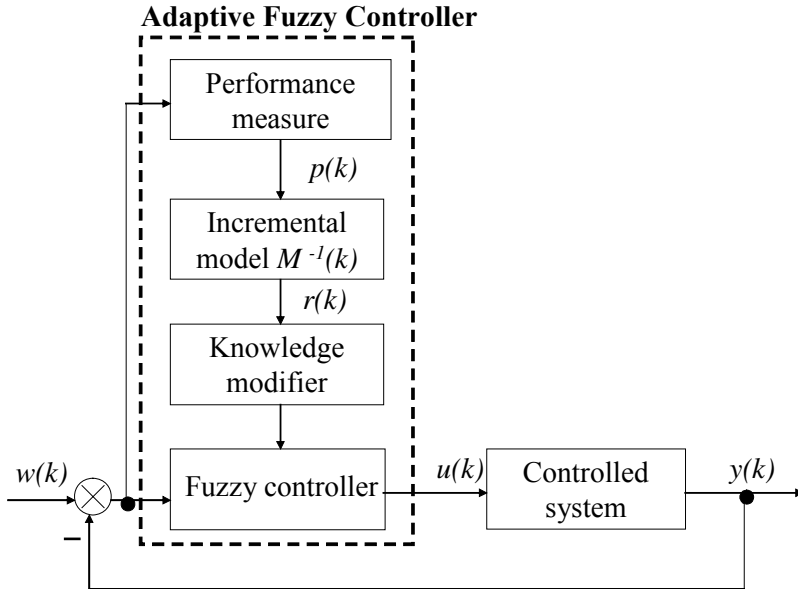

Figure 1

Structure of a self-organizing fuzzy logic controller

Control criteria are contented in the *performance measure block,* where the quality is evaluated by *p(k)*, which expresses the magnitude and direction of changes to be performed in the knowledge base of the controller. The basic design problem of AFC consists in the design of *M*, where for each time sample $t=K.T$ ($K=0,1, …$) a

simplified **incremental model** of the controlled system $M=J.T$ (J – Jacobian, T – sampling period) is computed. It represents a supplement to the original model and is analogous to the linear approximation of the first order differential equation or in other words to gradients, too. As Jacobian (1) is a determinant of all first derivatives of the system with n equations $f_1, …, f_n$ of n input variables $x_1, …, x_n$ it means J is equal to the determinant of the dynamics matrix, i.e. it is a numerical value describing all n gradients in the sense of a characteristic value:

$$J = \begin{vmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \dots & \dfrac{\partial f_1}{\partial x_n} \\ & & \dots & \\ \dfrac{\partial f_n}{\partial x_1} & \dfrac{\partial f_n}{\partial x_2} & \dots & \dfrac{\partial f_n}{\partial x_n} \end{vmatrix} \tag{1}$$

Now we need to transform this incremental description of a controlled system to the description of a controlling system, i.e. a controller. Considering the properties of the feed back connection we can see that $y(k) \approx e(k)$ ($w(k)$ is known). As inputs and outputs of a controlled system change to outputs and inputs of a controller, respectively we can get the controller description like the inverse function of $y(k) = f_M(u(k))$, i.e. the model of the controller is in the form $u(k) = f^{-1}{}_M(y(k))$. Because J is a number, then M^{-1} is the reverse value of $J.T$. The **reinforcement value** $r(k)$ is computed as $r(k) = M^{-1}.p(k)$ and represents the correction of the knowledge base.

Let the knowledge base in the time step k of such AFC be $R(k)$ and let its modification in next time step be $R(k+1)$. The general adaptation rule can be described like:

$$R(k+1) = (R(k) \cap \overline{R_{bad}}(k)) \cup R_{new}(k) \tag{2}$$

We see, firstly, the part of knowledge $R_{bad}(k)$ that caused the low quality control is removed from $R(k)$ and then it is completed by new knowledge R_{new}, which is corrected by $r(k)$. $R_{bad}(k)$ and R_{new} are computed as follows:

$$R_{bad}(k) = fuzz(x_1^*(k)) \, x \dots x \, fuzz(x_n^*(k)) \, x \, fuzz(u^*(k)), \tag{3}$$

$$R_{new}(k) = fuzz(x_1^*(k)) \, x \dots x \, fuzz(x_n^*(k)) \, x \, fuzz(u^*(k) + r(k)), \tag{4}$$

where $x_1, …, x_n$ are states of the controller, $u(k)$ is its output and * denotes these values are crisp (to protect possible misunderstanding). The only difference between $R_{bad}(k)$ and R_{new} is in the consequent part of IF-THEN rules, i.e. in the output, which confirms the role of $r(k)$ as a correction value. The implementation of the knowledge base adaptation can be either *rule-based* or *relation-based* and

will be explained in next sections. In the following section some properties of SOFLC will be described.

2.1 Advantages and Drawbacks of SOFLC

It seems to be reasonable to search methods that would minimize the deviation from the optimal state as quickly as possible. Therefore gradient-based methods should be the most convenient for the knowledge modification. In such a sense SOFLC is also a special form embedding this calculus since it utilizes Jacobian what can be seen very clearly in (1). The only fundamental difference between SOFLC and 'classical' gradient-descent methods (GDA) can be described as follows. SOFLC represents *gradient of behavior* and 'classical' gradient methods can be related to the *gradient of the knowledge base*. In other words, SOFLC directly calculates the derivative of the system behavior, i.e. its change and 'classical' gradient methods compute the change of the control error in dependence on the knowledge base parameters. From this reason there is a close relation between *gradient of behavior* and *gradient of the knowledge base* but no equivalence, rather resemblance or similarity ($\approx$).

Although GDA should be the fastest adaptation but two basic problems are related to it. Firstly, the error function $E(k)$ is unknown in advance and it may be of a complex shape with a number of local minima. It is very difficult in advance to estimate their number and possible place of the global minimum, i.e. optimal solution. Further, the absence of such estimation disables the determination of the learning factor value, too. If it is too small the convergence will be too slow and if it is too big there will be a risk the global minimum will be 'jumped over'. Secondly, there is possibility to minimize only one criterion – error function $E(k)$ but in the practice there are also other control criteria. SOFLC overcomes these problems partially and is more practice-oriented because it is able to involve other criteria, too. However, it is sensitive to external signals such as disturbances, noises and set-point changes [7, 8] because it is not able to distinguish whether the parameters of the controlled system are changed or an external signal entered the system. A negative effect can occur if the adaptation proceeds although it is not more necessary. In such a way some wrong changes in the knowledge base may be performed. The wrong understanding causes this state if e.g. an external error occurs and AFC will evaluate it as a parameter change. In [8] it is shown that adding some supervisory rules may solve this problem. A modification of SOFLC in the form of the so-called sliding mode control is made in [7] where not only the positions of the control error e(t) but also its change (the first derivative) is taken into consideration. This method needs complete knowledge about the states of the controlled system and proper design of the sliding hyper-plane. However, how to design a 'good' hyper-plane it is not solved in this approach. Further, the only criterion for the controller design is the control error. It is of course important that its value converges to zero but in many applications also another criteria may be

still more important. From this reason we proposed a modification of this method, which is discussed in section 4. Further, we proposed a hybrid structure merging both SOFLC as well as GDA to balance their properties. This structure is described in the section 3.

3   Rule-based Implementation

As already mentioned in section 2 knowledge base $R(k)$ can be described in two fundamental ways: *rule-based* and *relation-based*. In the first case there is a set of fuzzy IF-THEN rules and a set of definitions of linguistic terms in the form of membership functions. Let us denote such a set of rules in this section $R(k)$, too. $R(k)$ is the set of N_r fuzzy rules r_p ($p=1, …, N_r$) of n inputs and one output. Such a rule r_p represents the Cartesian product of these input/output variables and is also a fuzzy relation $R_p=A_{1,p} \ x \ … \ A_{n,p} \ x \ B_p$, where $A_{1,p}, …, A_{n,p}$ are linguistic values of $x_1, …, x_n$ for the p-th rule and B_p represents its output. The knowledge base R is then a union of such rules (fuzzy relations) and after substituting into (2) it will be changed to (5):

$$R(k+1) = \left[\bigcup_{p=1}^{N_r} \left(A_{1,p} \bigcap \overline{A_1^{bad}} \right) x \dots x \ A_{n,p} \ x \ B_p \right] \ \bigcup \ \dots \bigcup \ \left[\bigcup_{p=1}^{N_r} A_{1,p} \ x \dots x \left(A_{n,p} \bigcap \overline{A_n^{bad}} \right) x \ B_p \right] \ \bigcup$$

$$\bigcup \left[\bigcup_{p=1}^{N_r} A_{1,p} \ x \dots x \ A_{n,p} \ x \left(B_p \bigcap \overline{B^{bad}} \right) \right] \ \bigcup \ \left(\underbrace{A_1^{bad} \ x \dots x \ A_n^{bad} \ x \ B^{new}}_{R_{new}} \right) \tag{5}$$

$R_{bad}(k)$ can be a union of all previously fired rules, too. However, for the sake of simplicity we will consider only one rule with the greatest strength α and therefore $A_1^{bad} \ x \ … \ x \ A_n^{bad}$ is its premise. The reinforcement value $r(k)$ corrects only the consequent of such a rule and B^{new} is the fuzzified result of $u(k)+r(k)$, i.e. $fuzz(u(k)+r(k))$. The simplest fuzzification is in the form of singletons but in general, other forms are possible, too.

In (5) following drawbacks can be seen:

1   *Possibility to change only one rule in one step* - The adaptation process will be longer and it is possible the low control quality was not caused by the rule with the greatest strength but by another rules working as noise. In every case the convergence will be worse.

2   *Growth of the rule number* - Let us consider *4* rules with two inputs and one output then there will be *13* rules in next step, which leads to an enormous growth of rules in the step $k+2$, $k+3$, etc. ($N_r(k+1)=N_r(k).(n+1)+1$).

3   *Need of garbage collection* - Previous two points show us such filtering or 'garbage collection', i.e. removing useless rules, is necessary to prevent the computational complexity and to improve the adaptation convergence.

To utilize the advantages of GDA and relation-based implementation of SOFLC we proposed special hybrid connection of these two methods as seen in fig. 2 [19].

The adaptation process can be described in following steps:

1   Defining input and output variables.

2   Defining term sets for variables in the step 1.

3   Designing initial membership functions (not necessary).

4   Processing GDA until the prescribed threshold of the control error $e(k)$ is reached.

5   If $e(k)$ is under such a threshold then processing SOFLC otherwise jump (switch) to the step 4.
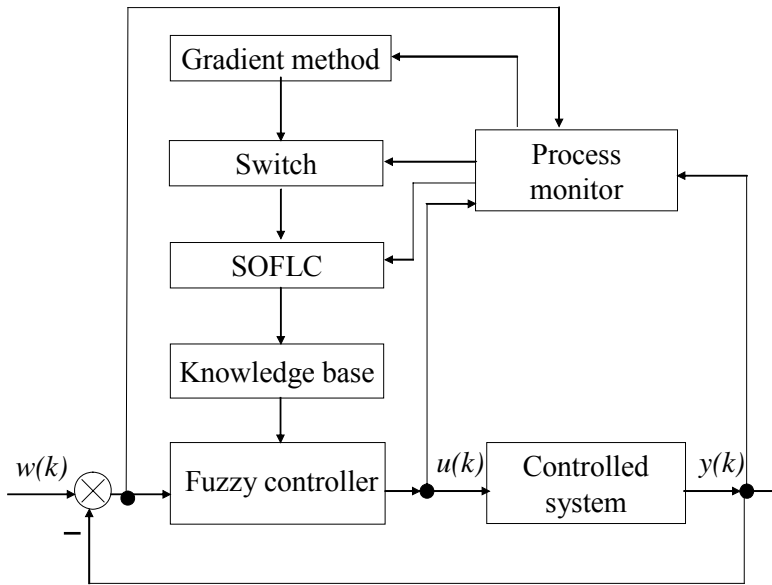

Figure 2

Hybrid control structure of a gradient-descent adaptation system and SOFLC

The main idea is that GDA is the fastest method if the threshold of the control error as the most important criterion is not too strict. In such a case we can choose a greater learning factor and speed up the adaptation. After this 'rough' adaptation we can switch the control to SOFLC to minimize the control error to be as small as possible and at this same time to include other criteria, too.

This hybrid control algorithm was implemented and tested on LEGO robots. Its results were compared also with a non-adaptive FC designed by a human operator. The control task was the so-called *parking problem*, i.e. to park a mobile robot at a given place and direction and was solved with and without obstacles. The process monitor (see Fig. 2) evaluates the parking process by two criteria: *parking error*

E_P - more important corresponding with the control error and *trajectory error E_T* - computed as division of the *real* trajectory length and *optimal* trajectory length. The optimal trajectory is the shortest distance between the robot and the goal. The first criterion is in the form:

$$E_P = \sqrt{(\phi_f - \phi)^2 + (x_f - x)^2 + (y_f - y)^2} \,, \tag{6}$$

where *(x, y)* are coordinates of the robot, ϕ is its position angle and *(x_f, y_f, ϕ_f)* are position and direction of the goal (parking place). Similar description is used for starting (initial) points, too. In the case of obstacles one additional criterion comes still into consideration - *number of impacts* on the obstacle.

The criterion predetermines also the structure of the IF-THEN rules, in our case three inputs *(x, y, ϕ)* and one output - change of the wheels angle for such a robot. The parking problem was solved with the help of a non-adaptive controller by 35 rules. It has no sense to observe the number of rules in the case of SOFLC because this number was very varying and there is almost no upper limit for their number. We need to consider there are unlimited numbers of combinations for intersections of fuzzy sets in (5). However, their number was considerably (several times) greater than for the non-adaptive controller. From this reason it was necessary to use a simple garbage collector, which minimizes the number of rules. It removes replaced and identical rules. If there are rules with identical premises but different consequents the older rule will be removed. Further, it is possible to improve its efficiency removing rules whose membership functions have little values of grade of membership or merging rules with similar premises.

Results of several experiments for different starting points (in parentheses) are depicted in Figures 3, 4 and 5.
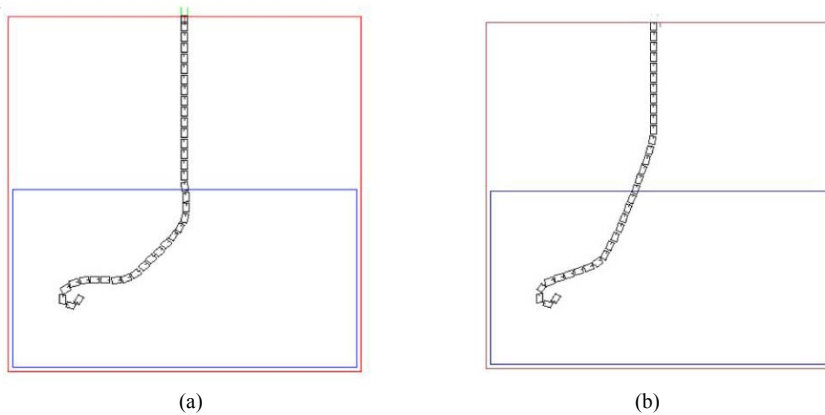

(a)   (b)

Figure 3

Comparison of trajectories for a non-adaptive FC (20, 80, 260) (a) and hybrid AFC (20, 80, 260) (b)

We can see that the first two criteria E_P and E_T are better fulfilled at a non-adaptive FC. There are two reasons. First, E_P and E_T are not totally independent. Both are quantitative and E_P influences E_T directly proportionally. If E_P increases then also the trajectory will be more different from the optimal length but the shape may be in spite of that of 'better quality', which is also this case. It can be seen especially at the obstacle avoidance Fig. 4 and 5. This assertion is supported by a smaller number of impacts at GIFC than at non-adaptive FC. Secondly, reinforced rules are fired till in next steps after the error already occurred and in such a way delay influences the efficiency of GIFC negatively. Shortening the sampling period T can eliminate this problem. There are only hardware limitations.
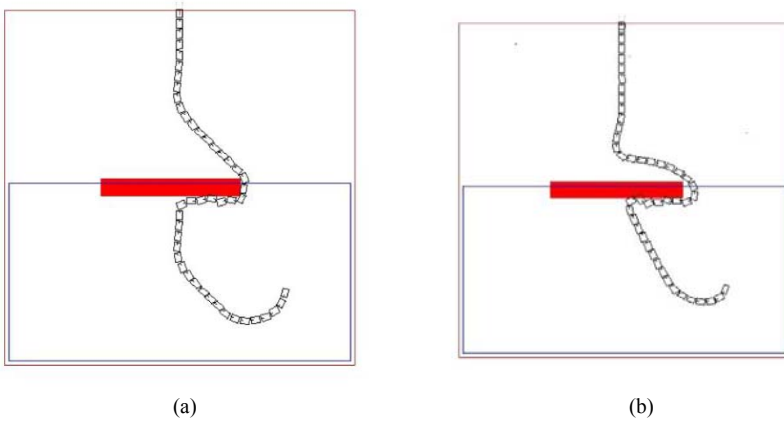

| (a) | (b) |

Figure 4

Comparison of trajectories with an obstacle for a non-adaptive FC (80, 80, 260) (a) and hybrid AFC (80, 80, 260) (b)

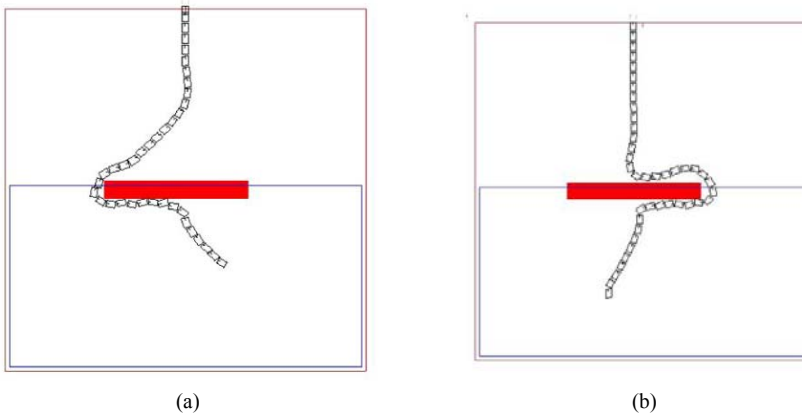| (a) | (b) |

Figure 5

Comparison of trajectories with an obstacle for a non-adaptive FC (60, 70, 150) (a) and hybrid AFC (40, 80, 110) (b)

4   Relation-based Implementation

This kind of implementation is considerably simpler than in the previous case. We will construct all three fuzzy relations $R(k)$, $R_{bad}(k)$ and $R_{new}(k)$ as described in (3) and (4). $R(k)$ can be set up in the initialization step as a zero matrix. In such a way we get n-dimensional cubes (in the case of two inputs and one output three-dimensional) where each element is characterized by the grade of membership to such a fuzzy relation as seen in Fig. 6.
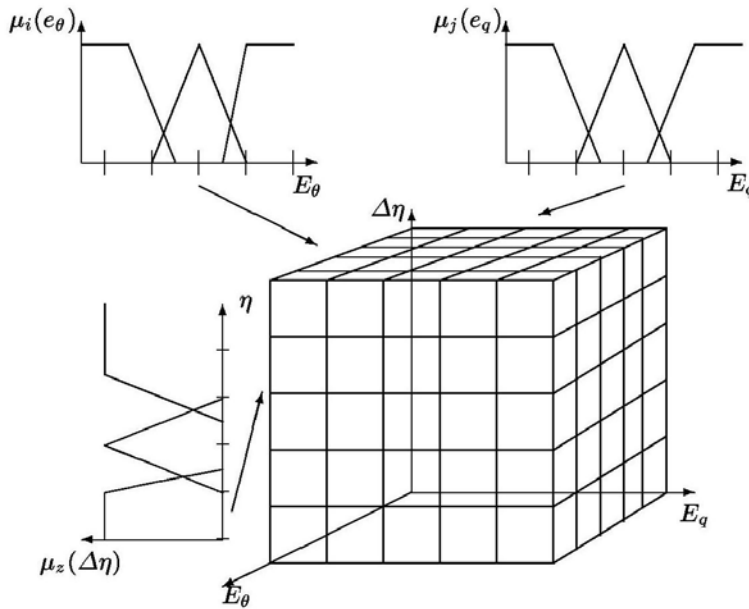

Figure 6
Structure of fuzzy relations for $R(k)$, $R(k+1)$, $Rbad(k)$ and $R_{new}(k)$

However, there are also two basic drawbacks:

1 *Loss of IF-THEN rules* - Knowledge representation in the form of fuzzy relations is not 'user-friendly' and it is not possible to make an unambiguous transform from a fuzzy relation to a fuzzy set (reverse transform is possible).

2 *Higher computational complexity* - To calculate next $R(k)$ it is needed to perform a complete set of all matrices calculations for each value from supports of input/output values. If the discretization of the support is dense (support has many values) then the size of such a matrix will grow enormously.

The relation-base implementation of SOFLC was used for the design of an automatic pilot [17]. For the sake of simplicity we will take into account only the longitudinal flight (plain $X \times Z$) and therefore we will control only the height of

the aircraft as seen in Fig. 7. The goal of control is to follow the prescribed rising trajectory, in other words, to keep the pitch angle equal to the angle of the prescribed trajectory. The basic description of an aircraft model resulting from motion equations by a fuzzy state model consists of four state quantities:

α - angle of attack (slope of the aircraft in the horizontal flight)

w - vertical velocity (projection of u into the vertical plane)

θ - pitch angle (angle between the longitudinal aircraft axis and earth)

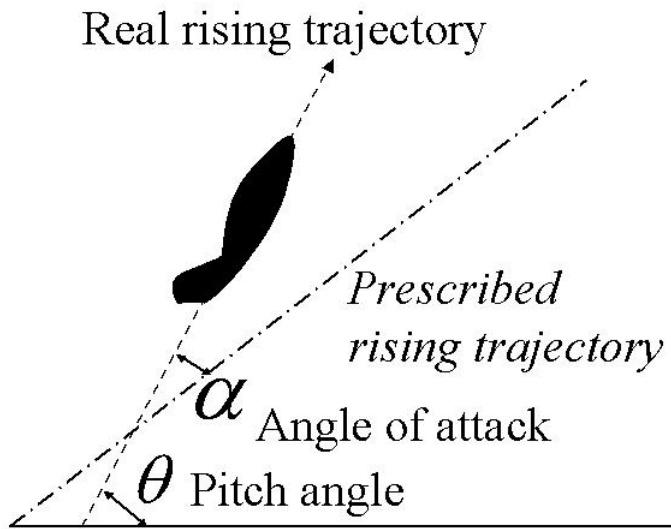q - pitch rate (derivative of θ)

Figure 7
State values of an aircraft for its control in the longitudinal plane

To overcome problems with the excessive sensitivity to external signals we will in contrast to [7] observe both the position of the performance measure *p(k)* and its trend, i.e. the first derivative $\dot{p}(k)$. The knowledge modifier (see Fig. 1) of our AFC is completed still by an *adaptation supervisor* where these rules are included. They are mostly problem-oriented and their parameters are dependent on the application but we can find some general knowledge, which can be described as follows. If we have a totally empty knowledge base we will let the adaptation (more correctly learning) in full processing without any limitations until it reaches a proper state of the controlled system (by defined criteria). Then we will stop the adaptation and will observe *p(k)* and $\dot{p}(k)$. If *p(k)* is not very significant and its change ($\dot{p}(k)$) is moving in an appropriate direction then we will not start the adaptation. This state indicates probably only an external error, which can be

eliminated by the controller without any change. It seems to be advantageous in many cases if we calculate the change of *p(k)* from a longer time interval.

The performance measure *p(k)* is defined in this case as the difference between the optimal dumping ratio λ_{opt} and the current one λ. If λ is high then the control is slow and with small oscillations. If λ is low then the control is fast but with high oscillations. The goal of the control is to minimize the performance measure, which is identical to the physical point of view, i.e. to stabilize the pitch angle ϕ at a certain value.

The results of some experiments are shown in Fig. 8 and Fig. 9. The main improvement is the minimization of oscillations so the dumping continuance form is almost aperiodic and the transition time remains the same. This fact enables a more comfortable flight as well as the increase of lifetime for mechanical parts of aircraft body (the most evidently in Fig. 9).
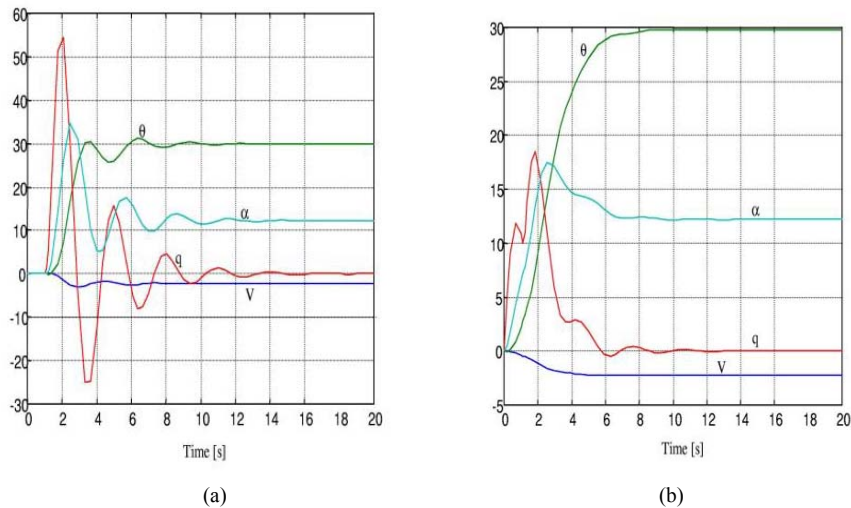

(a)   (b)

Figure 8

Responses of state quantities for an aircraft with a PI controller (a) and with AFC (b): α - angle of attack; *V* - air velocity; *q* - pitch rate; ϕ - pitch angle

A very important quantity is the angle of attack α that influences the flight stability and also the control process. Relative improvements in minimizing oscillations were also obtained when the flight altitude was changed (as an external error). The not fully smooth control is the tax the adaptation is limited by our adaptation supervisor. Its partially improvement can be reached by increasing the sampling frequency but this way enhances the computational efforts very enormously. Therefore it is important to find the balance between computational complexity and control quality.
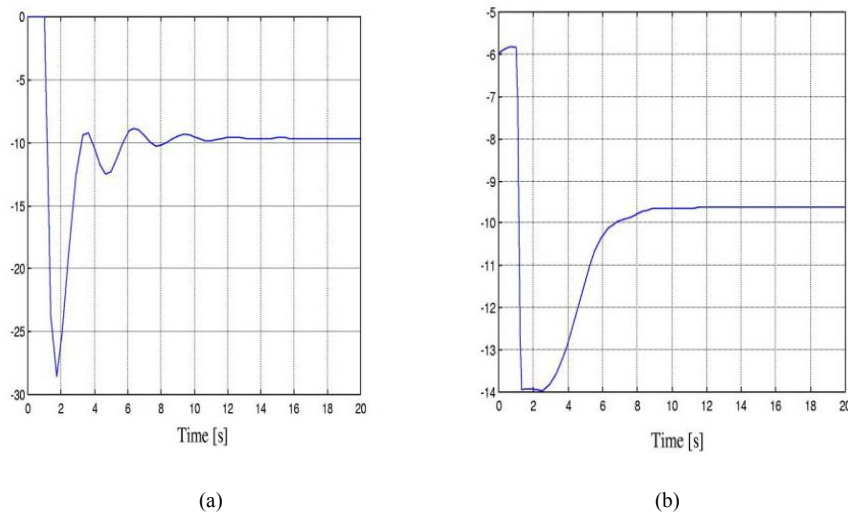
(a)   (b)

Figure 9

Responses of the actuator - elevator for an aircraft with a PI controller (a) and with AFC (b)

Conclusions

The principal advantage of both approaches is the substitution of a human expert in the establishment of the fundamental knowledge about the fuzzy controller, which is the most serious disadvantage of standard fuzzy systems. The designs presented enable fuzzy systems to be used for the control of systems characterized by great changes of parameters during their working. The second advantage is that we need to know only the dynamics of the controlled system and no input - output samples are necessary in advance. This fact enables the on-line adaptation and the set up of minimum parameters, which can lead to the decrease of the computational complexity. A hypothesis can be stated the rule-based implementation of SOFLC is more convenient for dynamical systems of higher order or with significant non-linearities but it has great demands on computational capacity.

References

[1]   P. Busaba, A. Ohsato, "Proposal of convex cone method for nonlinear optimization problems with linear constraints", Fuzzy workshop, Nagano, 25-26 October, 2001

[2]   W. L. Baker, J. A. Farrell, "An introduction to connectionist learning control system," IN: D. A. White, D. A. Sofge (Eds.), Handbook of Intelligent Control-Neural, Fuzzy and Adaptive Approaches, van Nostrand Reinhold Inc., New York, 1992

[3]   H. Bersini, J. P. Nordvik, A. Bonarini, "A simple direct adaptive fuzzy controller derived from its neural equivalent," see IEEE, 1993, pp. 345-350

[4]   F. Guély, P. Siarry, "Gradient descent method for optimizing various fuzzy rule bases," see IEEE, 1993, pp. 1241-1246

[5]   C. J. Harris, C. G. Moore, "Intelligent identification and control for autonomous guided vehicles using adaptive fuzzy-based algorithms," Engineering Applications of Artificial Intelligence 2, 1989, pp. 267-285

[6]   R. Jager, Fuzzy Logic In Control (PhD. thesis), Technical University of DELFT, Holland, 1995

[7]   Y. T. Kim, Z. Bien, "Robust self-learning fuzzy controller design for a class of nonlinear MIMO systems," Int. Journal Fuzzy Sets and Systems, Elsevier Publisher, Holland, N. 2, Vol. 111, 2000, pp. 17-135

[8]   M. Ma, Y. Zhang, G. Langholz, A. Kandel, "On direct construction of fuzzy systems," Int. Journal Fuzzy Sets and Systems, Elsevier Publisher, Holland, N. 1, Vol. 112, 2000, pp. 165-171

[9]   L. Madarász, "Intelligent Technologies and Their Applications in Complex Systems," University Press, Elfa, TU Košice, ISBN 80-8966-75, Slovakia, 2004, 348 pp.

[10]   H. R. Nauta Lemke, W. De-Zhao, "Fuzzy PID supervisor," IN: Proceedings of the 24th IEEE Conference on Decision and Control, Fort Lauderdale, Florida, USA, 1985

[11]   T. J. Procyk, E. H. Mamdani, "A linguistic self-organizing process controller," Automatica N. 15, 1979, pp. 15-30

[12]   Y. Shi, M. Mizumoto, "Some considerations on conventional neuro-fuzzy learning algorithms by gradient descent method," Int. Journal Fuzzy Sets and Systems, Elsevier Publisher, Holland, N. 1, Vol. 112, 2000, pp. 51-63

[13]   J. K. Tar, I. J. Rudas, J. F. Bitó, "Comparison of the Operation of the Centralized and the Decentralized Variants of a Soft Computing Based Adaptive Control," In: Proc. of Jubilee Conference Budapest Tech, September 4, ISBN 963 7154 31 0, pp. 331-342

[14]   J. K. Tar, A. Bencsik, J. F. Bitó, K. Jezernik, „Application of a New Family of Symplectic Transformations in the Adaptive Control of Mechanical Systems," In: Proc. Of the 2002 28th Annual Conference of the IEEE Industrial Electronics Society, Nov. 5-8 2002 Sevilla, Spain, Paper 001810, CD issue, ISBN 0-7803-7475-4, IEEE Catalogue N. 02CH37363C

[15]   J. K. Tar, I. J. Rudas, J. F. Bitó, L. Horváth, K. Kozlowski, "Analysis of the Effect of Backslash and Joint Acceleration Measurement Noise in the Adaptive Control of Electro-mechanical Systems," In: Proc. of the 2003 IEEE International Symposium on Industrial Electronics (ISIE 2003), June 9-12, 2003, Rio de Janeiro, Brasil, CD issue, file BF-000965.pdf, ISBN 0-7803-7912-8, IEEE Catalogue N. 03th8692

[16]   J. Vaščák, L. Madarász, I. J. Rudas, Similarity Relations in Diagnosis Fuzzy Systems; in: INES '99 – IEEE International Conference on Intelligent Engineering Systems, Stará Lesná (High Tatras), Slovakia, November 1-3 1999, pp. 347-352, ISBN 80-88964-25-3, ISSN 1562-5850

[17]   J. Vaščák, P. Kováčik, F. Betka, P. Sinčák, "Design of a Fuzzy Adaptive Autopilot," In: The State of the Art in Computational Intelligence, Proc. of the Euro-International Symposium on Computational Intelligence EISCI 2000, Košice, Slovakia, 2000, pp. 276-281, ISBN 3-7908-1322-2, ISSN 1615-3871

[18]   J. Vaščák, L. Madarász, "Similarity Relations in Diagnosis Fuzzy Systems," Journal of Advanced Computational Intelligence, Vol. 4, Fuji Press, ISBN 1343-0130, Japan, 2000, pp. 246-250

[19]   J. Vaščák, M. Mikloš, K. Hirota, "Hybrid Fuzzy Adaptive Control of LEGO Robots," Vol. 2, N. 1, International Journal of Fuzzy Logic and Intelligent Systems, Korea, March 2002, pp. 65-69

On the Polyhedral Graphs with Positive Combinatorial Curvature

Tamás Réti, Enikő Bitay, Zsolt Kosztolányi

Bánki Donát Faculty of Mechanical Engineering, Budapest Tech
Népszínház u. 8, H-1081 Budapest, Hungary
reti.tamas@bgk.bmf.hu

Abstract: The purpose of this article is to introduce a refinement of DeVos-Mohar conjecture in which the number of vertices of polyhedral graphs with positive combinatorial curvature, which are neither prisms, nor antiprisms, (PCC graphs) plays a significant role. According to the conjecture proposed by DeVos and Mohar, for the maximal vertex number Vmax of a PCC graph, the inequality VLB=120 ≤ Vmax ≤ VUB=3444 is fulfilled. In this paper we show that the lower bound VLB can be improved. The improved lower bound is VLB =138. It is also verified that there are no regular, vertex-homogenous PCC graphs with vertex number greater than 120. We conjecture that for PCC graphs the minimum value of combinatorial curvatures is not less than 1/380. If the conjecture is true this implies that the upper bound VUB is not greater than 760. Moreover, it is also conjectured that there are no PCC graphs having faces with side-number greater than 19, except two trivalent polyhedral graphs containing 20- and 22-sided faces, respectively.

Keywords: combinatorial curvature, planar graph, polyhedra, vertex corona

1   Introduction

If all of the vertices of a polyhedral graph G_P have positive combinatorial curvatures only, then G_P is said to be a graph with positive combinatorial curvature [1,2,3]. Let $\Re$ be the set of polyhedral graphs with positive combinatorial curvature, which does not contain the graphs of prisms and antiprisms. A graph included in $\Re$ is called a PCC graph. DeVos and Mohar proposed the following conjecture, which is still open: For the maximal vertex number V_{max} of PCC graphs in $\Re$, the inequality $120 \le V_{max} \le 3444$ is fulfilled [1]. In this paper we show that the conjectured lower bound ($V_{LB}=120$) can be improved. It will be verified that for any PCC graph, the maximum number of vertices is not less than 138.

2   Background: Definitions and Notation

Consider a polyhedral graph with vertex number V. The combinatorial curvature $\Phi(X_j)$ of an r-valent vertex X_j (j=1,2,…V) is defined as

$$\Phi(X_j) = 1 - \frac{r}{2} + \frac{1}{n_{j,1}} + \frac{1}{n_{j,2}}, ..., + \frac{1}{n_{j,r(j)}}$$

(1)

where $n_{j,1}, n_{j,2}, …n_{j,r(j)}$ are the side-numbers of r polygons incident with vertex X_j. It is known that for an arbitrary polyhedral graph the sum of combinatorial curvatures satisfies the relation [1,2,3]

$$\sum_{j=1}^{V} \Phi(X_j) = \chi$$

.   (2)

In Eq.(2) χ is the Euler-characteristic, which equals 2 for polyhedral graphs. The fundamental properties of PCC graphs can be summarized as follows:

i   The set of PCC graphs ($\Re$) is finite [1]. All prisms and antiprisms are represented by polyhedral graphs with positive combinatorial curvature, but their number is infinite.

ii   The number of vertices is not greater than 3444 [1].

iii The maximal number of edges incident to a vertex is less than 6. This implies that if a PCC graph has a vertex of valency 5, it contains at least four triangles, if it has a vertex of valency 4, then it contains at least one triangle.

iv   The graph of the snub icosidodecahedron with 60 vertices is the largest PCC graph which has vertices of valency 5, only.

From Eq.(2) it follows that the number of vertices V of a PCC graph can be easily estimated if we know the minimum value Φ_{min} of the possible combinatorial curvatures. In this case we have $V \le 2/\Phi_{min}$. It is obvious that the maximum value of $\Phi(X_j)$ is 1/2, this value is valid for the tetrahedron. A polyhedral graph is called a vertex homogenous graph (VH-graph), if its combinatorial curvatures are equal in each vertex (i.e. $\Phi(X_j)= 2/V$ is fulfilled for j=1,2,…V). For example, Platonic and Archimedean polyhedra, prisms and antiprisms are represented by VH graphs. It is obvious that every VH-graph which is neither a prism nor an antiprism is a PCC graph.

A polyhedral graph G_P is called a regular (ρ-regular) graph if all vertices of G_P have the same valency ρ, where ρ= 3,4 or 5. It is worth noting that among PCC graphs there can exist vertices with identical (equal) positive combinatorial curvatures, but with different numbers of valency. It is conjectured that every vertex homogenous graph is regular.

Consider a subset $\Re_S=\Re_S(n_1, n_2,\ldots n_k,\ldots n_K)$ of $\Re$ where n_k stands for the side number of faces, for which $3 \leq n_1 < n_2 < \ldots n_k, \ldots < n_K$, $k=1,2,\ldots K$, and $K \geq 1$ is a positive integer. This means that the number of possible face types is equal to K for any graphs in subset $\Re_S$. (In other words, all polyhedral graphs in $\Re_S$ have faces with K different side-numbers, exactly). In certain cases, $\Re_S$ is identical to the empty set, i.e. $\Re_S = \varnothing$. In case $K=1$, we have $\Re_S(n_1) = \varnothing$ if $n_1=4$ or $n_1 \geq 6$. The graph of dodecahedron is the unique PCC graph which is composed of pentagons only. For example, it is easy to see that $\Re_S(5,10) = \varnothing$, because there are no PCC graphs composed of 5- and 10-gons.

For set $\Re_S(n_1, n_2,\ldots n_k,\ldots n_K)$ we define the upper bound V_{US} on the maximum vertex number as follows

$$V_{US} = \left\{ \max V(G_S) \quad \middle| G_S \in \Re_S \right\}$$

(3)

In Eq.(3) $V(G_S)$ denotes the number of vertices of an arbitrary graph $G_S \in \Re_S$. If $\Re_S$ is identical to the empty set, then $V_{US} = 0$, by definition. It is obvious that inequality $V_{US} \leq 3444$ holds for any subset $\Re_S$. In other words, for an arbitrary PCC graph in $\Re_S$, the corresponding vertex number is not greater than 3444 [1].

Graphs $G_{S,1}, G_{S,2},\ldots G_{S,z},\ldots,G_{S,Z} \in \Re_S(n_1, n_2,\ldots n_k,\ldots n_K)$ are called maximal (more exactly maximal with respect to $\Re_S$) if equality $V_{US}=V(G_{S,1})= V(G_{S,1}),\ldots = V(G_{S,Z})$ is fullfilled for $z=1,2,\ldots Z$, where Z is a positive integer. Considering the polyhedral graphs consisting of triangles, there are only three vertex homogenous graphs in $\Re_S(3)$. Among these polyhedra only the tetrahedron and the icosahedron are represented by PCC graphs, the octahedron is a triangular antiprism. The graph of icosahedron with vertex number of 12 is the unique maximal PCC graph in $\Re_S(3)$. It is worth noting that in $\Re_S(3,4)$ there exist four distinct polyhedra represented by VH-graphs with vertex number of 24. All of them are characterized by maximal PCC graphs (the snub cube which has two different chiralities, the rhombicuboctahedron and the Miller polyhedron (called pseudo-rhombicuboctahedron) which is not Archimedean). In $\Re_S(5,6)$ there is only one VH-graph, this maximal PCC graph corresponds to the truncated icosahedron (the so-called Buckminster fullerene).

In **Fig. 1** the great rhombicosidodecahedron (truncated icosidodecahedron) is shown. It has 120 vertices, 180 edges and 62 faces of regular polygons: 30 squares, 20 hexagons and 12 decagons. The graph of the great rhombicosidodecahedron is a VH-graph, since values of combinatorial curvatures are equal for all vertices, $(1-3/2 +(1/4+1/6+1/10)= 1/60)$.
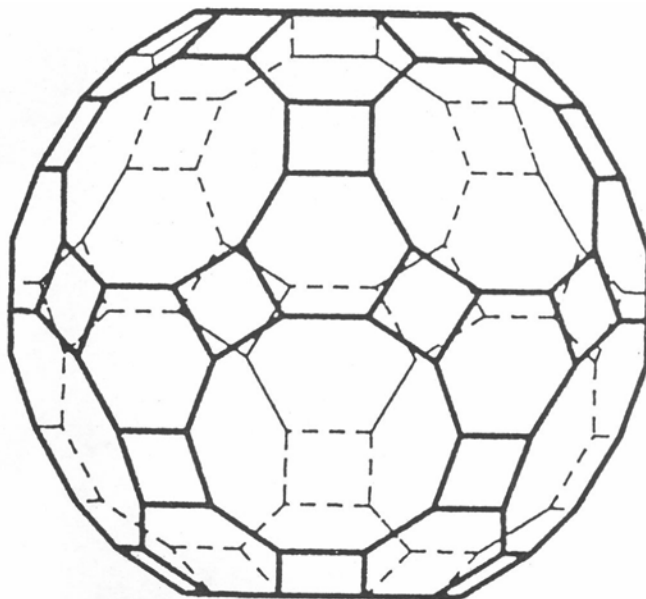
Figure 1

The great rhombicosidodecahedron

As far as the lower bound of maximum vertex number is concerned, the conjectured number of vertices for PCC graphs is equal to 120 [1]. This is identical to the vertex number of the great rhombicosidodecahedron [1]. The theoretical upper bound of maximum vertex number (V_{UB}= 3444) found by DeVos and Mohar corresponds to a vertex number of a "hypothetic" trivalent VH-graph composed only of 3-, 7- and 41-gons. We will show that there exist no trivalent PCC graphs composed of 3-, 7- and 41-gons for which the vertex numbers are equal to 3444. Moreover, it will be verified that there exist PCC graphs with vertex numbers greater than 120.

3 Results

The main results are represented by the following six theorems.

Theorem 1 There is no trivalent PCC graph in $\Re_S(3,7,b)$ if b > 11 positive integer.

Proof. The proof is based on computational results obtained by checking the fulfillment of necessary conditions given in the form of various equations and inequalities. We used the following three lemmas:

Lemma 1 Consider a trivalent polyhedral graph composed of α-, β- and γ-sided polygons. Let us denote by N_α , N_β and N_γ the corresponding number of faces. The

following well-known necessary conditions for the existence of such a graph (polyhedron) are easily derivable from Euler's formula:

$$\frac{1}{N_P}(\alpha N_\alpha + \beta N_\beta + \gamma N_\gamma) < 6$$

(4)

$$(6-\alpha)N_\alpha + (6-\beta)N_\beta + (6-\gamma)N_\gamma = 12$$

(5)

where $N_\alpha + N_\beta + N_\gamma = N_P$

Lemma 2 There are no edge-neighbor triangles in a trivalent polyhedral graph with vertex number greater than 4. (The exceptional case of V=4 corresponds to the graph representing a tetrahedron).

To prove this, let us suppose that there are two edge-neighbor triangles (i.e. two triangles with a common edge) in a trivalent polyhedral graph G_T. By deleting vertices X_A and X_C of the neighbor triangles we have separated graph components. (See **Fig. 2**)
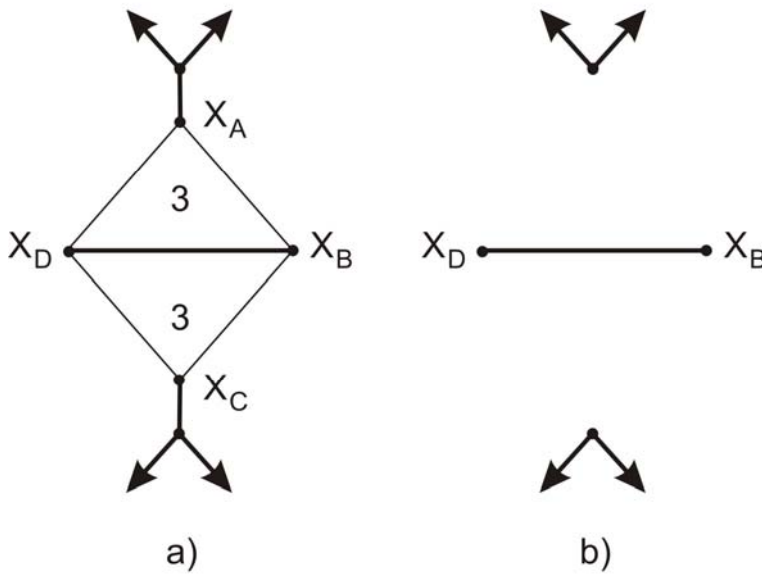

Figure 2

A 2-connected, trivalent, planar graph including edge-neighbor triangles

(a) The original graph. (b) The graph obtained by deleting vertices X_A and X_C.

It follows that G_T must be only 2-connected; consequently it is a non-polyhedral graph. This implies that all triangles are isolated in a trivalent polyhedral graph. As an example, the smallest 2-connected, planar, non-polyhedral, trivalent graph

is illustrated in **Fig. 3a**. It is worth noting that this graph can be represented by a non-convex polyhedron. (See. **Fig. 3b.**) (From the previous considerations it follows that there are no trivalent PCC graphs composed of 3- and 7-gons only.)
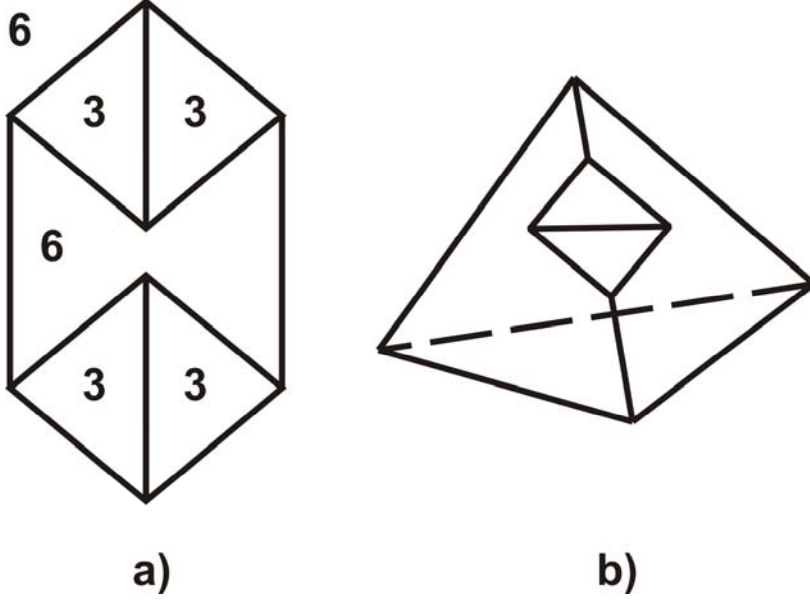

Figure 3

The smallest 2-connected, trivalent planar graph (a) and its 3-dimensional representation (b)

Lemma 3 Consider a trivalent polyhedral graph composed of α-, β- and γ-sided polygons. As it is illustrated in **Fig. 4**, by taking into consideration the types of vertex coronas, each vertex can be classified into 10 distinct classes denoted by $C_{\alpha,\alpha,\alpha}$, $C_{\alpha,\alpha,\beta}$,…,$C_{\gamma,\gamma,\gamma}$, respectively. Let V_1, V_2,…,V_{10} be the corresponding numbers of vertices belonging to different vertex configurations. It can be proved that [4]

$$\frac{1}{V}\sum_{i=1}^{10} V_i M_i = \frac{\langle n^2 \rangle}{\langle n \rangle}$$

(6)

In Eq.(4) $V = V_1 + V_2 + … + V_{10}$, quantities M_1, M_2, …,M_{10} are the corresponding vertex coordination numbers shown in **Fig. 4**, and

$$\langle n^k \rangle = \frac{1}{N_P}\left(\alpha^k N_\alpha + \beta^k N_\beta + \gamma^k N_\gamma\right)$$

(7)

is the kth moment of the side numbers of polygons [4]. It should be noted that an extended version of Eq.(6) is valid for 4- and 5-regular polyhedral graphs [4].
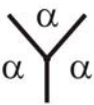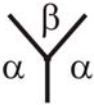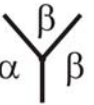
Figure 4

The ten possible vertex arrangements (vertex coronas) in a trivalent polyhedral graph

By using the three lemmas outlined above, the main steps of the proof are as follows:

1 Consider a subset $\Re_{SG}(3,a,b)$ of $\Re_S(3,a,b)$. We assume that $\Re_{SG}(3,a,b)$ contains trivalent PCC graphs consisting of 3-gons, a-gons and b-gons only, and for any graph of $\Re_{SG}(3,a,b)$ the following inequalities are fulfilled:

$$6 \le a < b \tag{8a}$$

$$\Phi(3,a,b) = 1 - \frac{3}{2} + \frac{1}{3} + \frac{1}{a} + \frac{1}{b} > 0 \tag{8b}$$

$$\Phi(3,b,b) = 1 - \frac{3}{2} + \frac{1}{3} + \frac{1}{b} + \frac{1}{b} \le 0 \tag{8c}$$

$$\Phi(a,a,a) = 1 - \frac{3}{2} + \frac{1}{a} + \frac{1}{a} + \frac{1}{a} \le 0 \tag{8d}$$

In the equations above, $\Phi(3,a,a)$, $\Phi(3,a,b)$, $\Phi(3,b,b)$ and $\Phi(a,a,a)$ denote the corresponding combinatorial curvatures. Every graph included in $\Re_{SG}$ has a positive combinatorial curvature in vertices of type of V(3,a,a,) and V(3,a,b), for which $\Phi(3,a,a) > \Phi(3,a,b) > 0$ is fulfilled.

It is easily seen that only the following PCC graphs satisfy the conditions formulated by Eqs.(8a-8d): all trivalent polyhedral graphs in $\Re_S(3,6,b)$ if b≥12, in $\Re_S(3,7,b)$ if 12≤b≤41, in $\Re_S(3,8,b)$ if 12≤ b≤ 23, in $\Re_S(3,9,b)$ if 12≤b≤ 17, in $\Re_S(3,10,b)$ if 12≤ b ≤ 14 and $\Re_S(3,11,12)$.

2   Let N_3, N_a and N_b be the number of 3-, a- and b-sided faces, respectively. Since $\mathfrak{R}_{SG}$ is a subset of trivalent polyhedral graphs, according to Lemma 1 we have

$$(3N_3 + aN_a + bN_b) < 6N \tag{9}$$

$$3N_3 + (6-a)N_a + (6-b)N_b = 12 \tag{10}$$

where $N = N_3 + N_a + N_b$.

3   From Eqs.(8a-8d) it follows that not only the triangles are isolated but all b-sided faces as well. (See **Fig. 5**)
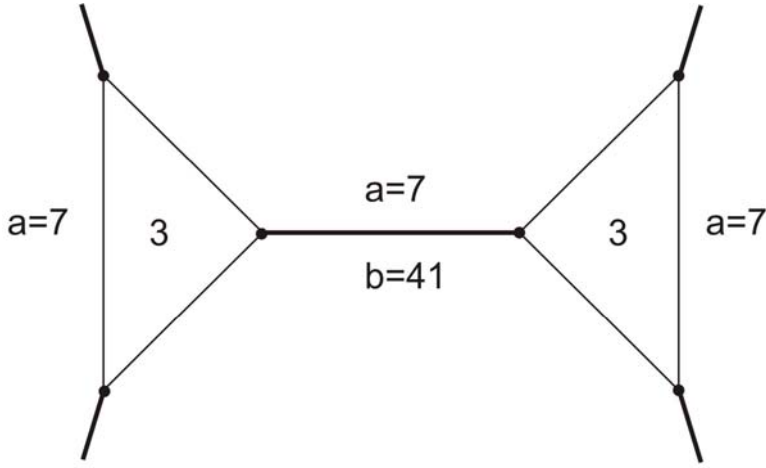
Figure 5

Isolated triangles in a trivalent PCC graph of $\mathfrak{R}_{SG}$ composed only of 3-,a- and b-sided polygons

Moreover, we can conclude that each vertex is incident with a vertex of an isolated triangle. Hence, for any graph of $\mathfrak{R}_{SG}$ we obtain

$$3N_3 = V \tag{11}$$

Since each vertex is incident with a vertex of an a-sided polygon as well, inequality

$$V \leq aN_a \tag{12}$$

yields. (**See Fig. 5**).

4   From Eqs.(11-12) we have

$$aN_a - V = aN_a - 3N_3 \geq 0 \tag{13}$$

Consequently,

$$\frac{aN_a}{3} \geq N_3$$

(14)

5   Since all triangles and all b-sided polygons are isolated, this implies that there is no polyhedral graph in $\mathfrak{R}_{SG}$ for which b is an odd integer. This can be easily verified on the basis of **Fig. 6**.
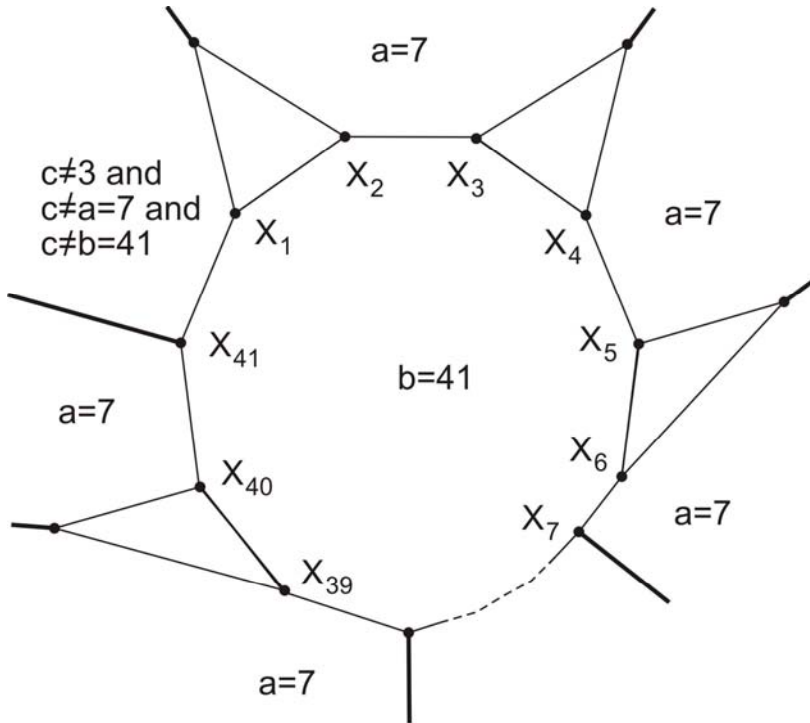
Figure 6

The neighborhood of b-sided faces in a PCC graph of $\mathfrak{R}_{SG}$ composed only of 3-,a- and b-sided polygons (for case: a=7 and b=41)

As we can conclude, the edge neighbors of b-sided polygons are or triangles either a-sided polygons, only. The sum of triangles and a-sided polygons should be equal to b, but two a-sided polygons having common edges with a b-sided one, cannot be edge-neighbors. Moreover, as it is demonstrated in **Fig. 6**, if b is an even number, then inequality

$$N_3 \geq \frac{b}{2} N_b$$

(15)

should be fulfilled. From Eqs.(11-15) we obtain

$$aN_a \geq 3N_3 = V \geq \frac{3b}{2}N_b$$

(16)

Inequality (16) gives a necessary condition for the existence of a graph in $\Re_{SG}$.

6 Because all triangles and b-sided polygons are isolated, and inequalities (12-16) are valid, this implies that there exist only two types of vertices (vertex coronas) which should be taken into account for graphs in $\Re_{SG}$. These are denoted by $C_{3,a,a}$ and $C_{3,a,b}$. In this case, $\alpha=3$, $\beta=a$ and $\gamma=b$ are the corresponding side-numbers of polygons incident with a common vertex. (See **Fig. 4**.)

7 According to Lemmas 2 and 3, as a particular case, from Eq.(6) it follows:

$$\frac{V(3,a,a)M(3,a,a) + V(3,a,b)M(3,a,b)}{V(3,a,a) + V(3,a,b)} = \frac{\langle n^2 \rangle}{\langle n \rangle}$$

(17)

where $V(3,a,a)$ and $V(3,b,b)$ are the numbers of vertices belonging to the vertex coronas $C_{3,a,a}$ and $C_{3,a,b}$ and $M(3,a,a)=(3+a+a)/3$ and $M(3,a,b)=(3+a+b)/3$ are the corresponding vertex coordination numbers. It is obvious that Eq.(17) represents a useful necessary condition for checking the existence of PCC graphs.

8 Equations (4-17) can be efficiently used to predict whether a particular subset of trivalent PCC graphs exists or not. More exactly, if there is no solution to Eqs. (4-17) for positive integers a, b, N_3, N_a and N_b we can conclude that this subset of PCC graphs is empty. Based on the concept outlined above, we checked by computer all possible cases concerning the existence of trivalent PCC graphs in $\Re_{SG}$. By performing computations, we verified, that there do not exist trivalent PCC graphs in $\Re_S(3,7,b)$ if b > 11. Consequently, there are no trivalent PCC graphs composed of 3-, 7- and 41-gons, only.

Theorem 2 There is no trivalent PCC graph in $\Re_S(3,8,b)$ if b > 22 positive integer. Moreover, if b=22 there exists a trivalent PCC graph G_x in $\Re_S(3,8,22)$ composed of 3-, 8- and 22-gons, for which $V(G_x) = 66$.

Proof. In this case, there are two types of vertex corona with positive combinatorial curvature (denoted by $C_{3,8,8}$ and $C_{3,8,b}$) which should be taken into account. The proof performed by computer was based on the same concept that was used in Theorem 1. We found that in a particular case when b=22, there exists a trivalent PCC graph composed of 3-, 8- and 22-gons for which $N_3=22$, $N_8=11$ and $N_{22}=2$. This graph denoted by G_{22} corresponds to a truncated 11-gonal prism. In **Fig. 7** the corresponding Schlegel diagram is shown.
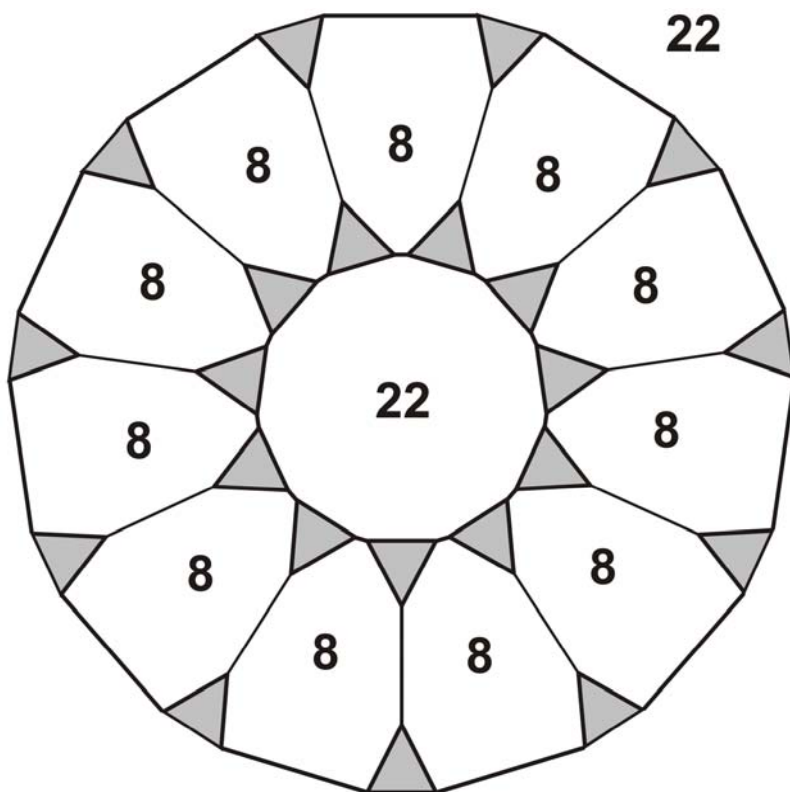
Figure 7
The Schlegel diagram of the PCC graph G_{22} (truncated 11-gonal prism)

We conjecture that the trivalent graph G_{22} is maximal with respect to set $\Re_S(3,8,22)$ where $V_{US} = 66$. It is supposed that G_{22} is the unique PCC graph containing 22-sided faces. Moreover, we assume that there exists only one PCC graph with 20-sided faces. This graph denoted by G_{20} which is composed of 20 triangles, 10 octagons and two 20-sided polygons, corresponds to a truncated 10-gonal prism. Based on previous considerations, the following conjecture can be formulated: There are no PCC graphs having faces with side-number greater than 19, except the two trivalent polyhedral graphs G_{20} and G_{22} containing 20- and 22-sided faces, respectively.

Theorem 3 There exists a trivalent PCC graph G_y in $\Re_S(4,6,11)$ for which $V(G_y) = 132$.

Proof. For trivalent graphs included in $\Re_S(4,6,11)$, we have from Euler's formula that $2N_4–5N_{11}=12$. Additionally, there are 5 types of vertex coronas with positive combinatorial curvature (denoted by $C_{4,4,4}$, $C_{4,4,6}$, $C_{4,4,11}$, $C_{4,6,,6}$ and $C_{4,6,11}$) which should be taken into consideration when using Eq.(6) for computational purposes.

In this case, $\alpha=4$, $\beta=6$ and $\gamma=11$ are the side-numbers of polygons which are incident with a common vertex. (See **Fig. 4**) For graphs in $\Re_S(4,6,11)$, the minimal value of combinatorial curvature is $\Phi_{min} = 1- 3/2 + (1/4+1/6+1/11)=1/132$. Theoretically, this implies that the possible maximum vertex number is not greater than 264. A computer search gave that the maximum vertex number of PCC graphs in $\Re_S(4,6,11)$ is less than 198.

In order to find the largest PCC graphs included in $\Re_S(4,6,11)$, we preselected seven "possible" subclasses of polyhedral graphs (candidates) whose existence could be expected. For the 7 subclasses denoted by G_A, G_B,….and G_G, the computed topological parameters (N_4, N_6, N_{11}, V(4,4,11), V(4,6,6), V(4,6,11) and V) are summarized in Table 1. Quantities V(4,4,11), V(4,6,6) and V(4,6,11) are the numbers of vertices belonging to the three vertex coronas denoted by $C_{4,4,11}$, $C_{4,6,6}$ and $C_{4,6,11}$, respectively. (A common feature of graphs included in the seven subclasses is that they do not contain vertex coronas of type $C_{4,4,4}$, and $C_{4,4,6}$, which implies that V(4,4,4)=V(4,4,6)=0.)

Subset of graphs	Topological parameters						
	N4	N6	N11	V(4,4,11)	V(4,6,6)	V(4,6,11)	V
GA	36	20	12	12	0	120	132
GB	36	26	12	0	12	132	144
GC	41	24	14	10	0	144	154
GD	41	29	14	0	10	154	164
GE	46	28	16	8	0	168	176
GF	46	32	16	0	8	176	184
GG	51	32	18	6	0	192	198

Table 1

Computed topological quantities for 7 subclasses of possible PCC graphs

From the computed results given in Table 1, it is clear that there do not exist trivalent PCC graphs in $\Re_S(4,6,11)$ with vertex number greater than 198. By analyzing the local structure of vertex coronas of graphs in Table 1, two cases should be taken into consideration.

CASE 1 concerns the subsets G_B, G_D and G_F in Table 1. Graphs in G_B, G_D and G_F have vertex coronas of type $C_{4,6,6}$ and $C_{4,6,11}$ with positive vertex numbers. It is easy to show that subsets G_B, G_D and G_F do not contain PCC graphs. The proof is based on the following observations. Each 11-gon is isolated, and each of them is surrounded by six 4-gons and five 6-gons, exactly. (There are no edge-neighbor hexagons since V(6,6,11)=0.) Consequently, among 6+5 neighbors of an arbitrary 11-gon there must exist two edge-neighbor squares. But, this is impossible because V(4,4,4)=V(4,4,6)=V(4,4,11) =0 is fulfilled.

CASE 2 concerns the subsets G_A, G_C, G_E and G_G in Table 1. Graphs in G_A, G_C, G_E and G_G have vertex coronas of type $C_{4,4,11}$ and $C_{4,6,11}$ with positive vertex

numbers. In this case, the following conclusions can be drawn: It is easily seen that every 11-gon and every hexagon are isolated. Additionally, each vertex is incident with a vertex of an 11-gon, this implies that $11N_{11}=V$.

In the following we suppose that all neighborhoods of every 11-gon are identical, that is, each 11-gon is surrounded by H hexagons and (11-H) squares exactly, where H is a positive integer not greater than five.

Figure 8 shows the four possible neighborhoods of an isolated hexagon in a hypothetical PCC graph. From this figure, we can conclude that among the edge-neighbors of a 6-sided polygon there exist always three squares and three 11-sided polygons, exactly (See Fig. 8a.)
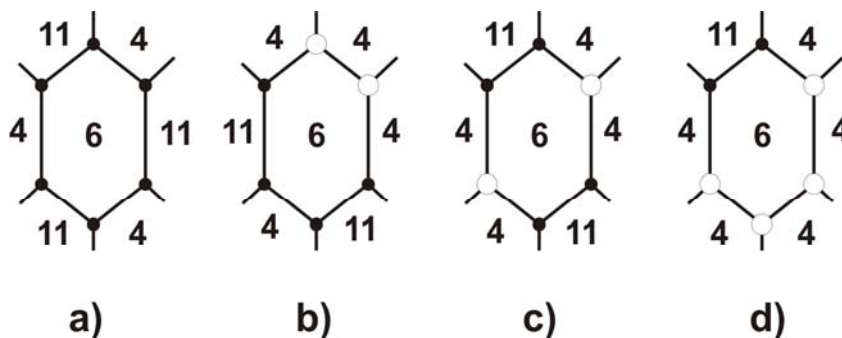
Figure 8

The possible edge-neighbor arrangements of a 6-sided polygon

The previous considerations imply that for the number of common edges, the following equalities should be fulfilled: $3N_6=HN_{11}$ and $2N_{11} = (11-H)N_4$. Since $N_4 = (12-5N_{11})/2$, we obtain that $H = 11 - 2N_4/N_{11} = 6 - 12/N_{11}$. From this latter equation we can determine the possible solutions for H. It can be stated that the proper solutions are: H=3 if $N_{11}=4$, H=4 if $N_{11}=6$ and H=5 if $N_{11}=12$.

From this it follows that in subset G_C, G_E and G_G there do not exist PCC graphs having identical neighborhoods for any 11-gons. It should be emphasized that the results obtained above are only valid, when it is supposed that each isolated 11-gon has identical neighborhood, that is, each 11-sided polygon is surrounded by H hexagons and (11-H) squares, exactly.

After analyzing the topological structure of vertex coronas, it was conjectured that only the existence of PCC graphs belonging to G_A could be expected (case H=5, $N_{11}=12$). It is easy to show that there exists a simple polyhedron which corresponds to a trivalent graph G_{AR} in subclass G_A. By transforming the great rhombicosidodecahedron we can generate a new polyhedron, which has the same topological parameters as graphs in G_A. This simple topological transformation of the great rhombicosidodecahedron can be performed by dividing six squares into two parts. As a result of this operation, we obtain 6*2+24=36 squares, 20

hexagons and 12 eleven-sided polygons. (See **Fig. 1**.) Consequently, the number of new vertices, faces and edges will be 132, 68 and 198, respectively. (As can be stated, each 11-gon is surrounded by H=5 hexagons and (11- H) =6 squares.)

The trivalent polyhedron contains two different types of vertices. The corresponding combinatorial curvatures are as follows: 1-3/2+(1/4+1/4+1/11) = 1/11 and 1- 3/2+(1/4+1/6+1/11)=1/132. It is conjectured that graph G_{AR} of this new polyhedron is maximal, that is G_{AR} with vertex number of 132 is the "largest trivalent PCC graph" in $\Re_S(4,6,11)$.

Theorem 4 There exists a non-cubic (non-trivalent) PCC graph which contains 138 vertices.

Figure 9
A non-cubic PCC graph with vertex number of 138

Proof. In **Fig. 9** the Schlegel diagram of this polyhedron is shown. It has 138 vertices, 219 edges and 83 faces (8 triangles, 45 squares, 24 pentagons and six 19-
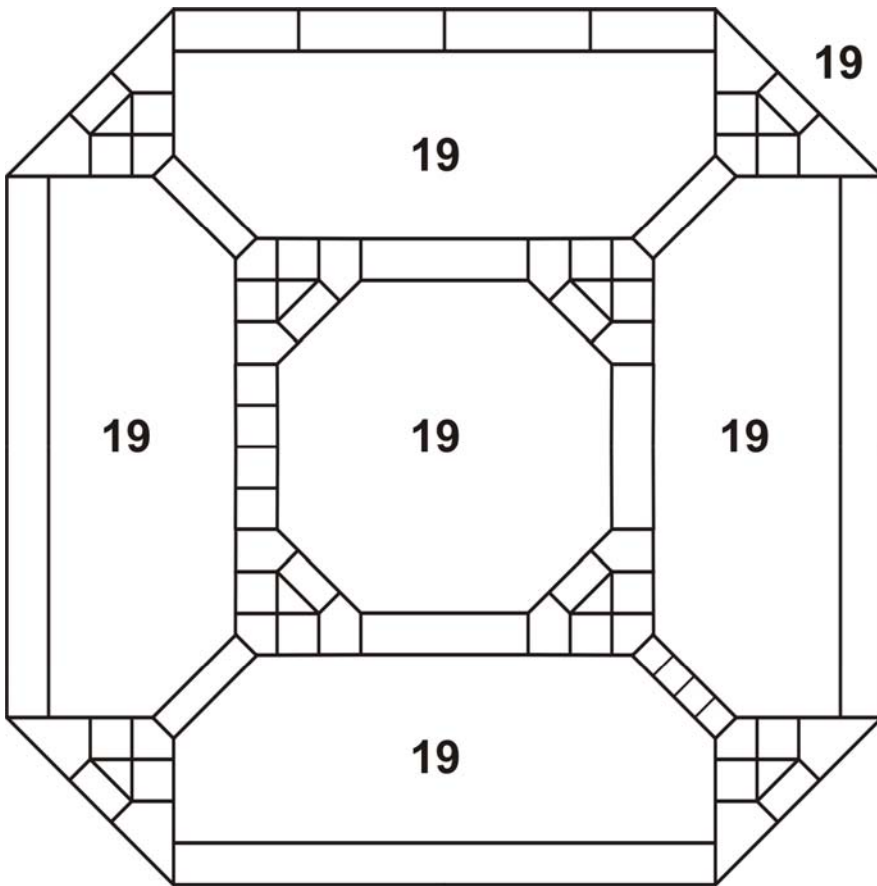
gons). As can be seen, there are 114 trivalent and 24 four-valent vertices. The minimum value of combinatorial curvatures is 1-3/2+ (1/4+1/5+1/19) = 1/380. Based on this observation, it is conjectured that for PCC graphs, the minimum value of combinatorial curvatures is not less than 1/380. If the conjecture is true, a simple consequence is that the upper bound V_{UB} of vertices is 760.

Theorem 5 There is no 3-regular, vertex homogenous PCC graph with vertex number greater than 120.

Proof. It is important to note that a number $0<\Phi_C \leq 1/4$ cannot be the constant positive combinatorial curvature of a VH graph, if $2/\Phi_C$ is not a positive integer. A consequence of this observation is that there are no 3-regular, vertex homogenous PCC graphs composed of 5- and 7-gons only, because $\Phi_C=\Phi(5,5,7)=1-3/2+(1/5+1/5+1/7)=3/70$, consequently $2/\Phi_C=140/3=46,667$.

Starting with the concept outlined in Ref.[1], consider a list of positive integers with $a_1 \leq....\leq a_L$. We say that a vertex X has a vertex pattern $V(a_1,...a_i...., a_L)$ if the faces incident with X may be ordered $f_1,...f_i...f_L$ so that f_i has size a_i for $1\leq i \leq L$. The complete list of the possible vertex patterns (V-patterns) is given in Ref.[1]. We define a vertex pattern $V(a_1,...a_i...., a_L)$ to be "bed" if there is no regular VH graphs which contain $V(a_1,...a_i...., a_L)$. By definition, a V-pattern which is not bed, is called "good". For example, considering 5-regular PCC graphs, V(3,3,3,3,5) is a good V-pattern. This belongs to the graph of the snub icosidodecahedron with 60 vertices which is the "largest" graph among 5-regular PCC graphs.

The candidates of 3-regular vertex-homogenous graphs can be classified into 5 groups which are denoted by $\prod_0, \prod_1, \prod_2, \prod_3$ and $\prod_4$, respectively. $\prod_0$ contains the monofaced, $\prod_1, \prod_2, \prod_3$ contains the bifaced, and $\prod_4$ includes the possible trifaced graphs, respectively.

Group $\prod_0$ includes only the tetrahedron and the dodecahedron, since the cube is a 4-sided prism. The graph of the dodecahedron is the largest PCC graph in Group $\prod_0$.

Group $\prod_1$ contains the following V-patterns: V(3,5,5), V(3,6,6), V(3,7,7), (3,8,8), V(3,9,9), V(3,10,10) and V(3,11,11). It is easy to show that V(3,11,11) is a bed V-pattern. It follows that the graph belonging to V(3,10,10) is considered as the largest 3-regular, vertex-homogenous graph in Group $\prod_1$. This graph represents the truncated dodecahedron with 60 vertices.

Group $\prod_2$ includes only three V-patterns: V(4,5,5), V(4,6,6) and V(4,7,7). It can be verified that V(4,5,5) and V(4,7,7) are bed V-patterns. The only good V-pattern is V(4,6,6) which corresponds to the truncated octahedron with 24 vertices.

Group $\prod_3$ contains five V-patterns: V(5,5,6), V(5,5,7), V(5,5,8), (5,5,9) and V(5,6,6). It is easy to see that the graph belonging to V(5,6,6) is considered as the

largest 3-regular, vertex-homogenous PCC graph in Group $\prod_3$. This is equivalent to the Buckminster fullerene with 60 vertices.

Group $\prod_4$ includes V-patterns represented by $V(c_1,c_2,c_3)$ where inequality $3\leq c_1<c_2<c_3$ holds. It is clear that $V(c_1,c_2,c_3)$ are good V-patterns only if c_1,c_2,c_3 are even integers. This condition is satisfied only for two particular cases: V(4,6,8) and V(4,6,10). The first one corresponds to the truncated cuboctahedron (rhombicuboctahedron) with 48 vertices, while the second one belongs to the great rhombicosidodecahedron with 120 vertices. We can conclude that the largest 3-regular, vertex homogenous PCC graph is represented by the great rhombicosidodecahedron.

Theorem 6 There is no 4-regular, vertex homogenous PCC graph with vertex number greater than 60.

Proof. For 4-regular PCC graphs the possible V-patterns are as follows: V(3,3,3,c) if $3\leq c$, V(3,3,4,c) if $4\leq c\leq 11$, V(3,3,5,c) if $5\leq c\leq 7$ and V(3,4,4,c) if $4\leq c\leq 5$. The graph of the small rhombicosidodecahedron composed of 20 triangles, 30 squares and 12 pentagons is a 4-regular VH graph with 60 vertices. This means that V(3,3,4,5) is a good V-pattern. For 4-regular, vertex homogenous PCC graphs the following relationships are valid:

$$V\Phi_P = 2 \tag{18a}$$

$$2E = 4V = \sum_k kN_k \tag{18b}$$

$$\sum_{k\geq 3}(4-k)N_k = 8 \tag{18c}$$

where Φ_P is the positive combinatorial curvature, E is the number of edges, and N_k is the number of k-sided faces for $k \geq 3$.

Among the possible vertex patterns there are only 4 types for which the vertex numbers of the corresponding 4-regular, VH graphs are greater than 60. These are: V(3,3,4,11), V(3,3,4,10), V(3,3,4,9) and V(3,3,5,7). We will verify that all of them are bed V-patterns. The proof is based on the following considerations. By definition, a k-sided face f_k is called an isolated face, if f_k has no k-sided edge-neighbor faces. It is easy to see that 4-regular, VH graphs characterized by vertex-patterns V(3,3,4,11), V(3,3,4,10), V(3,3,4,9) and V(3,3,5,7) cannot exist if all faces are isolated. We only need to prove that for graphs belonging to patterns V(3,3,4,11), V(3,3,4,10), V(3,3,4,9), V(3,3,5,7) the equality $E=3N_3$ holds. This implies that all faces are isolated. We suppose (for a contradiction) that V(3,3,4,11), V(3,3,4,10), V(3,3,4,9) and V(3,3,5,7) are good V-patterns. Using an extended version of Eq.(6) given in Ref.[4], we can conclude that the corresponding numbers of vertices are: 264, 120, 72 and 210, while the corresponding face numbers are: (N_3=176, N_4=66, N_{11}=24); (N_3=80, N_4=30,

N_{10}=12); (N_3=48, N_4=18, N_9=8) and (N_3=140, N_5=42, N_7=30), respectively. Consequently, we obtain $3N_3$=528=E, $4N_4$=$11N_{11}$=264=V for case V(3,3,4,11); $3N_3$=240=E, $4N_4$=$10N_{10}$=120=V for case V(3,3,4,10); $3N_3$=144=E, $4N_4$=$9N_9$=72=V for case V(3,3,4,9) and $3N_3$=420=E, $5N_5$=$7N_7$=210=V for case V(3,3,5,7). For all cases we have that E=$3N_3$ is fulfilled. We obtain a contradiction to the assumption that V(3,3,4,11), V(3,3,4,10), V(3,3,4,9) and V(3,3,5,7) are good V-patterns.

4   Final Remarks

The classification of polyhedral PCC graphs cannot be regarded complete as yet. In fact, work is in progress in this direction, and we hope that in subsequent publications an account on that will be given. To characterize quantitatively the geometric structure of PCC graphs, we introduced two topological shape factors Λ and Ω defined as

$$\Lambda(G) = \Phi_{max}(G) / \Phi_{min}(G) \tag{19}$$

and
$$\Omega(G) = \frac{380}{39} V(G)\Phi_{min}(G) \tag{20}$$

In Eqs.(19-20), for an arbitrary graph $G \in \Re$, V(G), $\Phi_{max}(G)$ and $\Phi_{min}(G)$ stand for the corresponding vertex number, the maximal and minimal combinatorial curvatures, respectively.
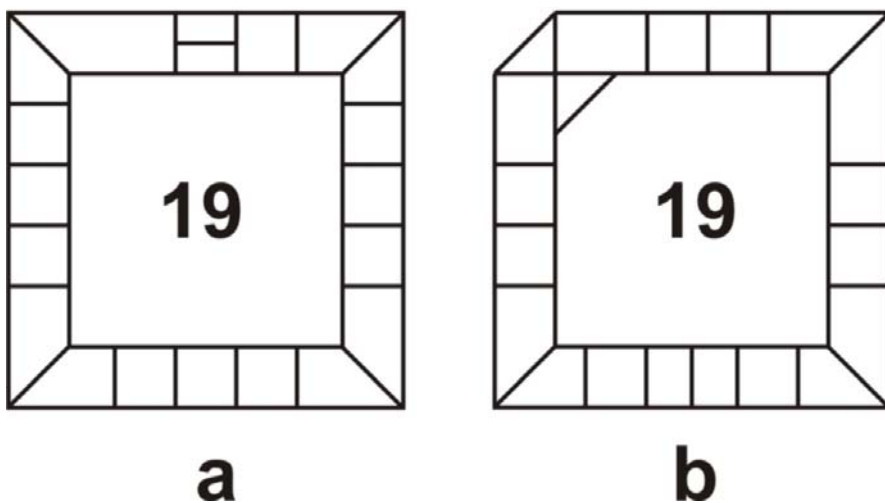

Figure 10
Non vertex-homogenous PCC graphs: a cubic graph with vertex number of 40 (a) and a non-cubic graph with vertex number of 39 (b)

Shape factors $\Lambda(G)$ and $\Omega(G)$ are finite positive numbers, for which relationships $\Lambda \geq 1$ and $\Omega \leq 2*380/39 = 19.487$ are fulfilled. Since $V(G)\Phi_{max}(G) \geq 2 \geq V(G)\Phi_{min}(G)$ we have

$$\Lambda(G)\Omega(G) \geq \frac{760}{39} \geq \frac{\Omega(G)}{\Lambda(G)} \tag{21}$$

It follows that $\Lambda=1$ and $\Omega=760/39=19.487$ if the graph G is vertex-homogenous. We conjecture that for any non vertex-homogeneous PCC graph, the following inequalities are valid:

$$1 < \Lambda(G) \leq 380/5 = 76 \tag{22}$$

and

$$1 \leq \Omega(G) \leq \frac{380*7}{39*4} = 17.0513 \tag{23}$$

If the conjecture represented by Eq.(22) is true, then the upper bound of Λ is sharp. It is reasonable to suppose that the upper bound (i.e. $\Lambda=76$) is valid only for the trivalent PCC graph shown in **Fig. 10a**. Moreover, it is conjectured that the lower and upper bounds of Ω are sharp. We suppose that equality $\Omega=1$ is fulfilled only for the non-cubic PCC graph depicted in **Fig. 10b**. For example, equality $\Omega = 17.0513$ holds for a 7-sided polyhedron where the minimum value of combinatorial curvatures is equal to $1-4/2+(1/3+1/3+1/3+1/4)=1/4$. This heptahedron has 7 vertices, 12 edges, its faces consist of 4 triangles and 3 quadrilaterals [5]. A common feature of PCC graphs shown in **Fig. 9** and **Fig. 10** is that the minimum values of combinatorial curvatures are identical: $1-3/2+1/4+1/5+1/19=1/380$.

Conclusions

According to the conjecture proposed by DeVos and Mohar, for the maximal vertex number V_{max} of a PCC graph (a polyhedral graph with positive combinatorial curvature, which is neither a prism, nor an antiprism), the inequality $120 \leq V_{max} \leq 3444$ is fulfilled. In this paper we verified that the lower bound (120) is not sharp, consequently, it can be improved. Based on theoretical investigations and numerical computations we obtained the following results:

i   For any PCC graph the maximum number of vertices is not less than 138.

ii   There is no trivalent PCC graph in $\mathfrak{R}_S(3,7,b)$ if $b > 11$ positive integer.

iii   There is no trivalent PCC graph in $\mathfrak{R}_S(3,8,b)$ if $b > 22$ positive integer. Moreover, if b=22 there exists a trivalent PCC graph G_x in $\mathfrak{R}_S(3,8,22)$ composed of 3-, 8- and 22-gons, for which $V(G_x)= 66$.

iv   There exists a trivalent PCC graph G_y in $\mathfrak{R}_S(4,6,11)$ for which $V(G_y)=132$.

v   There exists a non-cubic PCC graph which contains 138 vertices.

vi   There are no regular, vertex homogenous PCC graphs with vertex number greater than 120. The graph of the great rhombicosidodecahedron with 120 vertices is the largest regular, vertex homogenous PCC graph.

vii   It is conjectured that there are no PCC graphs having faces with side-number greater than 19, except the two trivalent polyhedral graphs G_{20} and G_{22} containing 20- and 22-sided faces, respectively.

viii   Moreover it is also conjectured that for PCC graphs the minimum value of positive combinatorial curvatures is not less than 1/380.

References

[1]   M. DeVos and B. Mohar: An analogue of the Descartes-Euler formula for infinite graphs and Higuchi's conjecture, submitted in 2004

[2]   Yu. Higuchi: Combinatorial Curvature for Planar Graphs, J. Graph Theory, Vol. **38** (2001) 220-229

[3]   Liang Sun and Xingxing Yu: Positively Curved Cubic Planar Graphs Are Finite, J. Graph Theory, submitted in 2004

[4]   T. Réti and K. Böröczky: Topological Characterization of 2D Cellular Systems, Materials Science Forum, Vols. **414-415** (2003) 471-482

[5]   P. J. Federico: Polyhedra with 4 to 8 faces, Geometriae Dedicata, Vol. **3** (1975) 469-481

Design of the Agent-based Intelligent Control System

Baltazár Frankovič, Viktor Oravec

Institute of Informatics, Slovak Academy of Sciences
Dúbravská 9, 84507 Bratislava, Slovakia
utrrfran@savba.sk, upsyviki@savba.sk

Abstract: This paper introduces the possibilities of multi-agent system application for the modeling and intelligent control in the case of coarse ceramics burning process. It consists of technological description of this process, its decomposition into agents and macro-model of the decision system. Then multi-agent system modeling tools, such as alternating-time temporal logic and alternating transition systems and their epistemic extensions, are presented.

Keywords: multi-agent system, decision system, macro model, coarse ceramic burning production process, design

1   Introduction

For the creation of the intelligent control system (IDC) of production processes, which includes the decision process, suitable tools must be developed. From this point of view the research in this area is oriented to the development of algorithms which use the principle of artificial intelligence. The algorithms should satisfy the following three conditions:

- there is a natural object (in technical practices it could be, for example, different types of technological processes) with the properties of using the perception, decision, reasoning, etc.

- there is an aim to create its duplicate as, for example, the model of the natural object,

- there is a possible way for the realization of the supposed aim (the implementation of the concrete process by man or by some algorithm realized on computer).

The solution of the above problems (evolutionary algorithm [9], neural network [9], decision process [5] [7] [10], game theory [8], multi-agent system [5] [6] [7]

etc.), is the subject of many publications. From our aspect it seems that for the creation of ICS for production system the MAS may have several advantages. MAS can be used to solve a complex problem as, for example:

▪ communication architecture provides the negotiation mechanism

▪ information architecture provides the framework for information modeling on negotiation.

The "mobile agent" based negotiation collects information and makes a decision for itself. It could be said that from logical/functional point of view, an agent-based distributed control system (IDCS) is a systematic network (within or without a hierarchy) of various local decision makers, which have independent knowledge sources such as database systems.

This advantage was, for us, the motivation to using the MAS, as the intelligent tool, for the solution of the decision process as a part of (IDCS) for COARSE CERAMIC BURNING PROCESS (CCBP).

The main scope of this paper is as follows: in section two the significant characteristics of MAS are described from the aspect of CCBP; in section three the CCBP technologies are formulated as an illustrative example of a multi-agent system. In section four a simple decision system was chosen as an application of multi-agent system. Then, the decomposition into agents is made. This section also contains macro-model of decision system. In section five, multi-agent design is discussed. In that section alternating-time temporal logic (ATL), alternating transition systems (ATS) and their extension are given. Section six contains the numerical results obtained from the verification of IDCS on the CCBP mathematical model.

2   The Significant Characteristics of MAS

Multi-agent system (MAS) may be considered as intelligent tool for the solution of such problems as planning, scheduling, decision making and control in the framework of production processes. The MAS approach seems to be the most feasible. It respects the complicated characteristics of the goal that we aim to achieve. There are some significant reasons that motivate us to choose the MAS approach to the solution of decision making, such as:

Modularity: Each agent is an autonomous module and can work without interventions of the external world. Each agent can have different capabilities or functionalities and through cooperation the agents are able to achieve a variety of goals.

Parallelism: The MAS approach enables to work in parallel. A complicated problem could be solved in an acceptable time by using a number of agents, e.g., gathering information from various resources allocated in different places.

Flexibility: The MAS approach is able to react in a flexible manner to each change in the environment. Through cooperation the agents can assist each other to compensate the lack of capability or knowledge. They can share information or own capacity to resolve a newly appeared situation, if one agent is not able to do so. Beside that, each intelligent agent can do reasoning about whom and when it has to cooperate with, in order to achieve the effective performance.

Of course, there are also some difficult questions associated with the MAS approach, e.g., which types of agents are needed, how many agents are optimal, what is a functionality of each agent, cooperation between agents, etc. We will have to successively deal with all these problems during developing this system, but in this paper we focus only on solving the problem how the agent system can assist the user in designing IDCS. We suppose that agents used in IDCS satisfied to the following properties:

- *Autonomy*: Each agent, as mentioned before, thinks and acts locally. It means that agent operates without direct interventions from other agents to achieve its own goals.

- *Social ability*: Agents can cooperate with other agents to achieve common goals.

- *Reactivity*: Agents react on changes in environment; it is needed to describe the negotiation process.

- *Pro-activeness*: Agents do more than response on events generated by environment, they can show goal-directed behavior.

3 Formulation of the Coarse Ceramics Burning Process Technologies – Illustrative Example

Our application is used to control temperature and gas flow in a tunnel kiln during coarse ceramics burning process [3]. The whole process is divided into three phases: heating, cooling and drying. Each product has to proceed through all phases of the process. The heating phase is the main one, where products are burned. After burning, the product proceeds into the cooling phase. When its temperature is the same as at the beginning of the heating phase, the drying phase can begin. This is the final stage, where residual humidity is evaporated.

Since big changes of coarse ceramics temperature could cause product damage, the kiln temperature is controlled in several zones. Each phase can be divided into

zones that differ from each other, i.e. the principle of each zone is the same, but the parameters are different. The number of zones in the heating and cooling phase can vary from time to time, but the drying phase has always two zones. Due to the absence of product temperature measuring, the system has first to control temperature of the kiln and then to wait till the product is heated up to certain temperature. For this approach each zone has to be divided into two stages. The first stage is the temperature changing stage, i.e. the system changes the kiln temperature from one level to another. During the second stage the product is heated to the kiln temperature particular for the given zone, i.e. the system remains in this stage for a certain time. The set of zones is defined by the kiln temperature limiting curve (Figure 1). An example of such curve is shown in the next picture.
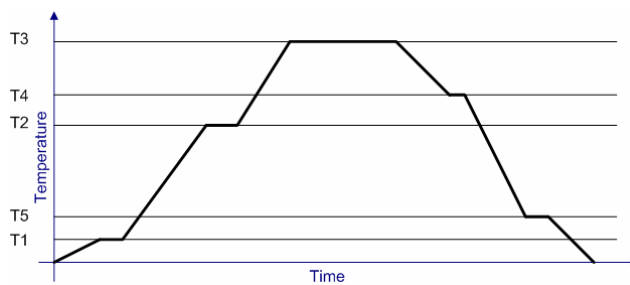
Figure 1
An example of limiting curve

The drying phase is always divided into two zones. During the first stage of the first zone the product is warmed up to the required temperature. Then residual humidity will be evaporated during the next stage and during the first stage of the second zone. During the third stage the product cools down.

Because of the kiln temperature limit curve, an adaptive algorithm with reference signal has to be used to control the temperature of the kiln. This adaptive algorithm is extended by the reference model algorithm, because of kiln parameters change. Due to changes of the kiln parameters, the kiln is "continuously" identified by some identification method.

Control of previously mentioned process is presently implemented with adaptive control algorithm an on-line identification [3]. This conventional approach works, but it is not so flexible as modern approach which is described in this paper. As mentioned in sections 1, 2 above, multi-agent system has several properties, which make the solution more flexible and intelligent. The conventional approach can be used for the described process only.

4   Design of a Decision System

Complete formalization of a decision model consists of two definitions:

- Definition of objective function – Objective function that has same meaning as fitness function in theory of genetic algorithms. Objective function helps decide which alternative in the decision process is the best. In our case, the objective function is the limiting curve.
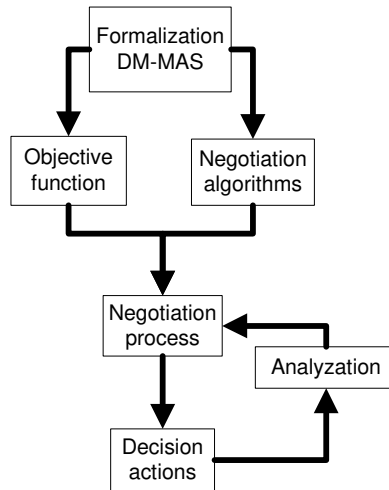
Figure 2
Design of decision system

- Definition of negotiation algorithm – The second definition in decision model formalization is the negotiation algorithm, which includes basic rules how to decide. This definition is included in the following three subsections. The first subsection, vertical cut, is the basic decomposition of the production process described in section 3. Vertical cut is something like domain analysis of the production process, i.e. basic entities are defined. The next subsection is the horizontal cut. More precise analysis is taken in the horizontal cut. Communication between entities is projected, negotiation algorithm is described. Finally, the last subsection present formal description of negotiation process.

4.1   Vertical Cut

The whole system is composed of five levels. The lowest level is the process level consisting of a *process model*, a *reference model* and a limit curve generation model (a *reference signal generator*). The next one is the level of control and

identification algorithms. Note that previous levels are not included in the decision system.

The third (lower decision) level is composed of two decision models (DM), *Stages* and *Identification*. The *Identification* DM does not belong to hierarchical decision system (HDS), i.e. it stands and acts alone in the whole decision system, and its decisions do not depend on decisions made by other DMs. This DM decides whether a process of identification will start or not. The *Stages* DM is the lowest part of our hierarchical decision system. It decides whether the process will advance to the first or second stage.

The second (middle decision) level consists of one decision model - *Zones*. This model decides to which zone the process proceeds. This DM is the highest model in the HDS that makes the decision useful for the control algorithm.

Finally we approached the third decision level. The highest level is composed of one decision model - *Phases*. This DM can be called the observation global supervisor, because it has no influence on the control algorithm and is used only for observation and recognition of the process phases.

4.2 Horizontal Cut

In the previous section the vertical cut of decision system was described. This horizontal cut shows the states inside the DM, and information flows between these nodes in an oriented graph $G = (D, I \cup H, T)$ [1], where D is a set of all *decision nodes* and it is the conjunction of decision node sets of each data model in HDS. I and H denote *inter-level information flow* and *information flow* inside the *same level*, respectively. T represents the event

$$T : (X) \rightarrow (D \times D); X \subseteq I \cup H \qquad (1)$$

and $T = T_C \cup T_{UC}$. $\qquad (2)$

The whole system behaves autonomously, so there is only one *controlled* event, namely is $T_C = f(H_{Start\,Pr\,oces}, u)$, where $u \in U$ is a control action generated by the object of an external interaction - *Start*. All other events are *uncontrollable* $T_{UC} = f(I, H)$.

---

[1] Note that for the above reasons Identification DM is excluded from next considerations
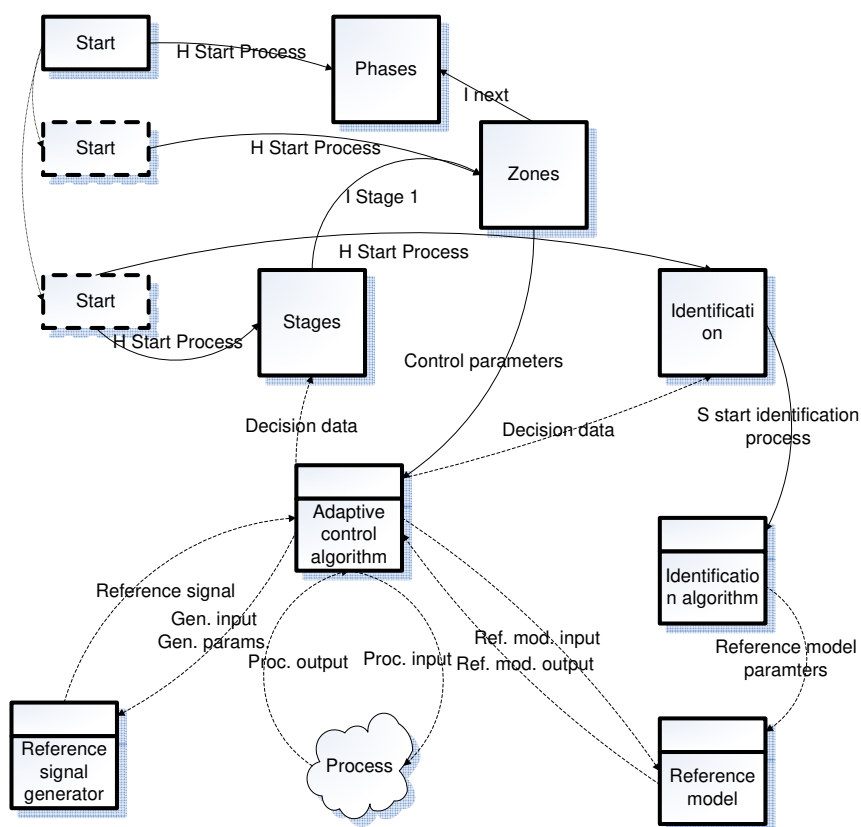
Figure 3
Vertical cut of decision system design
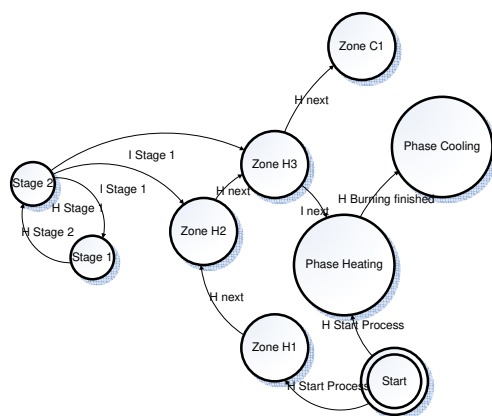
Figure 4
Part of design vertical cut

4.3   Formal Description of the Decision System

The previous considerations resulted in the following sets D, I, H, T, U :

$$D_P = \left\{ D_{heating}, D_{cooling}, D_{drying} \right\}, \tag{3}$$

$$D_Z = \left\{ D_{ZoneH1}, D_{ZoneH2}, ..., D_{ZoneHN_H}, D_{ZoneC1}, D_{ZoneC2}, ..., D_{ZoneCN_C}, D_{ZoneDC}, D_{ZoneDH} \right\} \tag{4}$$

$$D_S = \{ D_{Stage1}, D_{Stage2} \}, \tag{5}$$

$$I = \left\{ I_{Stage1}, I_{Next} \right\} \tag{6}$$

$$H = \left\{ H_{Start\,Pr\,oces}, H_{Stage2}, H_{Stage1}, H_{Next}, H_{HeatFinish}, H_{CoolFinish}, H_{DryFinish} \right\}, \tag{7}$$

$$S = \left\{ S_{StartIdent} \right\}, \tag{8}$$

$$U = \left\{ u \right\}. \tag{9}$$

where elements of D_P, D_Z and D_S are decision nodes of Phases DM, Zones DM and Stages DM, respectively. N_H, N_C stand for the total of decision nodes of the heating and cooling zones, respectively. S is a set of loop-back information and $S_{StartIdent}$ is the information that identification process has to be started. Set T is a set of events generated in the decision system, and each of them represents a mapping. Each of these mappings stands for relation between two decision nodes where the first is event generator and the other is event receiver.

$$T_1 \left[D_{heating}, D_{cooling} \right] \Rightarrow \exists H_{HeatFinish} \wedge \exists I_{Next} \tag{10}$$

$$T_2 \left[D_{cooling}, D_{drying} \right] \Rightarrow \exists H_{CoolFinish} \wedge \exists I_{Next} \tag{11}$$

$$T_3 \left[\left[D_{ZoneHj}, D_{ZoneHj+1} \right] \vee \left[D_{ZoneCm}, D_{ZoneCm+1} \right] \right] \Rightarrow \exists H_{Next} \wedge \neg \exists I_{Stage1}$$
$$j = 1, .., N_H - 1, k = 1, .., N_C - 1 \tag{12}$$

$$T_4 \left[\left[D_{ZoneHN_H}, D_{ZoneC1} \right] \vee \left[D_{ZoneHN_C}, D_{ZoneDH} \right] \right] \Rightarrow \exists H_{Next} \wedge \exists I_{Stage1} \tag{13}$$

$$T_5 \left[D_{Stage1}, D_{Stage2} \right] \Rightarrow \exists H_{Stage2} \tag{14}$$

$$T_6 \left[D_{Stage2}, D_{Stage1} \right] \Rightarrow \exists H_{Stage1} \tag{15}$$

5   Design of CCBP Multi-Agent System

In the previous sections, the problem of decision system design was transformed into that of multi-agent system design. Thus, let us take a look at multi-agent

system design. Multi-agent system can be modeled in various ways, but here an approach is described that was invented for this technology. We present alternating-time temporal logic (ATL) and alternating epistemic transition systems (ATS). In general, two models are used in MAS modeling. The first model is MAS' "behavioral" model, where its behavior is described by ATL formulas. ATL formula is a mathematical formula that represents single systems' behavior, such as "Whenever heating h is finished, then the system will proceed into cooling phase in the next step." This formula can be written in alternating-time temporal logic like this;

$$\langle\langle\ \rangle\rangle \ \square(\text{HeatingIsFinished} \rightarrow \langle\langle\ \rangle\rangle \circ \text{CoolingIsStarted}). \tag{16}$$

Note that in previous formula two propositions were used. Proposition HeatingIs Finished denotes that the system finishes the heating phase in the current step. Proposition CoolingIsStarted is true in the steps in which the system is in the cooling phase. The second model is MAS' "structural" model, where the structure of multi-agent system is captured in particular ATS. Parts of MAS structure are agents, states of agents, transitions between agents' states and propositions. Note that the set of propositions and the set of agents are the same in both models.

ATL and ATS are approaches which assume weak definition of agent, and this fits quite well for this application. But, if AI have to be modeled, epistemic extension of these approaches will have to be used. Epistemic extensions are alternating-time epistemic temporal logic and alternating epistemic transition systems. Agents' knowledge can be modeled in multi agent systems.

Using ATL and ATS or their epistemic extensions has several advantages. The most significant advantage is that the designer can use model checking algorithm to check the behavior of the designed transition system, i.e. to verify whether it is designed as intended. The designer's intention is described by set of ATL formulas.

6 Numerical Results from the Verification of Decision Process in CCBP Case

This section is intended to present practical results of our previous consideration. Experiments made in this section prove that our approaches are correct. Two decision systems can be found out in whole CCBP. The first one is the decision system of discrete event system (agents stages, zones and phases), let us label it as *discrete DS* (see section 4.1, 4.2) [4]. The second one is the decision system of adaptive control algorithm, *continuous DS* (6). Decisions made by discrete DS generate the following parameters:

- steepness of temperature change

- change of temperature (set point)

- time interval for stage 2

- error limit.

These parameters are different for each zone and are valid in whole zone, and are generated by agent Stages. Parameters are represented by loop-back information flow in the negotiation process. Inter-level information flow was represented by set of flows (4), but now it is better to use following function:

$$I : P \rightarrow M ,\tag{17}$$

where M is set of messages and P is set of parameters, i.e. the function is message generator for agents communication.[2] Note that the parameters are symbols, which have value, thus they are not constants. Whole continuous DS consists of eight agents; one of them is the agent Stages is also included in discrete DS (see section 4.1). Agents are divided into to two levels.[3] Higher level consists of Adaptive control algorithm agent (Agent 7) and the agent Stages. As shown in Figure 5, the Stages agent generates messages I(R), I(W), which consist of particular stage parameters. Agent 7 makes the decision. Decisions are dependent on loop-back information flow from lower level ($S(Y_{RM})$, $S(Y_{P})$, $S(Y_{NM})$). These decisions result in inter-level information flow, messages $I(U)$, respectively. This information is received by the agents Process (Agent 1), Model (Agent 2) and Reference Model (Agent 3). The first agent is the one which represents a process. It is precise identification of real process, i.e. in real application it is replaced by real process. The second one, Agent 2, is the agent which does not represent such precise identification of process as Agent 1. The last one, Agent 3, is the agent that represents a reference signal for adaptive algorithm. These three agents generate look-back information for Agent 7, as mentioned before. Look-back information flow can be represented as inter-level information flow by the function

$$S : P \rightarrow M .\tag{18}$$

Items of set P are generated by the following equations:

$$Y_{NM}(n+1) = 1.52359Y_{NM}(n) - 0.6362Y_{NM}(n-1) + 0.02476U(n)\tag{19}$$

$$Y_{RM}(n+1) = 0.95Y_{RM}(n) + 0.05U(n)\tag{20}$$

$$Y_{P}(n+1) = 1.5236Y_{P}(n) - 0.6364Y_{P}(n-1) + 0.0248U(n)\tag{21}$$

---

[2]   Inter-level information flow can be a unfinite set of elements, so it is better to describe it as function.

[3]   Agents are divided into levels with regard to: "Think globally, act locally".

$$U(n+1) = 40.387722(Y_{NM}(n) + 0.05W(n) - Y_P(n) + ...$$
$$+ 0.95Y_{RM}(n) - 1.52359YNM(n) + 0.6362Y_{NM}(n-1)) \tag{22}$$
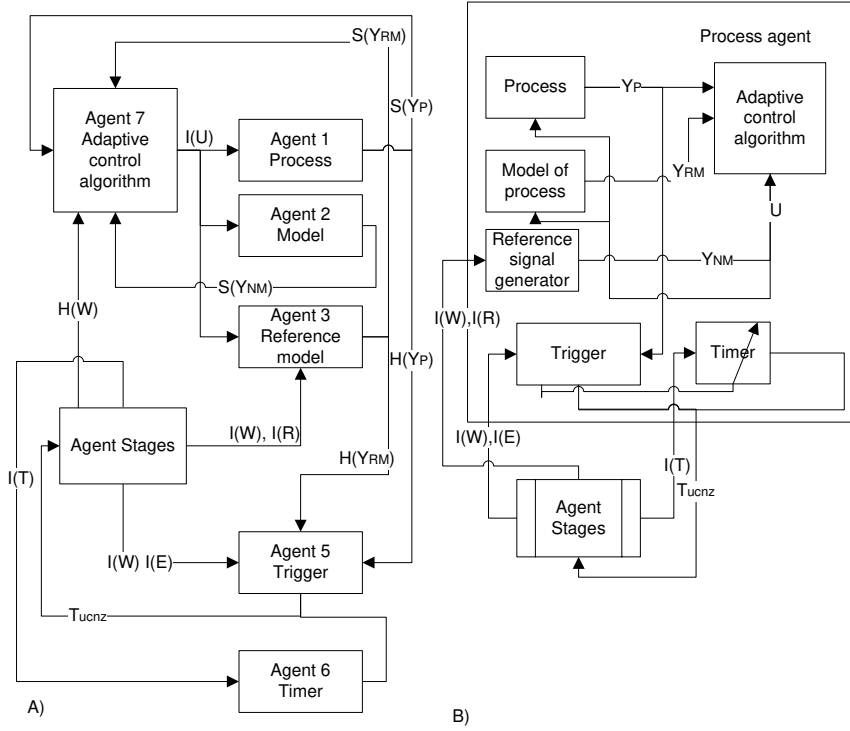
Figure 5

Description of agent based intelligent control system (A) and classical adaptive control system (B)

Two agents – Trigger (Agent 5) and Timer (Agent 6) are involved in continuous DS. These agents generate decision data through look-back information flow, as described in the scheme of vertical scheme (see section 4.1). This event represents the information that the Stage agent can proceed into next stage (1). The same event can be generated by two different agents, thus it has two definitions – one for agent 5, which decided upon level information flows, and one for agent 6, which decided upon time. Mathematical definition of the event is as follows:

$$T_{UCNZ}\left(H\left(Y_{RM}\right), H\left(Y_P\right)\right) = (D_5, D_{Stage1}) \tag{23}$$

and $T_{UCNZ}(I(T)) = (D_6, D_{Stage2})$, $\tag{24}$

where D_5 and D_6 represent the decisions nodes of agents 5 and 6. D_{Stage1}, D_{Stage2} have been defined in previous sections (4.3)(5).

Our experiment includes simulation of zone parameters setup is as follows:

Table 1
Values of zone's parameters

Parameter of particular zone	Value of parameter
steepness of temperature change	0.5 steps
change of temperature (set point)	40°C
time interval for stage 2	80 steps
error limit	0,23°C

Parameters setup results in simulation, which includes 184 simulation steps. First 104 simulation steps are concluded into stage and this is the transition part of the control process. Error limit is satisfied in step 104, thus the system proceeds into next stage and remains in it for next 80 steps. Values of signals are shown in graph.

Figure 6
Graphical representations of experiments results

One can analyze from Figure 6 that control process is stable throughout the whole zone. Quality of control is evaluated by sums of error's square (SES) between signals. SES's between signals are shown in the following tables. One can see that adaptive algorithm is quite good because its uses the model of process parameters

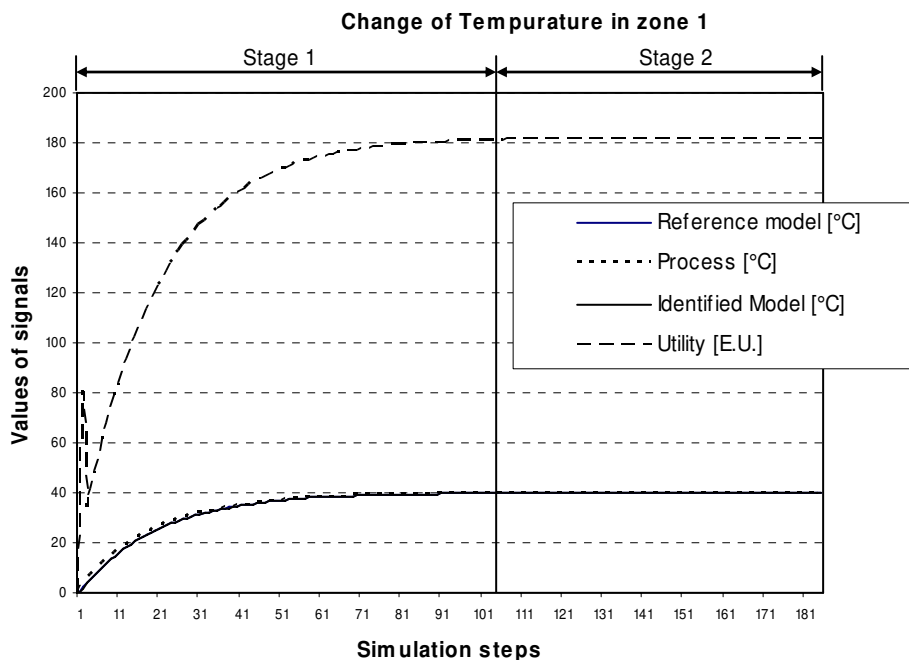to evaluate utility. The process output follows the reference signal, but with bigger error than the model output. This is caused by the absence of continuous identification in the experiment.

Table 2

Relations between system signals Process, Model and Reference model

Signals	SES
Reference model, Model of Process	0.0023159
Reference model, Process	40.72941
Model of Process, Process	41.27523

Knowledge from the decision process verification

The experiment shows that multi-agent system can be used to control continuous dynamic systems. This approach, switching parameters of control algorithm, is similar with gain scheduling well-known in non-linear theory, but multi-agent system are used here. Multi agent systems are more flexible in structure and behaviour, and can be easily advanced with knew features. For instance, agent process can make some reasoning about adaptive and identification algorithm, type of process model, and more. These features can be implemented as separate instances and process agent only switches between them. In another case, multi agent system structure is so flexible that it can be customized to another similar application.

Acknowledgement

Conclusion

The main scope of this paper was to introduce an illustrative example of an application for designing a multi-agent system. This paper presents an application of decision system. The decision system was decomposed into agents, modeled by macro model of control. However, this paper did not aim at designing a multi-agent system, although tools for designing such system were presented.

Macro model of control for a decision system is a quite simple and efficient approach to model decision system. It describes decision and data flow in the whole system. Decisions are modeled by events and data by response on events or by information flow. Then this model can be implemented in multi-agent technology, where agents communicate with each other by messages. Messages in multi-agent system represent events in macro model. Data flow in macro model can be represented by knowledge in multi-agent system. This is very crucial moment, because decision system can be modeled by multi-agent system; a multi-

agent system can be modeled and verified by help of alternating-time temporal epistemic logic and alternating epistemic transition logic.

References

[1]   M. Wooldridge, N. R. Jennings, "Intelligent Agents: Theory and Practice", Knowledge Engineering Review 10(2), 1995

[2]   M. Wooldridge, N. R. Jennings, "Pitfalls of Agent-Oriented Development", Agent '98: Proceedings of the Second International Conference on Autonomous Agents ACM Press, May 1998

[3]   B. Frankovič, „Adaptívne a učiace sa systémy riadenia", Veda, 1982

[4]   V Oravec., J. Fogel, „Cookbook of alternating time systems", 7th Scientific Conference on Electrical Engineering and Information Technologies for PhD. Students, 3 pages, February 2005

[5]   I. Budinská, T. T. Dang: A Case Based Reasoning in a Multi Agents Support System, In Proc. of the 6th International Scientific –Technical Conference, Process Control 2004, June 2004

[6]   I. Budinská, T. T. Dang: Ontology Utilization in MARABu – a Support System for Modelling, Simulation and Control Design, In Proc. of the International Conference on Intelligent Engineering Systems, INES 2004, September 2004

[7]   T. T. Dang: "Improving plan quality through agent coalitions", IEEE International Conference on Computational Cybernetics – ICCC'04, 6. pages, 2004

[8]   Sjoerd Druiven: "Knowledge Development in Games of Imperfect Information", Master Thesis, Institute for Knowledge and Agent Technology, University Maastricht, May 2002

[9]   Mařík V., Štěpánková O., Lažanský J., „Umělá inteligence (3)", Academia, 2001

[10]   Budinská I.: Generovanie trajektórie dynamickoho systému diskrétnych udalostí pomocou GSMP – Perturbačná analýza a jej použitie na optimalizáciu DEDS, Journal AT&P 8/97, 4 pages, august 1997

Material Transport with Air Jet

Dr. István Patkó

Budapest Tech
Doberdó út 6, H-1034 Budapest, Hungary
patko@bmf.hu

Abstract: In the field of industry, there are only a very few examples of material transport with air jet, and one of these is the air jet loom. In this weaving technology, the weft (the transversal yarn of the fabric) is shot by air jet. This paper will set up the mathematical model of yarn end movement. For a special case, I will specify a solution of the model.

1　Brief Description of Air Jet Looms

In air-jet looms, the weft is introduced into the shed opening by air flow.

The energy resulting from air pressure is converted into kinetic energy in the nozzle. The air leaving from the nozzle transfers its pulse to stationary air and slows down.

To this end, in order to achieve a larger rib width, V. Svaty developed in 1947 a confuser, which maintains air velocity in the shooting line.

The confuser drop wires are profiles narrowing in the direction of shoot, and they are of nearly circular cross section open at the top. These drop wires are fitted one behind the other as densely as possible. Therefore, they prevent in the shooting line the dispersion of air jet generated by the nozzle.

Fig. 1.1 shows the arrangement and design of the confuser drop wires applied in machines of the P type, as well as the arrangement schematic of weft intake. The nozzle (1) is secured to the machine frame, and the confuser drop wires (2) and the suction pipe (3) are fixed to the loose reed.

The confuser drop wires are profiles narrowing in the flow direction, and they have a conicity of 6°. These profiles (section A) may be made of metal (Fig. 1.1 a) or plastic (Fig. 1.1 b).

To be considered almost as a closed ring from the aspect of flow, a baffle plate of nearly circular cross section is placed on top of the latter.

In the top part – in comparison with the metal confuser – they substantially reduce the air outflow, and therefore the reduction of air jet velocity will be smaller in the direction of shoot in the confuser drop wires.

The slay (1) is oscillated by a specific drive mechanism, to make sure that during the shoot, the swinging motion of confuser drop wires does not possibly influence the conditions of flow. This is because in case the displacement of the confuser drop wires is large during the shoot, the air flow conditions are unfavourable from the aspect of introducing the weft into the shed, and hence the warps may reach into the inner space of the confuser.
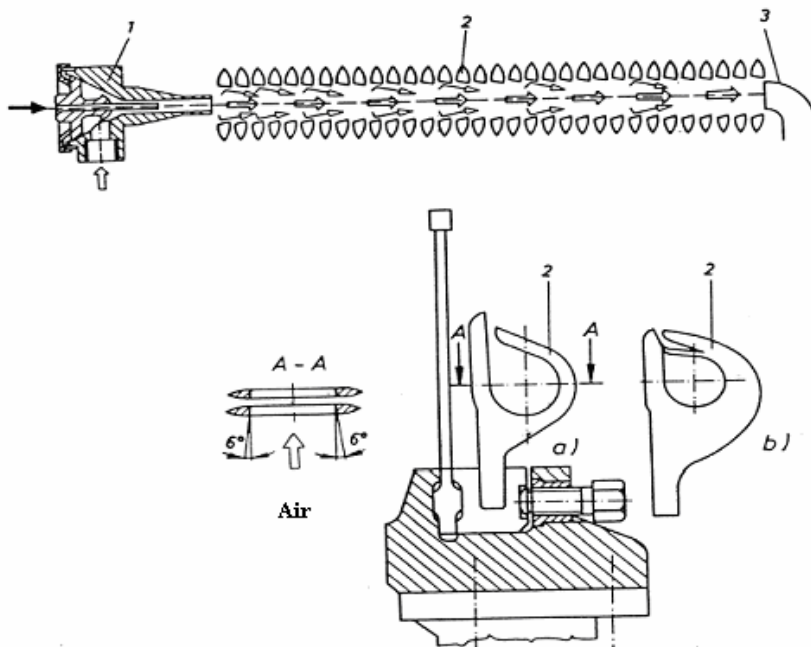
Figure 1.1

Arrangement and design of the confuser drop wires applied in a type P machine

a/   metal confuser and its fixing,   b/ plastic confuser

In a type P machine, the nozzle is secured to the machine frame, while the confuser drop wires swing with the rib (Fig. 1.2).

During the shoot, the nearly stationary position of the slay is ensured by an eccentric articulated movement. In machine P 165 mentioned as an example, the introduction of the weft is carried out in 0.08 sec, four times in a second.
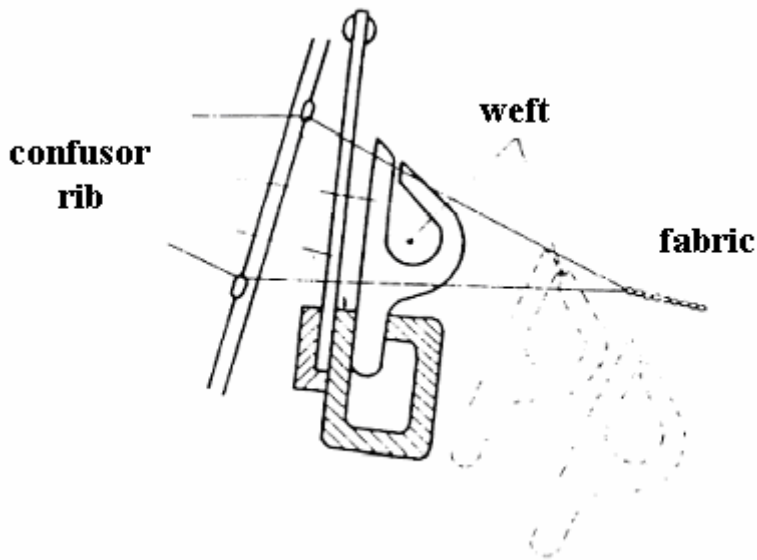
Figure 1.2
Movement of the confuser drop wires

By the application of the confuser drop wires, a rib width of b=165 cm was achieved. This is where the name of loom type P 165 comes from. The confuser drop wires cover 75 to 85% of the rib width.

The design of the loom nozzle is shown in Fig. 1.3, indicating the velocity patterns of the air leaving the nozzle, in addition to the weft.
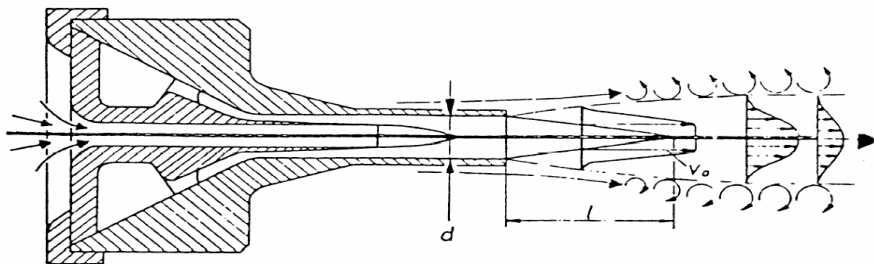
Figure 1.3
Velocity distributions evolving in the nozzle

In front of the nozzle there is a yarn box, the function of which is to store one shoot of yarn. The box is designed in a way that the yarn can be removed from it almost without resistance.

One type of these boxes is the pneumatic yarn box (Fig. 1.4).
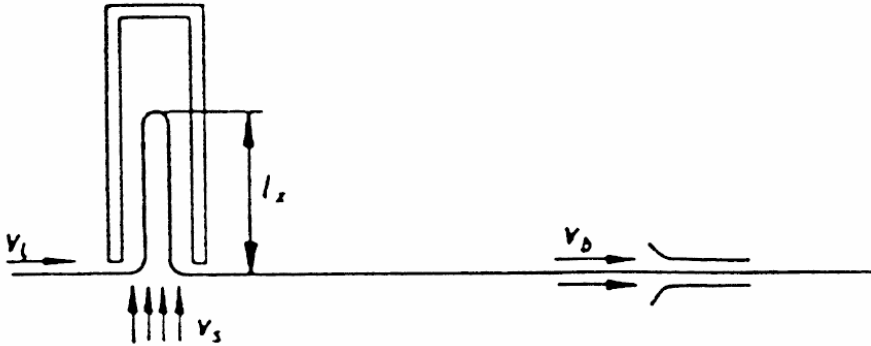
Figure 1.4
Pneumatic yarn box

The pneumatic yarn box has a simple design, and its stores the yarn of specified length in a tube with a slow airflow, in the form of a loop. For introducing the weft, depending on the structure of the yarn, compressed air of 1.5 to 3.0 bar pressure is required.

2 Relationship between Yarn and Air

The equilibrium of forces imposed on the yarn placed into the airflow:

$$\frac{d}{dt}I = \int_{x=0}^{1}dF_1 + F_s \tag{2.1}$$

where

    I: pulse of the yarn

    F_1:   force resulting from the liquid friction of yarn and air

    F_s: force resulting from the friction between yarn and a different solid body.

Hereinafter the friction force F_s will be disregarded, because the yarn comes out of the storing box almost without friction.

Inside the nozzle, the yarn proceeds in the yarn guide, and it is only exposed to the carrying air after leaving the nozzle.

The paper [33] deals with nozzles, in which the yarn proceeds along the nozzle over a distance of approx. 100 to 150mm, and the yarn is already exposed to the air within the nozzle. This paper only deals with the relationship between the flow outside the nozzle and the yarn. Furthermore, using the consideration as a point of

departure that the magnitude of friction force does not depend on the size of the surfaces in friction and that the weft intake is a quasi-stationary process, in addition to concentrating strictly on the relationship between the yarn and the air, the friction force F_s is disregarded.

Therefore, the model to be set up will not be comprehensive, but it will be appropriate for defining the basic characteristics of the yarn/air relationship evolving along the axis of intake.

Hence, the mechanical equilibrium describing the relationship between the weft of the pneumatic loom P165 and the weft intake flow is as follows:

$$\frac{d}{dt}I = \int_{x=0}^{1} dF_1 \tag{2.2}$$

The relationship between the yarn and the air is shown in Fig. 2.1.
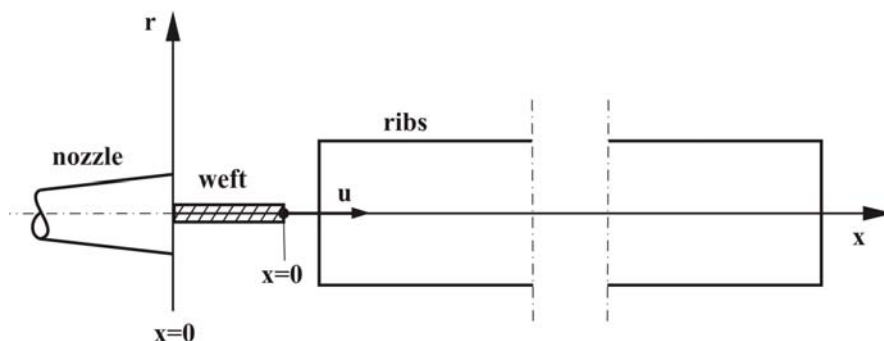
Figure 2.1

Relationship between yarn and air

u: yarn end velocity at the examined point

v: air velocity at the examined point

m: mass of yarn surrounded by air jet

ρ_l: air density

A: length of yarn protruding from the nozzle at the start of intake

C: yarn resistance factor

D_f: characteristic yarn diameter

ρ_f: yarn density

Using the symbols above, equation (2.2) is:

$$\frac{d}{dt}(mu) = \int\limits_{x=0}^{x} dF_f \tag{2.3}$$

where

$$F_f = \frac{\rho_1}{2}CD_f\pi(v-u)^2 x \tag{2.4}$$

with differential rates:

$$dF_f = \frac{\rho_1}{2}CD_f\pi(v-u)^2 dx \tag{2.5}$$

the yarn mass is:

$$m = \rho_f \cdot \frac{D_f^2 \pi}{4} x \tag{2.7}$$

substituting (2.7) and (2.6) into (2.2):

$$\frac{d}{dt}\left(\rho_f \cdot \frac{D_f^2 \pi}{4}xu\right) = \int\limits_{x=0}^{x}\frac{\rho_1}{2}CD_f\pi(v-u)^2 dx \tag{2.8}$$

by introducing the following symbols:

$$\rho_f \cdot \frac{D_f^2 \pi}{4} = Z \ \text{ and } \frac{\rho_1}{2}CD_f\pi = B \tag{2.9}$$

the new form of equation (2.8) is:

$$Z\frac{d}{dt}(xu) = B\int\limits_{x=0}^{x}(v-u)^2 dx \tag{2.10}$$

after performing the specified derivation

$$Z\left(u\frac{dx}{dt} + x\frac{du}{dt}\right) = B\int\limits_{x=0}^{x}(v-u)^2 dx \tag{2.11}$$

$$\frac{dx}{dt} = u$$

The equation describing the movement of yarn end is:

$$Z\left(u^2 + x\frac{du}{dt} \right) = B \int\limits_{x=0}^{x} (v - u)^2 dx \tag{2.12}$$

After making the equation dimensionless by the highest flow rate (V) prevailing in the entrance cross section of the drop wires, and by the length of protruding yarn end (A), the dimensionless kinetic equation of the yarn end is:

$$\left(\frac{u}{V}\right)^2 + \left(\frac{x}{A}\right)\frac{d\left(\frac{u}{V}\right)}{d\left(\frac{tV}{A}\right)} = \frac{AB}{Z} \int\limits_{\left(\frac{x}{A}\right)=0}^{\left(\frac{x}{A}\right)} \left[\left(\frac{v}{V}\right) - \left(\frac{u}{V}\right)\right]^2 d\left(\frac{x}{A}\right) \tag{2.12}$$

The dimensionless quantities featuring in (2.12) are:

$$\frac{u}{V} = u^* ; \qquad \frac{x}{A} = \xi ; \qquad \frac{v}{V} = v^* ; \qquad \frac{tV}{A} = T ; \qquad \frac{AB}{Z} = K$$

With the new symbols, the dimensionless kinetic equation is:

$$u^{*2} + \xi\frac{du^*}{dT} = K \int\limits_{\xi=0}^{\xi} \left(v^* - u^*\right)^2 d\xi \tag{2.13}$$

3   Solution of the Dimensionless Equation Describing the Yarn Movement

$$u^{*2} + \xi\frac{du^*}{dT} = K \int\limits_{\xi=0}^{\xi} \left(v^* - u^*\right)^2 d\xi \tag{3.1}$$

On the right-hand side of the kinetic equation, the integrated square of relative velocity is featured. Due to squaring, the square of relative velocity is always positive, which could lead to calculation troubles. In the course of a computer solution, in order to avoid sign problems resulting from eventual automatisms, it is advisable to modify (3.1) as follows:

$$u^{*2} + \xi\frac{du^*}{dT} = K \int\limits_{\xi=0}^{\xi} \left(v^* - u^*\right)\sqrt{\left(v^* - u^*\right)^2} d\xi \tag{3.2}$$

The dimensionless flow rate (v*) featuring in the equation (3.2) is only subject to the place, i.e.:

$$v^* = f(\xi)$$

while the yarn end velocity is subject to the time, too, i.e.:

$$u^* = f(T; \xi)$$

If it is assumed that the yarn behaves as a rigid body during its movement, when calculating the right-hand side integral in (3.2) $u^* \neq f(\xi)$ i.e. the yarn end velocity is not subject to the place, therefore

$$u^* = f(T)$$

in the course of solving (3.2), the initial condition is:

$$u^* = 0 ; \quad \xi = 1$$

that is

$$\frac{du^*}{dt} = K \int_{\xi=0}^{\xi} v^{*2} d\xi \tag{3.3}$$

By numeric integration from (3.2), at time T = 0, at place ξ= 1 (at the moment of starting the yarn), the initial acceleration $\dfrac{du^*}{dt}$ can be calculated. After a time ΔT has passed – during which the rate of acceleration is assumed to be constant – the position of the yarn end

$$\xi = 1 + \frac{1}{2} \frac{du^*}{dT} \Delta T^2 \tag{3.4}$$

and the velocity of the yarn end

$$u^* = \frac{du^*}{dT} \Delta T \tag{3.5}$$

can be calculated in a rough correlation.

And then, from (3.2), the rate of ξ acceleration applying to place $\dfrac{du^*}{dt}$ can be calculated by the following formula:

$$\frac{du^*}{dT} = \frac{K}{\xi} \int_{\xi=0}^{\xi} \left(v^* - u^*\right) \sqrt{\left(v^* - u^*\right)^2}\, d\xi - \frac{u^{*2}}{\xi} \tag{3.5}$$

the new position of the yarn end is:

$$\xi = \xi_{el\tilde{o}z\tilde{o}} + \frac{1}{2}\frac{du^*}{dT}\Delta T^2 + u^*_{el\tilde{o}z\tilde{o}}\Delta T \tag{3.7}$$

and here the velocity of the yarn end is:

$$u^* = u^*_{el\tilde{o}z\tilde{o}} + \frac{du^*}{dT}\Delta T \tag{3.8}$$

where the 'previous' index is a distinguishing symbol of the place of calculation directly preceding the actual calculation and of the velocity applying there.

In the course of a numeric solution, the equation must be solved by the rate $\Delta T =$ constant, and the change in the velocity of the yarn end subject to the place is to be determined:

$$u^* = f_1(\xi)$$

Next, by reducing the value ΔT, a new velocity distribution can be defined:

$$u^* = f_2(\xi)$$

The solution is to be repeated – always with a smaller ΔT – until the square sum of the deviations of the two solutions following each other will be smaller than a pre-specified rate (H). The value of constants required for the solution is: on the basis of mine measuring [ ].

$$K = \frac{AB}{2} = A\frac{\dfrac{\rho_1}{2}CD_f\pi}{\rho \cdot \dfrac{D_f^2\pi}{4}} = 2\frac{\rho_1}{\rho_f}\frac{AC}{D_f} \tag{3.9}$$

the rate of K is, if:

$\rho_l = 1.2\,kg/m3$

$\rho_f = 800\,kg/m3$

$D_f = 0.002$m

A= 0.02m

C= 0.3

V= 72m/s

$$K = 9 \ 10^{-3}$$

$$T = 20$$

$$H = 0.001$$

The solution is shown in diagram 3.1. The solution resulted in the rates $\Delta T = 6.5$ and H=0.00085. We have made the solution of (3.2) only up to the place (ξ) as long as the following situation prevailed

$$v^* \geq u^*$$

It is shown by diagram 7.1 that the yarn end velocity in the first quarter of the trajectory reaches its nearly constant value, the rate with which it covers most of the trajectory.

Furthermore, in the case of the value $\xi_k = \xi = 55$, the flow rate of the yarn end and of the transport air will be equal, that is:

$$v^* = u^*$$

In the range $1 < \xi = \xi_k$, the velocity of the yarn end is lower than that of the air, and therefore the trajectory of the yarn is definite and straight. We have solved equation (3.2) for this range only, that is:

$$\xi \leq \xi_k$$

In diagram 3.1 in range $\xi > \xi_k$, the rate u* has been shown for guidance only. In this range the yarn does not behave as a rigid body, and therefore the starting equation (2.2) is not suitable for describing the motion. By means of the model set up, the yarn end velocity can be calculated up to the rate of $\xi = \xi_k$.

In the interval $1 < \xi \leq \xi_k = 55$, the air velocity is higher than the yarn end velocity. Therefore, in this section the trajectory of yarn end is straight. In the range $\xi > \xi_k = 0.5$ the yarn end velocity is higher than that of the transport air.

Therefore, in this section of the intake, the trajectory of yarn end cannot be defined as it only depends on spontaneous circumstances. It frequently happens in this range that the yarn end 'comes back' i.e. it is looped. If the suction aperture at the end of the rib is unable to straighten this loop, a weaving defect occurs. If the weaving width is shorter than the critical distance (ξ_k) no weaving defect is generated due to flow technology reasons.
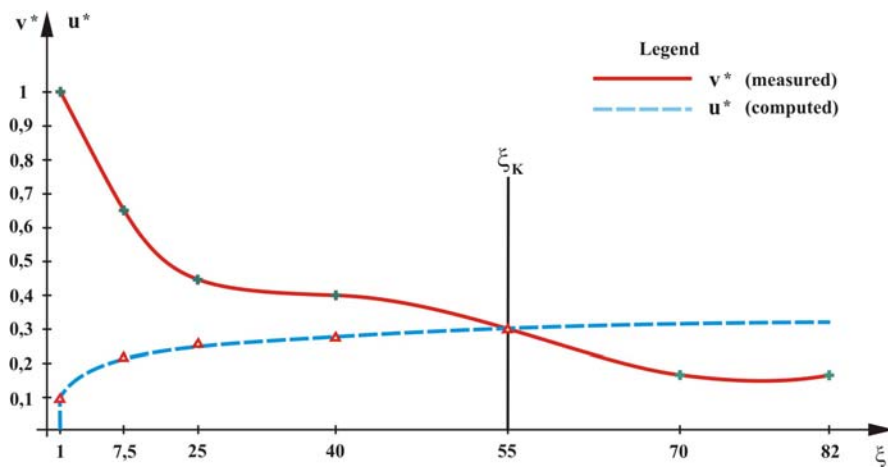
Diagram 3.1

The weaving width of the machine P165 is 165 cm, which corresponds to ξ= 82.5 in our dimensionless system. Consequently, on looms P 165, with technological data associated with K = 0.009 and with the measured – by me - function $v^* = f(\xi)$, the weaving defect resulting from the weft intake may not be (fully) avoided. The only way to exclude a weaving defect is raising the critical distance to above the weaving width (fabric width). There are two possibilities to do so:

-   changing the rate of parameter K,

-   $v^* = f(\xi)$ changing the function relationship.

4   Impact of K on Weft Movement

According to (3.3), the material and flow characteristics of parts contributing to the movement feature in the dimensionless K, like the density of air, the density, diameter and air resistance factor of the yarn, as well as the length (A) of the yarn protruding from the nozzle in a stationary position ($v^* = 0$).

If A = 0 then K = 0, i.e. the yarn may not be started.

The relationship between the initial acceleration valid at the moment of starting the yarn and K according to (3.3) is the following:

$$\frac{du}{dT} = KQ \tag{4.1}$$

Where the result of integration specified by Q is constant.

By increasing the K rate, the initial acceleration of the yarn end can be increased, but at the same time the risk that the yarn end will break off increases. In the calculation shown in diagram 3.1, at $\xi = 1$, the rate of initial acceleration is, which represents with the assumed values the initial acceleration

$$\frac{du}{dt} = 2566 \frac{m}{s^2}$$

of the yarn end.

5   Impact of Air Velocity (V*) on Weft Movement

According to diagram 3.1, the critical distance (ξ_k) increases, if the air velocity (v*) rises along the shooting line.

Increasing the air velocity may be provided by:

-    increasing the supply pressure

-    by designing an appropriate nozzle shape.

Increasing the supply pressure may only be imagined within a short interval, and the extent of increase heavily depends on the material characteristics of the weft.

This is because in case the supply pressure is higher than permissible, the yarn will be torn off the nozzle.

According to the description, increasing the flow rate is only possible with an appropriate nozzle shape (Laval-tube).

Conclusions

I investigated the critical distance (ξ_k) of weaving width. I speified the impact of K – wich was founded by me – and the dimensionless air velocity (v*) on weft movement.

References

[1]   Patkó I.: Lamellák közötti áramlás tulajdonságainak meghatározása, Kandidártusi disszertáció, Budapest, 1994

[2]   KMF Gépészeti Tanszék: A P165 tipusú szövőgép vetülékbeviteli folyamatának fejlesztése, Kutatási jelentés, 1982

[3]   Szabó R.: Szövőgépek, Műszaki Könyvkiadó, 1976

[4]   Alther R.: Automatische Optimierung des Schusseintrages beim, Luftdüsenweben, Kandidátusi disszertáció, ETH, Zürich, 1993

[5]   Lünenschloss J., Wahhoud F. J.: Das Eintragsverhalten verschidener Filamentgarme beim, Idustriellen Luftweben, Textil Praxis Int, 1985

The Kinetics of Cellulose Grafting with Vinyl Acetate Monomer

Éva Borbély

associated professor

Dept. of Packaging and Paper Technology, Budapest Tech
Doberdó út 6, H-1034 Budapest, Hungary
borbely.endrene@rkk.bmf.hu

Abstract: Cellulose is a natural raw material recurring in a great quantity. The demand to use it more and more widely is increasing. The production of cellulose derivates started as early as the 19th century, however the modification of these materials meant the breaking up the fibrous structure, which made their use more difficult in paper industry. The modified cellulose made by graft copolymerization, however, keeps its fibrous character, which provides a great advantage regarding its use. Grafting of industrial cellulose pulp with vinyl-acetate allows for the production of grafted wood cellulose fibres that have a thermoplastic layer on their surface. The binder fibre (fibrid) produced in this way can be excellently used for producing synthetic papers.

In the first part of my experiments I dealt with choosing the parameters of graft copolymerization which are best suited to various uses and after that I studied the dependence of graft reaction on the composition and properties of industrial cellulose applied. The selection of the suitable reaction parameters was followed by the study of reaction speed and activation energy. I have stated that the gross reaction of grafting industrial cellulose with vinyl-acetate monomer is a second order reaction, which is proven by the fact that the invert of the momentary monomer concentration of the system plotted against time is a linear function. The rise of the curves, that is, the reaction speed increases when the temperature in the range of 293–323 K is increasing, while the average activation energy decreases.

Keywords: cellulose, copolymerization, vinyl acetate, reaction parameters, binder fibre, reaction speed, activation energy.

1   Introduction

The behaviour of paper made on a paper machine by planking the fibres – despite the development of ingredients used in paper industry – depends first of all on the physical and chemical properties of cellulose fibres. With the advance of paper-

synthetics combinations, however, several disadvantageous properties of paper due to the behaviour of fibres can be excluded.

There are two basic methods for the completion of paper-synthetics systems:

- the paper or cellulose is associated with finished synthetics or semi-finished synthetic products. Such products are: plywood paper, paper covered with plastics, paper coated with warmish, and paper treated on its surface or inside. For these combinations foil, latex, synthetic rosin hardening with heat, synthetic dispersions, synthetic solutions and synthetic hot-melts can be used.

- the other way to produce paper-synthetics systems is the modification of cellulose itself. This modification can be done by esterification, alkalization, nucleophil substitution, oxidation, or copolymerization.

As the cellulose itself has a polymer structure, it can only produce block or graft copolymers. The synthesis of cellulose block copolymers cannot be widely applied for the chemical modification of cellulose, since it essentially changes the physical structure of cellulose.

The chemical modification of cellulose fibre by graft copolymerization is of great importance because, by grafting a relatively small amount of monomers, this natural macromolecular material can be provided with a number of new advantageous properties. Graft copolymerization makes it possible to produce industrial cellulose fibres which can be successfully applied for making special, marked papers, and it can be also used as binder fibres of synthetic papers.

2   Background to the Research

Cellulose fibre is a natural raw material recurring in a great quantity. The demand to use it more and more widely is increasing. The production of cellulose by-products started as early as the 19th century, however the modification of these materials meant the breaking up the fibrous structure, which made their use more difficult in paper industry. The modified cellulose made by graft copolymerization, however, keeps its fibrous character, which provides a great advantage regarding its use.

Analysing the literature on the techniques of graft copolymerization, we can state that initiation by redox systems for initiating the reaction is the technique that can be carried out in the easiest way in industry. Among these techniques the best seems to be the one in which the role of a reducing component is acted by the cellulose itself, and the oxidizing component is a metal of changing valence.

In most cases the grafting reaction was carried out on regenerated cellulose pulp, fibres and films /1,2,3,4/. In some cases cellulose derivates, above all cellulose

esters and ethers were used. The high price of cellulose and cotton cellulose pulp makes it necessary that a cheaper industrial-technical cellulose fibre should be examined on its grafting ability. The grafting efficiency of industrial cellulose is affected, to a great extent, by the lignin left behind in the cellulose pulp during exploration or bleaching. This lignin acts as an inhibitor in the grafting reaction of the cellulose pulp, since it retards the formation of macro-radicals required for the beginning of grafting. According to the publications it was stated in the publications studied that the greater amount the lignin the greater the induction period of the reaction. /5,6/.

Few papers deal with the effect of the specific surface of the cellulose pulp on grafting /7,8/. I have found no papers dealing with the effects of the incrusting materials of industrial cellulose, the content and origin of hemi cellulose on graft copolymerization.

In Hungary experiments have been going on in the Research Institute of Paper Industry in connection with the grafting of cellulose pulp in paper industry, and experiments related to graft copolymerization in textile industry have been carried on in the Research Institute of Textile Industry, since 1966.

At the Paper Industry Department of the Technical College of Light Industry, which is the predecessor of the Budapest Tech, Rejtő Sándor Faculty of Light Industry, we have been dealing with grafting of industrial cellulose pulp with vinyl monomers since 1974.

3 Effect of Temperature, Reaction Time, Monomer - and Initiator Concentration, Liquor Ratio, and the Composition and Properties of Industrial Cellulose on Grafting Efficiency

In my first experiments I dealt with choosing the parameters of graft copolymerization which are best suited to various uses. Using an initiation method, which has already been studied in literature, that is using Ce(IV) salt, I grafted vinyl-acetate on bleached cellulose and examined the dependence of graft yield on the conditions of reaction, such as, temperature, reaction time, concentration of monomer and initiator, and liquor ratio /9/.

In the case of sulphate-cellulose fibre, which I examined, the reaction parameters of binder fibre, best suited to practical use are:

- temperature: 323 K

- reaction time: 40 minutes

- monomer concentration: 1 mol/dm^3

- initiator concentration: $2 \cdot 10^{-3}$ mol/dm^3

- liquor ratio: 100 cm^3/g cellulose

In the following part of my investigations I studied the dependence of graft reaction on the composition and properties of industrial cellulose applied. I examined the effect of the lignin content, hemi cellulose content and the specific surface of industrial cellulose on graft copolymerization.

I have stated that when grafting industrial cellulose with vinyl-acetate, a sufficient yield can be reached only if the lignin content of cellulose fibre is less than 2%. The hemi cellulose content of industrial cellulose fibre increases the yield of graft copolymerization, the grafting efficiency of the alpha cellulose obtained is lower than that of industrial cellulose containing hemi cellulose too. This can be explained by the greater reactivity and better availability of hemi cellulose. In the case of the examined reaction, the increase of the specific surface also increases – up to a given limit – the reaction yield. My experiments have shown that a further increase of the pulping time results in a decrease of yield because, in the case of an identical initiator concentration, while the specific surface rapidly increases, the number of radicals per unit surface decreases.

4 Study the Kinetics of Grafting Reaction

The selection of the suitable reaction parameters was followed by the study of reaction speed and activation energy /9/.

For studying the kinetics of grafting reaction I carried out the grafting at temperatures 293, 303, 313 and 323 K. The measured grafting efficiency is shown in Table I and Figure 1.

Reaction time (min) / Temperature (K)	293	303	313	323
5	1,98%	4,3%	6,5%	7,6%
10	4,5%	9,2%	16,4%	21,2%
20	9,3%	19,5%	36,9%	52,3%
30	22,6%	55,9%	108,9%	166,3%
40	36,2%	98,0%	181,7%	274,5%
60	61,9%	161,4%	253,1%	366,0%

Table I

Effect of reaction time and temperature on grafting efficiency
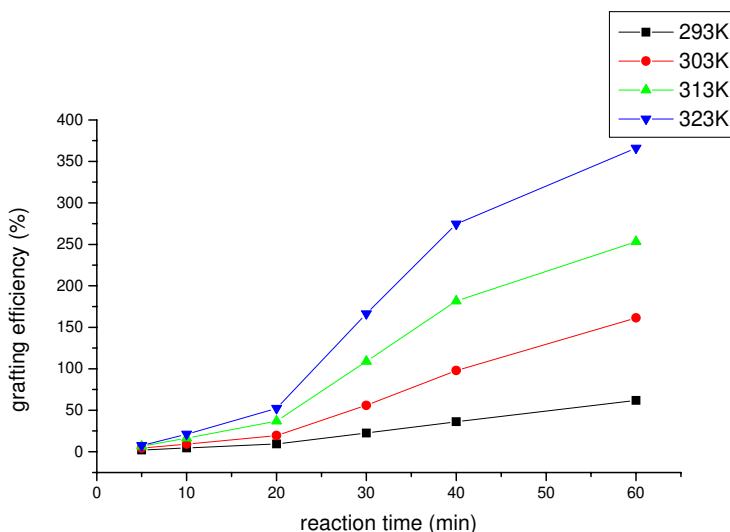
Figure 1

Effect of reaction time and temperature on grafting efficiency. Monomer concentration: 1,0 mól/dm^3, initiator concentration: 2.10^{-3} mól/dm^3, liquor ratio: 100 cm^3/g cellulose

From the results I calculated the average activation energy values for each temperature, using the Arrhenius equation (1), which is well suited to approximate calculations:

$$E = 2{,}3 \cdot R \cdot \frac{T_1 \cdot T_2}{T_2 - T_1} \cdot \lg \frac{k_2}{k_1} \qquad \text{(KJ/mol)} \tag{1}$$

where: T_1 = lower temperature (K)

T_2 = higher temperature (K)

k_1 = reaction speed constant at the lower temperature (mol/dm^3.min)

k_2 = reaction speed constant at the higher temperature (mol/dm^3.min)

The calculation of the reaction rate constants (2) is the following:

$$k_\tau = \frac{c_0 - c_\tau}{\tau \cdot c_0 \cdot c_\tau} \tag{2}$$

where: τ = reaction time (min)

c_0 = initial monomer concentration (mol/dm^3)

c_τ = monomer concentration after τ reaction time (mol/dm^3)

The results have shown that – according to my expectations – when the temperature increases, the average activation energy of the reaction decreases (Table II).

Reaction time (min)	T = 293-303 K (kJ/mól)	T = 303-313 K (kJ/mól)	T = 313-323 K (kJ/mól)
5	$E_5 = 57{,}2863$	$E_5 = 26{,}5167$	$E_5 = 21{,}1128$
10	$E_{10} = 55{,}2730$	$E_{10} = 24{,}4425$	$E_{10} = 23{,}0555$
20	$E_{20} = 58{,}9075$	$E_{20} = 47{,}7685$	$E_{20} = 26{,}0562$
30	$E_{30} = 63{,}4572$	$E_{30} = 57{,}9796$	$E_{30} = 41{,}8575$
40	$E_{40} = 78{,}7772$	$E_{40} = 58{,}2355$	$E_{40} = 41{,}8463$
60	$E_{60} = 79{,}1305$	$E_{60} = 58{,}9343$	$E_{60} = 46{,}4898$

Table II

The average activation energy values

As it has already been mentioned, the reaction goes on in a complicated heterogeneous system thus, apart from the kinetics of the chemical reaction, certain macro-kinetic factors affecting the rate of material transmission may also have an effect on the gross rate of the reaction. From the results of the experiments on parameters I have calculated the gross rate of the reaction in the range of 20–60 reaction time after the induction period (Figure 2, 3, 4, 5), and by analysing the curves, I have defined the equations for the changes of rate in time.

These are:

at temperature 293 K. $v = 2{,}32114 \cdot 10^{-5} \, t^2 - 0{,}00269 \, t + 0{,}09354$

at temperature 303 K. $v = 2{,}45545 \cdot 10^{-5} \, t^2 - 0{,}00283 \, t + 0{,}09528$

at temperature 313 K. $v = 2{,}66932 \cdot 10^{-5} \, t^2 - 0{,}00298 \, t + 0{,}09543$

at temperature 323 K. $v = 3{,}04886 \cdot 10^{-5} \, t^2 - 0{,}00336 \, t + 0{,}10166$
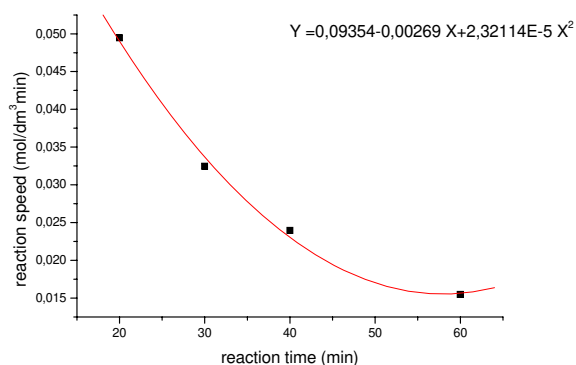
Figure 2

The change in reaction rate with increasing reaction time at temperature 293 K
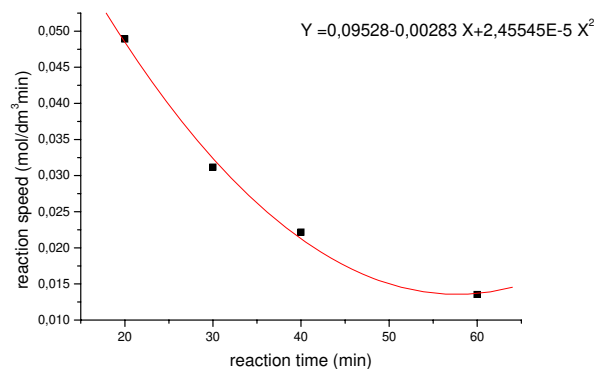
Figure 3

The change in reaction rate with increasing reaction time at temperature 303 K
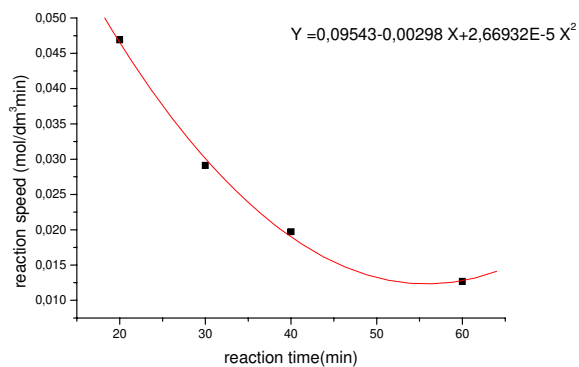
Figure 4

The change in reaction rate with increasing reaction time at temperature 313 K
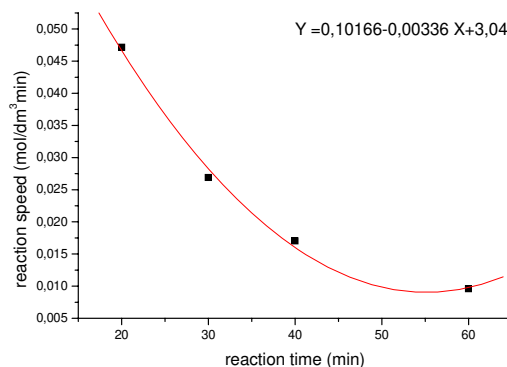
Figure 5

The change in reaction rate with increasing reaction time at temperature 323 K

When the invert of the momentary monomer concentration is plotted against reaction time, it can be stated that the function is a linear one at those four temperatures studied, which proves the fact that the gross reaction rate is a second order reaction (Figure 6, 7, 8, 9).

Figure 6

The change in the inverse of the momentary monomer concentration reaction with increasing reaction time at temperature 293 K

The equation for the change of rate in time at temperature 293 K:

$Y = 0,97714 \, (\pm \, 5,40641 \cdot 10^{-4}) + 0,00167 \, (\pm \, 1,3412 \cdot 10^{-5}) \, X$ $R = 0,99994$
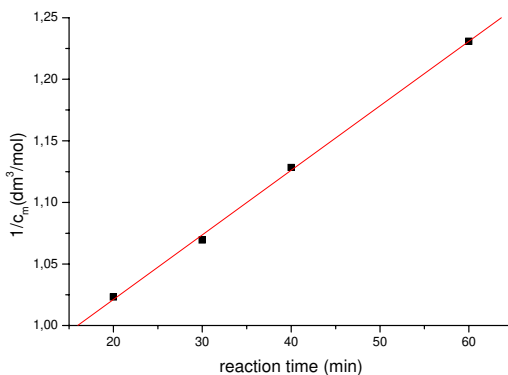
Figure 7
The change in the inverse of the momentary monomer concentration reaction with increasing reaction
time at temperature 303 K

The equation for the change of rate in time at temperature 303 K:

$$Y = 0{,}9164 \ (\pm \ 5{,}06{\cdot}10^{-3}) + 0{,}00524 \ (\pm \ 1{,}2541{\cdot}10^{-4}) \ X \qquad R = 0{,}99943$$
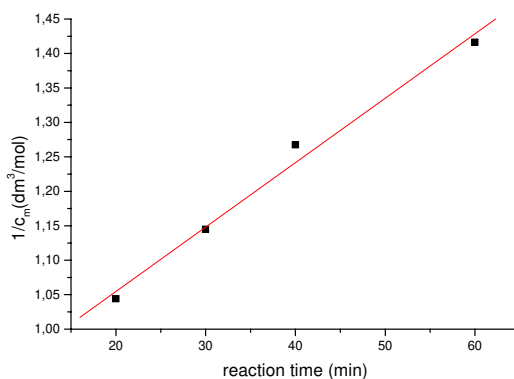
Figure 8
The change in the inverse of the momentary monomer concentration reaction with increasing reaction
time at temperature 313 K

The equation for the change of rate in time at temperature 313 K:

$$Y = 0{,}86759 \ (\pm \ 2{,}963{\cdot}10^{-2}) + 0{,}00935 \ (\pm \ 7{,}34973{\cdot}10^{-4}) \ X \qquad R = 0{,}99388$$
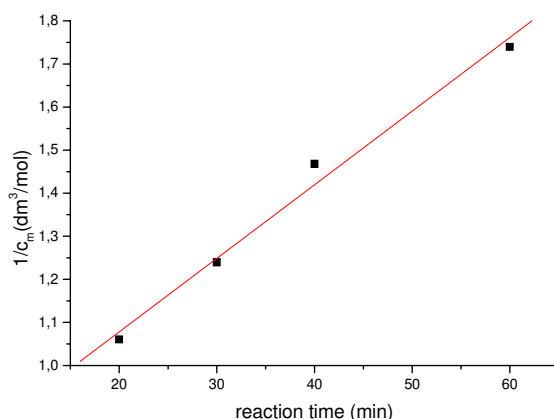
Figure 9

The change in the inverse of the momentary monomer concentration reaction with increasing reaction time at temperature 323 K

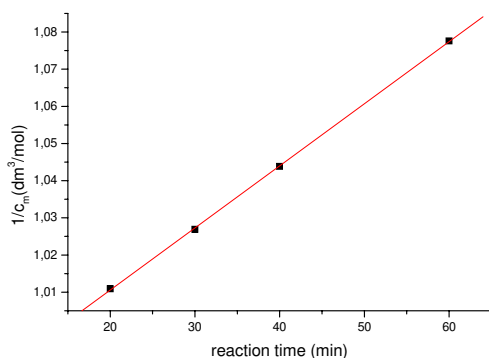The equation for the change of rate in time at temperature 323 K:

$$Y = 0{,}73603\ (\pm\ 5{,}478 \cdot 10^{-2}) + 0{,}01709\ (\pm\ 1{,}360 \cdot 10^{-3})\ X \qquad R = 0{,}99374$$

Conslusions

On the basis of my investigations I have stated that the grafting of industrial cellulose fibre with vinyl-acetate monomer, cerium(IV)-ammonium-sulphate initiator can be carried out. When studying the kinetics of the process I have stated that the gross reaction of grafting industrial cellulose with vinyl-acetate monomer is a second order reaction, which is proven by the fact that the invert of the momentary monomer concentration of the system plotted against time is a linear function. The reaction rate increases when the temperature in the range of 293-323 K is increasing, while the average activation energy decreases.

References

[1]   See, E. G., Bains, M. S.: J. Poly. Sci. 1972, 37, p 125

[2]   Gupta, K. C., Sahoo, S.: J. Appl. Polym. Sci. 2001, 79, p 767

[3]   Flaque, C., Rodrigo, L. C., Ribes-greus, A.: J. Appl. Polym. Sci. 2000, 76, p. 326

[4]   Abdel-Razik, E. A.: J. Photochem Photobio. 1993, 73, p 53

[5]   Zahran, M. K., Mahmoud, R. I.: J. Appl. Polym. Sci. 2003, 87, p. 1879

[6]   Okiemen, F. E., Idehen, K. I.: J. Appl. Polym. Sci. 1989, 37, p 1253

[7]   Phillips, R. B., Quere, J., Guiroy, G. Tappi. 1972, 55, p 858

[8]   Erdelyi, J. Thesis, Budapest, 1993

[9]   Borbely, E. Thesis, Sopron, 2004

Route Elimination Heuristic for Vehicle Routing Problem with Time Windows

Sándor Csiszár

Department of Microelectronics and Technology, Budapest Tech
Tavaszmező u. 17, H-1084 Budapest, Hungary
csiszar.sandor@kvk.bmf.hu

Abstract: The paper deals with the design of a route elimination (RE) algorithm for the vehicle routing problem with time windows (VRPTW). The problem has two objectives, one of them is the minimal number of routes the other is the minimal cost. To cope with these objectives effectively two-phase solutions are often suggested in the relevant literature. In the first phase the main focus is the route elimination, in the second one it is the cost reduction. The algorithm described here is a part of a complete VRPWT study. The method was developed by studying the graph behaviour during the route elimination. For this purpose a model -called "Magic Bricks" was developed. The computation results on the Solomon problem set show that the developed algorithm is competitive with the best ones.

Keywords: Vehicle routing problem, Time windows, Tabu search, Transportation

1 Introduction, Problem Definition

The Vehicle Routing Problem (VRP) has a rich literature, so in this paper only a short introduction of the topic is given. VRP is well known combinatorial NP hard problem having several industrial realizations: VRP with time windows, multi-depot, split delivery or other similar problems such as Travelling Salesman Problem, Bin packing, or Job-shop scheduling. Whether the subject of transportation is the raw-material supply to manufacturers or distribution of end products to vendors a vitally important question – next to the in time delivery and quality of the transportation – is the cost of the logistics service. To solve these problems effectively good logistics management is required including vehicle routing optimization. The lowest number of routes is primarily important because it determines the numbers of vehicles applied, consequently influences the investment and fix cost of the company. The second priority is the minima of the total travel distance. There are studies where the second objective is the minimum schedule time when quick and in time service is more important then the travel

distance. Exact mathematical formulation of VRP can be found in [1]. Problem characteristics are as follows:

- the number of customers, their demand, delivery time windows, service time, customer positions – coordinates – and vehicle capacities are given,

- distances between customers and depot are determined by the Euclidean distances,

- all vehicles start from and arrives at the depot,

- all customers can be visited only once,

- the capacity of vehicles is maximized and uniform and must not be violated,

- the service must be started within the given time window of the customer,

- vehicle travel time constraint is given by the depot time window.

As we know, the application of exact methods in the VRP problem solving is quite limited because of the combinatorial "explosion". During the decades different successful metaheuristics have been developed, for instance Simulated Annealing, Evolutionary Algorithms, and Tabu Search (TS) etc. If we analyse the TS we must admit that despite its indisputable success it has problems in special cases when the route elimination goes together with considerable cost increment. Normally the object function of the TS is designed for finding cheaper solutions. Depending on the length of the tabu list the algorithm is able to reveal new regions. We can in the meantime change the object function and the length of the tabu-list but despite of these techniques it is difficult for the pure TS algorithm to get out from "deep valleys", so the chance for eliminating a route is quite limited and the search is basically guided by the second priority objective. This topic is detailed in [2]. To leave such kind of deep valleys we have to find effective oscillation – sometimes it is called diversification – methods. In the route elimination respect – although it is the primary objective – the pure TS loses to other – lately developed – metaheuristics first of all hybrid metaheuristics [3]. The purpose of this part of the research and this article is to develop an effective route elimination phase.

The remaining part of the paper is structured as follows. Section 2 describes the "Magic Bricks" model and its consequences, Section 3 and 4 explains the developed route elimination procedure while Section 5 is about the computational results on route reduction and finally Section 6 is about experiments conclusions and future plans.

2   The "Magic Bricks" (MB) Model

If we want to study the features of the graph during the search we have to find an appropriate model. The graph itself is not suitable for that because all the

information about the graph is in the nodes and we know that it is impossible or at least not worthy– according to our present knowledge and computers – to reveal all the relationships within the nodes if the number of nodes is above 50. Suppose that we have an initial solution. Let the width of a brick the distance – cost – between two nodes on any route and the waiting time is the gap between the bricks. Similarly a single route can be considered a row of bricks in the wall and the whole number of routes would create a wall. Now the objective of VRP can be redrafted: rebuild the wall to get primarily smaller wall – with fewer routes – and secondly try to reduce the length of the brick-rows. We can easily recognize the unique behaviour of this wall, because if we swap any two bricks in the wall each of them changes its width and maybe the following gaps as well. Moreover, not only the two bricks but their predecessors in the rows change similarly, because their neighbours are changed. So swapping two bricks at least four brick widths will change and in bad situation many gaps are affected. From this respect this wall is exactly as complicated as the graph itself, but if we think of the route elimination we can identify its requirements more clearly. If we select a row for elimination:

a)   the bricks have to be inserted into the gaps or make series of changes to find an appropriate place for a certain brick,

b)   if possible move certain bricks forward or backward (effective in case of wide time windows).

2.1   Consequences of the Model

The point (a) can be satisfied easier if the bricks are narrower – consequently the gaps are wider – that means the chance for successful insertion from a lower cost graph state is better. It must be noted that in case of wide time windows the total waiting time is usually low and the mentioned effect is not significant. This recognition does not mean that from a local minimum it is easy to eliminate a route – the low cost is not a sufficient condition. It means only if we could wander many low cost solutions we could increase the chance of eliminating a route. Based on this idea a new route elimination (RE) procedure was developed where a continuous cost control is applied. It must be emphasized that owing to the continuous cost control the search is inclinable to clog, that is why a fundamental question is how to ensure the a continuous diversification and the cost control parallel. As far as point (b) is concerned, it is realistic in case of wide time windows. This seemingly increases the chance for the route elimination and it is true if the elimination can be achieved from many graph states (solutions), but at the same time the wide time windows are increasing the complexity of the search. The success depends on which of the above mentioned effects is stronger.

2.2 Checking the Model in Practice

Many computation trials were made to check this model on the Solomon problems, although trials were possible only on those problem instances where the initial number of routes were more then the ever found best one. Explanation below supports the concept. Figure 1 shows how the "flexibility" of the graph is changing during the route elimination and increasing after cost reduction. On the vertical axis for instance the numbers of possible insertions are indicated. These numbers can be obtained the following way: select two adjacent nodes and try all possible insertions excluded the selected one and the depot. Summarize these numbers for each possible pairs on the routes. The charts show a strong increment after the cost reduction especially in the insertion numbers.

Table 1 shows the formation of the route elimination data with and without cost reduction based on 25 cycles for each.
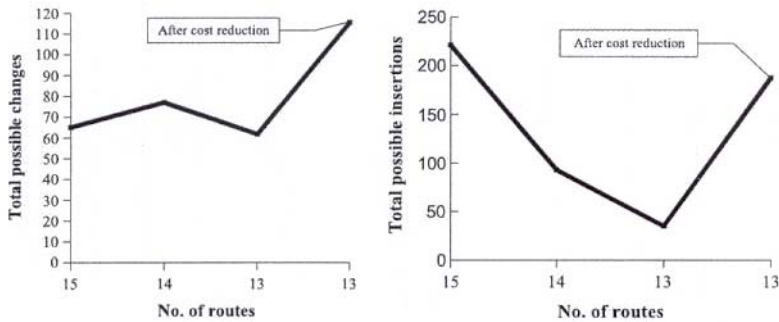
Figure 1

Average insertions and changes on R103 versus number of routes

25 search cycles were made in each case. From low cost solution also the successful ratio was higher and the number of necessary cycles was lower. The computation trials supported the idea to try this concept at the design of a new route elimination algorithm. At the application we must bear in mind two things. One of them is the extra time needs for the cost control the other one is the earlier mentioned clogging problem – or to avoid that a careful diversification is needed.

Notations used in the following part of the article:

r_{max} :   Depot distance of the farthest customer,
ItNo :   Actual number of iterations of the route elimination,
n :   Number of customers,
N_v :   Number of vehicles (routes),
N_r :   Number of customers on the actual route,
C :   Total cost,
$\bar{r}$:   Average customers distance,

i :   Customer identifier, serial number,

r_i :   Depot distance of the customer i,

w_i :   Waiting time at customer i,

Remark: Not explained equations and parameters in the following part of the article derived from computation run experiences.

	Normal search	After cost reduction
Ratio of success	72%	84%
Necessary cycles	349	296

Table 1

Success ration of route elimination with and without cost reduction

3   Main Features of RE

The solution is based on depth-first search. Depth of the search tree depends on the average time constraint: $\overline{TC} = 1/n \sum_{i=1}^{n} TC_i$. In this equation the time constraint factor of customer i is: $TC_i = (t_{li} - t_{ei})/(t_{l0} - t_{e0})$, where t_{ei} and t_{li} are the earliest and latest times to start the service at customer i (depot is costumer 0).

- IF $\left(\overline{TC} \le 0.45\right)$ THEN depth=8

  ELSE IF $\left(\overline{TC} > 0.45\right)$ and $\left(\overline{TC} \le 0.6\right)$ THEN depth=7

  ELSE depth=6

- Depth-first search is executed within given cost limits, otherwise the expensive insertions and changes increase the total cost, damaging the first condition detailed in the MB model. The only exception is the last customer provided all the previous insertions were successful, in this case the cost limit is not considered.

- Until no unsuccessful insertion happens in case of failure a repair procedure is activated after a couple of cost reduction steps.

- After the route elimination, if a limited number of customers remain unrouted – and their time constraint factor and depot distance satisfy certain criteria – a post search is taking place.

- The whole process is guided by Tabu Search for keeping the total cost down and controlling diversification.

- Successful insertions are registered. These data are used for three purposes:

  - route selection for elimination,

  - insertion sequence for depth-first search,

  - diversification made by the TS,

- In case of unsuccessful route elimination, the route – if certain criteria are satisfied – filled up in order to draw away customers from other routes.

4   Detailed Description of the RE Procedure

4.1   Route Selection for Elimination

Three types of root selections are used in the method. The first one selects according to the number of customers on the route (the shorter routes are preferred). The second one takes into account also the insertion frequency of the customers – that can not be used at the beginning of the search – and 65-35% weighting is applied by the following equation:

$$selCrit = \min\left[0.65 N_r N_v / n + 0.35\left(1 - N_v / ItNo\, N_r \sum ins\right)\right] \tag{1}$$

In equation (1) $\sum ins$ is the total successful insertions of customers on the given route. It is important here to compare only relative quantities such as N_r / N_v. The third one selects by the route selection frequency. This latest one prefers those routes that are selected rarely. The route selection is controlled by the block management unit (Figure 4) and its purpose is to ensure the right balance between the diversification and the selection criteria.

4.2   Route Filling up

If the route elimination was not successful and only a few customers remained unrouted – less then $(0.2 \cdots 0.3)\, n / N_v$ – then it seems to be rational to fill the route up as much as possible in order to draw off customers from the other routes and at the same time to increase diversification. The filling up is done by combining parameters in the insertion equation:

$$C = \left[\frac{TC_i}{\overline{TC}}\right]^{\omega}\left[\alpha\left(r_{ik} + r_{ki} - r_{ij}\right) + (1-\alpha)\left(w_a - w_b\right)\right] + \lambda r_k \tag{2}$$

Equation (2) is a modified version of cost equation used at insertion heuristics that takes the time window constraint into account. Detailed description can be found in [4].

4.3   Depth-first Search

If a certain route is selected for elimination all the customers are tried to be inserted somewhere onto other routes. Depth-first search was applied because it effectively supports diversification – an important objective as it was stated earlier. The first task is to determine the insertion sequence for the customers on the selected route. At the beginning of the search the following method is used. If there is a customer whose time constraint factor is lower then 0.1 then this customer is selected otherwise a weighting is used considering the depot distance and the TC factor of the customers:

$$selCrit = r_i / r_{max} \sqrt{\overline{TC} / TC_i} \tag{3}$$

If enough insertion data are available customers with the least successful insertion are tried first. After a costumer has been selected for insertion it is tried first to insert to any possible place with a reasonable cost limit: $(2 \cdots 2.6)\overline{r}$. The purpose of this limit is to avoid drastic cost increment that would hinder further insertions. If this insertion fails try 3-Opt insertions provided the time windows are wide enough. During the initial solution a "learning process is made" the successful intra route 3-Opt insertions are registered and if the success ratio reaches a certain percent the 3-Opt reordering is used – Figure 2 – otherwise not. On Figure 2 a continuous line shows the route before change while a dashed line after that.
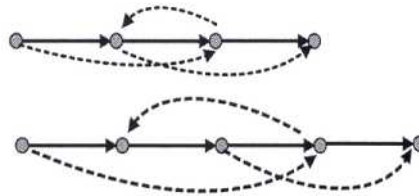
Figure 2
Intra route 3-Opt exchanges

If 3-Opt insertions fail try "Intelligent Reordering" suggested by O. Braysy (detailed description can be found in [5]). The main idea of Intelligent Reordering is to try the insertion of customers to an infeasible – previously unsuccessful but promising – place, the insertion cost calculation is made by:

$$C = \alpha(r_{ik} + r_{ki} - r_{ij}) + (1 - \alpha)(w_a - w_b) + \lambda r_k \tag{4}$$

In Equation $\lambda = 0, \quad \alpha = 0.5$. Select the most promising insertion place using Equation (4) where the customer is inserted to. Then the window violated

customer to be looked for. Finally the window violation is resolved by either reordering customers before the window violation or moving customers from the preceding positions to a new position that follows the violated customer. The number of infeasible trials can be decided by the user. If none of the trials of the given customer were successful then compute all the possible swaps of the customer with the earlier mentioned cost limit and try the whole procedure with the replaced customer. The evolution of maybe circles must be blocked by storing all the already executed swaps on the suitable list. The depth of the search gives the number of consecutive swaps as it was described at the main features of the algorithm. If the insertion is unsuccessful and so far no other unsuccessful insertion has been made a repair algorithm is initiated.

4.4   Repair Algorithm

First the graph has to be modified by the TS algorithm in order to reduce the maybe increased cost and to diversify. (Diversification is detailed at the search management.) After that in a user given angle (+/- 40°) at both sides around the unsuccessfully inserted customer all the routes must be identified in two times 40° sector. Try to combine these routes every possible way according to Figure 3 and at each route combination try the depth-first search again.
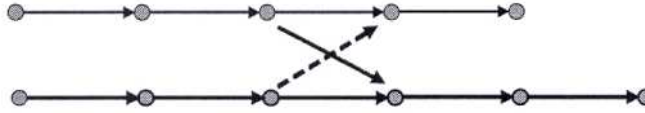
Figure 3
Route combination for repair algorithm

4.5   Post Search

If a reasonable number of customers remained unrouted at the end of the depth-first search a post search is executed. A limited number of customers is allowed for the post search: $\max Cust = 2 + Trunc(0.1n/N_v)$. If this criterion is satisfied further investigation is made, because it would not make sense to spend time if there are customers among the remaining ones they have no or very few successful insertions (see data management). At the beginning of the search – no data available – a similar decision is made based on the TC factors and the depot distances of the remained customers. The total cycle number of the post search depends on the outlook for the success:

$$\max Cycle = \min\{40, Round[B_c \max(0.3, \min TC_r)/\max(0.5, r_r)]\}$$

Here B_c is the post search basic cycle time, $minTC_r$ is the minimum time constraint factor of the remained customers and r_r is maximum relative depot distance of the

remained customers. Between the post-search cycles there are oscillations, those are identical to that one applied at depth-first search.

4.6   Search Data Management, Data Collection and Processing

During the route elimination procedure a list is used to prevent evolution of circles in customer exchange, additionally the successful insertion frequency and the number of route elimination trials per route are registered and processed later. The insertion frequency is used for three purposes:

- route selection for elimination – already described,

- deciding insertion sequence – also discussed,

- diversification by TS, it is done by the search management.

At the initiation of each search block the customer move frequency data – used by the Tabu Search – are adjusted to 100 as a starting number. In RE procedure TS and the route elimination are sequentially running. If a successful insertion occurs at depth-first search also the customer move frequency of Tabu Search is modified in order to move those customers that are not successful at the insertion. This way the Tabu Search finds their move cheaper and prefers their move to reveal new regions for the depth-first search. As it is known the TS penalises frequently moving customers. This is the basic idea of this route elimination concept. This process must be controlled because after a while the graph would turn into an expensive state that would be disadvantageous for the search – according to the MB model. The Search management checks regularly the total cost and compares it to the initial cost. If the relative cost increment is higher then 1.1 – or the user defined value – then the customer move frequency data of TS are readjusted to 100. The 100 value of the adjusted move frequency must be in accordance with the block cycle to get reasonable cost and diversification ratio. See Figure (4).

5   Computation Results

The worked out RE algorithm is written on Delphi platform by dynamic memory programming and was tested on the Solomon Problem Set on 1.7 GHz computer. A maximum search time of 30 minutes and fix configuration parameters were used. There are results in the literature with variable configuration parameters also, but at this research it was not applied. Fix parameters were used for the number of cycles in the block management (c_1=3) and (c_2=5). The cost limit used in the depth-first search and at the cost control cycle was $(2\cdots2.6)\bar{r}$.

In the literature slightly different comparisons are used. Usually the best result is selected from a given (5 - 18 runs). At this study an average value (10 runs) was applied. In this comparison the algorithm gave the ever found best results in the primary objective and proves to be the best one in the fix configuration parameter category. Similar result was achieved by J. Homberger and H. Gehring with variable parameters.
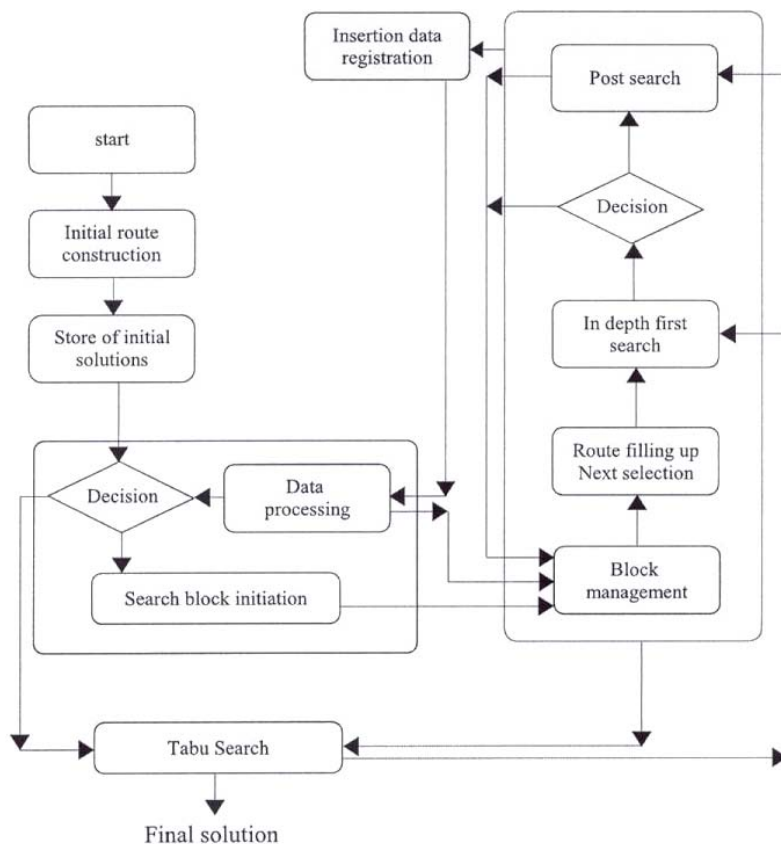
Figure 4
Block scheme of Route Elimination

In 96% of the total 560 runs the best number of routes was found. The computation time of the whole search process, the initial route construction and the second search phase were registered. The computation time was less then half of the lately developed best solutions. It must be noted that the computation time is the less comparable characteristic of the algorithm because it depends not only on the design of the algorithm itself but on the technical data of the computers (RAM, Processor etc.) and programming language, nevertheless the significant computation time difference can not be explained purely by differences in the

computers.At problem R207 the best result was found in 50% of the runs, at R211 in 40%. At problem R104 and R112 the best result was found in 20% of the runs. In Table 2-4 the average number of vehicles (*MNV*) and the best number of vehicles (Best *NV*) of 10 runs can be seen. The updated best results are available on the website: http://www.sintef.no/static/am/opti/projects/top/vrp/bknown.html

Problem	$M\,N\,V$	Best $M\,N\,V$	Problem	$M\,N\,V$	Best $M\,N\,V$
R101	19	19	R201	4	4
R102	17	17	R202	3	3
R103	13	13	R203	3	3
R104	9.8	9	R204	2.1	2
R105	14	14	R205	3	3
R106	12	12	R206	3	3
R107	10	10	R207	2.5	2
R108	9	9	R208	2	2
R109	11	11	R209	3	3
R110	10	10	R210	3	3
R111	10	10	R211	2.6	2
R112	9.8	9	-	-	-
Average	12.05	11.92	Average	2.84	2.73

Table 2
Computational results, R10X, R20X

Problem	$M\,N\,V$	Best $M\,N\,V$	Problem	$M\,N\,V$	Best $M\,N\,V$
C101	10	10	C201	3	3
C102	10	10	C202	3	3
C103	10	10	C203	3	3
C104	10	10	C204	3	3
C105	10	10	C205	3	3
C106	10	10	C206	3	3
C107	10	10	C207	3	3
C108	10	10	C208	3	3
C109	10	10	-	-	-
Average	10	10	Average	3	3

Table 3
Computational results, C10X, C20X

Problem	$M\,N\,V$	Best $M\,N\,V$	Problem	$M\,N\,V$	Best $M\,N\,V$
RC101	14	14	RC201	4	4
RC102	12	12	RC202	3	3
RC103	11	11	RC203	3	3
RC104	10	10	RC204	3	3
RC105	13	13	RC205	4	4
RC106	11	11	RC206	3	3
RC107	11	11	RC207	3	3
RC108	10	10	RC208	3	3
Average	11.5	11.5	Average	3.25	3.25

Table 4
Computational results, RC10X, RC20X

Conclusion, Future Plans

The worked out method is a part of a research work devoted to VRPTW. The Route Elimination algorithm is based on the described MB model. One of the objectives of the study was to find out if the search can be guided by a general property – total cost – of the graph. The precondition of good realization was to wander the low cost solutions and find one where the depth-first search is effective from. For further development a possible way could be to use a simpler cost control algorithm to further reduce the computation time. For larger number of customers (above 200) more sophisticated intelligence is needed in order to decide on the application of deep search because of its time needs.

Acknowledgement

References

[1] O. Bräysy, W. Dullaert: A fast evolutionary metaheuristic for the vehicle routing problem with time windows, Int. J. AI Tools 12 (2003), pp. 153-172

[2] J. Homberger, H. Gehring: A two-phase metaheuristic for the vehicle routing problem with time windows, European Journal of Operation Research 162 (2005), pp. 220-228

[3] O. Bräysy and M. Gendreau: Tabu Search Heuristics for the Vehicle Routing Problem with Time Windows. Sociedad de Estadística e Investigatión Operativa, Madrid, Spain (December, 2002)

[4] S. Csiszár: Initial route construction for Vehicle Routing Problem with Time Windows, XXIInd International Conference "Science in Practice", Schweinfurt (2005)

[5] O. Bräysy: A reactive variable neighbourhood search for the Vehicle-routing problem with time windows. Informs Journal on Computing 15 (2003), pp. 347-368

[6] A. Van Breedam: A parametric analysis of heuristics for the vehicle routing problem with side-constraints, European Journal of Operation Reports 137 (2002), pp. 348-370

[7] G. Clarke, J. W. Wright: Scheduling of Vehicles from a Central Depot to a number of delivery poing. Operation Research 12 (1964), pp. 568-581

[8]   M. M. Solomon: Algorithm for the Vehicle Routing and Scheduling Problem with Time Window Constraints, Operation Research 35 (1987), pp. 254-265

[9]   N. Christofides, A. Mingozzi, P. Toth: The vehicle routing problem, Combinatorial optimization, Chichester, Wiley (1979), pp. 315-338

[10]   K. Altinkemer, B. Gavish: Parallel savings based heuristic for the delivery problem, Operations Research 39 (1991), pp. 456-469

[11]   G. Laporte: The vehicle routing problem: An overview of exact and approximate algorithms, European Journal of Operational Research 59 (1992), pp. 345-358

[12]   E. Taillard: Parallel iterative search methods for vehicle routing problems, Networks 23 (1993), pp. 661-672

Maintenance and Rehabilitation Systems of Infrastructures Management

András Bakó, Kornélia Ambrus-Somogyi

Budapest Tech
Doberdó út 6, H-1034 Budapest, Hungary
bako@bmf.hu, a_somogyi.kornelia@nik.bmf.hu

Abstract: The road network is one the most important element of the infrastructure system. The available budget for the rehabilitation and maintenance usually not enough for holding the system in a certain condition level its whole lifetime.

The paper briefly discusses some elements (monitoring of trial sections, asset value calculation) and several models (urban and motorway PMS, network level, multi-stage highway PMS, BMS). The formal construction of a combined PMS-BMS system is also presented where Markovian type deterioration process is supported and can be solved by Linear Programming.

Keywords: management system, infrastructure management, BMS, PMS, combined optimization model

1   Introduction

The solution of the theoretical and practical problems of Maintenance and Rehabilitation Systems of infrastructure management, as well as the establishment of working systems seem to be an up-to-date task in our age. Several publications were made also in the field of transportation engineering.

In Hungary, the ministry responsible for highway transportation has revealed the importance of the completion of a Transport Asset Management System, and initiated the establishment of several system elements and models in infrastructure management.

The paper briefly discusses some elements (monitoring of trial sections, asset value calculation) and several models (urban and motorway PMS, network level, multi-stage highway PMS, BMS). The theoretical background of a combined PMS-BMS model, as well some basic information of Asset Management is also presented.

2 Analysis of the Condition Monitoring Data of Trial Sections

Although the systematic (yearly) sufficiency rating provides a kind of information about the performance of our national highway network of 30 000 km total length, it can not result in reliable performance models due to the relative infrequency of the measurement of some condition parameters (bearing capacity, unevenness etc). That is why, in 1991 some 60 trial (experimental) sections of 500 m length each were selected of the national highway network. They represent the typical (flexible, superflexible-macadam type, semi rigid) pavement structure types, the characteristic (0-1500, 1501-3000, min. 3001 vehicle unit/day) traffic volumes and typical (max 3%, 4-5% and min. 6%) subgrade CBR-values. The trial sections were classified into 14 road section categories. Four condition parameters (bearing capacity, surface defects, unevenness, rut depth) are yearly monitored. The time-series data of the trial sections within a road section category are used to determine the relevant performance models. The regression functions best fitted to the measuring points constitute the performance models (the generalised deterioration curves) for each road section category and condition parameter. The performance models can be the function of pavement age (since the construction or the past rehabilitation) or the traffic passed on the section [5,6].

The already 12-year monitoring of trial sections allowed evaluating the actual condition improving effect of various road maintenance techniques (e.g. pavement strengthening, thin asphalt laying, surface dressing). More importantly, the life cycle after intervention is also investigated to compare its performance with that of the earlier life cycle immediately after the new construction.

All of the information gained during trial section monitoring are used to the establishment of more and more reliable PMS models [4].

3 Asset Value Calculation of the Road Network

An asset management system badly needs reliable information about the asset value. That is why, in Hungary the systematic evaluation of the gross and the net values of the whole national highway network has been performed since 1981. The gross value means the value of a new facility (reproduction value), while the net value is the difference between the gross value and the depreciation meanwhile that is the actual asset value.

The Hungarian asset value calculation method considers physical wear (condition worsening) and economic depreciation (due to the continuous technical development).

The following average life times were taken into account in the net value calculation:

  ♦   earthworks (subgrades)   90 years,

  ♦   pavement structure   36 years,

  ♦   other road elements   45 years,

  ♦   bridges   60 years.

The residual asset value was considered as 10% of gross value for each road element.

The so-called net factor of pavement structure can be calculated in accordance with Table 1 using the 5-grade notes of various condition parameters. (Note 1 is the best condition level, while note 5 the worst one).

Condition parameter notes	Net factors for				
	bearing capacity	surface defects	unevenness	rut depth	surface texture
1	1,00	1,00	1,00	1,00	1,00
2	0,89	0,94	0,88	0,96	0,98
3	0,78	0,87	0,75	0,92	0,96
4	0,66	0,80	0,50	0,87	0,94
5	0,53	0,72	0,43	0,81	0,91

Table 1

Net factors of various condition parameter notes

It can be seen from the above table that the following weights were applied for the condition parameters: unevenness: bearing capacity: surface defects: rut depth: texture = 6 : 5 : 3 : 2 : 1.

The net factors for earthworks and other road elements can be deducted from that of pavement structure [5].

The level of economic depreciation is characterised by the actual traffic safety situation of the road section during the past 3 years, the capacity factors and the actual vehicle operating costs compared to the optimum one.

The net factor for bridges depends on the following factors: bridge age, traffic volume, and use of ice-melting salt, technical condition level and economic depreciation. Linear yearly depreciation factors (%) are applied.

Table II shows the changing of the %-ratios of gross and net asset values between 1981 and 2001. The values for roads, bridges and the whole road network are presented separately.

Year	Roads			Bridges			Road network		
	Gross (G)	Net (N)	N/G	Gross (G)	Net (N)	N/G	Gross (G)	Net (N)	N/G
	value (10^9 HUF)		(%)	value (10^9 HUF)		(%)	value (10^9 HUF)		(%)
1981	195.9	130.6	66.7	16.8	12.7	75.6	212.7	143.3	67.4
1986	279.8	172.7	61.7	25.8	18.2	70.5	305.6	190.9	62.5
1990	1535.7	922.7	60.1	131.3	74.7	56.9	1667.0	997.4	59.8
2000	5802.0	3623.6	62.5	300.8	173.9	55.8	6113.8	3797.5	62.1
2002	6412.3	3984.6	62.1	331.3	180.9	546	6743.6	4165.6	61.8

Table 2

Asset values of Hungarian roads and bridges between 1981 and 2002

The time-series data of N/G (%) asset value ratios proves that the financial means available between 1981 and 1990 were much below the level needed for the asset preservation of the national road network. While the funds for road construction, rehabilitation and maintenance during the period between 1990 and 2000 permitted a slight improvement of N/G (%), although it did not attain the 1981-level yet. Lately, the worsening of this ratio has been registered again.

4   Urban Road PMS Model

A project level Pavement Management System was developed in Hungary for urban road networks in the 90's [3]. The following tasks were aimed to be solved:

♦ optimum scheduling of highway maintenance activities,

♦ selection of the optimum maintenance techniques for certain highway sections,

♦ project ranking in a certain funds limit.

An analysis period of 10 years was selected, using rolling-type planning technique.

The main part of the model is the ranking subsystem, which produces a priority list for the interventions based on the results of cost-benefit analysis. In order to estimate the costs, performance models and optimum intervention techniques were determined for each condition variant. Three grades of the following condition parameters were applied: bearing capacity, unevenness, surface defects. Various performance models (deterioration processes) were selected as a function of initial condition, traffic size and pavement type. Also the yearly condition grades are forecasted for the 10-year investigation period if the optimum intervention is performed. Similar condition grades are forecasted for the patching needs for the period if only local interventions are performed. Using these kinds of information,

the total costs of optimum and non-optimum variants for every section can be calculated. The benefits of the interventions are considered by the help of the changing in the vehicle operating costs (actually its main element the fuel consumption cost). Finnish measurement results [7, 8] were utilized in the estimation of fuel consumption cost reduction. The idea of efficiency calculation is that the costs of optimum interventions are compared to the extra vehicle operating costs due to the postponement of this optimum technique. The resulting cost/benefit ratios of various projects supply the intervention ranking order.

5   Motorway Pavement Management Model

In 1995-96, the model of a special Motorway Pavement Management System (APMS) was developed in Hungary [2].

The APMS has the following main tasks: determination of the rehabilitation activities and costs on the Hungarian motorway (expressway) system,

♦   priority ranking of interventions on a basis of the national economy in case of limited funds,

♦   pavement performance prediction at a given funds level.

A priority ranking type model was selected which can be used for both network level and project level cases. Three pavement types, three-five condition parameters (depending on the pavement type), four intervention types, two traffic categories were applied. An investigation period of 25 years was chosen.

Expert elicitation was used to determine the most important technical, economic and organizational parameters.

Two main strategies are considered. One systematically selects the recommended type of intervention, and the other exclusively applies routine maintenance. Their long-term consequences are compared, and the results supply the basis for project ranking.

Deterioration tables, intervention tables, intervention cost tables and user cost tables were compiled and used.

The optimum criterion is the lowest national economy cost (agency + user costs) during the whole investigation period.

The output of the system is a list of expressway sections where repair is recommended, together with the types of intervention ranked in a descending order of benefit/cost parameters. It permits every action where the benefit expressed in net present value exceeds its costs.

The model can be used for the estimation of the required funds, for funds split (distribution) and for the calculation of economic loss due to the delay of economically efficient projects.

6   Network Level Multi-Stage PMS-Model

The first single-stage network level optimisation model was developed in Hungary in the late 1980's. Its multi-stage version, the PMMS, was created in 1991.

In spite of its mathematical "elegance", the model had several "infantile disorders" which resulted from the number of intervention types, which was too low (only 3), and from some problems related to the deterioration and the cost model.

The road administration needed quick and practical results. That is why the Finnish HIPS model (see Männistö) was chosen, because the experiences gained over several years were available.

The recently developed HUPMS-model [3] has been created using the optimisation procedure of MPMS and the model structure of HIPS [1], which forms one component of the combined pavement/bridge model.

The main features of the model are:

- several (a maximum of 10) time periods (stages);

- 2 pavement types (asphalt concrete and asphalt macadam);

- 3 traffic categories;

- 4 condition parameters (unevenness, bearing capacity, rut depth, surface defects);

- combined target function;

- a maximum of 8 intervention types.

In the long-term model, the optimum solution is sought for the distribution of pavement condition in the network which can be attained after the optimum interventions; it is the Markov-stable condition. The target function is the minimum of the sum of agency and user costs (i.e. social total optimum).

For asphalt concrete roads the following traffic categories were chosen: 0-1500; 1501-6000; min 6001 pcu/day. The traffic categories for asphalt macadam roads were: 0-500; 501-1000; min 1001 pcu/day.

Interventions possible for asphalt concrete roads are: routine maintenance, patching, rut repair, surface dressing, thin asphalt layers, asphalt overlay, and reconstruction. The interventions for asphalt macadam roads are: routine

maintenance, patching, surface dressing, profile repair, asphalt overlay, reconstruction.

The Markov transition probability matrices are used for determine the deterioration processes. The mathematical model contains of several conditions which are related to the Markov stability, the initial and latter years condition state distribution, proportions of the different condition states, cost bounds etc.

In the case of the multi-stage model, one of the objectives could be to reach a stable model result by means of an approximation over a period of several years. The number of time periods is generally 10, and the model gives the necessary interventions in each period. The same designations are applied in the mathematical formulation of the multi-stage model as were selected for the single-stage model, except for the additional period index t, with values 1,2-10. The objective is the weighted combination of the intervention and the user cost.

7   Bridge Management Systems

Several elements of a bridge management system had been developed and implemented in Hungary already in the 70's and 80's, but it was only in 1993 that the Ministry of Transport decided to adapt the American PONTIS bridge management system to Hungarian conditions. Since then, the Bridge Management Task Group has investigated the original model and agreed to adapt it (bridge elements, actions, unit costs, deterioration matrices etc). Several trial runs have taken place using the PONTIS-H(ungary) models, the results of which could be utilized by the decision makers in several bridge management problems. The first PONTIS-type inspection of the whole bridge stock (some 6000 projects) was completed in 1999. Control surveys were done and are planned for the future to enhance the reliability of this activity. Several further development activities of the system are planned.

- it can manage the whole country-wide bridge stock using comprehensive, uniform and reliable data base,

- it is a valuable decision-supporting tool at network level, although the system has several project features, as well,

- it covers not only the MR&R needs but also the reconstruction ones,

- the PONTIS-type bridge inspection provides detailed and reliable information about the condition state distribution of the bridge stock,

- the bridge deterioration is treated on a high mathematical level,

- its optimisation procedure is based on the analysis of actions, costs and environments,

- its outputs are clear and readily utilizable by the user.

At the same time, Hungary seemed to be appropriate for the PONTIS adaptation due to

- the bridge management elements already available,

- detailed inspection methodology applied,

- need of and decision about the clearance of MR&R backlog,

reliable bridge data bank.

When making the preparatory activities for the use of PONTIS under Hungarian conditions – after having decided about the bridge elements, the condition states and the environment – the transition probability matrices related to the partly new bridge elements had to be developed. These matrix elements reflect the deterioration of the elements in the case of "do nothing" and the improvement of the bridge element in the case of "do something".

In addition to the actual adaptation activities, the members of Hungarian Bridge management Committee have revealed several limitations and drawbacks of the American PONTIS, and the possibility of their solution is considered when developing the new PONTIS-H model. These problems are as follows:
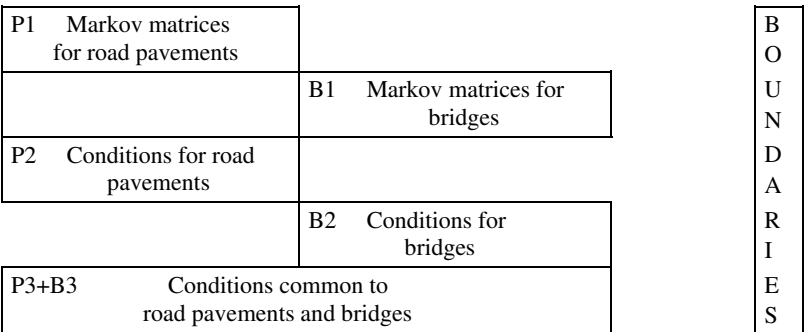
- Reduction of bridge element number (rare or low cost elements)

- Managing of interconnected bridge elements (bridge deck – waterproofing – slab; main girder - open expansion joint – bearing – abutment; bridge footpath structural elements – footpath surfacing).

- Managing of standard bridge spans. A model containing standardized bridge spans is much more realistic than the traditional one with separate bridge elements. In the Hungarian model, bridges are described using standard bridge spans. The network-level optimisation utilizes these spans, the actions are given as a function of condition state for the bridge spans over a period of 30 years for the model.

- Total optimisation (HUBMS) model. The original American PONTIS optimises each bridge element separately. It is evident mathematically that the optimum for each element and the optimum when all bridge elements are optimised in a single model provides different results with the same constraints. Since the model has common constraints (e.g., cost constraint) – excluding the Markov-stable model – in the model the total optimum should be aimed at. The HUBMS (Hungarian Bridge Management System) optimises all the "simple" and interrelated bridge elements simultaneously for a 30-year period.

8   Combined PMS-BMS Model

In the majority of countries – including Hungary – the PMS (Pavement Management System) and BMS (Bridge Management System) operate independently. However, their interdependence is obvious since the bridge surfacing constitutes part of the road pavement. Very often their financial sources are also identical (e.g. Road Funds) contributing to the need for more or less common management. Both PMS and BMS apply the same concept and application of system technology and require a system output function that can be optimised in relation to the benefits and costs.

In Hungary, both the adapted PONTIS-H Bridge management System and the HIPS-HUPMS network level Pavement Management System are based on the use of Markov transition probability matrices. As a result, their identical structure allows the joint optimisation of both systems. This activity is especially important when the aim is the distribution of the funds available between the two infrastructure elements (road pavements and bridges).

The mathematical-engineering model of this BMS-PMS(PBMS), [2] common model has already been completed. Its implementation is planned for the near future.

P1 Markov matrices for road pavements		B
	B1 Markov matrices for bridges	O
P2 Conditions for road pavements		U N
	B2 Conditions for bridges	D A
P3+B3 Conditions common to road pavements and bridges		R I E S

Target functions

P4+B4   User costs   MIN!

P5+B5   Intervention costs   MIN!

P6+B6   Weighted intervention and user costs   MIN!

Figure 1

Combined model of management of road pavement and bridges

The structure of this model is presented in Figure 1. It has two columns. The first one (P1 and P2) contains the elements of the HUPMS model. In the right-hand

column the relevant BMS conditions can be seen (B1 and B2). The PBMS model can also have common conditions, for example relating to the annual sum, which is commonly available (P3 and B3).

The objective is the sum of the object functions of pavement and bridge models. The object can be here the minimisation of the intervention costs (P4 + B4), the minimisation of user costs (P5+B5) or the weighted sum of these costs (P6 + B6). By varying the weight of the parts, any arbitrary combination of the target function can be produced, for example, the minimisation of the sum of road (pavement) user costs and bridge intervention costs.

9   The Basic of Asset Management System

The PMS, BMS and its combination lead the way for Asset Management System (AMS).

It has several definitions, one of the best is that of FHWA [9]:

Asset Management (AM) is a business process and a decision-making framework that covers an extended time horizon, draw from economics as well as engineering, and considers a broad range assets. The AM approach incorporates the economic assessment of trade-offs among alternative investment options and uses this information to help make cost-effective investment decisions. Thus, asset management provides a framework for handling both short -and long-range planning.

The main elements of highway AM are: pavements, bridges, tunnels, guardrail, signs, barriers, lighting, other equipments and facilities ([9]).

Some other elements could be contacted to the system: equipments, vehicles, materials, human resources, and buildings. The municipal AM contains of several other elements which are related to mostly the infrastructure: sewer, emergency services, electricity, water, garbage collection, recycling, drainage, park facilities, traffic signals, signs, and markings, buildings, refineries, parks and recreation arias, airport.

The realisation of AM needs several subsystems:

- monitoring of subsystems,

- life-cycle cost analysis,

- asset value,

- analysis of maintenance and rehabilitation actions,

- information system management,

-   condition assessment and performance modelling,

-   optimisation methods (max. benefits min. cost),

-   economic analysis (costs and benefits),

-   others.

The main questions which arise at different administrative levels of technical and economic decision making are as follows:

-   how much money is required for AM elements pavement maintenance and rehabilitation?

-   which AM elements condition distribution would be expected if the sum mentioned above were to be available?

-   what consequences would be expected if this sum were not available?

-   what consequences would be expected if the maintenance expenditures were significantly increased?

-   what are the optimum times for maintenance actions?

-   what consequences would be expected from the delay of necessary actions?

-   who benefits and who loses, to what extent, etc?

Conclusions

As mentioned earlier, several subsystems exist already in Hungary in the field of Transport Asset management. The systematic monitoring has begun more than a decade ago. The asset value calculation related to bridges and roads is also performed regularly. We have urban, motorway and highway PMS systems, as well.

A combined PMS-BMS model is also completed. The generalization of this model system is under development.

The first version of the model family consist of the following parts:

-   the exact mathematical model (e.g. BMS + PMS),

-   normative model for some other elements,

-   cost/benefit types models.

Literature

[1]   Bakó, A.: Optimization Model for Higway pavement Maintenance, Proceedings of ISAPT II, 2004, Kuala Lumpur, Malaysia, 1-23

[2]   Bakó, A., Csicsely-Tarpay M., Gáspár L. and Szakos P.: The Development and Application of a Combined Highway Pavement Management System

in Hungary, 4th International Conference on Managing Pavements, Durban (South Africa), 1998, Proceedings Vol. 3, pp. 1091-1105

[3]   Bakó, A., Klafszki, E., Szántay T., Gáspár, L.: Optimization Techniques for Planning Highway Pavement Improvements, Annal of Operations Research 58(1995) 55-66

[4]   Gáspár, L. et al.: FORMAT Fully Optimised Road Maintenance. Project funded by the European Community under the „Competitive and Sustainable Growth Programme". Final Technical Report. March 2005, 113 p

[5]   Gáspár, L., Karoliny, M.: Bottom Ash Embankments and Their Long-Term Performance, The International Journal of Pavement Engineering & Asphalt Technology, 6(1), 2005, 27-43

[6]   Hudson, W. R., Hudson, S. W., Pavement Management System, Proceedings of the 3rd International Conference on Managing Pavement, Sant Antonio, 1999, Texas, 99-111

[7]   Männistö, V., Tapio, R.: Infrastructure Management System, Transp. Res. Rec. 1445, TRB 1994, 132-138

[8]   Talvitie A. and Olsonen R.: Selecting Asphalt Concrete Condition States for Finland's Highways, 67th Annual Meeting of the TRB, Washington, D.C., 1988, pp. 1-38

[9]   What is Asset Management, FHWA, Office of Asset Management, July 27, 2001