Agents' Coalitions Based on a Dynamic Programming Approach

Baltazar Frankovič, Tung-T. Dang, Ivana Budinská

Institute of Informatics, Slovak Academy of Sciences
Dúbravská c. 9, 84237 Bratislava, Slovakia
E-mails: utrrfran@savba.sk, utrrbudi@savba.sk

Abstract: Cooperation among agents is a very actual problem, which is a focus of lots papers published recently. The main purpose of cooperation is to improve each agent's execution or whole execution of agents' society and it depends strongly on concrete kinds of agents and a situation in which the agents are situated. Moreover, agents' intelligence and their relationship also influence lots to cooperation process, mainly, when the agents' society is huge or composed from a variety different agents. In this paper we focus to a problem how to create optimal coalitions among the given agents, where a coalition is a one of possible forms for cooperation and in which the common goal has highest priority for all the agents creating this coalition. Further, we will introduce methods, which are able to guarantee sub-optimal solutions by reducing a search space and estimate the domain, where the optimal solution possibly could be. In the end we will discuss about a problem of creating coalition with more parameters.

Keywords: Agents, Cooperation, Coalition, Optimization and Multi-Agent Systems

1 Introduction

In a lot of recent papers an agent has been introduced. The agent may be software, hardware, robot or another, but one of the relevant differences between the agent approach and the well-known traditional ones is an agent's ability to migrate from one place to another, to exchange its experiences with other ones in purpose to improve knowledge, skill, etc. The agent does not have to work only alone, in that case, the final result depends only on agent's capabilities and given tasks, but moreover it can cooperate with another ones if it is able. Cooperation may be executed in a variety forms, depends on which kinds of the agent are and which capabilities each agent can possess. If the agents are radically different, then, their cooperation may be very complicate. From that reason it is necessary to restrict and clearly define exact areas where the agents can cooperate, main objects of cooperation, methods and a common protocol for agents' communication. These main features, however, are very important for formulation of agents' cooperation, but unfortunately, in general we cannot specify them exactly in advance when we do not know which kinds of the agent will be implemented to and which functions each agent will have. Another essential property inside of agents' cooperation is behavior of each agent. Agent's behavior may be simply predictable – it means we can describe, approximate or explain it by exact mathematical formulations, or completely stochastic – belong to this category such agents as for example: a dealer in stock exchange, a manager of a factory, etc. Cooperation between various kinds of agents is also complicated problem, if is considered that the agents' characteristics are very pragmatic. Every kind of the agent may have different capabilities of reasoning, inferring and evaluating, moreover, via combination with variety of knowledge or experiences' basis, then, their behavior perhaps becomes unimaginable and the final result of their cooperation may grow to immeasurable dimensions [12], [13]. Another property of agent's behavior, which possible influences a lot to cooperation process with another ones, is an agent's capability to improve its knowledge's basis during cooperation. They are known several kinds of such agents, for example: an adaptive agent, which is able to analyze its environment (including behavior of another agents working in the same environment) and act (or react) so as to achieve the best results for itself. Next kind of such agent is a learning agent, which, in a difference of the previous kind, sequentially learns from behaviors of another agents and improves its knowledge's basis to enable to meet its goals. A quality of a learning process can be reflected via actions that the agent takes in cooperation process and their results the agent receives for them. A learning process may last as long as the received results satisfy or come near agent's goals; even though, the process of learning is strongly dependent on duration of agents' wiliness to join cooperation with another ones and intelligence of each agent [4], [7], [8]. The creation of coalition and coalition structure of agents is one of possible technical way for formulation agents' cooperation. A coalition of agents essentially may be considered in the framework of multi-agent-system (MAS) as a group of such agents, which are willing to cooperate with all another members in this group and their common cooperative activities aim to reach the optimum of the given criteria, but on other hand, methods for cooperation are collected by these members via negotiation among themselves and they depend on concrete situations. The optimum, which the members in coalition try to reach, does not have to be always a global optimal solution, but in most cases only is optimal from any point of view (criterion) or a Nash optimal solution (Nash optimal solution is defined in [2], [9]). From this reason the coalition formation includes the activities as follows:

- Generation of coalition structure: to create or distribute the agents to individual sets for special problem solution.
- Solving the optimization problem in each coalition and a coalition structure, too.
- Distribution of the obtained requirements to each coalition.

There exists a lot of possible ways to search for and create an optimal coalition – mainly via negotiation, or sequentially search all possible variants - and several of potential methods that we shall to use for finding the solution of such formulated problem are as for example: an application of automat theory, game theory, genetic algorithms or another. The application field of this approach may be a variety and very wide, it is applicable in resolving planning problems in manufacturing [9], [10]. Another applicable field is in marketing for enterprises, where to maximize own profit each manager (considered as one agent) can try to cooperate with another ones [6], [11]. Coalition in this case may have another effect to eliminate execution of agents not belonging to the same coalition. Finding a solution for the above-introduced problem has been the subject of many publications, which tried to obtain a feasible way to resolve a general problem of creating coalition [1], [2], [3]. As a final note, a cooperation problem in general has many open questions to research; one of them is how to create the optimal structure so as the agents cooperate together with the maximal final effects. The focus of this paper is to find an optimal coalition structure of agents, where each agent and coalition has an own value expressed by its criterion function. The optimal coalition structure must guarantee that the solution is within permitted bounds for each agent and moreover the performance of whole system (involving all agents' performance) will be optimal. In this paper are presented some algorithms (computation methods), which are able to reduce a searching process in comparison with traditional approaches until the final solution will be found.

2 General Formulations for Agents' Coalition

Initially, let us introduce some general notations that we will use in the rest of this paper. Afterwards we will show some characteristics related with agents' coalitions and some basic approach, which are able to find the global optimal solution.

2.1 General Notifications

Let $\{A_1,...,A_n\}$ be a set of *n* agents, and I denote a set of index, I ={1,..,*n*}. In some parts of this paper possible occurs a remark $i \in I$, it means the agent A_i from this set. For simplicities we shall use a note I and a mark $K \subseteq I$ denotes a subset created by the agents, whose index belongs to a set K. Let denote $\Re_+^s$ as an *s*-dimensional set of nonnegative real values, which is used to assess an agent's execution. Each option in this set represents one parameter of agent's execution, e.g. cost, payoff, effectiveness, etc.

Definition 1: An *agent's value* is a function mapping from a set of all possible results that the agent can achieve (independently or via cooperation) to $\Re_+^s$.

Further notations are: for $\forall\ i \in I, K \subseteq I$,

– q_i^* denotes a value if the agent A_i works independently,

– q_i^K denotes a value if the agent A_i joins coalition K,

– f_i denotes an expected value of the agent A_i.

$f_i = q_i^*$ if the agent A_i works independently, or

$= q_i^K$ if the agent A_i joins the coalition K.

– F_K denotes a value of a set of agents $K \subseteq I$ and is defined as:

$$F_K = \sum_{i \in K} c_i * f_i \,.$$

where $\forall i \in K$, c_i is a parameter, which is used to express the priority of the agent A_i in this society (since some agents might be more important than other ones depending on concrete applications, therefore these parameters might be various) and $\sum_{i=1}^{k} c_i = const$. Generally, all the agents have the same priority; therefore, the function F_K can be defined as:

$$F_K = \sum_{i \in K} f_i \tag{1}$$

In this case $\forall i \in [1, n]$, $c_i = 1$.

In practice, each agent can know what it will get after joining any coalition until the discussion with other ones (about what every member within this coalition has to do regarding their possibilities, aims and willingness, and consequently which results each member can achieve) within this coalition finishes. As a result, in order to know all the values q_i^K the agents need another phase to negotiate with other ones about each possible coalition. It is clear that this process requires very exhausted work, and that is not just a focus of this paper, therefore to simplify let assume that these values are defined in advance or taken from historic database. Certainly, the more these values are defined, the larger is a probability to achieve the optimal solution. If the agents' society is too large, and that results an agent cannot explore every possible coalition, in such case all the values $q_i^K |_{i \in K}$, where the coalition K has not been explored, are defined to be zero. In the next part we will show some important definitions and characteristics of the agents' coalition.

2.2 General Definitions and Properties of Agents' Coalitions

Initially, let consider a following important property of the optimal coalition: for each set $K \subseteq I$ is valid that the set K is an optimal coalition if every experiment trying to separate the set K to a set of smaller coalitions will decrease a value F_K (defined by an equation (1). This property is derived from the assumption, in which the agents are always willing to remain in the large coalition, if they know that breaking this one to smaller sets will bring the whole party worse results.

Definition 2: A set $K \subseteq I$ is an optimal coalition if

$$\sum_{i \in K} f_i \geq \sum_{K_j} \sum_{i \in K_j} f_i \,, \qquad (2)$$

Where is valid $K = \bigcup_j K_j$.

Every agent, in fact, is an optimal coalition itself, because it is impossible to divide it to smaller subsets. From definition 1 the following theorem is deduced.

Theorem 1: Let be any structure of a set I for which the value F_I achieves maximum, then, such structure must be only composed from the optimal coalitions.

Proof: Let a set I be divided to *m* subsets like coalitions.

$$\mathrm{I} = \bigcup_{j=1}^{m} K_j \text{, and } \forall i \neq j \in [1, m] \text{ is valid } K_i \cap K_j = \{\varnothing\}.$$

If any subset $K_{i \in [1,m]}$ is not an optimal coalition, then, from the definition 1 it results: here must exist such a structure for organization a set K_i, $K_i = \bigcup_r K_r^i$, in which a value F_{K_i} is bigger than an actual value. Afterwards a structure composed as $I = \{ \bigcup_{j=1, j \neq i}^{m} K_j \} \bigcup_r K_r^i$ will have a bigger value than the actual one. Confrontation. ف

Further, we will introduce some notations that are useful for later applications.

Definition 3: Let S_i denote a set of coalitions, which can bring the agent A_i at least the same or better results than when it works independently.

$$S_i = \{K \subseteq I \mid q_i^K \geq q_i^*\} \tag{3}$$

In order to understand it let's take a simple example:

Example 1: Let be three agents $\{A_1, A_2, A_3\}$ and the values that each agent might get by joining coalitions or working alone be given in the Table 1.

On a basis of the values shown in this table and according to the definition 3 the agent A_1 will prefer to choose coalitions (except itself): $\{A_1, A_2\}$ and $\{A_1, A_2, A_3\}$, since in these ones it could gain more than when it works alone (f_1=6 and 8 in comparison with f_1=3 when it works alone), then the set S_1 will be {{1}, {1,2}, {1,2,3}}. Similarly it might be possible to calculate that S_2={{2}, {1,2,3}} and S_3={{3}, {2,3}, {1,2,3}}.

Table 1
Expected values - example with three agents

	f_1	f_2	f_3
$\{A_1\},\{A_2\},\{A_3\}$	3	8	9
$\{A_1,A_2\},\{A_3\}$	6	7	9
$\{A_1,A_3\},\{A_2\}$	2	8	5
$\{A_1\},\{A_2,A_3\}$	3	7	12
$\{A_1,A_2,A_3\}$	8	12	9

Definition 4: Let S_K denote a set of coalitions, which can bring at least the same or better results for all agents belonging to a set K than when they work independently.

$$S_K = \{Ko \subseteq I \mid \forall\, i \in K,\ q_i^{Ko} \geq q_i^*\} \tag{4}$$

Example 2: In the previous example, if a set K ={1,2}, then, $S_{\{1,2\}}$ ={1,2,3}, etc.

From these definitions the following results are derived:

Lemma 1:

1. If $K_1 \subset K_2$, then, $S_{K_1} \supset S_{K_2}$.
2. $S_{K_1 \cup K_2} \subset S_{K_1} \cap S_{K_2}$.
3. $S_{K_1} \cap S_{K_2} \subset S_{K_1 \cap K_2}$.

$$K \in \bigcap_{i \in K} S_i \tag{5}$$

Proof: (see in [1])

Consequences 1:

1. $S_{K_1 \cup K_2} \subset S_{K_1 \cap K_2}$
2. $S_{K_1 \cup \ldots \cup K_m} \subset S_{K_1} \bigcap \ldots \bigcap S_{K_m}$,

   where $K_1, \ldots, K_m \subseteq I$.

Proof:

1. The consequence 1.1 is followed directly from the lemmas 1.2 and 1.3.
2. The consequence 1.2 could be proved by successively applying the lemma 1.2.

   $$S_{K_1 \cup \ldots \cup K_m} = S_{(K_1 \cup \ldots K_{m-1}) \cup K_m} \subset S_{(K_1 \cup \ldots K_{m-1})} \bigcap S_{K_m} \ldots \subset S_{K_1} \bigcap \ldots \bigcap S_{K_m}.$$

Certainly, we want to attempt to implement all possible features of the real agent's behavior into our model, however, a mathematical model cannot copy exactly and completely agent's behavior, but one property has to be considered and it is also very often happened in practice. It is that, every agent is willing to join coalition with another ones if and only if this coalition brings it at least the same or better results than when it works independently. Furthermore, a coalition might exist if all the agents creating this one are willing to join.

Assumption 1: the agent A_i joins coalition $K \Leftrightarrow (q_i^K \geq q_i^*)$.

Assumption 2: a set $K \subseteq$ I is an *acceptable coalition* $\Leftrightarrow \forall i \in K : (q_i^K \geq q_i^*)$

For simplicities we will assume that these properties will be valid in the whole rest of this paper. With these assumptions it is possible to deduce a following theorem.

Theorem 2: A set $K \subseteq$ I is an *acceptable coalition* (does not have to be optimal) if and only if:

Proof:

1. If K is an *acceptable coalition*, it means $\forall\ i \in K$ it is valid: $q_i^K \geq q_i^*$. Afterwards, $K \in S_i$ for $\forall\ i \in K$. A result from this is $K \in \{ \bigcap_{i \in K} S_i \}$.
2. If $K \in \bigcap_{i \in K} S_i$, it means that the coalition K can bring at least the same or better results for all the agents belonging to this coalition $\Rightarrow$ a set K can be an *acceptable coalition* (does not have to be optimal). ⊔

Consequence 2: If $\bigcap_{\alpha_i \in I, i=1,..,m} S_{\alpha_i} = \{\varnothing\}$, then these agents $A_{\alpha_1},\ldots,A_{\alpha_m}$ cannot be in the same acceptable coalition. ڤ

The proof is derived directly from the theorem 2.

Example 3: In the previous example, since $S_1 \bigcap S_2 \bigcap S_3 = \{1,2,3\}$, therefore, the set $\{1,2,3\}$ is a unique possible coalition that all the agents are willing to join.

The main task in this paper that we want to resolve is: to find such a structure of a set I to maximize a function F_I defined by (1). Another way to say: we want to find such a structure $I = \bigcup_{i=1}^{m} K_i$, where $\forall$ i, j $\in$ [1,m] $K_i \cap K_j = \{\varnothing\}$ and $F_I = \sum_{i=1}^{n} f_i$ =max. To resolve completely this task and to find a maximal value of a function F_I it is necessary to search the whole space of all possible structures of a set I. The complex searching of all possible structures of a set I can be executed in a variety ways, but a complexity of those approaches is the same for every method, because every variant have to be examined and for arbitrary *n* the problem is known to be NP-hard, thus, it is necessary to turn to heuristic methods to search, which are computationally efficient but which might guarantee only sub-optimal solutions from any point of view, for example: *tabu search, branch and bound, genetic algorithms*, etc. Generally, a quality of those methods strongly depends on how long the process has repeated and a quality of an initial choice. On the other hand, each agent presents one independent object with own goals and intelligence, and consequently the agent's execution (or choice) will aim to the best results for the agent; therefore, naturally each agent might poll such coalitions, in which its benefit is the best from its point of view.

Furthermore, next property of such formulated problem that is easy to verify is that the mentioned problem might be resolved by a *recursive approach*, sequentially for cases with 1 to *n* agents, where each new agent will try to join any formulated coalition, afterwards a result of the criterion function F_I will be compared with the best one so far. This approach can guarantee the complex searching of the whole space and the final solution certainly will be the globally optimal one. But, as we have mentioned above, agents' choices might influence the searching process a lot, because on a basis of these choices lots of coalitions might not be necessary to be examined and searching space could be reduced too much. In the following section we shall introduce an approach based on a principle supporting complex searching and concurrently regarding agents' choices.

3 A General Dynamic Programming Scheme

For a set $K \subseteq I$ it is assumed: a set K can be decomposed to an amount of independent nonempty subsets. $K = \bigcup_{j=1,..,m} K_j$, where, $\forall$ i, j$\in$[1,m], $K_i \cap K_j = \{\varnothing\}$. Certainly $m \leq |K|$. Let us define a function Q_K as follows:

$$Q_K = \arg\max_{\{K_j\}} \{\sum_{K_j} \sum_{i \in K_j} f_i\} \tag{6}$$

Or by other words, Q_K is a maximal sum of all agents' expected values, which are creating the set K, among all possible structures of the set K.

Furthermore, by comparison an equation (1) and (6) it is possible to derive that

$Q_K = \max\{F_K\}$ for $\forall\ K \subseteq I$. Therefore, the main task mentioned in the previous section might be transformed to the task to find a value Q_I and such a structure, which achieves this value.

Because each agent has only two alternatives: to work alone or to join any coalition. From this reason we shall present a general dynamic programming method for resolving problem Q_I. We consider n stages, where in each stage one agent is added and after n-th stage the whole problem will be resolved. For j=1,..,n-1 we denote $I_j = \{1,..j\}$, where every set I_j means we will search an optimal solution (an optimal structure) for j agents $\{A_1,\ldots,A_j\}$, which guarantees a maximal value F_{I_j} (=Q_{I_j}). Since each set I_j can be obtained by adding the agent A_j to the set I_{j-1}, we might have the following recursive equation:

$$Q_{I_j} = \arg\max_K \{(Q_{I_{j-1}} + q_j^*), (Q_{I_{j-1}\setminus K} + Q_{K\cup j})\} \tag{7}$$

Where $K \subseteq I_{j-1}$ and j={1,..,n-1}, $I_0 = \{\varnothing\}$.

Further, for each j={1,..,n-1}, if we denote

$$h_j = \arg\max_{K \subseteq I_{j-1}} \{Q_{I_{j-1}\setminus K} + Q_{K\cup j}\} \tag{8}$$

The equation (7) could be rewritten as:

$$Q_{I_j} = \max\{(Q_{I_{j-1}} + q_j^*), h_j\} \tag{9}$$

On the basis of these recursive equations we can, in principle, find an optimal solution for Q_I by the following backward scheme:

- Sub-problem in stage *n*: Search a value h_n in (8) by examining all possible coalitions that the agent A_n is willing to join with using theorem 2 and consequence 2. After that, replace them to (9) to obtain a value Q_I .
- Sub-problem in stage *n-1*: Analogously as above: it is necessary to examine those coalitions that the agent A_{n-1} is willing to join and do not involve the agent A_n , etc.
- Sub-problem in stage 1: $Q_{I_1} = q_1^*$.

Now we get:

Theorem 3: By solving sub-problem in *n* stages we will obtain the optimal structure for the given set I and a maximal value of F_I .

Proof: (outline) Since by adding an agent A_n to a set I_{n-1} we obtain the set I , and from these equations (7), (8) it is easy to verify that calculating of a value Q_I can be made from the known values in the previous stages from the first to (*n*-1), otherwise, these values have been calculated for optimal cases, in which a structure of each subset was optimal. Consequently, a value Q_I from an equation (9) has to be optimal one.

From an equation (8) it is possible to contend that the main complication in the general method presented above is that the process of calculating a function $h_{j\in[1,..,n-1]}$ is huge and complicated, essentially with large values of *n*. Because, theoretically, the new added agent is able to join every coalition (if there are $j_{j=1,..,n-1}$ agents in the previous set, then, the new $j+1$ agent could join one of 2^j possible coalitions), therefore dimensions of this problem will grow with an exponential speed when *n* is increased.

To improve the previous approach to computable dimensions, the first step that we need to do is to reduce an amount of coalitions that the new added agent can join. To do it, it is necessary to use agents' intelligence and their attributes. Because each agent can own different mechanisms to poll coalitions that it is willing to join. The second step is: applying theorem 2 and consequent 2 to remove coalitions that the new added agent cannot join and leave only the *acceptable coalitions,* which are accepted by all the agents creating these ones. Afterward, an equation (8) could be resolvable. Certainly, the results of the first step will hardly depend on each agent's choices even though, these choices might be executed by different ways and they also might be dependent on another agents' decisions too. In the next section we will discuss this problem.

4 Creating the Optimal Structure by Sequential Conceding

As we have mentioned above, reducing an amount of potential coalitions that the new added agent could join is a necessary thing, which enables to overcome huge complexities of an equation (8). We assume that, these expected values ($q_{i\in I}^{K\subseteq I}$), which each agent is able to achieve by joining coalition with another ones, are random and known for all the agents creating this coalition. Moreover, any continue function does not exist, which is able to approximate exactly these values (because if it exists, this problem could be solved by different ways as for example: searching global maximum or minimum of any multi-parameters' function). Furthermore, for simplicities we will omit a negotiation part to achieve these values $q_{i\in I}^{K\subseteq I}$ and assume that they are known.

The first step mentioned in the Section 3: to reduce an amount of coalitions that the new added agent could join, might be made in the following principal:

Because for each agent, a coalition, in which its expected value is maximal of all, will have the highest priority in its choice. But, on other hand, with such choice of each agent, they will with a very high probability never reach an optimal structure. Therefore, each agent will have to concede, however, a new question: which agent has to concede and how much is a further complicated problem, because the agent could prefer one coalition before another from its criteria or on the basis of its deliberations it might consider its choices as the best for its execution. In this paper we shall propose three methods, in which the agents are assumed to have the same aim to achieve the global optimal executions of the whole agents' society. Before that, we have to define some important specifications:

Definition 5: for each agent $i \in I$, let $v_i^0 = \max_{K\subseteq I}\{q_i^K\}$ be a maximal expected value that the agent A_i can receive and for $\forall 0 < \alpha \le v_i^0$ let $\Omega_\alpha^i = \{K \subseteq I \mid q_i^K \ge \alpha\}$ be a set of coalitions in which the expected value for the agent A_i is more or equal than α. After these preparations we can now formulate our algorithms:

First, each agent states its aim equal a maximal value that it can achieve in all coalitions and collects only these coalitions in which its expected value is more or equal this value. After that, the agents try together to find a solution among these sets, if they cannot find any solution, all the agents sequentially have to concede by decreasing its aim. In an algorithm 1, all the agents sequentially decrease their aim but with the same ratio (ratio between a new aim and present one), in an algorithm 2, a difference between a new aim and the present one is the same for every agent. In an algorithm 3, we consider that in each turn only one agent decreases its aim and it always collects such coalitions, in which its expected value is maximal between remained ones.

An algorithm 1: (based on an *equal ratio* of conceding)

Step 1:

a) $\forall\ i \in I$ Choose coefficient c=1 and $\alpha_i = c * v_i^0$.

b) $\forall\ i \in I$ Search a set $\Omega^i_{\alpha_i}$.

Step 2:

a) Using a dynamic programming scheme presented in section 3 to search a sub-optimal solution among sets $\Omega^i_{\alpha_i}$, $i \in I$.

b) If a solution is not found, decrease a coefficient c and repeat a step 1.

c) If any solution is found, this process is stopped.

To effectively resolve a step 1.b, at first each agent sorts out all possible coalitions accordingly a value q_i^K, $K \subseteq I$ (since each agent could join 2^{n-1} possible coalitions, then, this procedure has a complexity $\cong O(2^{n-1} * (n-1) * \log 2)$). Afterwards, in every time when the coefficient c is decreased, each agent can choose immediately from its sorted coalitions feasible ones, which satisfy a condition for creating a set Ω^i_α, $\forall\ i \in I$. To solve a step 2.a we could apply the approach based on a dynamic programming scheme presented in a Section 3.

Similarly, we shall propose further algorithm based on equal regressions of each agent in a searching process.

An algorithm 2: (based on an *equal regression* of conceding)

Step 1:

a) $\forall\ i \in I$ Choose coefficient c=0 and $\alpha_i = v_i^0 - c$.

b) $\forall\ i \in I$ Search a set $\Omega^i_{\alpha_i}$.

Step 2:

a) Search a sub-optimal solution among sets $\Omega^i_{\alpha_i}$, $i \in I$.

b) If a solution is not found, increase a coefficient *c* and repeat a step 1.

c) If any solution is found, this process is stopped.

Differences between these both algorithms are not very significant; the results obtained by applying them depend on values q_i^K , for example: when values q_i^K are significantly different (a value $\text{var}(q_i^K)_{K\subseteq I}$ is high) for every $i \in I$, then, the first algorithm could achieve better results than the second one, but in another case when these values are nearly similar, the results could be contrary. To improve these algorithms, we shall present the third one, which is modified from the previous methods and differs from them in such point that in each turn only one agent tries to decrease its aim and how much this agent decreases, it depends on its decision. Certainly, if each agent chooses a different way to decide how much it wants to concede, then, this process might be very interesting and immediately occur lots of problems necessary to resolve. For short framework of this paper we should omit these related problems as agent's decision, negotiation or further problems as selecting agent to concede, etc.

In the next method, each agent sequentially will modify its aim in turn and we assume that each will choose immediately the best coalitions from a set of the remained ones. For the next algorithm we define some notations:

Let $\Psi_i = \{\varphi_i^j\}_{j=1}^{\lambda_i}$ be a sorted set of values q_i^K for the agent A_i (from top to down). λ_i is a sum of possible coalitions that the agent A_i could join - $\max(\lambda_i) = 2^{n-1}$. Certainly it is valid: $\forall i \in I$, $\varphi_i^1 = v_i^0$. Instead of coefficient c we will use only α_i for simplicity.

After preparations we can present the following algorithm:

An algorithm 3: (modified from the *algorithm 1 and 2*)

Step 1:

a) $\forall\ i \in I$ Choose coefficient $\alpha_i = \varphi_i^1 = v_i^0$.

b) $\forall\ i \in I$ Search a set $\Omega_{\alpha_i}^i$.

Step 2:

$i = 1$.

a) Search a sub-optimal solution among sets $\Omega_{\alpha_j}^j$, for all $j \in I$.

b) If a solution is not found,

1. decrease a coefficient α_i to: $\alpha_i = \max\{\varphi_i^k \mid \varphi_i^k < \alpha_i\}$,

2. update a set $\Omega_{\alpha_i}^i$,

3. repeat a step 2.a,

4. if $i < n$, then $i = i + 1$, else $i = 1$.

c) If any solution is found, this process is stopped.

Comparison between these algorithms is implemented in an illustrate example with 11 agents and shown in following graphs and a number of steps of the searching process to achieve the final solution is depicted in a Table 2.

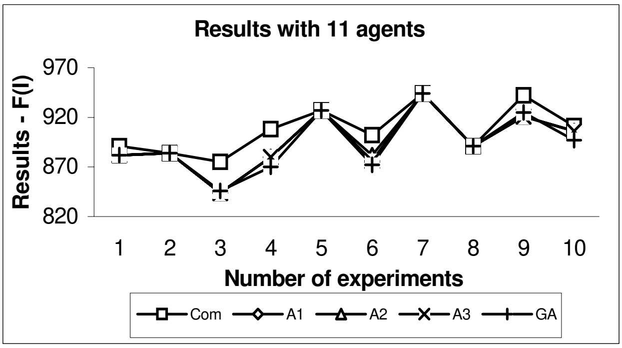

Figure 1
Results of searching max(F_I) with 11 agents

Notation: in figure 1 and table 2: *com* = complex searching, *GA* = genetic algorithm, *A1* = algorithm 1, *A2* = algorithm 2, *A3* = algorithm 3.

In these experiments we have bounded that each value q_i^K lies in an interval [0,100] and is randomly generated. From the obtained results it is possible to conclude that the algorithm presented above have achieved average very good results (average 95 percent of the optimal results, in some cases they even achieved the optimal results) and for a short time (see Table 2). The difference between these algorithms is not so significant, in some examples one algorithm could achieve better results than another ones, but both algorithms 1 and 2 are more appropriate for parallel working, since after each round each agent could immediately calculate its value α_i and does not have to wait until its turn. On the other hand, for an arbitrary set of values $\{q_i^K\}$, the algorithm 3 generally produced less complicated sets $\Omega_{\alpha_i}^i|_{i=1,..,n}$ than both first ones, since after each turn only a small number of possible coalitions is added, not massively as in the algorithms 1 and 2. Moreover, the algorithm 3 could be extended and modified for the case when each agent should choose a different way to calculate its value α_i. In this graph, the results of complex searching always have the maximal values. Another problem that may appear in this graph is that in several points is only one pattern; it means more results have the same value in this point.

In Table 2 is shown how many steps the searching program has repeated until the final result is found. These values may be used also to appoint to a time-complexity of each method to achieve the final results. For a genetic algorithm we fixed that the program will finish after certain number of steps.

Table 2
A case with 11 agents - complexity for searching max(F_I)

Order	Complex searching	GA	A1	A2	A3
1	859868	8500	134	173	238
2	900848	8500	112	113	189
3	1025352	8500	359	190	522
4	839476	8500	76	47	135
5	864455	8500	144	125	228
6	904389	8000	86	77	154
7	926803	8000	57	36	89
8	865845	8000	12	24	20
9	886832	8000	164	103	287
10	1112554	8000	152	168	260

The experiment has been repeated with more agents and from the obtained results we can see that: with a small number of agents (up to 8 agents) the complex searching approach can be applied, because a complexity is acceptable and all another reducing algorithms give sufficiently large differences. For cases with more agents (9 agents and more), all reducing methods have an omissible low complexity in comparison with using the complex searching approach or a genetic algorithm; nevertheless the results of these algorithms achieve just about the best results obtained by searching complex space. Better results we can reach by modifying the step 2.b.1 in the algorithm 3, instead of adding only coalitions, which have a value φ_i^k or q_i^K respectively immediately after a coefficient α_i, we can attach more ones at once. By this way a number of satisfied solutions will increase and a chance to reach the best solution will be bigger. Otherwise, for a case with 11 agents, a number of steps that the searching program repeated are about 1 million and this number will continuously increase with an exponential speed, therefore, from a practical point of view it is not possible to apply the complex searching method but more effectively is to use one of some presented reducing mechanisms, which give also average optimal results.

Conclusions

In this paper we have introduced one of especially actual problems that is how to create an optimal structure of agents' coalitions. Characteristics and conditions related creation coalitions have been pointed and proved. Further, we have introduced some methods, which are addressed to find sub-optimal solutions. Our methods have been formulated on a basis of modifying dynamic programming scheme for searching the global optimal solution and on other hand we have considered also to use agents' intelligence and their properties of independent deliberation to reduce useless coalitions, which are omitted in a searching process. These algorithms that have been presented here, then, were implemented for concrete examples and compared with another heuristic algorithm as genetic algorithm or the worst case when all variants are examined. From some declared problems in a Section 4 we see several features, that we should continue to resolve in future works, that are: a problem to create optimal structure for agents' coalitions with more parameters, and more kinds of agents may be connected in a cooperation process. Another problem is implementation of agents' negotiation abilities to improve cooperation process and approximate closer real situations. The last possible focus in the future is to find an optimal manner for gathering input information, because this problem has very important effect to the solving process and the achieved results respectively.

Acknowledgement

The work presented in the paper was supported by following projects:

- APVV LPP-0231-06 ; VEGA 2/7101/27

References

[1] Dang, T-Tung, Frankovic, B., Budinska, I.: Agents Coalition in Coordination Process. *In XV IFAC World Congress b'02*, Barcelona, 2002

[2] Wooldridge, M., Bussmann, S., Klosterberg, M.: Production Sequencing as Negotiation. In *Proceedings of the First International Conference on the Practical Application of Agents and Multi-Agent Technology (PAAM '97)*, 709-726, 1996

[3] Sandholm, T., Larson, K., Andersson, M., Shehory, O., Tohmé, F.: Coalition Structure Generation with Worst Case Guarantees. In *Artificial Intelligence* 111, 209-238, 1999

[4] Stone, P., Littman, M. L., Singh, S., Kearns, M.: An Adaptive Autonomous Bidding Agent. In *Journal of Artificial Intelligence Research* 15, 189-206, 2001

[5] Cohen, W., Schapire, R., Singer, Y.: Learning to Order Things. In *Journal of Artificial Intelligence Research* 10, 243-270, 1999

[6] Cantamessa, M., Villa, A.: Negotiation Models and Production Planning for Virtual Enterprises. *In Proceeding of The IFAC Workshop on Manufacturing, Modeling, Management and Control – MIM 2001, Prague, Czech Republic,* 1-5, 2001

[7] Liebrman, H., Nardi, B. A., Wright, D. J.: Training Agents to Recognize Text by Example. In *Autonomous Agents and Multi-Agent Systems*, 4, 79-92, 2001

[8] Gmytrasiewicz, P. J., Durfee, E. H.: Rational Communication in Multi-Agent Environments. In *Autonomous Agents and Multi-Agent Systems*, 4, 233–272, 2001

[9] Sandholm, T., Sikka, S., Norden, S.: Algorithms for Optimizing Leveled Commitment Contracts. *In International Joint Conference on Artificial Intelligence (IJCAI)*, 535-540, Stockholm, Sweden, 1999

[10] Odubiyi, J. B., Kocur, D. J., Weinstein, S. M.: A Scalable Agent-based Information Retrieval Engine. *In Proceedings of the First Annual Conference on Autonomous Agents*. California USA, ACM Press / ACM SIGART, 292-299, 1997

[11] Wellman, M. P.: A Market-oriented Programming Environment and its ication to Distributed Multicommodity Flow Problems. *In Journal of Artificial Intelligence Research* 1, 1-23, 1993

[12] Maturana, F., Shen, W., Hong, M., Norrie, D. H.: Multi-Agent Architectures for Concurrent Design and Manufacturing. *In Proceedings of IASTED International Conference on Artificial Intelligence and Soft Computing*, Banff, Canada, 355-359, 1997

[13] Frankovič, B., Dang, T. Tung: Cooperating Agents for Planning and Scheduling. *In Proceeding of The IFAC Workshop on Manufacturing, Modeling, Management and Control – MIM 2001, Prague, Czech Republic,* 06-11, 2001

Composition and Thickness of RF Sputtered Amorphous Silicon Alloy Films

Tivadar Lohner[1], Miklós Serényi[1], Dipak K. Basa[2], Nquyen Quoc Khánh[1], Ákos Nemcsics[1,3,*], Péter Petrik[1], Péter Turmezei[3]

1) Research Institute for Technical Physics and Materials Science, P.O.Box 49, H-1525 Budapest, Hungary

2) Department of Physics, Utkal University, Bhubaneswar-751004, India

3) Institute for Microelectronics and Technology, Budapest Tech, Tavaszmező utca 15-17, H-1084 Budapest, Hungary

*) Corresponding author: e-mail: nemcsics.akos@kvk.bmf.hu

Abstract: Because the composition and the thickness of the thin films are very important for the fabrication of the devices, in this study we have undertaken the determination of the composition and the thickness of the RF sputtered amorphous silicon alloy thin films deposited at room temperature under very different preparation conditions by using various techniques. Incorporation of argon is demonstrated in the room temperature deposited films and the thickness of the films measured by different methods such as Rutherford backscattering, spectroscopicellipsometry and step-profiler are found to be in reasonable agreement with each other.

Keywords: amorphous Si alloys, Sputtering, RBS, Ellipsometry

1 Introduction

Amorphous silicon and silicon based binary alloys have generated a lot of interest both fundamentally and technologically and have assumed great importance in the fabrication of electronic, opto-electronic, photo-voltaic as well as sub-micron devices [1-17]. Low temperature deposition and the precise control of the composition as well as the thickness of the desired films are very much essential for the fabrication of the devices. Low temperature deposition of the thin films is necessary to eliminate the problems like increased probability of defect formation and dopant redistribution associated with high temperature processing. Accordingly, low temperature techniques such as plasma enhanced chemical vapor deposition (PECVD), various sputtering techniques (such as DC, RF, etc.) are being used for the purpose. Because the independent control of the constituent elements of the film is possible in sputtering, therefore, sputtering techniques are more desirable [9,18-20] for the deposition of the thin films of interest. Further, the precise control of composition and thickness of the thin films are necessary as the material properties of the films and the performance of the devices depends strongly on composition as well as on the thickness of the films [21-26].

In this work we have undertaken the determination of the composition and the thickness of room temperature deposited RF sputtered amorphous silicon and its alloy films, deposited under very different preparation conditions as well as gas ambient. The composition of the films have been obtained using Rutherford backscattering (RBS) while the thickness of the films are determined using RBS, spectroscopic ellipsometry (SE) and step-profiler. Incorporation of argon in the films deposited at room temperature is demonstrated and the thickness of the studied films as determined from very different techniques such as RBS, SE and step-profiler, are found to be in reasonable agreement with each other.

2 Experiment

The amorphous silicon alloy films studied here were prepared in a commercial high vacuum RF sputtering system (Leybold Z400). The RF sputtered deposition system was pumped by a combination of a mechanical and a turbomolecular pump to obtain a base pressure lower than $5 \cdot 10^{-5}$ Pa. A liquid nitrogen cold trap was used to reduce the partial pressure of water vapor in the residual gas. Different targets were used for depositing different varieties of films. High quality polycrystalline silicon was used as the target to prepare amorphous silicon films while high quality polycrystalline silicon and germanium as well as polycrystalline silicon and SiGe were used as target to deposit amorphous silicon germanium thin films. The details regarding the targets as well as the gas ambient is shown in Table 1. The targets were coupled to a RF generator (13.56 MHz) via the network for impedance matching between the generator and its load. Further, there is provision for mounting the substrate on a heated stage located 50 mm away from the target. The studied films were deposited at room temperature onto the crystalline silicon substrate and the bias of substrate was maintained at -50 V. Sputtering was done using high purity (99.999%) argon gas for the unhydrogenated samples while a mixture of high purity argon and hydrogen were used for the hydrogenated samples. Argon as well as hydrogen gas were fed continuously into the chamber by means of electronically controlled mass flow controller. The partial pressure of argon was maintained between 2-3 Pa while that of hydrogen was around 0.03 Pa. Pressure fluctuations, which influence the sputtering rate remarkably, were observed to be absent. A DC wall potential of 1500 V was applied on the targets. With these said parameters sputtering rates of 7 nm/min could be achieved. Before sputtering the substrates were cleaned by wet chemical process.

To determine the composition of the studied film, RBS measurements were performed using He ion beam of 2 MeV energy in a scattering chamber with a two-axis goniometer connected to a 5 MeV Van de Graaff accelerator. In the scattering chamber the vacuum was maintained to be better than $1x10^{-4}$ Pa. To reduce the hydrocarbon deposition, liquid N_2 cold traps were used along the beam path and around the wall of the chamber. Backscattered He^+ ions were detected using an ORTEC surface barrier Si detector mounted in Cornell geometry at scattering angle of 16.5^o. The beam current was measured by a transmission Faraday cup and the composition of the studied films were obtained by using the program developed by Kotai [27].

Because SE measurement is a precise, fast and non destructive technique, it was used to measure the thickness of the studied films. The ellipsometry data were obtained from a SOPRA spectroscopic ellipsometer of rotating analyzer type. All measurements were performed in air at room temperature for the wavelength ranging from 250-850 nm in steps of 5 nm at the angle of incidence of 70.2^o. The evolution of the SE spectra were performed by WVASE 32 software [28].

The thickness of the studied films were measured directly by the alpha-step-profiling equipment (Tencor 100). A small region of the sputtered surface was masked by the sample fixing apparatus and the step between the sputtered and non-sputtered region was scanned by the profiler and the thickness of the film was measured directly.

3 Results and Discussion

In this study we have investigated the composition and the thickness of the five different types of room temperature deposited samples, prepared under very different preparation conditions. Amorphous silicon films were deposited on Si substrates in argon ambient from a polycrystalline silicon target (#84) while another class of amorphous silicon films, contaminated with carbon (from the substrate kapton), were also deposited on Si substrate in argon ambient from a polycrystalline Si target (#126). Amorphous silicon-germanium films were deposited on Si substrate in argon ambient from poly-crystalline Si and Ge target (#155). Hydrogenated amorphous silicon germanium films were deposited on Si substrate in argon and hydrogen ambient from polycrystalline silicon and germanium targets (#136). Amorphous silicon-germanium films were deposited on a SiO_2/Si substrate in argon ambient from polycrystalline Si and SiGe targets (#71).

The composition and the thickness of the films were obtained from RBS spectra using RBX program developed by Kotai [27]. The composition and the thickness of the films obtained from the RBS analysis are shown in Table 1. This table also contain the thickness of the films as obtained from spectroscopic ellipsometry as well as measured directly from step-profiler.

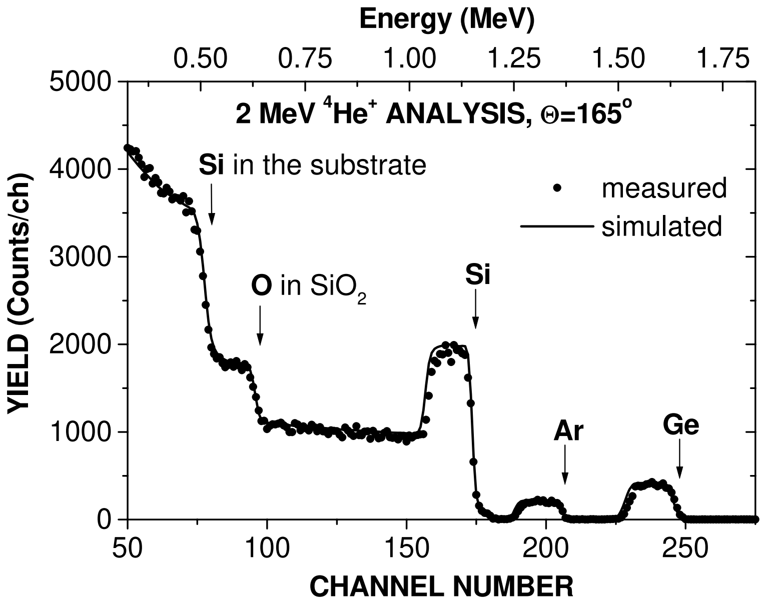

Figure 1
A typical RBS Spectra of amorphous silicon germanium film deposited on SiO_2/Si substrate along with the fit

Fig. 1 exhibits the typical RBS spectra of an amorphous silicon germanium thin film deposited on SiO_2/Si substrate (#71). The peaks arising from Si, Ge and Ar of the amorphous silicon germanium film are found to be well separated and each peak is further observed to have a relatively flat plateau indicating homogeneous distribution of elements in the film. Because the amorphous silicon germanium film (sample #71) was deposited on SiO_2/Si substrate not only oxygen edge from SiO_2 layer but also Si edge from the silicon substrate were also observed in the RBS spectra. Simulating the RBS spectra using RBX program, the areal density of the buried SiO_2 layer was estimated and using the density of SiO_2, the thickness of the buried SiO_2 layer is determined to be 10000 Å. The thickness of the said buried SiO_2 layer is also determined to be 10460 ±8 Å from the spectroscopic ellipsometry (SE), indicating reasonable agreement between the thickness as measured by RBS and SE. Similarly from the RBS analysis the thickness of the amorphous silicon-germanium layer for the sample #71 is found to be 2310 Å while the composition of Si, Ge and Ar are found to be 90.5%, 3.5% and 6% respectively. The composition and thickness of all the studied samples obtained from RBS analysis is shown in Table 1.

Table 1

sample #	Growth Parameters Target Gas	Thickness (A) Step Profile	RBS	SE	Composition Si/Ge/Ar/C
#84	p-Si Ar	1800	1800	1745	93.5/0/6.5/0
#126	p-Si Ar	5200	5000	-	84.7/0/6.8/8.5
#155	p-Si&Ge Ar	4000	3919	-	73.8/22.4/3.8/0
#136	Si&Ge Ar+H_2	10000	9961	-	73/22/5/0
#71	p-Si&SiGe Ar	2500	2310	2236	90.5/3.5/6/0

Because SE measurement is very effective and useful in determining the thickness of the films, we have also used SE to determine the thickness of an amorphous silicon germanium film (sample #71) as well as an amorphous silicon film (#84) for comparison with other measurements. The evaluation of the SE spectra have been done using WVASE 32 software. To simulate the SE spectra of the sample #71, the components used are air, surface roughness, semiconductor a-SiGe layer, silicon-dioxide layer and single crystalline Si substrate respectively. The roughness layer has been treated as a mixture of the layer material (a-SiGe) and 50% void. The effective medium approximation has been used for the purpose. The measured and the simulated spectra of the sample #71 is shown in Fig. 2. The thickness of the roughness layer, the amorphous SiGe layer and the silicon-dioxide layer are found to be 30, 2236 and 10460 Å respectively. Further, to simulate the the SE spectra of the sample #84, the components used are air, roughness layer, a-Si layer, single crystalline Si substrate respectively. The measured as well as the simulated SE spectra of the sample #84 are shown in Fig. 3. The thickness of the roughness layer and amorphous silicon are found to be 10 and 1745 Å respectively. The quality of the fit (mean square error (MSE)) for the samples #71 and #84 are found to be 11.2 and 9.2 respectively. The thickness of the two films (#71 and #84) as determined from SE spectra are also listed in Table 1. The thickness of the a-Si layer as well as the a-SiGe layer, determined from SE measurements and other measurements are found to be in reasonable agreement with each other.

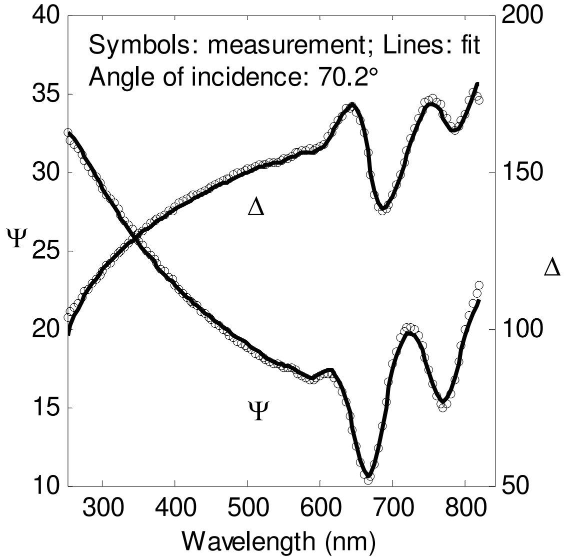

Figure 2
Measured (symbols) and simulated (solid line) ellipsometric spectra of amorphous silicon germanium film as a function of wavelength

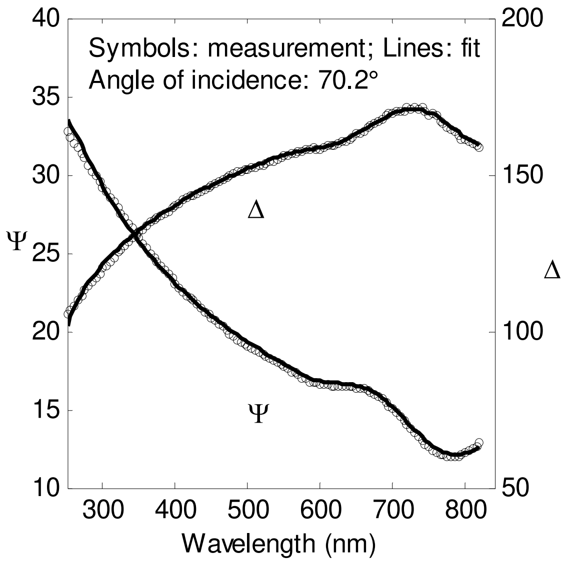

Figure 3
Measured (symbols) and simulated (solid line) ellipsometric spectra of amorphous silicon film as a function of wavelength

Conclusion

In this study we have deposited five different type of amorphous silicon based thin films, prepared at room temperature under very different deposition conditions and are also characterized by widely different techniques. The principal conclusion of the study are:

1 The RBS spectra and the composition analysis of the studied films clearly reveal the inclusion of Argon in the films deposited at room temperature.

2 The thickness of the studied films as measured directly by step-profiler and determined from RBS as well as SE analysis, are found to be in reasonable agreement with each other.

Acknowledgements

This work was partly supported by the Hungarian National Scientific Research Fund (OTKA) through Grant No TO30421, TO47011, TO43704 which are gratefully acknowledged by the authors. The authors are thankful to M. Janko for her help in sample preparation.

References

[1] S. Aljishi, Jin Shu, L. Ley, S. Wagner, Phys. Rev. Lett. 65, 629 (1990)

[2] T. Unold, J. D. Cohen, C. M. Fortmann, Appl. Phys. Lett. 64, 1714 (1994)

[3] B. D. Chapman, S. W. Han, G. T. Seidler, E. A.Stern, J. D.Cohen, S. Guha, J. Yang, J. Appl. Phys. 92, 801 (2002)

[4] H. Kanno, A. Kenjo, T. Sadoh, M. Miyao, Jpn. J. Appl. Phys. 45 (#5B)4351 (2006)

[5] D. K. Basa, Thin Solid Films 406, 75 (2002)

[6] D. K. Basa Solid State Commun. 118, 535 (2001)

[7] P. Servati, Y. Vygranenko, A. Nathan, J. Appl. Phys. 96, 7578 (2004)

[8] K. Mui, D. K. Basa, F. W. Smith, R. Corderman, Phys. Rev. B 35, 8089 (1987)

[9] A. V. Vasin, S. P. Kolesnik, A. A. Konchits, V. I. Kushnirenko, V. S. Lysenko, A. N. Nazarov, A. V. Rusavsky, S. Ashok, J. Appl. Phys. 99, 113520 (2006)

[10] D. K. Basa, F. W. Smith, Thin Solid Films 192, 121 (1990)

[11] T. Kobayashi, Y. Nakano, M. Ogawa, R. Hashimoto, S. Kamikawa, Y. Itoh, J. Vac. Sci. Technol. A 24, 223 (2006)

[12] D. K. Basa, M. Bose, D. N. Bose, J. Appl. Phys. 87, 4324 (2000)

[13] K. D. MacKenzie, J. H. Burnett, J. R. Eggert, Y. M. Li, W. Paul, Phys. Rev. B 38, 6120 (1988)

[14] K. D. MacKenzie, J. R. Eggert, D. J. Leopold, Y. M. Li, S. Lin, W. Paul, Phys. Rev. B 31, 2198 (1985)

[15] P. Wickboldt, D. Pang, W. Paul, J. H.Chen, F. Zhong, C. C. Chen, J. D. Cohen, D. L. Williamson, J. Appl. Phys. 81 (1997) 6252

[16] E. Maruyama, S. Okamoto, A. Terakawa, W. Shinohara, M. Tanaka, S. Kiyama, Solar Energy Materials & Solar Cells 74 (2002) 339

[17] K. H. Jun, J. K. Rath, R. E. I.Schropp, Solar Energy Materials & Solar Cells 74 (2002) 357

[18] R. W. Collins, W. Paul, Phys. Rev. B 25, 5257 (1982)

[19] M. Serenyi, J. Betko, A. Nemcsics, N. Q. Khanh, D. K. Basa, M. Morvic, Microelectronics Reliability 45, 1252 (2005)

[20] T. Maruyama, H. Fujii, J. Electrochem. Soc. 147, 363 (2000)

[21] K. Mui, D. K. Basa, R. Corderman, F. W. Smith, J. Non-Cryst. Solids 97-98, 999 (1987)

[22] M. Bose, D. N. Bose, D. K. Basa, Material Letters 62, 417 (2002)

[23] M. Bose, D. K. Basa, D. N. Bose, J. Vac. Sci. Technol. A 19, 41 (2001)

[24] M. Bose, D. K. Basa, D. N. Bose, Materials Letters 48, 336 (2001)

[25] C. Summonte, R. Rizzoli, M. Servidori, M. Milita, S. Nicoletti, M. Bianconi, A. Desalvo, D. Lencinella, J. Appl. Phys. 96, 3998 (2004)

[26] K. Yokota, Y. Miyoshi, M. Saoyami, Jpn. J. Appl. Phys. 45 (#2A), 860 (2006)

[27] E. Kotai, Proc. 14^{th} Int. Conf. on Appl. Accelerators in Res. Ind. Nov 6-9, 1996, Denton, Texas USA, (AIP Press, New York 1997, Eds. J. L. Duggan and I. L. Morgan) p. 631

[28] J. A. Woollam Co Inc 645 M Street, Suite 102, Lincoin, NE 68508, U.S.A

Computational Environment of Software Agents[1]

Martin Tomášek

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice
Letná 9, 042 00 Košice, Slovakia
E-mail: martin.tomasek@tuke.sk

Abstract: Presented process calculus for software agent communication and mobility can be used to express distributed computational environment and mobile code applications in general. Agents are abstraction of the functional part of the system architecture and they are modeled as process terms. Agent actions model interactions within the distributed environment: local/remote communication and mobility. Places are abstraction of the single computational environment where the agents are evaluated and where interactions take place. Distributed environment is modeled as a parallel composition of places where each place is evolving asynchronously. Operational semantics defines rules to describe behavior within the distributed environment and provides a guideline for implementations. Via a series of examples we show that mobile code applications can be naturally modeled.

Keywords: software agent, multi-agent system, mobile code, communication, process calculus, mobile agent

1 Introduction

Mobile agent [1] is an autonomous program that decides which places of the distributed application visits and what operations uses there. Distributed systems based on mobile agents are more flexible than static ones: they support mobile users and can reduce network bandwidth [2]. It means the user just sends an agent then disconnects from network and finally receives the agent with result upon new connection.

Formal description and specification of such systems is very important for modeling and successful implementation of the application. If we think of most important system characteristics, we identify communication and mobility as a key point. There are a lot of techniques to describe mobile processes and

1 This paper was supported by the grant Nr. 1/0176/03 of the Slovak Grant Agency.

communication in existence. Very powerful tools for describing parallelism, communication and mobility are process algebras [3] and other formal techniques [4].

In this paper we present process calculus to describe mobile agents and their communication strategies. We provide basic abstraction of the distributed system and its parts and we define syntactic and semantics rules for modeling mobile applications. At the end we provide a formal description of three mobile code paradigms to illustrate the flexibility and expressiveness of the presented abstraction. Some very typical applications that implement code mobility are showed too.

2 Abstraction of the Software Architecture

We can identify three main entities from the abstraction of distributed system architecture: agents, interactions and places.

Agents are abstraction of the functional part of the system. They are evaluated in distributed computational environment and they are performing basic actions in their evolution.

Interactions are events presented between two agents or more agents in the computational environment. Basic agent actions are communication and mobility.

Places are abstraction of distributed computational environment. Whole distributed system is a set of places. Each place consists of agents and they are evaluated there. Interactions between agents can rise within one place or between two or more places.

3 Abstract Syntax of the System

We define terms of process algebra for modeling agents that can interact by performing three basic actions (read, write and move). The agents are modeled as process terms. The constructions for building agent terms are taken from Milner's CCS [5] and π-calculus [6, 7] and correspond to basic notions of process algebras [3].

Distributed system is defined as a parallel composition of independent places within a network. Each place is represented by its name and an agent term defining agents located inside the place. We define operator $||$ for parallel composition of places and its notion is very similar to $|$ operator for parallel composition of agents.

Abstract syntax of the calculus is following:

$\alpha ::=$		(actions)
$\mid$	x	(perform name)
$\mid$	$\mathbf{r}_p(x)$	(read name)
$\mid$	$\mathbf{w}_p(y)$	(write name)
$\mid$	$\mathbf{m}_p(Q)$	(move agent)
$P ::=$		(agents)
$\mid$	$\mathbf{0}$	(inactivity)
$\mid$	$\alpha.P$	(action composition)
$\mid$	$P_1 \mid P_2$	(parallel composition)
$\mid$	$P_1 + P_2$	(choice)
$\mid$	$A\langle \tilde{x} \rangle$	(agent invocation)
$S ::=$		(system)
$\mid$	$[P]_p$	(place)
$\mid$	$S_1 \parallel S_2$	(system composition)

Symbols $x, y, p, \ldots$ are called names and $\mathcal{N}$ is the set of all names. Names are an abstraction of manipulated data within agent interactions. Abbreviation $\tilde{x}$ is a sequence of names and $\{\tilde{x}\}$ is a set of names in $\tilde{x}$.

Symbol α denominates the actions provided by the agents. Action x performs an operation represented by name x. Action $\mathbf{r}_p(x)$ reads a name that was sent by another agent to place p and stores it in name x. Action $\mathbf{w}_p(x)$ outputs name x in place named as p. Action $\mathbf{m}_p(P)$ moves agent term P to the place p and the term P is computed there.

Agents are defined as process terms very similar way as in other standard calculi and they are denominated as $P, Q, \ldots$ symbols. The inactivity $\mathbf{0}$ defines an agent with no activity. Term $\alpha.P$ is an action composition and its notion is that when an action α is performed the term continues as P. Parallel composition $P_1 \mid P_2$ defines two independent agents P_1 and P_2 that can be computed in parallel. Agent term $P_1 + P_2$ is nondeterministic choice where an agent can be computed either as P_1 or P_2.

We assume that each agent abstraction A is defined by equation $A(\tilde{x}) \overset{def}{=} P_A$ where all free names of P_A are contained in $\tilde{x}$. Process abstraction is then a term without free names while $A(\tilde{x})$ binds names of $\tilde{x}$. Agent invocation $A\langle\tilde{y}\rangle$ is then the use of P_A term where all occasions of names from $\tilde{x}$ are substituted by $\tilde{y}$.

The distributed system is composed of places. Place $[P]_p$ is defined by its name p and agent term P which is computed inside the place. System $S_1 \parallel S_2$ is parallel composition of independent places in S_1 and S_2. Given a system S, we assume the existence of function *sites* which returns the set of places of S. The composition $S_1 \parallel S_2$ is defined only if $sites(S_1) \cap sites(S_2) = \emptyset$, thus we can consider a system just as a set of disjunctive places.

4 Semantics of the System

Presented semantics describes possible evolution of agents, places and whole distributed system without providing the actual allocation of processes and names. We will define operational semantics of the system in a notion of evaluating of the actions.

4.1 Semantics of Software Agents

The rules of agent semantics describe the evolution of an agent. We present labeled transition $P \xrightarrow{\alpha} P'$ where agent P' is derived from agent P by performing action α. Structural rules of the agent semantics are following:

$$\mathbf{r}_p(x).P \xrightarrow{\mathbf{r}_p(x)} P \tag{A1}$$

$$\mathbf{w}_p(x).P \xrightarrow{\mathbf{w}_p(x)} P \tag{A2}$$

$$\mathbf{m}_p(Q).P \xrightarrow{\mathbf{m}_p(Q)} P \tag{A3}$$

$$\frac{P \xrightarrow{\alpha} P'}{P + Q \xrightarrow{\alpha} P'} \tag{A4}$$

$$\frac{P \xrightarrow{\alpha} P'}{Q + P \xrightarrow{\alpha} P'} \tag{A5}$$

$$\frac{P \xrightarrow{\alpha} P'}{P \mid Q \xrightarrow{\alpha} P' \mid Q} \tag{A6}$$

$$\frac{P \xrightarrow{\alpha} P'}{Q \mid P \xrightarrow{\alpha} Q \mid P'} \tag{A7}$$

$$\frac{P\{\tilde{y}/\tilde{x}\} \xrightarrow{\alpha} P'}{A\langle\tilde{y}\rangle \xrightarrow{\alpha} P'} \quad A(\tilde{x}) \overset{def}{=} P \tag{A8}$$

Rules (A1), (A2) and (A3) describe how the actions are evaluated by agents. Rules (A4) and (A5) describe behavior of nondeterministic composition of agents, while rules (A6) and (A7) describe semantics of parallel composition of agents. Last rule (A8) describes invocation of agent named A .

We will use the standard notion $P\{\tilde{y}/\tilde{x}\}$ to indicate the simultaneous of any free occurrence of $x \in \{\tilde{x}\}$ with corresponding $y \in \{\tilde{y}\}$ in P .

4.2 Semantics of Distributed System

Semantics of the distributed system is defined by reduction relation ($\rightarrow$) rules which present basic computational paradigm for agent interactions within the system and evolution of the system. In addition the structural congruence ($\equiv$) is defined for the system semantics. Reduction rules are following:

$$\frac{P \xrightarrow{\mathbf{m}_p(Q)} P'}{[P]_p \rightarrow [P' \mid Q]_p} \tag{S1}$$

$$\frac{P_1 \xrightarrow{\mathbf{m}_{p2}(Q)} P_1'}{[P_1]_{p_1} \parallel [P_2]_{p_2} \rightarrow [P_1']_{p_1} \parallel [P_2 \mid Q]_{p_2}} \tag{S2}$$

$$\frac{P_1 \xrightarrow{\mathbf{r}_p(x)} P_1' \quad P_2 \xrightarrow{\mathbf{w}_p(y)} P_2'}{[P_1 \mid P_2]_p \rightarrow [P_1' \mid P_2']_p} \tag{S3}$$

$$\frac{P_1 \xrightarrow{\mathbf{r}_{p1}(x)} P_1' \quad P_2 \xrightarrow{\mathbf{w}_{p1}(y)} P_2'}{[P_1]_{p_1} \parallel [P_2]_{p_2} \rightarrow [P_1'\{y/x\}]_{p_1} \parallel [P_2']_{p_2}} \tag{S4}$$

$$\frac{[P_1]_p \rightarrow [P_1']_p}{[P_1 \mid P_2]_p \rightarrow [P_1' \mid P_2]_p} \tag{S5}$$

$$\frac{S_1 \rightarrow S_1' \quad sites(S_1) \cap sites(S_2) = \emptyset}{S_1 \parallel S_2 \rightarrow S_1' \parallel S_2} \tag{S6}$$

$$\frac{S \equiv S_1 \quad S_1 \rightarrow S_2 \quad S_2 \equiv S'}{S \rightarrow S'} \tag{S7}$$

Reduction rules clearly distinct between local and remote interactions performed by located agents and provide a formal model to guide the implementation.

Rule (S1) describes movement and evaluation of an agent. Agent P evaluates the agent Q at the same place. Agent Q is running in parallel with agents located at place p. Rule (S2) is very similar to the rule (S1) while agent P moves the agent Q to another place p_2 where it is evaluated in parallel with existing agents (P_2) there.

Rule (S3) describes synchronous communication between two agents located at the same place. The communication is synchronized when both peers want to interact (read or write) within the same place. It means two communication actions $\mathbf{r}_p(x)$ and $\mathbf{w}_{p'}(y)$ will interact when $p = p'$ and then the name y will be substituted for all occurrences of name x in term followed by $\mathbf{r}_p(x)$ prefix. Rule (S4) is very similar to rule (S3) while communicating agents are located on different places.

Rule (S5) describes asynchronous evolution of subcomponents of the place. It means each site of the system is working autonomously.

How the reduction behaves in presence of operator of parallel composition of places is defined by rule (S6).

The reduction behaves with respect to structural congruence as we can see in rule (S7). Structural congruence is defined following way:

$$S_1 \parallel S_2 \equiv S_2 \parallel S_1 \qquad \text{(C1)}$$

$$(S_1 \parallel S_2) \parallel S_3 \equiv S_1 \parallel (S_2 \parallel S_3) \qquad \text{(C2)}$$

The rule (C1) shows the operator $\parallel$ is commutative and rule (C2) shows the operator $\parallel$ is associative.

5 Expressing Mobile Code Applications

According to the classification proposed in [8], we can single out three paradigms, apart from the traditional client-server paradigm, which are largely used to build mobile code applications:

- remote evaluation,
- code on demand and
- mobile agent.

However we think of distributed systems based on mobile agents, our model of communication and mobility can describe all three programming paradigms. Now we will show expression of the three mobile code paradigms and some practical examples of mobile code applications.

5.1 Mobile Code Paradigms

5.1.1 Remote Evaluation

Remote evaluation is performed when a client sends a piece of code to the server and server evaluates the code and client can get the results back from the server.

We define term $Client$ that sends a request for remote evaluation to the $Server$'s place s. Request consists of a code P and a name of the $Client$'s place c. Then the $Client$ reads the result into the name y and continues as C.

Term $Server$ reads the request from his local place s. The received code is stored in name x and the name of $Client$'s place is stored in name p. Then the code in x is evaluated and the result r is sent back to the $Client$'s place. The $Server$ is performing an independent work in S.

We define the following terms where the whole system defined by term $System$ is a parallel composition of $Server$'s place and $Client$'s place:

$$Client \stackrel{def}{=} \mathbf{w}_s(P,c).\mathbf{r}_c(y).C$$

$$Server \stackrel{def}{=} \mathbf{r}_s(x,p).x.\mathbf{w}_p(r) \mid S$$

$$System = [Client]_c \parallel [Server]_s$$

5.1.2 Code on Demand

Code on demand describes the situation where a client wants to perform a code that is presented by the server. Client asks for a code and server sends it to the client where it can be evaluated.

We define term $Client$ that sends a request to the $Server$'s place s. The request consists of a name of the $Client$'s place c Then the $Client$ reads the code from local place into the name x. Finally the code is evaluated and $Client$ continues as C.

Term $Server$ reads the request from his local place s. The received name of $Client$'s place is stored in p. Then the $Server$ sends a code P to the $Client$'s place. The $Server$ is performing an independent work in S.

We define the following terms where the whole system defined by term $System$ is a parallel composition of $Server$'s place and $Client$'s place:

$$Client \stackrel{def}{=} \mathbf{w}_s(c).\mathbf{r}_c(x).x.C$$
$$Server \stackrel{def}{=} \mathbf{r}_s(p).\mathbf{w}_p(P) \mid S$$
$$System = [Client]_c \parallel [Server]_s$$

5.1.3 Mobile Agent

Mobile agent is a paradigm where an autonomous code (agent) is sent from the client to the server. By autonomous we mean that the client and server do not need to synchronize the agent invocation and the agent is running independently and concurrently within the server's place.

We define an abstraction $Agent(\tilde{x})$ of a mobile agent and term $Client$ is moving the agent to the $Server$'s place s .

Term $Server$ is performing its independent work S and it is able to receive the agent which is then running in parallel with other $Server$'s actions in its local place s .

We define the following terms where the whole system defined by term $System$ is a parallel composition of $Server$'s place and $Client$'s place:

$$Agent(\tilde{x}) \stackrel{def}{=} P$$
$$Client \stackrel{def}{=} \mathbf{m}_s(Agent\langle\tilde{z}\rangle).C$$
$$Server \stackrel{def}{=} S$$
$$System = [Client]_c \parallel [Server]_s$$

5.2 Examples

5.2.1 Remote Procedure Call

This example shows that we are able to model very traditional mobile code application that is performing remote procedure call.

A client sends a request to a server and waits for response. The request consists of procedure name and its real parameters that should be performed by a server and the address of the client's place where to send a result.

Term $Client$ sends the request with name of the procedure $Proc$, its real parameters $\tilde{z}$ and the name of the $Client$'s place c to the $Server$'s place s . Term $Server$ reads from its local place s the request into the x (name of the procedure), $\tilde{y}$ (parameters of the procedure) and p (name of the *Client* 's place). Then in parallel it runs the *Server* recursively and continues as procedure stored in x with $\tilde{y}$ parameters. When procedure x is finished the result r is sent back to the *Client* 's place which name is stored in p.

The whole distributed system is defined in term *System* where *Client* 's place and *Server* 's place are computed in parallel:

$$Proc(\tilde{x}) \stackrel{def}{=} P$$
$$Client \stackrel{def}{=} \mathbf{w}_s(Proc, \tilde{z}, c).\mathbf{r}_c(y).C$$
$$Server \stackrel{def}{=} \mathbf{r}_s(x, \tilde{y}, p).(Server \mid x\langle\tilde{y}\rangle.\mathbf{w}_p(r))$$
$$System = [Client]_c \parallel [Server]_s$$

5.2.2 Dynamic Data Gethering

This example shows the model of simple mobile agent system. We define a mobile agent, which travels from place to place and searches for information.

A user defined by term *User* needs additional information on a data represented by name z. *User* launches mobile agent *Seeker* that dynamically travels among nodes looking for result information r in distributed database and stores it in y. If the information is found it is sent back to the *User* otherwise the *Seeker* continues in next place. *User* is waiting for the result and in parallel it continues with other independent work U.

Agent $Seeker(x,h,l)$, where x is searched information, h is home place of the *User* and l is local place of the agent, is reading the data from the local place l. It reads either searching result or the name of the next place where to search. In the first case it sends the result stored in y back to the *User*. In he second case it moves a new instance of the agent to the new place p and ends

The whole system is defined in term *System* where each independent place is sending either result information or the next place for searching:

$$User \stackrel{def}{=} \mathbf{m}_{p_1}(Seeker\langle z,u,p_1\rangle).(\mathbf{r}_u(y) \mid U)$$
$$Seeker(x,h,l) \stackrel{def}{=} \mathbf{r}_l(x,y).\mathbf{w}_h(y) + \\ +\mathbf{r}_l(x,p).\mathbf{m}_p(Seeker\langle x,h,p\rangle)$$
$$System = [User]_u \parallel [\mathbf{w}_{p_1}(z,r) + \mathbf{w}_{p_1}(z,p_i)]_{p_1} \parallel \cdots \\ \cdots \parallel [\mathbf{w}_{p_n}(z,r) + \mathbf{w}_{p_n}(x,p_i)]_{p_n}$$

Conclusions and Future Work

Modeling rules presented in the paper seem to be very suitable tool for formal description of distributed systems based on agent technology and technology of mobile code. The formal semantics is useful for discussing the design of modeled application and provides guidelines for its implementations in programming languages.

Primitive actions defined in the model present communication and mobility as key interactions for mobile agents. Abstraction of places, their parallel composition and performing interactions within places are very natural for distributed system architectures. These approaches in our model differ from very general π-calculus and ambient calculus [9].

Security properties of distributed system are also very important area and research on presented apparatus continues in this field [10, 11]. For example, presence of typing information [12, 13] within the names can provide privacy and security properties. In addition implementation of spi-calculus [14] primitives can add usage of secure communication protocols to the model.

We also work on multi-agent system platform [15] where mobile agents can work together to solve the common tasks. We use these models to define and to make verification of communication schemes [16, 17] for mobile agents coordination and cooperation within the multi-agent environment.

References

[1] Jennings, N. R., Wooldridge, M. J.: Applications of Intelligent Agents. In N. R. Jennings, M. J. Wooldridge (eds.): Agent Technology: Foundations, Applications, and Markets, Springer, Berlin, Heidelberg, New York, 1998, pp. 3-28

[2] Harrison, C. G., Chess, D. M., Kershenbaum, A.: Mobile Agents: Are They a Good Idea? Technical report, IBM Research Division, T. J. Watson Research Center, March 1995

[3] Baeten, J. C. M., Weijland, W. P.: Process Algebra. Cambridge University Press, Cambridge, New York, Port Chester, Melbourne, Sydney, 1990

[4] Šimoňák, S., Hudák, Š.: Petri Net Semantics for ACP Terms. Vol. 4, No. 1, Košice, 2004, pp. 55-59

[5] Milner, R.: Communication and Concurrency. Prentice Hall, 1989

[6] Milner, R., Parrow, J., Walker, D.: A Calculus of Mobile Processes, Part I and II, Information and Computation, 100, 1992, pp. 1-77

[7] Milner, R.: Communicating and Mobile Systems: the -Calculus. Cambridge University Press, Cambridge, New York, Melbourne, 1999

[8] Fuggetta, A., Picco, G. P., Vigna, G.: Understanding Code Mobility. Software Engineering, 24(5), May 1998, pp. 342-361

[9] Cardeli, L., Gordon, A. D.: Mobile ambients. Theoretical Computer Science 240, 2000, pp. 177-213

[10] Tomášek, M.: Specification Issues of Communication and Code Mobility. Acta Electrotechnica et Informatica, Vol. 4, No. 3, pp. 56-60, Košice, 2004

[11] Tomášek, M.: Vyjadrenie dynamiky mobilných programov. PhD. Thesis, Technical University of Košice, Košice, 2004 (in Slovak)

[12] Pierce, B., Sangiorgi, D.: Typing and Subtyping for Mobile Processes. In Proceedings of LICS '93, IEEE Press, 1993

[13] Kobayashi, N., Pierce, B., Turner, D.: Linearity and the -Calculus. In Proceedings of POPL '96, 1996

[14] Adabi, M., Gordon, A. D.: A Calculus for Cryptographic Protocols: the Spi-Calculus. In Proceedings of the Fourth ACM Conference on Computer and Communications Security, ACM Press, April 1997, pp. 36-47

[15] Krokavec, M., Paralič, M., Andričík, M., Tomášek, M.: Mobile Agent Based Distributed Programming. In D. Kocur, A. Filasová, J. Šaliga (eds.): 3rd Internal Scientific Conference of the Faculty of Electrical Engineering and Informatics (3rd ISC'2003), pp. 37-38, Košice, Slovakia, May 28, 2003

[16] Tomášek, M.: Concepts for Mobile Agents Interaction. In M. Jelšina, J. Kollar (eds.): Proceedings of Scientific Conference with International Participation Computer Engeneering and Informatics '99 (CEI '99), Košice - Herľany, Slovakia, October 14-15, 1999, pp. 259-264

[17] Tomášek, M.: Distributed System Based on Technology of Mobile Agents. Acta Electrotechnica et Informatica, Vol. 1, No. 1, Košice, 2001, pp. 55-60

Student Assessment of Desirable Technical Skills: A Correspondence Analysis Approach

András Farkas, Viktor Nagy

Faculty of Economics, Budapest Tech
Tavaszmező út 17, H-1084 Budapest, Hungary
E-mails: farkas.andras@kgk.bmf.hu, nagy.viktor@kgk.bmf.hu

Abstract. The importance rating of the technical skills and their related knowledge of industrial engineering and business management graduates are developed. A survey was conducted at the Faculty of Economics of the Budapest Tech to measure and evaluate 17 technical skills required by employers. The study utilized questionnaires to rate and rank these skills based on student assessment. The research was supported by a multivariate statistical method referred to as correspondence analysis. Details of the mathematical background of this method and the interpretation of the findings are also discussed.

Keywords: correspondence analysis, survey method, importance ranking

1 Introduction

Higher education needs to be aware of the changing nature of the workplace and of the requirements of employing organizations. Universities/Colleges need to be responsive to these changes and demands. Institutions have a great responsibility to its principal stakeholders (students) to equip them with more than a profound knowledge of an academic subject area. Therefore, they should continually examine their curriculum in order to enhance the educational experience and achieve continuous quality improvement. To assure a viable curriculum, some academic departments are developing closer contacts with their stakeholders (employers). They are also attempting to adapt curriculum to meet the changing needs and skills required by employers. As they begin to develop closer ties with industry, the academic community becomes more aware of the needs of the business community. The intent is to close the gap in understanding between the academic and the business communities.

At the Faculty of Economics of the Budapest Tech, this business outreach includes developing strategic alliances and partnerships with local companies, surveying employers and receiving input from students. In the latter respect we have focused on to survey the employment of our graduates thus far and, also, we have extended our investigations to the entire school [6], [7]. One of the central questions of our interest is as to whether is it more beneficial to emphasize technical skills and knowledge over general skills, or vise versa. In addition, we wonder whether gaps are existing between perceptions of students and the business environment regarding the value of certain skills and traits. When we have been evaluating qualities relevant for securing employment in our fields of studies we have identified three categories based on whether the item relates more to a qualitative, quantitative, or personal attribute. We called them competencies of graduates expected by the employers [6].

In this paper we are only concerned with one of these categories. This group in that graduates need to be successful at work can be called as quantitative attributes which tend to deal more with technology- or discipline-based knowledge. We will refer to these issues as *technical skills*. They include knowledge and application capabilities in computer use, programming languages, database management, optimization and the major areas of managerial accounting, finance and operations management.

The main purpose of this study is to find the technical skills and the major areas of knowledge that are sought by employers for our industrial engineering and business management students entering the job market. Student assessment for the importance of these attributes will be made based on representative samples of student bodies. Student perceptions about the importance of these skills must be comparable to those of employers in order to measure the actual performance of our graduates as well.

In Section 2 a detailed description of the correspondence analysis method will be presented, in Section 3 the steps of determining the importance of the technical skills of our engineering and economics students at the Budapest Tech will be given, and, finally, the major results will be presented.

2 Correspondence Analysis Technique

This study relies on descriptive empirical research by using quantitative methods for observable phenomena. The tools discussed involve collecting survey information and summarizing data. The term survey means the process of looking at the problem in its entirety. For our goals, the survey is a study that uses questionnaires and interviews to discover descriptive characteristics of the problem. To give more insight into the characteristics of our data sets descriptive statistical measures and some graphical displays are applied. Levels of measurement are ordinal and interval.

To analyze the relationships between a great number of variables in a subtle way the multivariate statistical method called correspondence analysis can be used. In this Section an overview of this multivariate statistical method is presented. For a comprehensive description of this technique we refer to the classic texts by Greenacre [10] and Benzecri [2]. Cross-tabulation of categorical data is perhaps the most commonly encountered simple form of analysis in applied research. Correspondence analysis (CA) is a descriptive technique of factoring categorical variables and displaying them in a property space which maps their association in two or more dimensions. It can be used where the contingency table approach is less effective due to the very complex large tables. So CA starts with tabular data, usually two-way cross-classifications. A measure of distance between any two points, is defined, where points are values (categories) of the discrete variables. CA employs chi-square distances to calculate the dissimilarity (or similarity) between the frequencies in each cell of the table. The cross-tabulation table of frequencies is then standardized, so that the relative frequencies across all cells sum to unity. The goal of such an analysis is to represent the entries in the table of relative frequencies in terms of the distances between particular rows and/or columns in a low-dimensional space, in usually two dimensions. This is achieved by factoring the basic structure (through singular value decomposition) of the chi-square distance matrix, resulting in a set of row vectors, column vectors and singular values [10]. In this step, the procedure requires the researcher to choose between different methods of normalization. Finally, the CA scales the vectors to create scores for each participant and each variable. These scores are plotted in a visual display, by showing which category values are close together. The interpretive strength of CA lies with its representation of low-dimensional solutions in these graphical displays, which permit the researcher to make comparisons between row variables and/or between column variables (after principal normalization) and between row and column variables (after symmetrical normalization) in their relative placement in shared low-dimensional space.

The differences between the methods of creating the graphical representations has caused debate in the literature about the most appropriate choice of normalization and methods of interpretation of the visual display [9], [11], [16]. Gabriel [9] calculated goodness-of-fit for the various forms of graphical representation available in CA. He concluded that researchers who have a specific interest in actual magnitudes of difference between row variables, or between column variables should choose the appropriate row or column principal normalization. However, researchers whose interest lies in comparing the general orientation of row points and column points in terms of the relative magnitude of the distance, rather than visualizing actual magnitudes, are well served by the symmetrical normalization option. The symmetric biplot, in addition to its optimal fit of the data, proportionally fits the form and the variance almost optimally and is an excellent candidate for general usage, unless one requires representation of the actual magnitudes [9, p. 435].

Examples of applications of CA can be found in various fields of interest, e.g., in medical research [11], for students' and teachers' cognitions about good teachers [1], for higher education institution image [15], for personalities [16], and in marketing research [2].

The formal description of the mathematics of CA is now presented (see the excellent book of Greenacre [10] for more details). Suppose that we are given a matrix **N** of non-negative numbers with I rows and J columns, such that its row and column sums are non-zero. Elements of this matrix are denoted by n_{ij}, i=1,…,I, j=1,…,J. The *correspondence matrix* **P** is defined as the elements n_{ij} divided by the grand total of **N**. The vectors of row and column sums of **P** are denoted by **r** and **c** respectively and the diagonal matrices of these sums by $\mathbf{D}_r$ and $\mathbf{D}_c$ respectively. Thus, the data matrix and the correspondence matrix are defined as

$$\mathbf{N} \equiv [n_{ij}],\ n_{ij} \geq 0, \quad \text{and} \quad \mathbf{P} \equiv (1/n_{..})\mathbf{N}, \text{ where } n_{..} = \mathbf{e}^{\mathrm{T}}\mathbf{Ne}, \text{ and } n = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}, \quad (1)$$

where $\mathbf{e}^{\mathrm{T}}\equiv[1,\ldots,1]^{\mathrm{T}}$ denotes an I-vector or a J-vector of ones respectively. The row and the column sums are:

$$\mathbf{r} \equiv \mathbf{Pe} \quad \text{and} \quad \mathbf{c} \equiv \mathbf{P}^{\mathrm{T}}\mathbf{e}, \text{ where } r_i > 0,\ i = 1,\ldots,I, \text{ and } c_j > 0,\ j = 1,\ldots,J, \quad (2)$$

and the diagonal matrices are defined as

$$\mathbf{D}_r \equiv \mathrm{diag}(\mathbf{r}) \quad \text{and} \quad \mathbf{D}_c \equiv \mathrm{diag}(\mathbf{c}). \quad (3)$$

The row and column *profiles* of **P** are defined as the vectors of rows and columns of **P** divided by their respective sums. The matrices of the row and the column profiles are:

$$\mathbf{R} \equiv \mathbf{D}_r^{-1}\mathbf{P} \equiv \begin{bmatrix} \breve{\mathbf{r}}_1^{\mathrm{T}} \\ . \\ . \\ . \\ \breve{\mathbf{r}}_I^{\mathrm{T}} \end{bmatrix} \quad \text{and} \quad \mathbf{C} \equiv \mathbf{D}_c^{-1}\mathbf{P}^{\mathrm{T}} \equiv \begin{bmatrix} \breve{\mathbf{c}}_1^{\mathrm{T}} \\ . \\ . \\ . \\ \breve{\mathbf{c}}_J^{\mathrm{T}} \end{bmatrix}. \quad (4)$$

Both the row profiles $\breve{\mathbf{r}}_i$, i=1,…,I, and the column profiles $\breve{\mathbf{c}}_j$, j=1,…,J, are written in the rows of **R** and **C** respectively. These profiles are clearly identical to the rows and columns of **N** divided by their respective sums, just as **r** and **c** are identical to the row and column sums of **N** divided by the grand total $n_{..}$.

The row and column profiles define two clouds of points in respective J- and I-dimensional weighted Euclidean spaces. The row cloud is the set of *points*, i.e. the I row profiles $\breve{\mathbf{r}}_1$,…, $\breve{\mathbf{r}}_I$ in J-dimensional space, which has the *masses*, i.e. the I elements of **r** and its *metric* is the weighted Euclidean with dimension weights defined by the inverses of the elements of **c** (chi-square metric), that is $\mathbf{D}_c^{-1}$. Accordingly, the column cloud is the set of the *points*, i.e. the J column profiles

$\breve{\mathbf{c}}_1,\ldots,\breve{\mathbf{c}}_J$ in I-dimensional space, which has the *masses*, i.e. the J elements of **c** and its *metric* is the weighted Euclidean with dimension weights defined by the inverses of the elements of **r** (chi-square metric), that is $\mathbf{D}_r^{-1}$.

The *centroids* of the row and column clouds in their respective spaces are **c** and **r** respectively:

$$\mathbf{c} = \mathbf{R}^{\mathrm{T}}\mathbf{r} \quad \text{and} \quad \mathbf{r} = \mathbf{C}^{\mathrm{T}}\mathbf{c}. \tag{5}$$

The overall spatial variation of each cloud of points is quantified by their *total inertia*, that is the weighted sum of squared distances from the points to their respective centroids, the masses and the metric. They are for the row points and the column points, respectively:

$$in(I) = \sum_i r_i(\breve{\mathbf{r}}_i - \mathbf{c})^{\mathrm{T}}\mathbf{D}_c^{-1}(\breve{\mathbf{r}}_i - \mathbf{c}), \qquad in(J) = \sum_j c_j(\breve{\mathbf{c}}_j - \mathbf{r})^{\mathrm{T}}\mathbf{D}_r^{-1}(\breve{\mathbf{c}}_j - \mathbf{r}), \tag{6}$$

i.e.

$$in(I) = \mathrm{trace}[\mathbf{D}_r(\mathbf{R} - \mathbf{e}\mathbf{c}^{\mathrm{T}})\mathbf{D}_c^{-1}(\mathbf{R} - \mathbf{e}\mathbf{c}^{\mathrm{T}})^{\mathrm{T}}], \; in(J) = \mathrm{trace}[\mathbf{D}_c(\mathbf{C} - \mathbf{e}\mathbf{r}^{\mathrm{T}})\mathbf{D}_r^{-1}(\mathbf{C} - \mathbf{e}\mathbf{r}^{\mathrm{T}})^{\mathrm{T}}]. \tag{7}$$

The total inertia is the same in both clouds and is equal to the mean-square contingency coefficient calculated on N, that is the chi-square statistic for independence divided by the grand total $n_{..}$ (calculated as if **N** were a contingency table):

$$in(I) = in(J) = \sum_i \sum_j \frac{(p_{ij} - r_i c_j)^2}{r_i c_j} = \frac{\chi^2}{n_{..}} = \mathrm{trace}[\mathbf{D}_r^{-1}(\mathbf{P} - \mathbf{r}\mathbf{c}^{\mathrm{T}})\mathbf{D}_c^{-1}(\mathbf{P} - \mathbf{r}\mathbf{c}^{\mathrm{T}})^{\mathrm{T}}], \tag{8}$$

where

$$\chi^2 \equiv \sum_i \sum_j \frac{(n_{ij} - e_{ij})^2}{e_{ij}},$$

and $e_{ij} \equiv n_{i.}n_{.j}/n_{..}$, the 'expected' value in the (i,j) cell of the matrix based on the row and column marginals $n_{i.}$ and $n_{.j}$.

The respective K^*-dimensional subspaces of the row and column clouds which are closest to the points in terms of weighted sum of squared distances are defined by the K^* right and left (generalized) singular vectors respectively of $\mathbf{P}-\mathbf{r}\mathbf{c}^{\mathrm{T}}$, in the metrics $\mathbf{D}_r^{-1}$ and $\mathbf{D}_c^{-1}$, corresponding to the K^* largest singular values. In other words the right and left singular vectors define the principal axes of the row and column clouds respectively.

Let the generalized SVD of $\mathbf{P}-\mathbf{r}\mathbf{c}^{\mathrm{T}}$ be:

$$\mathbf{P} - \mathbf{r}\mathbf{c}^{\mathrm{T}} = \mathbf{A}\mathbf{D}_{\mu}\mathbf{B}^{\mathrm{T}} \quad \text{where} \quad \mathbf{A}^{\mathrm{T}}\mathbf{D}_r^{-1}\mathbf{A} = \mathbf{B}^{\mathrm{T}}\mathbf{D}_c^{-1}\mathbf{B} = \mathbf{I}, \tag{9}$$

And for the singular values: $\mu_1 \geq \ldots \geq \mu_K > 0$, then the columns of **A** and **B** define the principal axes of the column and row clouds respectively. (In (9) matrix **I** denotes the matrix of identity.)

The respective co-ordinates of the row and column profiles with respect to their own principal axes (i.e. the *principal co-ordinates*) are related to the principal axes of the other cloud of profiles by simple rescaling. Let

$$\mathbf{F}_{I\times K} \equiv (\mathbf{D}_r^{-1}\mathbf{P} - \mathbf{e}\mathbf{c}^{\mathrm{T}})_{I\times J}\,\mathbf{D}_c^{-1}{}_{J\times J}\,\mathbf{B}_{J\times K} \text{ and } \mathbf{G}_{J\times K} \equiv (\mathbf{D}_c^{-1}\mathbf{P}^{\mathrm{T}} - \mathbf{e}\mathbf{r}^{\mathrm{T}})_{J\times I}\,\mathbf{D}_r^{-1}{}_{I\times I}\,\mathbf{A}_{I\times K} \quad (10)$$

be the co-ordinates of the row and the column profiles with respect to principal axes **B** (in the chi-square metric $\mathbf{D}_c^{-1}$) and **A** (in the chi-square metric $\mathbf{D}_r^{-1}$) respectively. Then

$$\mathbf{F} = \mathbf{D}_r^{-1}\mathbf{A}\mathbf{D}_\mu \quad \text{and} \quad \mathbf{G} = \mathbf{D}_c^{-1}\mathbf{B}\mathbf{D}_\mu . \quad (11)$$

Expressions (11) define the co-ordinates of the row and column profiles with respect to all the principal axes (the co-ordinates of individual points are contained in the rows of **F** and **G**). The co-ordinates of the points with respect to an optimal K^*-dimensional subspace are contained in the rows of the first K^* columns of **F** and **G**.

With respect to the principal axes, the respective clouds of rows and column profiles have centroids at the origin, e.g. the centroid of rows of **F** is $\mathbf{r}^{\mathrm{T}}\mathbf{F}=\mathbf{0}^{\mathrm{T}}$. The weighted sum of squares of the points' co-ordinates (i.e. the weighted variance or inertia) along the *k*th principal axis in each cloud is equal to μ_k^2, which will be denoted by λ_k and called the *k*th *principal inertia*, e.g. for the row cloud it is: $\mathbf{F}^{\mathrm{T}}\mathbf{D}_r\mathbf{F}=\mathbf{D}_\mu^2\equiv\mathbf{D}_\lambda$. In correspondence analysis the centering of the row and column profiles is a symmetric operation which removes exactly one dimension from the original spaces of these profiles. This is embodied in the property that the SVD of the "uncentered" matrix **P** consists of the SVD of the centered matrix $\mathbf{P}-\mathbf{r}\mathbf{c}^{\mathrm{T}}$. The columns of **F** and of **G** are the eigenvectors of the respective matrices **RC** and **CR**, standardized by $\mathbf{F}^{\mathrm{T}}\mathbf{D}_r\mathbf{F}=\mathbf{D}_\lambda= \mathbf{D}_\mu^2$.

The first one or two vectors of **F** and **G** are usually taken into consideration (one or two dimensions are displayed which have the highest contributions to the variance). Then, either the columns or the rows are plotted separately, or they are plotted simultaneously for constructing a *biplot* to find possible associations between the row and the column variables. Transitions between the columns and the rows might be done easily in the following way:

$$\mathbf{F} = \mathbf{D}_r^{-1}\mathbf{P}\mathbf{G}\mathbf{D}_\mu^{-1} = \mathbf{R}\mathbf{G}\,\mathbf{D}_\mu^{-1} \quad \text{and} \quad \mathbf{G} = \mathbf{D}_c^{-1}\mathbf{P}^{\mathrm{T}}\mathbf{F}\mathbf{D}_\mu^{-1} = \mathbf{C}\mathbf{F}\mathbf{D}_\mu^{-1} . \quad (12)$$

Expressions (12) can be used directly for the addition of supplementary rows and/or columns.

3 Exploration and Analysis of the Technical Skills

There are innumerable studies that have produced lists of graduate attributes desired by employers. Some of these studies sophisticated in attempting to prioritize these attributes in terms of their importance for employers [5]. The various studies have produced useful indicators of the kinds of skills, abilities and knowledge that graduates need. Increasingly, they are showing enormous similarities across discipline boundaries. A set of 'transferable' skills including communication, team-working, problem-solving, leadership, self-confidence and willingness to learn and flexibility are widely required. The set of such skills are often called graduate competencies. The reader may find definitions and the notion of this term in the book [14].

Analytic studies of the labor market have found, however, that there is a skills gap between the labor requirements of an industrially developed society and the outputs from the educational system. One feature of current skills shortage is the widespread lack of important generic skills and social skills such as quality assurance skills, problem-solving skills, learning efficiency, flexibility and communication skills. These are in addition to shortages of critical scientific and technological skills. Harvey with Green [13] argues that the perceived skills gap occurs for four reasons. First, that education is a 'once-and-for-all' activity, which ignores the need for life-long learning and skills updating. Second, a lack of communication is experienced between higher education and commerce and industry. Third, a strong indifference and inconsistency of industrialists exist in identifying what they want. Forth, there is a belief of threat to academic autonomy and freedom posed by closer links to commerce and industry.

This research has focused on the recent process of education at the Faculty of Economics (KGK) of the Budapest Tech. In particular, we have investigated the basic technical skills and the related knowledge of our students expected by employers when they enter employment. We considered both of our two undergraduate (BA and BSc) degree programs. Since we also wish to carry out a profound analysis within the specialized braches of our programs we divided our students into four categories. The abbreviations used throughout this paper are:

BM = Full-time BA students in Business Management.

IES = Full-time BSc students in Industrial Engineering with a major in Systems' Control.

IEG = Full-time BSc students in Industrial Engineering in different areas (e.g. quality management) of General Management.

PTP = Part-time BA/BSc program students in each of the above streams.

To make our findings enable for international comparisons we have chosen the questionnaire that was developed for similar programs in the United States [8]. The items of the technical skills are as follows:

A = Word processing (Word, etc.)

B = Spreadsheets (Excel, etc.)
C = Databases
D = Operating systems design
E = Project management
F = Computer software and programming languages
G = Inventory management
H = Logistics (transportation, distribution, warehousing, suppliers)
I = Quality management
J = Resource planning & control
K = Web designing, IT, Internet operations
L = Telecommunication
M = Quantitative analysis (statistics, optimization, etc.)
N = Managerial accounting (budgeting, break-even, cost controlling)
O = Finance (balance sheet, cost-benefit, cash-flow, investments)
P = Marketing & market research (sales, behaviour, etc.)
Q = Entrepreneurship

Students were asked to rate from 1 to 7 on a seven point ascending scale these 17 technical skills according to their importance. These competencies are required on entering employment and also for their professional career. The use of such Likert-scales is common in applied research. They simply gauge the degree to which there is agreement or disagreement with statements to reflect clear positions on an issue and represent a desirable goal, a transition from ordinal scales to interval scales [17]. The verbal interpretation of the scale used in the questionnaire is given below:

1 Not at all important
2 Scarcely important
3 Slightly important
4 Moderately important
5 Usually important
6 Significantly important
7 Extremely important

The survey was conducted for the each of the four student groups. The filled questionnaire forms have been exposed to severe inspections to filter them for possible non-sampling errors (data tabulating, coding, etc.). The sample sizes are, in turn:

BM (56) + IES (41) + IEG (59) + PTP (86) = KGK TOTAL (242).

We can ascertain that considering the total number of full-time and part-time students currently completing their studies at the KGK they are well represented by the sample sizes both for the total and for the partitioned (component) samples.

Completing such a questionnaire study involves the measurement of the internal consistency of the measure as a standard requirement. This is called reliability of the measure. Reliability refers to the property of a measurement instrument that causes it to give similar results for similar inputs. The commonly used measure is Cronbach's α based on the average inter-item correlation [4]. We carried out a reliability analysis for both the total and the partitioned (component) samples. Reaffirming our suspicion, the responses for each item are not correlated (inter-item correlations are below 0.30, except 10-15 pairs of approx. 5,000 ones with values of around 0.50). Moreover, there is a positive correlation between almost every pair of items, which we expected. These results show the independence of the items and the correct interpretability of the scales for the respondents. As concerns the coefficients of Cronbach's alpha, they are: for the total sample 0.820, and for the partitioned samples: 0.768, 0.803, 0.839 and 0.818, respectively. These coefficients are close enough to 1 (perfect reliability), indicating strong internal consistency among items within each factor. These findings validate the use of the various sets of questions to represent each factor of the study.

In Table 1, the weighted arithmetic means (mean scores) and the importance rankings (given by their rank numbers) of the student groups as well as the aggregated results are presented. The weighted standard deviations are ranging from 1.110 to 1.555 for the total KGK evaluation and from 1.046 to 1.877 for the group assessments. These latter measures represent a rather moderate variability in the respondent's judgments. Concerning the rather homogenous opinion of the student body of the KGK it is not surprising how close the scores of the single items are to each other for the different groups, and, similarly in their magnitudes. Regarding the total scores, the range of the weighted mean scores for the KGK is relatively small: R=5.87–3.26=2.61, whereas the maximum of the ranges for the student groups is for the IES: R=6.07–2.85=3.22.

Table 1

Technical skills	BM		IES		IEG		PTP		TOTAL	
	Mean	R	Mean	R	Mean	R	Mean	R	Mean	R
A word processing	5.80	5	5.88	2	5.51	4	5.63	2	5.68	3
B spreadsheet	5.88	4	5.73	3	5.61	3	5.63	2	5.70	2
C database	4.89	16	4.88	12	4.81	14	4.87	7	4.86	13
D operating syst.	5.07	13	4.59	14	4.83	13	4.79	11	4.83	14
E project man.	5.50	8	4.63	13	5.03	12	4.79	11	4.99	11
F comp. software	3.48	17	2.85	17	3.49	17	3.14	17	3.26	17
G inventory man.	5.07	13	4.56	15	4.46	16	4.51	16	4.64	16
H logistics	5.50	8	4.95	10	5.07	11	4.71	14	5.02	10

I quality man.	5.25	10	5.02	9	5.37	8	4.83	9	5.09	7
J res. planning	5.34	9	4.88	12	5.08	10	4.85	8	5.02	10
K web, IT	5.02	14	5.05	8	5.44	5	4.60	15	4.98	12
L telecommunic.	5.16	11	5.17	7	5.42	6	4.73	12	5.07	8
M quant. analysis	4.93	15	4.32	16	4.58	15	4.72	13	4.67	15
N man accounting	5.64	6	5.24	6	5.31	9	5.15	6	5.32	6
O finance	6.18	1	5.49	5	5.37	8	5.15	6	5.50	5
P marketing	6.09	2	5.54	4	5.86	2	5.24	4	5.64	4
Q entrepreneurship	6.07	3	6.07	1	5.93	1	5.60	3	5.87	1
Grand mean	5.35		4.99		5.13		4.89		5.07	

For the aggregated student body of the KGK we displayed the size of the weighted averages assessed for importance in Figure 1. Here, the 17 technical skills appear in an ascending order. This bar chart exhibits that the students ranked Entrepreneurship (Q), Spreadsheets (B), Word processing (A) and Marketing/Market Research (P) to be the most important and Computer Software/Programming Languages (F) as being the least important technical skills. These findings are not unexpected at all and quite similar to those of obtained for the American students. Just at the top of the list, one of these skills is remarkable, namely the Entrepreneurship (Q). This dominant view of the students is probably due to the particular Hungarian economic environment.

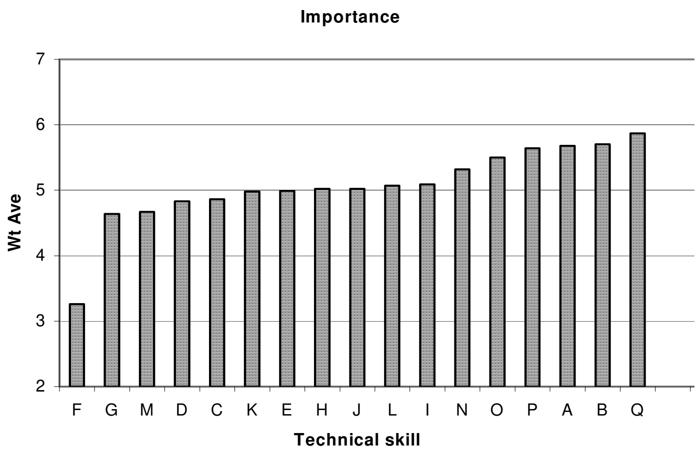

Figure 1
Priority ranking of the weighted mean scores of the KGK

The different student groups agree in their importance priority to a great extent. This statement is well supported by the numerical results when the Spearman's coefficients of rank correlation (interpreted on an ordinal scale) are computed.

These statistical measures can be generated from the rank numbers of the student ratings given in Table 1. They are ranging from ρ=0.698 (between IEG & PTP) to ρ=0.940 (between IES & IEG). Latter coefficient value reflects the very high degree of agreement in the rankings within the industrial engineering students. The IES group ranking is the closest to the aggregated KGK ranking: ρ=0.956. Notice also that there are some ties in the rankings as it can be seen in Table 1.

For the strength of the linear relationship s between the magnitudes of the weighted mean scores the Pearson's product moment coefficients of correlation are calculated (interval scale). These coefficients are even higher than those of the rank correlations and ranging from r=0.871 (between IEG & PTP) to r=0.951 (between IES & PTP). The strongest relationship in the magnitudes of the scores can be found between the IES group and the KGK average: r=0.984.

Now we consider the aggregated sample (KGK) and carry out a statistical analysis by applying the method of correspondence analysis. First the frequency of the responses (matrix **N**) is presented in the contingency table (see Table 2).

Table 2
Contingency Table

ITEMS	Importance							
Technical skills	**1**	**2**	**3**	**4**	**5**	**6**	**7**	Active Margin
Word processing	0	4	10	27	57	64	80	242
Spreadsheets	0	2	3	31	65	70	71	242
Databases	1	8	30	46	82	49	26	242
Operat. systems	3	4	37	46	74	53	25	242
Project mgmt	5	7	26	46	60	60	38	242
Comp. software	25	58	62	47	31	15	4	242
Inventory mgmt	5	9	40	51	68	48	21	242
Logistics	3	6	21	43	85	48	36	242
Quality mgmt	2	3	27	37	78	60	35	242
Res. planning	4	6	11	51	84	59	27	242
Web, IT	3	14	28	48	44	57	48	242
Telecommunic.	5	11	20	43	56	60	47	242
Quant. analysis	6	12	38	52	59	49	27	242
Man. accounting	4	7	14	34	61	68	54	242
Finance	2	6	16	30	53	61	74	242
Marketing	0	5	9	31	51	73	73	242
Entrepreneurship	2	4	8	18	37	81	92	242
Active Margin	69	166	400	681	1045	975	778	4114

In Table 2, one may think of the 7 column values in each row of this cross-tabulation as coordinates in a 7-dimensional space. Due to the enormous difficulties in computing the (Euclidean) distances between the row/column points, we want to find a lower-dimensional space, in which to position the row/column points in a manner that retains all, or almost all of the original information about the differences. Using the frequency values as the necessary input, the correspondence analysis procedure has been run. Then a step-by-step analysis of the statistics follows as is described below:

In terms of the *significance of dependencies* the value of the chi-square statistic is χ^2= 925.794, which at a stated level of α=0.05 indicates a significant dependency between the rows (attributes) and the columns (contributions) (p= 0.000).

The next step concerns the *dimensionality of the solution*, i.e. to determining the appropriate number of dimensions to use in the solution. This requires the analysis of the singular values, percentages of inertia explained, the cumulative percentages and the contribution to the overall chi-square value. The dimensions are extracted so as to maximize the distances between the row or column points, and successive dimensions (which are orthogonal to each other) will explain less and less of the overall chi-square value. In our problem the first and the second axes account for 72.8% and 21.6% of the inertia, respectively, i.e. a cumulative total of 94.4%. This retention of the solution is very high, as if the data were purely random with no significant dependencies, the average axis should account for 100/(7-1)=16.7% of the inertia and 100/(17-1)=6.25% in terms of the rows. Since the third axis accounts for only 3.2% of the inertia, hence a 2-dimensional solution is powerful enough to be used.

It may well occur that not all of the rows or columns are equally well represented. Determining the *quality of representation* of a particular row or column provides additional richness to the interpretation of the relationships in the contingency table. The quality of representation is evaluated by the sum of the squared correlations of that row and column over the n dimensions. In our problem all score categories and technical skills are well represented in the two dimensions (degree of association between a particular column and a particular axis).

As concerns the *compatibility of row and column coordinates* it is customary to summarize the row and column coordinates in a single plot. In such plots, however, one can only interpret the distances between row points, and the distances between column points, but not the distances between row points and column points. But our concern is to interpret the meaning of the dimensions extracted in the low-dimensional solution and the placement of the technical skills relative to those dimensions. We, therefore, utilize a symmetrical normalization for the graphical representations.

The major result of our analysis is shown in Figure 2. This biplot is a perceptual map containing almost 95% of the original information in two dimensions. In addition, as supplementary points, we displayed the student group averages in the same plot to be capable of making a subtle analysis of the results (a complex evaluation of the curriculum and the academic content of its constituting courses with respect to each technical skill and knowledge based area). From this chart, one may be readily determine that which are the most important technical skills, how the relative magnitudes look like for the distances between the skills and what sort of associations are among these student competencies.

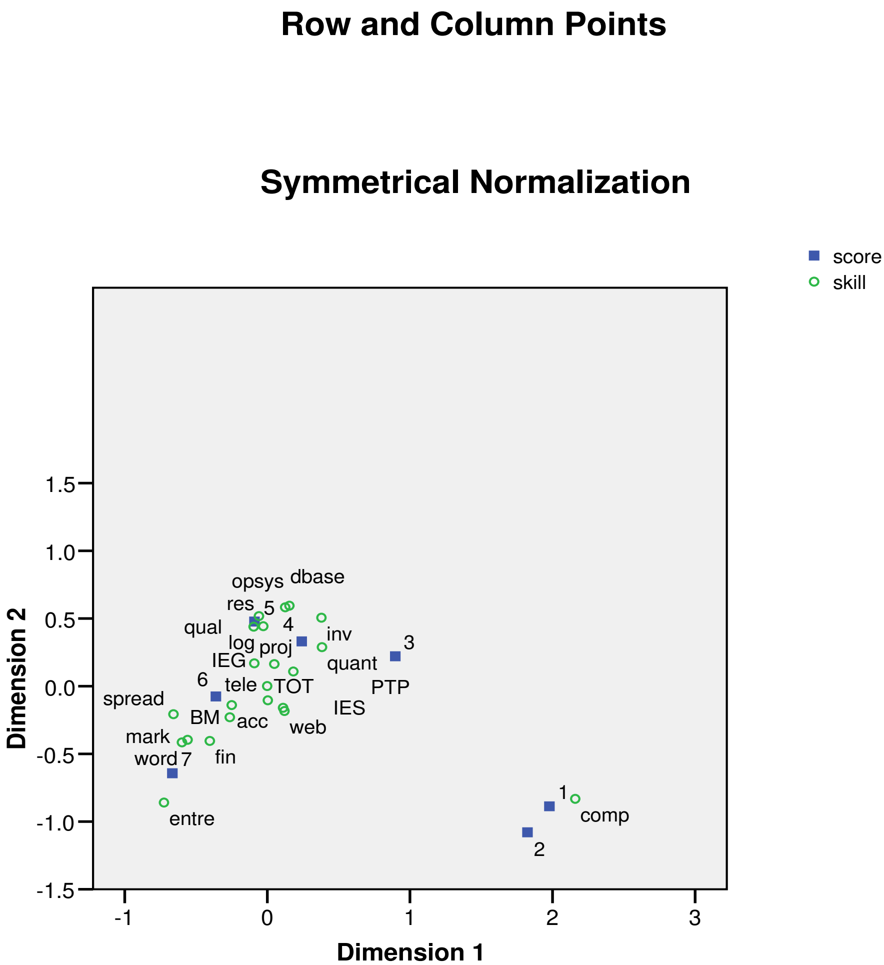

Figure 2

Symmetric biplot of the importance of the technical skills

It is also seen in Figure 2 that business management (BM) students rated the technical skills closest to the maximum grade (7) on the average by giving them the highest importance among the student groups. This result conforms perfectly to that of given in Table 1 (see there the grand means).

Conclusion

In this paper the technical skills and the related knowledge of Budapest Tech graduates have been evaluated with respect to importance. Efforts must be made by the school to minimize any gaps between our engineering and economics students' perceptions of marketable skills and actual skills expected by employers. The results of this study furnish a good basis for supervising and fur ther developing the present curricula into this direction. Correspondence analysis has proven a versatile and easily implemented tool in handling categorical data and, also, in detecting and explaining relationships among this complex phenomenon. Further investigations will reflect to the employer ratings of importance expected of graduate employees. This task, however, is subject to future research.

Acknowledgement

The authors express their thanks to Csaba Iván for his valuable help in gathering and preparing the data for the analyses.

References

[1] Beishuizen, J. J., Hof, E., Putten, C. M. v., Bouwmeester, S., Ascher, J. J., Students' and Teachers' Cognitions about Good Teachers. British Journal of Educational Psychology, **71**, (2001), 185-201

[2] Bendixen, M., A Practical Guide to the Use of Correspondence Analysis in Marketing Research. (electronic version). Marketing Research On-Line, 2003, (1996), 16-38

[3] Benzecri, J.-P., Correspondence Analysis Handbook. New York: Marcel Dekker. 1992

[4] Cronbach, L. I., Coefficient Alpha and the Internal Structure of Tests. Psychometrica, **16**, (1951), 297-334

[5] Diplomával a munkaerőpiacon. [szerk. Györgyi Zoltán] (2006) (With a Degree on the Labor Market). Tanulmány. NKTH projekt. PH Felsőoktatási Kutatóintézet. Budapest, 2006, p. 128 (in Hungarian)

[6] Farkas, A., On the Competency Profiles of Graduates. Proceedings of the 5th International Conference on Management, Enterprise and Benchmarking, Budapest, Hungary, June 1-2, 2007, Budapest Tech, Hungary, (2007), 153-166

[7] Farkas, A., Competitiveness of Graduates on the Job Market, Proceedings of the Symposium for Young Researchers in Business Sciences, Budapest, Hungary, November 16, 2007, Budapest Tech, Hungary, (2007), 9-18

[8] Gabric, D., McFadden, K I., Student and Employer Perceptions of Desirable Entry-Level Operation Management Skills. Journal of Education for Business, **16**, (2001), 1-10

[9] Gabriel, K. R., Goodness of Fit of Biplots and Correspondence Analysis. Biometrika, **89**, (2002), 423-436

[10] Greenacre, M. J., Theory and Applications of Correspondence Analysis. Academic Press. London, 1984

[11] Greenacre, M. J., Correspondence Analysis in Medical Research. Statistical Methods in Medical Research, **1**, (1992), 97-117

[12] Greenacre, M., Correspondence Analysis and its Interpretation. In: M. Greenacre, J. Blasius (Eds.), Correspondence Analysis in the Social Sciences, pp. 3-22, London: Academic Press, 1994

[13] Harvey, L. with Green, D., Employer Satisfaction. Birmingham, Higher Education Quality Council QHE, 1994

[14] Henczi, L., Zöllei, K., Kompetencia menedzsment. (Competence management). Perfekt Kiadó, Budapest, 2007 (in Hungarian)

[15] Ivy, J., Higher Education Institution Image: a Correspondence Analysis Approach. The International Journal of Educational Management, **15**, (2001), 276-282

[16] Nishisato, S., Elements of Dual Scaling: An Introduction to Practical Data Analysis. Hillsdale, NJ: Erlbaum, 1994

[17] Torgerson, W. S., Theory and Methods of Scaling. New York, Wiley, 1967

Computation of Rolling Stand Parameters by Genetic Algorithm

František Ďurovský, Ladislav Zboray, Želmíra Ferková

Department of Electrical Drives and Mechatronics
Technical University of Košice
Letná 9, 042 00 Košice, Slovakia
e-mail: frantisek.durovsky@tuke.sk

Abstract: Mathematical model of rolling process is used at cold mill rolling on tandem mills in metallurgy. The model goal is to analyse rolling process according to process data measured on the mill and get immeasurable variables necessary for rolling control and optimal mill pre-set for next rolled coil. The values obtained by model are used as references for superimposed technology controllers (thickness, speed, tension, etc.) as well. Considering wide steel strip assortment (different initial and final thickness, different hardness), and fluctuation of tandem mill parameters (change of friction coefficient, work rolls abrasion, temperature fluctuation, etc.) the exact analysis of tandem is complicated. The paper deals with an identification of friction coefficient on a single rolling mill stand by a genetic algorithm. Mathematical description of tandem mill stand is based on the modified Bland-Ford model. Results are presented in graphical form.

Keywords: cold rolling, Bland-Ford model, mathematical model, genetic algorithm

1 Introduction

Cold rolling of steel is very complex process. Knowledge of conditions in the roll bite is essential to achieve a good quality of final production. Some of its parameters may be determined exactly by measurement on the mill stand (geometrical dimensions, rolling force, front and back tensions, roll velocity), another only approximately by a suitable mathematical model (e.g. hardening of the processed material, friction coefficient between rolls and strip, strip temperature and flatness). There are some types of mathematical models for cold steel rolling. Off-line models are used for post processing analysis of rolling condition at a mill stand. The time for data processing is not critical in this case. Hence more detailed models and complicated time consuming computational method can be used. On-line models are used for real-time control of rolling. The model issues have to be at disposal pending the current coil rolling. Therefore the simplified but faster models are applied. The on-line models are used for mill pre-setting too. This type of model is often called torque-force model. The model computes friction coefficient, steel hardening, roll force and torque according to real data measured on the mill. The main identified values are friction and steel hardening. The friction coefficient depends on lubrication and on the state of rolls abrasion. Steel hardening is affected by chemical composition and previous treatment of material. Both variables strongly influence rolling force and torque on mill stand. Incorrect mill pre-setting causes overloading or ineffective exploitation of mill drive; unbalanced power distribution over the mill drives in tandem or strip sliding in some mill stand. Hence, the optimal tandem pre-setting is changed from coil to coil.

Classical Bland-Ford model [1], [2] is suitable for on-line modelling of rolling process and it is applied in industry as well. In this model an identification of pressure shape along the roll bite is essential for exact computation of the roll force. A roll pressure p_x reaches its maximum value in the neutral plane, where it creates so called ‘friction hill’ (Fig. 1b). The rolling force is then computed as the area under roll pressure curve. Since computation of this integral is complicated and time consuming, the researchers have sought different methods. One approach replaces the complex computation of roll pressure by a constant value – average roll pressure p_{avg} [3], [4] or by simplified models with substitute functions [5]. A common disadvantage of those methods is their validation for limited rolling conditions only. The roll force computation is also complicated by the fact that it depends on deformed rolls radius (i.e. rolling arc length) and vice versa. At analytical approach a right value of rolling force and deformed rolls radius are computed by iteration and multiple re-computation of rolling model.

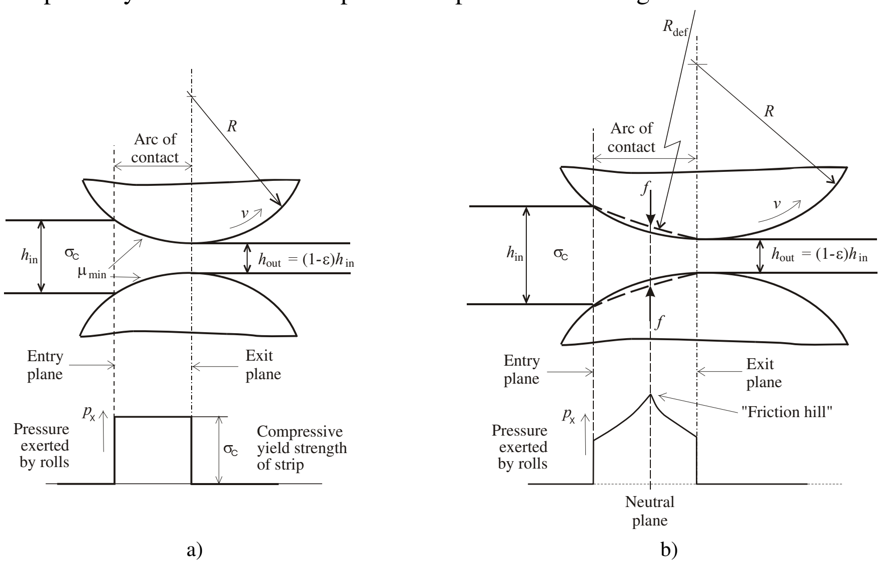

Figure 1
Rolling with a) minimum coefficient of friction and b) with coefficient greater than minimal one (friction hill presence)

f	specific rolling force for plastic deformation
h_{in} , h_{out}	strip thickness at entry and exit of the roll bite, respectively
p_x	rolling pressure
R	roll radius
R_{def}	deformed roll radius
v	circumferential velocity of roll
σ_c	Compressive yield strength
ψ	angular coordinate of the strip element in the roll bite
μ_{min}	minimum coefficient of friction
ε	reduction

A different approach has appeared during the last decade. The Finite Element Method (FEM) is applied to analysis of roll bite conditions. This model is more precise then above-mentioned ones but it requires a lot of computational time. Therefore it is not suitable for direct on-line control. In this case the on-line controller is based on Artificial Neural Network (ANN) [6] and model is used for off-line analysis and ANN controller training only.

Mathematical model of rolling process applied in the paper is based on a modified Bland-Ford model [5]. The modified model unlike classical one considers also the elastic compression of strip at input and elastic recovery of strip at output of the roll bite, respectively, and gets more precise results. The most important values to be identified by model are friction coefficient and steel hardening. Presented paper shows a new approach to friction coefficient identification by application of genetic algorithm.

2 Torque-Force Mathematical Model of Rolling Process

The neutral plane placement in rolling bite depends on many factors: friction coefficient (it depends on state of rolls and strip surfaces and used lubrication), rolls dimensions, desired reduction and front/back tensions. Friction between rolls and strip is unavoidable for transfer of deformation energy. At very small friction a tangential rolls velocity is smaller than that of strip and material slides over the rolls. On the contrary, too great friction needs higher rolling force and motor power, and causes a higher heat production in the roll bite. Rolling with high friction is used in special cases of material treatment, e.g. temper rolling. In other cases a good lubrication is necessary. There is a minimum value of friction coefficient μ_{min} that enables required strip reduction (Fig. 1a). The expression for μ_{min} follows from equality of horizontal forces in the roll bite [7]

$$\mu_{\min} \cong \left(1 + \frac{\sigma_{in} - (1-\varepsilon)\sigma_{out}}{\varepsilon\sigma_c}\right)\sqrt{\frac{h_{in}\varepsilon}{4R}} . \tag{1}$$

If friction coefficient exceeds μ_{min}, the pressure in roll bite will vary and create friction hill. The neutral plane position is characterised by equal circumferential speed of rolls and strip. The placement of neutral plane is influenced by friction coefficient, back/front tension and deformed roll radius. At cold rolling the neutral plane is kept near the exit plane of roll bite.

Steel hardening occurs at cold rolling. A strip hardness is describes by the compressive yield strength σ_c and it increases proportionally with reduction ε. Some types of equations can describe this effect. Strip yield strength according to cubic polynomial [8] was applied in this paper:

$$\sigma_c = f(\varepsilon) = A\varepsilon^3 + B\varepsilon^2 + C\varepsilon + D \qquad (2)$$

where A, B, C, D are coefficients derived through statistical regression of data got by laboratory test on steel samples. Samples were cut from processed material between rolling stands. Initial value of yield strength (coefficient D) depends on steel chemistry and hot roll coiling temperature. The example of yield strength shape applied in the paper is shown in Fig. 2.

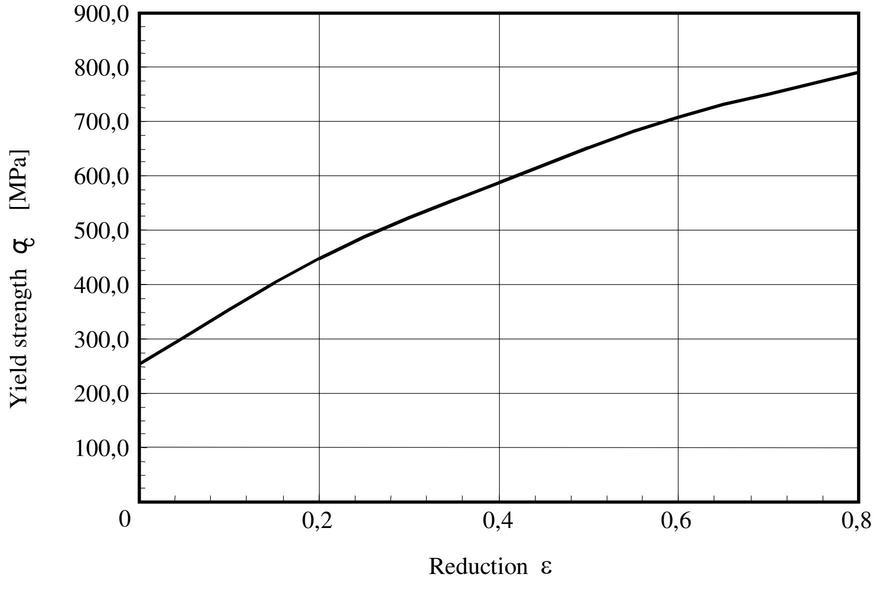

Figure 2

Steel yield strength shape

(Applied polynomial: $\sigma_c = 313{,}11\varepsilon^3 - 1024{,}5\varepsilon^2 + 1340{,}5\varepsilon + 256{,}65$)

As was mentioned above, the mathematical model of rolling applied in the paper is based on a modified Bland-Ford model. Classical model describes the plastic deformation in roll bite (zones II and III in Fig. 3) only. Modified model considers also the elastic deformation of strip at input (zone I) and output (zone IV) of the roll bite, respectively (see Fig. 3). The introduced model is valid under following assumptions:

- A friction coefficient is constant within the whole roll bite,
- A contact arc between rolls and strip remains circular after roll deformation,
- The strip perpendicular sections remain straight within the whole plastic deformation zone.

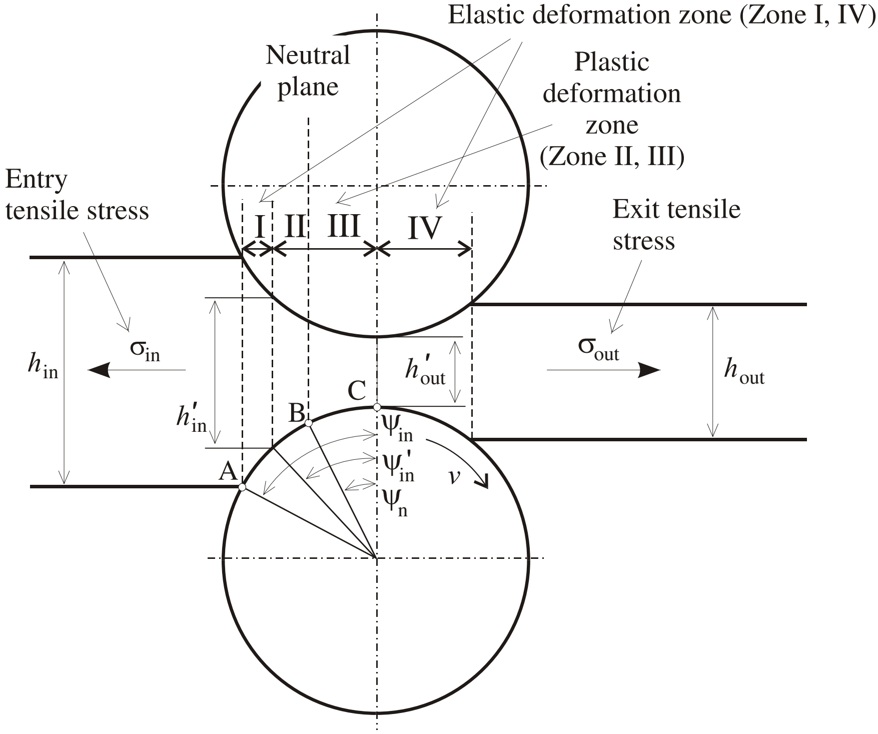

Figure 3

Conditions in roll bite with considering the elastic compression zone and elastic recovery zone

h_{in}, h_{out}	strip thickness at entry and exit of the roll bite, respectively
h'_{in}, h'_{out}	strip thickness at entry and exit of roll bite plastic zone, respectively
v	circumferential velocity of roll
σ_{in}, σ_{out}	back and front tensile stress, respectively
ψ_{in}	angular coordinate of the strip element in the roll bite relating to the entry plane of roll bite
ψ'_{in}	angular coordinate of the strip element in the roll bite relating to the entry plane of plastic deformation zone
ψ_n	angular coordinate of the strip element in the roll bite relating to a neutral plane

Reader can find the entire derivation of modified Bland-Ford model in [5]. Only the final equations will be presented here. Specific rolling force (per unit width of strip) f_{roll} is

$$f_{roll} = f + f_{e\_in} + f_{e\_out} \ . \tag{3}$$

The specific rolling force in the plastic deformation zone f will be obtained by integration of specific roll pressure p_x along the whole arc of contact between strip and work rolls. Specific roll pressure at stand entry (zone II) is given by following equation

$$p_x^{input} = k_x \frac{h_x}{h'_{in}}\left(1 - \frac{\sigma'_{in}}{k_{in}}\right) e^{\mu\left(H'_{in} - H_x\right)} \ . \tag{4}$$

Similarly specific roll pressure at stand exit (zone III) is determined by

$$p_x^{output} = k_x \frac{h_x}{h'_{out}}\left(1 - \frac{\sigma'_{out}}{k_{out}}\right) e^{\mu H_x} \ . \tag{5}$$

Expression H_x depends on a roll bite geometry. Its value in any place of contact arc equals

$$H_x = 2\sqrt{\frac{R_{def}}{h'_{out}}}\, arctg\left(\psi_x \sqrt{\frac{R_{def}}{h'_{out}}}\right). \tag{6}$$

H'_{in} is obtained by substitution of value ψ'_{in} in (6).

Angle Ψ_n corresponding to the neutral plane follows from equality of p_x in equations (4) and (5)

$$\psi_n = \sqrt{\frac{h'_{out}}{R_{def}}}\, tg\left(\frac{H_n}{2}\sqrt{\frac{h'_{out}}{R_{def}}}\right) \tag{7}$$

where H_n in the neutral plane is

$$H_n = \frac{H'_{in}}{2} - \frac{1}{2\mu}\ln\left[\frac{h'_{in}}{h'_{out}} \frac{\left(1 - \frac{\sigma'_{out}}{k_{out}}\right)}{\left(1 - \frac{\sigma'_{in}}{k_{in}}\right)}\right] \ . \tag{8}$$

Specific force f for plastic deformation is then given by equation

$$f = R_{def}\int_{\psi_n}^{\psi'_{in}} p_x^{input}\, d\psi_x + R_{def}\int_{0}^{\psi_n} p_x^{output}\, d\psi_x \tag{9}$$

After substitution of (4) and (5) into (9) we get

$$f = R_{def} \int_{\psi_n}^{\psi'_{in}} k_x \frac{h_x}{h'_{in}} \left(1 - \frac{\sigma'_{in}}{k_{in}}\right) e^{\mu\left(H'_{in} - H_x\right)} d\psi_x + R_{def} \int_0^{\psi_n} k_x \frac{h_x}{h'_{out}} \left(1 - \frac{\sigma'_{out}}{k_{out}}\right) e^{\mu H_x} d\psi_x \ . \tag{10}$$

The k_x represents a strain resistance and is defined as

$$k_x \doteq 1{,}15\sigma_{cx} \ , \tag{11}$$

where σ_{cx} is assigned according to (2) for element of strip at position given by x coordinate. Total specific rolling force also includes forces for elastic deformation at entry $f_{e\text{-}in}$ (zone I) and exit $f_{e\text{-}out}$ (zone IV) of stand (see Fig. 3)

$$f_{e\_in} = \frac{1-\upsilon^2}{4.E}.h_{in}.\sqrt{\frac{R_{def}}{h_{in} - h_{out}}}.\left(k_{in} - \sigma_{in}\right)^2 \tag{12}$$

$$f_{e\_out} = \frac{2}{3}\sqrt{\frac{1-\upsilon^2}{E}.R_{def}.h_{out}.\left(k_{out} - \sigma_{out}\right)^3} \ . \tag{13}$$

A rolling force causes deformation of processed steel and working rolls as well. Therefore a radius of deformed working rolls has to be considered. Deformed radius according [5] is defined as

$$R_{def} = R\left(1 + \frac{16.\left(1-\upsilon^2\right).f}{\pi.E.\left(\sqrt{\left(h_{in} - h_{out}\right) + \Delta h_t + \Delta h_2} + \sqrt{\Delta h_2}\right)^2}\right) \tag{14}$$

where

$$\Delta h_2 = \frac{1-\upsilon^2}{E}.h_{out}.\left(k_{out} - \sigma_{out}\right)$$

$$\Delta h_t = \frac{\upsilon\left(1-\upsilon\right)}{E}.\left(h_{out}.\sigma_{out} - h_{in}.\sigma_{in}\right).$$

As could be seen in equations (9) and (14) a roll force depends on a deformed roll radius and vice versa. True values of both parameters are got by iteration. The model main issue is then actual friction coefficient. Its value is used for prediction of roll force and roll torque for next rolled coil. The critical point of iterative computation is right choice of initial conditions especially initial coefficient of friction, steel hardness and roll radius. Improper choice can cause divergence of solution and model gives no results.

With right values of friction and deformed roll radius a rolling torque could be computed. This parameter is important for checking current load of rolling motor and prediction of intended one. Knowledge of predicted load torque enables changes in rolling strategy (distribution of strip reduction and mill drives load over the tandem, respectively). There are some equations for roll torque calculation. One of common use is as follows

$$M = RR_{def}\left(\int_{0}^{\psi_{in}} p_x \psi_x d\psi_x + \frac{\sigma_{in}h_{in} - \sigma_{out}h_{out}}{2R_{def}}\right). \tag{15}$$

Models used for tandem mill pre-set require fast computation methods. Values identified by the model from current rolled coil are used immediately for tandem pre-set of the next rolled coil. If the model is satisfactory quick the obtained values may be used for control of some variables on the current rolled coil.

Entire control of rolling process also includes a thermal and strip flatness models, which are not subjects of this article.

3 Application of Genetic Algorithms

Genetic algorithm falls into evolutionary algorithms group like evolutionary programming, simulated annealing, etc. [9]. It works on Darwinian principle of natural selection where biological individuals of population strife for survival and possible reproduction. Their success depends on adaptation ability. Genetic algorithm (GA) presents a stochastic optimisation method of multi parameter functions with more extremals.

GA is characterized by following properties:

- It provides wider limits of application in comparison with classical optimisation methods, because it is applicable for non-linear, discontinuous, time delay containing and similar systems. The main constraint is need of satisfactory time for off-line computation.
- It does not work with local parameters of the optimisation process but with global chromosome structure, within which parameters are coded.
- It applies searching methods that do not require information about computation control, but only definition of the objective function. However, additional information (structure, parameter interval, initial values) may substantially decrease the time necessary for finding the optimal solution.
- Statistical transition rules are used for searching process control.

GA randomly chooses for computation the parameters falling into admissible range. The parameters combinations that suit the best to the desired solution are used in next computations. The others including divergent ones are rejected. The interval of admissible parameter combination is gradually reduced by this way.

The following approach was applied in the presented paper. Input variables for GA were

- Measured rolling force f_{roll_act} ,
- Hardening shape $\sigma_c = f(\varepsilon)$ according to known grade of rolled steel (Fig. 2),
- Input thickness h_{in} ,
- Desired thickness reduction ε ,
- Measured back and front tension value σ_{in} and σ_{out}, respectively.

A change in steel hardness and friction coefficient has the same influence on rolling force. Hence, it is not possible to identify the right value of both variables by data measured on one stand only. Therefore, the steel hardness was chosen as a fix value and friction coefficient was identified. The fitness function was defined as a minimal difference between computed roll force f_{roll} and its measured value f_{roll_act}. Exactness of the searched friction coefficient μ was increased by hybrid approach, i.e. after approximate solution by genetic algorithm a direct method such as hill climbing was applied. Further, GA enables to check the parameter combination out of searched range by mutation.

In technical praxis, the parameter values fall into known range. The fact enables to set initial intervals more precisely and increase the computational speed by this way.

Table 1 shows an example of values applied in computation for one working point (point A in Fig. 5). The values match to the 1st stand of five-stand tandem mill.

Table 1

Example of values applied in the computation

Specific total roll force f_{roll}	833,7 kN/m
Roll radius R	0,25 m
Deformed roll radius R_{def}	0,344 m
Input thickness h_{in}	0,002 m
Output thickness h_{out}	0,001522 m
Reduction ε	0,239
Initial steel hardness	256,56 MPa
Back tensile stress σ_{in}	23,4 MPa
Front tensile stress σ_{out}	125,2 MPa
Angular coordinate of neutral plane Ψ_n	0,0162 rad
Rolling torque M	28,302*10^3 Nm
Identified friction coefficient μ	0,103

Curve of a specific rolling pressure p_x over the roll bite angle Ψ_x is shown in Fig. 4. A presence of friction hill is evident.

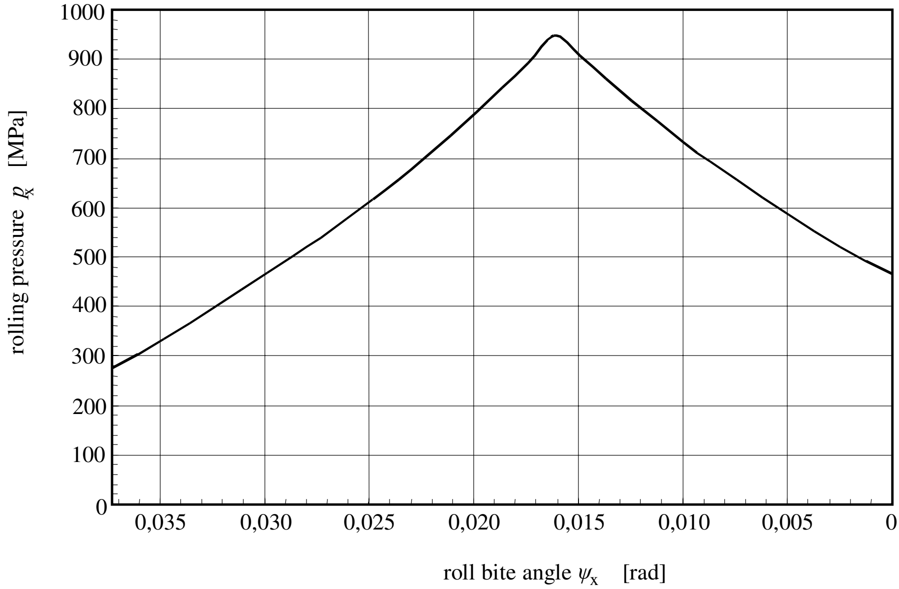

Figure 4
Specific rolling pressure p_x as a function of roll bite angle Ψ_x

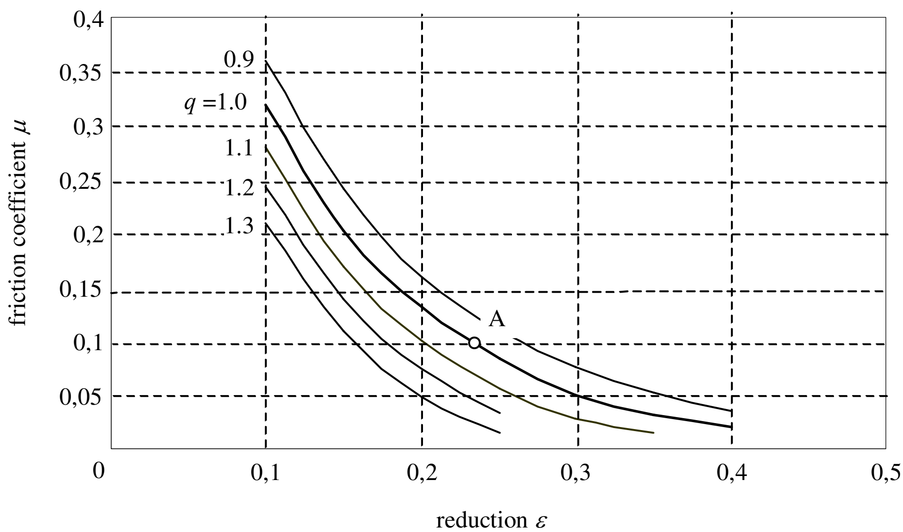

Figure 5
Friction coefficient μ as a function of reduction ε and steel hardening q (q in p.u.)

On the real rolling mill stand the friction coefficient, steel hardness and reduction could vary in certain interval. Fig. 5 shows the parameter combinations for the same rolling force and back/front tension. The friction coefficient was calculated for steel hardening values in the range from 90% to 130% of nominal value. In Fig. 5 the steel hardening is expressed by parameter q (in p.u.) with reference value σ_c (according to parameters in Fig. 2).

Fig. 5 shows conditions that could be attended at next roll coil. It gives a possible pre-setting of strip reduction on the roll stand for identified friction and different steel hardness.

Conclusion

Application of genetic algorithm instead of analytical calculation offers relatively simple and fast way to determination of important rolling process parameters. In the paper, the relations amongst rolling force, tensions, steel hardening and friction coefficient at cold steel rolling has been studied. GA was used to find an actual value of friction coefficient between rolls and steel strip on the current rolled coil. The results of analysis could be used for adaptation of rolling strategy for next rolled coil. The computations were performed on PC equipped by Athlon 2,5 GHz and Genetic Algorithm Toolbox of Matlab. The time elapsed for off-line model computation for one rolling mill stand was less then 50 seconds. However, the presented approach may be applied for analysis of more rolling mill stands. Approximately one minute of computational time could be reached with optimised algorithm for full tandem mill what is acceptable for mill presetting in real conditions.

Acknowledgment

This work was supported by the VEGA Project 1/2177/05 “Intelligent mechatronic systems”.

Nomenclature

E	modulus of working rolls elasticity
f	function
f	specific rolling force for plastic deformation (per unit width of strip)
f_{roll}	total specific rolling force (per unit width of strip)
f_{e_in}, f_{e_out}	specific rolling force for elastic deformation (per unit width of strip) at entry and exit of roll bite, respectively
h_x	thickness of strip element in the roll gap
h_{in}, h_{out}	strip thickness at entry and exit of the roll bite, respectively
h'_{in}, h'_{out}	strip thickness at entry and exit of roll bite plastic zone, respectively
H	expression given by roll bite geometry
k	strain resistance
M	roll torque
p_{avg}	average roll pressure
p	specific roll pressure
p_x^{input}, p_x^{output}	specific roll pressure in a strip element in the roll gap at entry and exit zone, respectively

q	parameter
R	radius of working rolls
R_{def}	radius of deformed working rolls
v	circumferential velocity of rolls
ε	strip thickness reduction
μ	coefficient of friction
σ_c	compressive yield strength
σ_{in} , σ_{out}	back and front tensile stress in strip, respectively
ψ	angular coordinate of the strip element in the roll bite
υ	Poisson's ratio

Indices

$_x$	distance along roll bite in the direction of rolling
$_{in}$	relating to the entry plane of the roll bite
$_{out}$	relating to the exit plane of the roll bite
$_n$	relating to the neutral plane
´	relating to the entry or exit planes of plastic deformation zone

References

[1] D. R. Bland-H. Ford: The Calculation of Roll Force and Torque in Cold Strip Rolling with Tensions. Proc. Inst. Mech. Eng. 1948, Vol. 159, pp. 144-153

[2] W. L. Roberts: Cold Rolling of Steel. Marcel Dekker, New York and Basel, 1978, pp. 243-253

[3] A. I. Tselikov: Theory of Rolling Force Computation in Rolling Mills. Metallurgia Moskva, 1962 (in Russian)

[4] K. Stýblo: Machinery Equipment of Rolling Mills and Drawing Mills I. ES VŠB Ostrava. 1970 (in Czech)

[5] C. C. Y. Lin, M. Atkinson: Comparison of 3 Cold Mill Rolling Force Models: The Modified Bland-Ford Model, The Roberts Model and the Stone Model. Modelling of Metal Rolling Processes. 2nd International conference London, 1996, pp. 478-485

[6] O. Wiklund: Modelling of the Temper Rolling Process Using Finite Elements and Artificial Neural Networks. Modelling of Metal Rolling Processes. 2nd International conference London, 1996, pp. 400-410

[7] W. L. Roberts: Computing the Coefficient of Friction in the Roll Bite from Mill Data. Blast Furnace and Steel Plant, June 1967, pp. 499-508

[8] V. Machek: Cold Roll Thin Steel Strips and Sheets. SNTL Praha 1987 (in Czech)

[9] Z. Michalowicz: Genetic Algorithms + Data Structures = Evolution Programs. Springer Verlag 1995

Number Archetypes and "Background" Control Theory Concerning the Fine Structure Constant[1]

Péter Várlaki

Budapest University of Technology and Economics
Széchenyi István University
Bertalan Lajos utca 2, H-1111 Budapest, Hungary
varlaki@kme.bme.hu

László Nádai, József Bokor

Computer and Automation Research Institute of Hungarian Academy of Sciences
Kende utca 13-17, H-1111 Budapest, Hungary
⟨nadai⟩|⟨bokor⟩@sztaki.hu

Abstract: In this paper we analyze in detail the central role of number '137', the so-called Fine Structure Constant in the collaboration of Pauli and Jung. First, we present the fascination or the obsession of Pauli for the interpretation of number '137'. Second, we treat the spontaneous messages originating from unconscious concerning number '137' in the well-known dreams of Pauli. We restrict our investigations to the dreams containing the especially important formulae of Fine Structure Constant ($4\pi^3 + \pi^2 + \pi$), and also that containing the so-called background models of mathematical control systems. Third, we shortly mention four of the numerous synchronicities arising during the Pauli–Jung collaboration.

1 Introduction

The co-operation and correspondence of Wolfgang PAULI and Carl Gustav JUNG in physical, philosophical and psychological topics have led many authors from the 60ties onwards to discuss and analyze these issues. A special feature of these discussions is that they are directed, apart from philosophical and physical aspects, to the interpretation of the dreams of Pauli by Jung. This statement of ours is

[1] Extended version of a lecture given at *12th International Conference on Intelligent Engineering Systems*, February 25-29, 2008, Miami, Florida.

based on Jung’s book *Psychology and Alchemy* [14] where he analyzes, apart from 4 explicitly religious dreams, 72 other dreams that all share a pattern presumably from alchemy and Cabbala.[2] (He analyzes three out of the four religious dreams in the *Terry Lectures* at Yale in 1937 [13].)

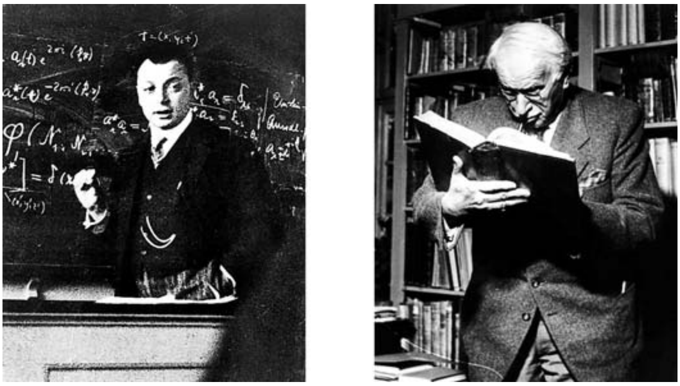

Figure 1
Wolfgang Pauli and Carl Gustav Jung

The *World Clock* vision is of special importance here because of its complex dynamical symmetry structure which gives an appropriate basis both to physical and psychological interpretations as well as to hermeneutics. We mention just three names of those authors who yielded analysis of this vision: W. Byers-Brown, D. F. Peat [36], van Erkelens [41].

We can associate to these dreams the other series of dreams uncovered and published in the Pauli–Jung correspondence (see [27], in German 1992, in English 2002), whose patterns, language, and dynamic shows a close tie to those analyzed in the publication mentioned above. The date of the last dream mentioned in the correspondence is August, 1957. The World Clock vision (1932) plays a central role in the correspondence; one of its paraphrases is communicated to Jung by Pauli in 1955. The quarter-of-a-century correspondence of the two thinkers puts emphasis on this dream.

Another significant aspect of the analysis of dreams and visions is the concept of background processes introduced by Pauli. According to Pauli, these background processes have an impact not only on the development of scientific concepts and ideas, but also on the so-called *Weltanschauung* approaches.

---

2 There is the well known medieval mystical pattern (4 + 72) – the name of God with 4 or 72 letters – which is echoed in Jung’s writing, probably quite unintentionally. It functions like pure synchronicity.

In this paper we analyze in detail the hidden governing role of number '137', the so-called *Fine Structure Constant*[3] in the collaboration of Pauli and Jung. The analysis covers four different areas.

1 First, we present the fascination or the obsession of Pauli for the *interpretation of number '137'*. Here we consider primarily his theoretical or philosophical writings instead of his professional papers; and we quote five especially important extracts word for word.

2 Second, we treat the spontaneous messages originating from unconscious concerning number '137' in the well-known dreams of Pauli [14], [27]. They are so numerous that even to comment them would require a whole supplementary book. So we restrict our investigations to the dreams containing the especially important *formulae* of Fine Structure Constant ($4\pi^3 + \pi^2 + \pi$) [42], and also containing the so-called *background models* of mathematical control systems.

   Remark. It is astonishing that neither Pauli, nor Jung have ever thought of interpreting number '137' in the dreams of Pauli, however, from their correspondence [27] it can be proved without any doubt that Jung had red at least four of the five cited studies.

3 Third, we deal with the comparison of the two kinds of "observations" (external and internal) concerning the computation and interpretation of the fine structure constant. The epistemological aspect of the *two observations problem* will also be briefly touched upon.

4 Fourth, we shortly mention some of the numerous interesting and significant *synchronicities* arising during the Pauli–Jung collaboration.

To anticipate, we shall stress that our approach is basically *empirical* and *heuristic* and it concerns the questions of discovery rather than that of philosophical legitimating.

2 The Fine Structure Constant and Pauli

2.1 Number '137' – the Fine Structure Constant

The concept of fine structure constant (*Feinbaukonstante*) was introduced in the early 1910s by A. SOMMERFELD explaining the spectral lines of hydrogen atom radiation, i.e. the fine structure of hydrogen spectrum. Indirectly this "innovation" played an important role in the development of the Bohr–Sommerfeld atom modeling framework [40].

[3] In fact $1/\alpha = 137.035\ 999$ is the *inverse* Fine Structure Constant [38].

Definition 1. Fine structure constant is a ratio which characterizes the "amount" of the electromagnetic (mutual) effect (independent of the selection of the dimension) and can be found in the description of the fine structure of the hydrogen spectrum:

$$\alpha = \frac{e^2}{4\pi\varepsilon_0 hc}$$

where e is the elementary charge of electron, c is the speed of light, h is the Planck-constant and ε_0 is the vacuum permittivity.

On the other hand, the value of fine structure constant can be calculated from the direct spectroscopic measurement of hydrogen spectrum too, without knowing the concrete values of the above non-dimensionless fundamental constants. However, the specific "value of the concept" obtained from the spectroscopic measurements depends upon the type (and accuracy) of the concrete quantum-electrodynamic model, which can describe the fine and hyperfine structure of hydrogen spectrum very naturally.

Consequently, despite the fact that it cannot be directly determined from measurement, it can be considered from a "hermeneutic point of view" as an empirical concept "interpreted minimally", i.e. on the basis of the three most important constants of physics, and (at the same time) defines the fine (and hyperfine) structure of hydrogen spectrum.

Properly, in the center of the arising question we can find basically the "true" or "best" value of the fine structure constant [38], because the values "obtained" from the above two possible numerical determinations are not equal to each other. It is natural because we can measure "precisely" neither the speed of light, nor the Planck constant and the elementary charge of electron, furthermore, the accuracy of spectroscopic measurements of hydrogen-spectrum are also very limited not to mention the "higher members" of the intermediate calculations. Therefore, after 1910 the demand has arisen, beside the definitive calculations and spectroscopic measurements, to determine the "true value" of the fine structure constant on a geometric basis, first of all using the value of π.

The intellectual and spiritual challenge of fine structure constant for Pauli can be characterized by the following "emotional impression" of Richard FEYMANN:

"There is a most profound and beautiful question associated with the observed coupling constant, e, the amplitude for a real electron to emit or absorb a real photon. It is a simple number that has been experimentally determined to be close to –0.08542455. (My physicist friends won't recognize this number, because they like to remember it as the inverse of its square: about 137.03597 with about an uncertainty of about 2 in the last decimal place. It has been a mystery ever since it was discovered more than fifty years ago, and all good theoretical physicists put this number up on their wall and worry about it.) Immediately you would like to

know where this number for a coupling comes from: is it related to π or perhaps to the base of natural logarithms? Nobody knows. It's one of the greatest damn mysteries of physics: a magic number that comes to us with no understanding by man. You might say the »hand of God« wrote that number, and »we don't know how He pushed his pencil.« We know what kind of a dance to do experimentally to measure this number very accurately, but we don't know what kind of dance to do on the computer to make this number come out, without putting it in secretly!"

2.2 Pauli on the Fine Structure Constant

As mentioned above the concept of fine structure constant (*Feinbaukonstante*) was introduced in the early 1910s by A. SOMMERFELD explaining the spectral lines of hydrogen atom radiation. In an essay[4] Pauli wrote appreciation of Sommerfeld as follows:

"It must however not be forgotten that here, on account of the well-known divergences in the results following that here, on account of the well-known divergences in the results following from the quantum theory of wave fields, we already find ourselves outside the range of a logically closed theory, and are once more reduced to guessing the correct final formulae. The smallness of these new effects is a consequence of the smallness of the so-called fine structure constant, which is often linked with Sommerfeld's name, since its fundamental significance first came clearly to light through his theory of 1916 of the fine structure of hydrogen spectra. The theoretical interpretation of its numerical value is one of the most important unsolved problems of atomic physics." [31]

While the other fundamental physical constants of Nature are all immensely small or enormously large, this fine structure constant ≈ 137 (precisely unknown value) turns out to be a human sized number. The number '137' and its (possible) place in the scale of the universe particularly fascinated and at the same time puzzled Pauli and continues to challenge the physicists today as well [36].

"On the other hand the law of conservation of electric charge occupies a fundamental and equally important place side by side with the laws of conservation of energy and momentum. There are some interesting attempts at classical field theories which, by using a formally more unified representation of the connection of electromagnetic and gravitational fields, unite the law of conservation of electric charge with the laws of conservation of energy and momentum into a single structure, consisting of five equations. But so far these theories have no natural connection with quantum theory, and are unable to interpret the additional fundamental property of charge, namely that it is atomic. By this latter property we mean

4 "Sommerfeld's Contributions to Quantum Theory." Dedicated to A. Sommerfeld on his 80th birthday on 5th December, 1948. – First published in *Die Naturwissenschaften,* 35, 129 (1948).

the fact that every electric charge occurring in nature is always a (positive or negative) integral multiple of the elementary electric charge

$$e = 4.8 \times 10^{-10} \quad e.s.u.$$

This constant of nature has not yet found its appropriate place beside the constants c, h and κ. A new formulation of quantum theory would be satisfactory only if by an interpretation of the numerical value of the dimensionless number

$$\frac{hc}{2\pi e^2} = 136.8 \pm 0.2 \,.$$

In one of his most interesting, philosophical writings[5] dated from 1936 Pauli wrote as follows:

It were to oppose the atomic nature of charge to the classical law of its conservation as its quantum-theoretical correlate in the same way as it opposes as complementary the laws of conservation of momentum and energy to the description in space-time. Nor do we yet know if such a future theory will or will not modify the connection between the notion of charge and that of space-time, which is so characteristic of theories available at present." [29]

In one of his last essays[6] Pauli set a high importance on fine structure constant:

"One of the most assured empirical results of physics is the atomistic structure of electric charge. Charge values are integral multiples of a fundamental unit, the electric elementary quantum, from which, along with the quantum of action and velocity of light, one can from a dimensionless number, 137.04. To reach this result one requires a considerable part of the classical theory of electricity. In the 17th century, for instance, when it was not known how to measure electric charges and how they are defined quantitatively, this empirical result could never have been obtained and formulated. But we are unable to understand or explain the above number." [34]

As mentioned, in the centre of the arising question we can find basically the "true" or "best" value of the fine structure constant because the values "obtained" from the above two possible numerical determinations are not equal to each other. It is natural because we can measure "precisely" neither the speed of light, nor the Planck constant and the elementary charge of electron, furthermore, the accuracy of spectroscopic measurements of hydrogen spectrum are also very limited not to mention the "higher members" of the intermediate calculations.

---

5 "Space, Time and Causality in Modern Physics." Expanded version of a lecture to the Philosophical Society in Zurich in November 1934. From *Scientia*, 59, 65-76 (1936).

6 "Phenomenon and Physical Reality." *Dialectica*, 11 (March 15, 1957), pp. 35-48. Introduction to a Symposium on the occasion of the International Congress of Philosophers in Zürich, 1954.

In his *Nobel Lecture*[7] Pauli summarizes the epistemological conclusion:

"From the point of view of logic, my report on 'Exclusion principle and quantum mechanics' has no conclusion. I believe that it will only be possible to write the conclusion if a theory will be established which will determine the value of the fine structure constant and will thus explain the atomistic structure of electricity, which is such an essential quality of all atomic sources of electric fields actually occurring in nature." [30]

As also mentioned above, after 1910 the demand has arisen, beside the definitive calculations and spectroscopic measurements, to determine the "true value" of the fine structure constant on a geometric basis, first of all using the value of π. This concept in a certain measure was intuitively advanced by EINSTEIN.[8]

"Inside physics in the proper sense we are well aware that the present edifice of quantum mechanics is still far from its final form, but, on the contrary, leaves problems open which Einstein considered already long ago. In his previously cited paper of 1909 [4b], he tresses the importance of Jeans' remark that the elementary electric charge e, with the help of the velocity of light c, determines the constant e^2/c which is of the same dimension as the quantum of action h (thus aiming at the now well known fine structure constant $2\pi e^2/hc$). He emphasized (1.c., p. 192) "that the elementary quantum of electricity e is a stranger in Maxwell-Lorentz' electrodynamics" and expressed the hope that "the same modification of the theory which contains the elementary quantum e as a consequence, will also have as a consequence the quantum structure of radiation." The reverse of this statement certainly turned out to be not true, since the new quantum theory of radiation and matter does not have the value of the elementary electric charge as a consequence, so that the latter is still a stranger in quantum mechanics too.

The theoretical determination of the fine structure constant is certainly the most important of the unsolved problems of modern physics. We believe that any regression to the ideas of classical physics (as, for instance, to this goal. To reach it, we shall, presumably, have to pay with further revolutionary changes of the fundamental concepts of physics with a still farther digression from the concepts of the classical theories." [32]

---

7 "Exclusion Principle and Quantum Mechanics." *Nobel Lecture* in English, delivered at Stockholm on December 13, 1946 (Stockholm 1948).

8 "Einstein's Contribution to Quantum Theory." First published in P. A. Schilpp: *Albert Einstein: Philosopher Scientist.* The Library of Living Philosophers, Vol. 7, Evanston, III. 1949, pp. 149-160. The present version is translated from the German edition, *Albert Einstein als Philosoph und Naturforscher*, ed. P. A. Schilpp, Stuttgart (1955), pp. 74-84, which is believed to be Pauli's original wording.

2.3 Our Formula and Interpretation for the Fine Structure Constant

Without knowing the "accepted", probably, most accurate two values (considering just the 137.03… value of FSC), the following formula for the general (synchronistic) definition of the fine structure constant was proposed [12]:

$$\alpha^{-1} = 4\pi^3 + \pi^2 + \pi^1 = \pi\left(4\pi^2 + \pi + 1\right) = 137.036\,303\,7\ldots = \alpha^{-1}(\pi) \qquad (1)$$

It can be seen that this formula is simple, general, self-expressive and aesthetically also neat. Furthermore, besides π, sufficiently (according to certain alchemical and traditionally hermeneutical rules) it consists only of the first four integer numbers. The first three numbers (as powers) have some "generative characteristics" but the fourth one (4) with certain topological characteristics (as a multiple) also meets the usual "symbolic demands".

Therefore it is able to symbolize the completeness or perfectness according to the mentioned Caballo-Alchemistic and hermeneutical principles. On the other hand, the first three integer numbers appear in generative way as powers of π, while the fourth one, the '4' hints at a topological structure (as a multiple) satisfying the usual Jungian interpretations as well [14].

The generative geometric structure number version of the fine structure constant can ensure a rather *unique* possibility of hermeneutical interpretation through the tetragonal substitution (interpretation) of π, a slightly similar to the classical alchemical problem of "quadratura circuli". The quaternary substitutive interpretation of $\pi \sim 4$ or $\pi \sim 2$ numbers rewriting into the expression of $\alpha^{-1}(\pi)$ the following natural (integer) structure numbers can be obtained.

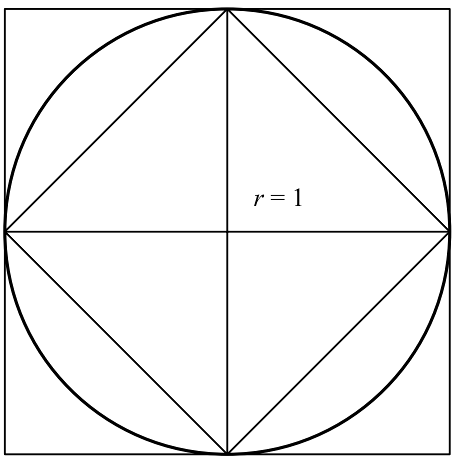

Figure 2
A geometric illustration for problem of "quadratura circuli"

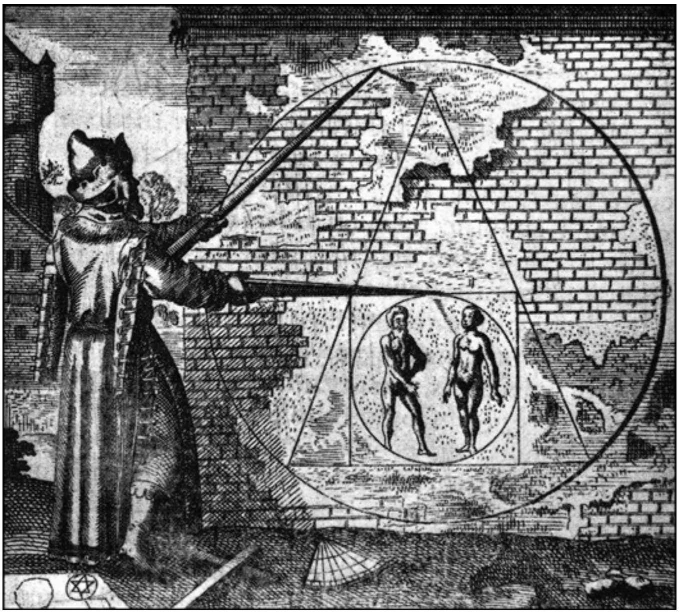

Figure 3
Squaring the circle: two makes the two sexes one whole. (Maier, *Scrutinium chymicum*, 1687)

$$\alpha^{-1}(4) = 4 \cdot 4^3 + 4^2 + 4^1 = 4 \cdot 64 + 16 + 4 = 256 + 16 + 4$$
$$\alpha^{-1}(2) = 4 \cdot 2^3 + 2^2 + 2^1 = 4 \cdot 8 + 4 + 2 = 32 + 4 + 2$$
$$\frac{\alpha^{-1}(4)}{2} - 1 = 137 < \alpha^{-1}(\pi) < 138 = \frac{\alpha^{-1}(4)}{2}$$

The essence of the interpretations is that the tetragonal substitution of π – according to the schema shown in Fig. 2 – can be only the $\pi \sim 4$ as outside the square-measure and $\pi \sim 2$ as the inside square-measure of the "generative circle" with unit radius [44, 46].

3 Background Control Theory

The *Kalmanian* view of systems and their models [20], [9], besides the usual dichotomic reality and description/model "attitude" use trichomic paradigm. It means that besides the (in itself unknowable) "reality" – which is characterized by a certain kind of measurement data – the theory postulates hypostatized *system classes* (using a neutral language of conception), furthermore, constitutes, as a third entity, the model classes related to the former classes of systems.

Our intention is to demonstrate the “background” forms of mathematical system theory both in Pauli’s dreams and in the writings of Pauli and Jung, with a special emphasis on the controlling/regulating principle. While Jung intuitively hypostasize the self-regulating psyche, and stress on the regulating principles, therefore one can perceive a kind of *geometric system theory* with the adoption of the significant concept of ‘projection’. Pauli instead emphasizes the role of iso- and homomorphic relations, algebraic groups and automorphisms that is he intuitively formulates the cognitive basis of background *algebraic system theory*.

“As a consequence of the rationalistic attitude of scientists since the eighteenth century, the background processes that accompany the development of the natural sciences, although present as always and of decisive effect, remained to a large extent unheeded, that is to say, confined to the unconscious. On the other hand, in the Middle Ages down to the beginning of modern times, we have no natural science in the present-day sense but merely the pre-scientific stage, just mentioned, of a magical-symbolical description of nature. This, of course, is also to be found in alchemy, the psychological significance of which has been the subject of intensive investigation by C. G. Jung. My attention was therefore directed especially to the seventeenth century, when, as the fruit of a great intellectual effort, a truly scientific way of thinking, quite new at the time, grew out of the nourishing soil of a magical-animistic conception of nature.

I was well aware, as a pupil of Sommerfeld’s, how these Pythagorean elements appearing in Kepler retrain their vitality even today. That ancient spiritual ‘dynamics’ of number is still active, which was formerly expressed in the ancient doctrine of the Pythagoreans that number are the origin of all things and as harmonies represent unity in multiplicity.” [35]

Pauli postulates a transcendental background, an “informational cosmos” (symbolically returning to the “concept” of *Anima Mundi* – see e.g. [48]) which is the (probabilistic) basis of both physics and psychology, or the subject of both theological and mythological ways of cognition; in which the physical, mental, and spiritual layers create a unity on the basis of a transcendental informational and control language, cf. [37]. This “background” (from which the white noise “steps out”) is called by the authors *Pleromatics* (after the Greek word *πλήρωμα*), and “contains” only symmetry structures and related abstract fields of random variables or fields of contingency without any direct information about the space-time continuum.

Pauli’s theory on background cognitive processes is clearly introduced in his famous study on archetypal influences in Kepler’s scientific discoveries:

“Their agreement with the »primordial images« or archetypes introduced into modern psychology by C. G. Jung and functioning as »instincts of imagination« is very extensive. When modern psychology brings proof to show that all understanding is a long-drawn-out process initiated by processes in the unconscious long before the content of consciousness can be rationally formulated, it has di-

rected attention again to the preconscious, archaic level of cognition. On this level the place of clear concepts is taken by images with strong emotional content, not thought out but beheld, as it were, while being painted." [19, 35]

So, the human recognition concerning the above processes – according to Pauli – formulates dynamical control systems of informational type, where one can not preclude the possibility of personal connections of dynamical systems.

"Furthermore, although I have no objection the existence of relatively constant psychic contents that survive the personal ego, it must always be borne in mind that we have no way of knowing what these contents are actually like 'as such'. All we can observe is their effect on other living people, whose spiritual level and whose personal unconscious crucially influence the way these contents actually manifest themselves." [27]

These can be discussed as a relatively natural and artificial competence in the form of mythological, theological, psychological and biological recognition. The presence of regulating (control) systems can be shown in all of them on the level of both natural and artificial competence. This can be traced in on the level of natural competence in mythology and theosophy as the directing dynamics of the pleromatic world, e.g. in the sephirotic system of the Cabbala. A good example of artificial competence regarding this is Paul RICOEUR's "control-theoretical" hypothesis regarding the magical–mythological ideas, the hermeneutics of meaning, and the Jewish–Catholic religion [39].

Natural competence in psychology, just like artificial competence, appears in less reflected experiments too. The concept of the archetype considered as regulator can be seen in Jung's hermeneutics, tightly connected with the concept of the self-regulating psyche. Quoting Pauli:

"The concept 'archetype' in Jung's psychology, and of its transformation from the original meaning of 'primordial image' to that of an irrepresentable (unanschauliches) structural element of the unconscious, a regulator, which organizes representations (Vorstellungen). Personally I see in this the first indications of the recognition of ordering principles, which are neutral in respect of the distinction psychical-physical, but which, in contrast with the concretistic psycho-physical unified language of ancient alchemy are ideal and abstract, that is, of their very nature irrepresentable (unanschaaulich). Thus the great difficulties and paradoxes in the problem of observation appear clearly. These changes in the ideas of the unconscious show that while still far from having been definitively worked out from the logical side, they are the expression of a line of research in course of development." [35]

In physics artificial competence appears in the control-theoretical and informatical interpretation of Schrödinger's equation, at the same time, it can be discovered in Pauli's dreams, reflections and conscious works corresponding to background control.

Pauli's aim was to establish connection with those cognitive and emotional unconscious processes bearing a transcendental-type and maybe personality which are in tight connection with the evolution of natural and artificial competence of mankind in these fields in a constant interaction. In this aim, new in its own reflection, an idea of central importance which can really be observed in the modern scientific concept in the Jung–Pauli relation.

According to both Jung and Pauli the number archetype is '137' above all, and in tight connection with it the *interpretation* of the number '137'; and in connection with this the mathematical system theory, as well as the use of the control-theoretical paradigm.

In Pauli's view it is the number '137', the value of fine structure constant which surpasses the present quantum theory. As Pauli's former assistant Professor Charles P. ENZ formulated, it takes us beyond physics, and, leaving the separate observers behind, transforms the thinking and feeling scientist confronting with former unconscious processes into the active role of the partaking observer.

"My feeling is that the common ground shared by physics and psychology does not lie in the parallelism of the formation of concepts but rather in »that ancient spiritual 'dynamics'« of numbers that you point out on p. 295. The archetypal numinousity of number expresses itself on the one hand in Pythagorean, Gnostic, and Cabbalistic (Gematria!) speculation, and on the other hand in the arithmetical method of the mantic procedures…"

"Even mathematicians cannot agree among themselves as to whether numbers have been discovered or invented, a fact that finds its counterpart in the modern dilemma of whether the archetype is acquired or is innate. (In my view, both are true.) »In the Olympian host, Number eternally reigns« is a valuable acknowledgment from mathematicians as to the numinousity of number." [27]

Pauli explains several times in his letters and essays that there are at least three different type of mediator "languages" of cognition corresponding to usual (e.g. scientific discovery) and unusual (e.g. active imagination, dreams) ways of knowledge transition:

- theological/metaphysical language;
- physical/symbolic language; and
- language of the psyche or analytical psychology.

Besides the above he postulates a fourth symbolic mediator language that is based on *number archetypes*, but he considers this language "unknown" so far. These four languages can be related to the four ontical layers of Jung–Pauli: spiritual, psychical, material, and transcendent.

In addition to the competence and language of the three fields mentioned (mythological, theological, and physical background theory) the fourth is in our opinion not a future "neutral" new language connected to the number archetype, but an artistic competence (and language) in the widest sense: the *aesthetical category*. The mentioned fourth neutral language appears in the focus as a fifth one, with the number archetype, with the number '137', and maybe with the paradigm of the background system and control theory. This scheme corresponds to the four functions of the conscious, where the central, not as a fifth, but as the entity unifying the other four there is the *transcendent function* according to Jung. In his answering letter Jung wrote that:

"Your explanation of the consciousness quaternio is interesting and, I would say correct. This is also where the 'origin' and primordial home of the number is probably to be found. At any rate where it begins to make its presence felt." [27]

Pauli intuitively grasped the significance of perspective mathematical system and control theory. In one of his important letters to M. FIERZ (Jan. 1948) he wrote:

"The ordering and regulating (controlling) must be placed beyond the difference between psychical and physical as known as Plato ideas are something of the concepts and also something of the natural forces." [48]

K. V. LAURIKAINEN summarized Pauli's ordering and regulating concepts in the following way:

"Pauli understood this ordering and regulating as the common source for science and religion. With the aid of intuition we can come in contact with it in the depths in the psyche, but we can come to studying the logical orders and changes in the nature."

In another place Laurikainen discussed the Pauli approach to the concept of the synchronicity:

"Idea of synchronicity […] is idea that noncausal (acausal) events would be controlled by some kind of regular correspondence." [48]

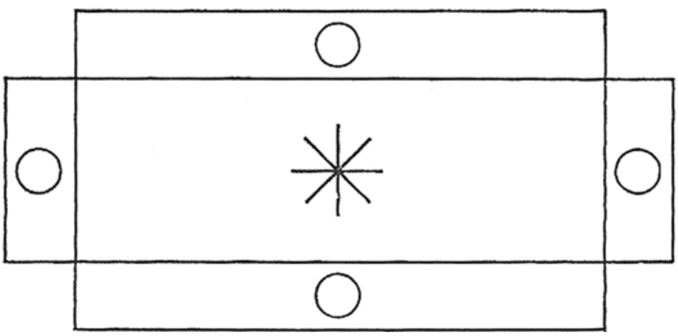

Figure 4

Dream of the four rectangles forming a geometric quaternion [17]

4 The Dream of World Clock as an "Algebraic" Archetype of Control Systems

The *Dream of World Clock* essentially contracts the dream of the four-eyed bear (dream No. 39) with the dream of the four rectangles forming a geometric quaternion (dream No. 51, see Fig. 4).

The dream of two rectangles which we may derive the 32 + 4 + 2 structure from, has a further interesting dynamic feature. Namely, people move clockwise around the four colored waters in the middle of the four derived rectangles. In the centre, however, people move anti-clockwise around the asterisk. This *contrary motion* of the centre already refers to the connection between the energy and the controlling systems. The etymology of the English word 'control' is in complete consonance with the background control image of Pauli's dream. The word 'control' originates from the medieval let in expression 'contra rotulare' being the linguistic mapping of the dream image mentioned above. This feature synchronistically connects the four colors with the World Clock dream, the number '137' and the concept of control together with its dynamic system. The delineation reaches the English world control trough the English-French 'contreroller' and the French 'controlle'.

Jung rather frequently speaks about his patient's dreams containing symmetry structures which are very similar to the "Ezekiel pattern" without any knowledge of the vision of Ezekiel. He remarks that despite its significance in the spiritual history of the Christian-Jewish world, even among the highly educated people there is an almost complete ignorance "in this field". Jung classified this kind of dreams as Mandala ones, which can be experienced as certain psychic synthesis of a psychoid transcendental background and real psychic "foreground". Furthermore, the appearance of the possible synchronistic phenomena was interpreted by him as the parapsychological equivalent of this transgressive totality (completeness) experience [46].

"I have always been particularly interested to see how people, if left to their own devices and not informed about the history of the symbol, would interpret it to themselves. I was careful, therefore not to disturb them with my own opinions and as a rule I discovered that people took it to symbolize themselves or rather something in themselves. They left it as belonging intimately to themselves as a sort of Creative background, a life-producing sun in the depths of the unconscious mind. Though it was easy to see that it was often almost a replica of Ezekiel's vision, it was very rare that people recognized the analogy, even when they knew the vision -which knowledge, by the way, is pretty rare nowadays."

Pauli probably was influenced by the concept of "creative background" from Jung's *Terry Lectures* (1937).

Unfortunately, we have no possibility here to give even a short survey about the known analysis and interpretations for the symmetry structures of the vision of

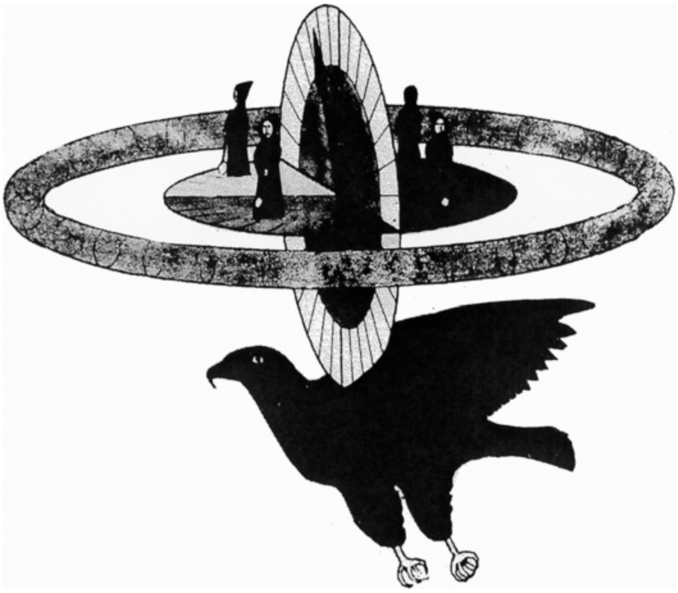

Figure 5
The World Clock.
An impression generated by W. Byers-Brown based on accounts of Pauli's dream [17].

Ezekiel. It would demand a separate book… However, in the following we shortly outline the background control system "interpretation" of the vision, and the 4 + 16 + 256 system representations. In this rare interpretation one counts the four Cherubims standing at the four corners of the Chariot of God. All the Cherubims have 4 × 4 body-parts (4 faces, 4 wings, 4 hands, and 4 legs) according to the four natures (man, lion, bull, and eagle). This system of 4 + 16 + 256 can be related to the 39th dream of Pauli [46]:

"Dreamer is falling into the abyss. At the bottom there is a bear whose eyes gleam alternately in four colours: red yellow green and blue. Actually it has four eyes that change into four lights."

Inasmuch the bear symbolizes north in mythology, as well as in astronomy and the Chariot of God arrives from north on the sky in the vision. The algebraic variation system of the four lights of the four eyes can be related in a natural way to the 4 + 16 + 256 system interpreted in the previous paragraphs. According to the parity conception of Pauli (right-left sides of space), the above structure can be simplified into 128 + 8 + 1. That is, the dream of Pauli connects the Merkabah vision with the fine structure constant – without any knowledge of mythology [46].

In the dream of four rectangles (dream No. 51, see Fig. 4) beside the four colors we can identify 32 geometric elements (12 corners, 16 lines and 4 rectangles). It is originally formed from two basic rectangles, so the structure 32 + 4 + 2 is valid.

The exact text of the dream of World Clock is the following:

"There is a vertical and a horizontal circle, having a common centre. This is the world clock. It is supported by the black bird.

The vertical circle is a blue disc with a white border divided into 4 × 8 = 32 partitions. A pointer rotates upon it.

The horizontal circle consists of four colours. On it stand four little men with pendulums, and round about it is laid the ring that was once dark and is now golden (formerly carried by the children).

The »clock« has three rhythms or pulses:

1) The small pulse: the pointer on the blue vertical disc advances by 1/32.

2) The middle pulse: one complete revolution of the pointer. At the same time the horizontal circle advances by 1/32.

3) The great pulse: 32 middle pulses are equal to one revolution of the golden ring."

The rotations define three rhythms that are in an order of powers: 1/32, 1, 32, 32^2. The archetype number of rotation (or the circle) is naturally the π. In the spatial structure – as Jung analyzes the dream – number 4 dominates over number 3 of temporality. The rotation itself symbolizes Time, in this manner the spatial and temporal structure of the Dream of World Clock is symbolically isomorphic with the formula of the fine structure constant $\alpha^{-1} = 4\pi^3 + \pi^2 + \pi$.

Namely, the dream can be considered as a central algebraic (system) archetype, and at the same time, as an automorphic mapping of the cardinal number archetype in the symbolic system of the dream.

From the Pauli–Jung letters [27] it is clear that the black bird is supported by the female symbolic figure of *anima*, where anima is suited with number '7'. In this way the black bird (1), the rotating discs and ring (3), and the figure of anima (7) altogether associates to number '137'.

Furthermore, the temporal structure (rotation scale) itself is double 32, and the spatial structure is also evidently 2 × (32 + 4) (the four little men with pendulums – horizontal disc, and four cardinal positions of pointers – vertical disc), that altogether results in 1 + 2 × 32 + 2 × 36 = 137.

The dreams No. 39 and 51 naturally and evidently contained in the structure of World Clock vision (four little men, four colors create an isomorphic map with the four eyes and four colors of the bear). Therefore, the three power-like rhythms with the four space-like quaternio (like a multiplicator) is a natural and evident isomorphic map of the $4\pi^3 + \pi^2 + \pi = 137{,}036\ldots$ Consequently, the structure of the World Clock vision (together with the dreams No. 39 and 51) is a perfect isomorphic structure of the above formula of fine structure constant and its discussed isomorphic interpretations (if π ~ 4 or π ~ 2). Jung often refers to the old familiar saying: *on the Olympus numbers reign*. Following this pattern, we regard the supreme ruler ("controller") of number archetypes the specific formula introduced for fine-structure constant.

Remark. The interpretation of the four wheels in the Ezekiel vision with the (then usual) 8 spokes can ensure a structure number of '32 + 4 + 2' beside the above '256 + 16 + 4' one. Since the number of the Hebrew word for wheel (which is a fundamental concept in the tradition of Ezekiel vision) 137 (אופן = 137), so the isomorphy with the structure of World Clock vision and the formula $4\pi^3 + \pi^2 + \pi$ can be considered as a complete one. The interpretation of three rhythms and space quaternio also can be easily detected from the structure of the Ezekiel vision. (Other similar historical examples were mentioned in our earlier works [44, 45, 46]). From the point of view of the depth psychology the "numbers" 137 and $4\pi^3 + \pi^2 + \pi$ are the twin number archetypes of the Self (Selbst).

Table 1
The role of fine structure constant in topological and dynamical structure of the hydrogen atom

Combination	*Atomic property*	*Characteristic length (m)*
$\alpha^3/4\pi R_\infty$	The classical electron radius, r_e	$2.8179380(70) \times 10^{-15}$
$\alpha^2/2R_\infty$	The Compton wavelength of the electron, λ_e	$2.4263089(40) \times 10^{-12}$
$\alpha/4\pi R_\infty$	The Bohr radius of the hydrogen atom, a_0	$5.2917706(44) \times 10^{-11}$
$1/R_\infty$	The reduced wavelength of hydrogen radiation	$9.11267034(83) \times 10^{-8}$

5 Two Kinds of 'Observations'

5.1 The Interpretation of the Five Concepts of FSC in Physics

On the vertical axes of the Pauli–Jung quaternio we can see the concept of spatial-temporal continuum and the indestructible energy. The fine structure constant (FSC) appears as a connecting constant, just like in connection with electro photon. Its value can be taken as a wavelength. It plays a vital role in the spatial-temporal structure of the hydrogen atom since it matches to the ratio of the reduced radiation and the *Compton wavelength*. The same ratio is determined between the *Bohr radius* and the elementary radius of the electron. The hydrogen atom gives the measure of the component of the relativistic energy during the energy emission and can easily be measured from the spectrum of the hydrogen atom by well-known spectroscopic measurements. These two determination conceptions stand in a complementary relation to each other. The horizontal axes illustrate the contingent (synchronistic) and causal concept of FSC.

EDDINGTON's model already mentioned matches the causal descriptive mode in which the four variables of the spatial temporal continuum gives 16 equations where the number of the independent variables arranged in a matrix is 137 $= (16^2 - 16)/2 + 16 + 1$. This approach seems to be synchronistic from the perspective of classical physics while it seems causal in character from the perspective of modern quantum physics. As appose to the real physical phenomena of the previous axis it is obvious that we do not have concepts emerging from immediate experience but an intuitively appealing mathematical (or meta-mathematical) interpretation leading out from physics. Although it starts out from physical thought, the result – properly speaking – is not physical, but it is the concept of a background language. This solution is near to Pauli's mental world since he often makes references to the problem of FSC and his explanation lies outside quantum physics. Pauli applies the same line of thought to the Einsteinien criticism against the quantum theory. According to him these lifelike questions are outside physics (*In Sinne des Lebens betrachtet*), or, in other words, these questions can be interpreted and approached in a wider framework.

We have suggested the $4\pi^3 + \pi^2 + \pi$ formula to the concept of FSC. This fulfills the claim of maximal simplicity, and also it ensures the value exactly matching the measurements. Simplicity is fulfilled by the four integer and the π appearing in the equation in the simplest way. The connection opens the possibility of interpretations on the basis of approaching the circle by quadrates and so it gives the expressions 256 + 16 + 4, and 32 + 4 + 2. The former also reflects the Eddington approach.

This way the determination of FSC of the horizontal axis is complementary too, but also mutually containing. In a similar way the contingency approach and the causal approach are complementary. Obviously this connection also leads out from physics although the measured values within the error limit and is derived from the π and it can be interpreted only in the wider conceptual framework called life realization by Pauli. We put into the centre unifying concept in contrast to the Pauli–Jung idea. Jung did the same in the case of the four orientation function, where the unifying transcendent function is in the centre. In our case the central concept is the controlling of the world and the concept unifies the information cosmos and the controlling world. With regard to FSC the central concept is intimately connected to '137' as it appears in the Dirac hypothesis and his calculation according to which the electron trajectories are stable up to the nucleus containing maximum 137 protons. (As we could see '137' is seen as the structure number of mathematical controlling systems.) Thus we could get to the unique number '137' in five different ways.

Remark: From the point of view of the depth psychology the number '137' is the number archetype of Self (Selbst).

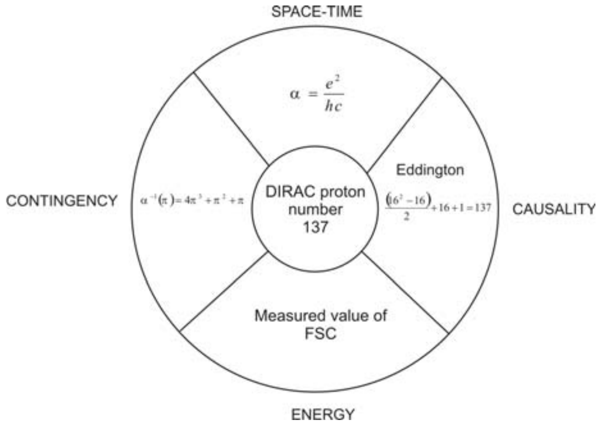

Figure 7

"Ontical versions" of the interpretation of the fine structure constant

5.2 The Interpretation of the Five Concepts of FSC in the World Clock Dream

The contingency interpretation of FSC appears with the three temporal rhythms and the spatial quaternio in the dream. This interpretation is not ad hoc, since Jung gave the same summary about the dream. The rotation consisting in a power-like 32 rhythm metric corresponds to the triple: π, π quadrate, π cube. The rotation whose basic rhythm has length π and whose shape is a *sin* or *cos* time function matches well to the elements of formula (1). The semi-wave with length π is information without redundancy.

That is the rhythm of rotation is isomorphic with π, and the operation of rising to power is isomorphic with the mutual embedding of the three rhythms. The spatial quaternio is connected to the three rhythms of time. Thus this connection is isomorphic with the multiplication of four and the π. These considerations are supported by Pauli's later dreams showing oscillation (vibration) curves [27]. The dream interpretation of causal FSC can be arrived at by taking the four small men with pendulum and the four spatial directions changing permanently with temporal rotation. In this case we get a 16 by 16 system on the basis of the symbolic interpretation of the spatial temporal quaternio, which is isomorphic with the Eddington model concerning the identification of the number of independent variables.

We can derive the permanent symbolic equivalent of the FSC determined by spectroscopic measurements from the dream of bears having four eyes in four colors. In this description the four (basic) colors randomly alternate with the four eyes making a "stochastic process", whose (virtual) spectrum is the symbolic equivalent of the color spectrum. As we have seen, the order with a 256 + 16 + 4 structure also leads to the number '137', as to the number of its independent variables. The dynamic picture of the colored bear with four eyes is isomorphic with the four different little men and the four colors. Since Pauli's dream succeeds Eddington's model with two years it can, at best be called cryptomnesic symbolization.

The spatial-temporal symbolic interpretation of FSC can be understood with quadrate of '32'. Namely, '32' in the contingency formula of FSC symbolize '137'. Thus it matches to the ratio of the Compton wavelength and the reduced radiation wavelength of the hydrogen atom. In the spatial interpretation the same applies to the ratio of the elementary radius of the electron and the Bohr radius. Thus the World Clock dream can be considered as a symbolic isomorphic representation of the spatial temporal structure of the hydrogen atom. Here again we have '137' in the centre whose spatial structure 2×36 and temporal structure 2×32 plus the unique entity (the black bird), altogether give 137.

According to Jung's concise interpretation of World Clock dream:

"Thus circle and quaternity on the one side and the threefold rhythm on the other side interpenetrate each other so that the one is also contained in the other." [13, p. 88]

Remark: The relationship between the two observations can be explained with the help of the $4\pi^3 + \pi^2 + \pi$ formula. Number '137' is not entirely sufficient in itself, since it does not directly explain the value of the physical measurements. The $4\pi^3 + \pi^2 + \pi$ formula is a productive concept, being reliable guide to leading out the interpreter from physics and directing him to the hermeneutics of Pauli's life "realization". Meanwhile it gives an obvious explanation of the three physical interpretations. It also reflects the Eddington model and shows its legitimacy. Similarly, it leads to the control theoretical interpretation which is central importance. The discovery, rather than the construction of the system and control theoretic approach is the discovery in a Popperian sense of the laws of nature just like the laws of physics are discovered by the physicists. As the quotations demonstrated Pauli has come very near to recognizing the significance of the information ordering principals. He was explicit about it, but there were not available modern mathematical and control theoretic "operators" to support his point. As we have mentioned Jung talks about self-regulating psyche from the 10s, and conceives the archetypes, together with Pauli, as an ordering and regulating system (entities).

There arises the question whether we are entitled to establish a connection between these two different observations. The epistemology of our enterprise cannot be easily settled. Either we can appeal to analogies and afford analogical explanation; the retort could be that it is a mere contingency that we have the associations which are not upheld by the objective underlying phenomena. In this case we can defend our approach by its satisfying the minimum constraint of coherency, albeit we are fully conscious of the possibility of alternative coherent explanations. Or, the unification has an objective character; in this case we can appeal to the so-called inference to the best explanation (IBE) which justifies us to accept a theory provided it yields the best possible explanation of the phenomena (Harman, *The Thought*, 1960). Thus, our formula and interpretation satisfies either the less ambitious criterion of coherency, or, the more ambitious one of inference to the best Explanation.

Remark. The relevance of the two observations is this: In the first case we are related to the detached observer, in the sense of Pauli, in the second case we are related to the undetached observer and his experiences concerning the synchronicities. In the latter the "validity condition" depends upon the density and quality of the appropriate synchronistic events.

Remark. The synchronistic feature of the dream and the four plus one scheme interpretation is unambiguously shown by the fact that in the *Collected Works* of Jung (Psychology and Alchemy) the number of the footnote referring to the dream is astonishingly '137'.

The spontaneous feeling of Pauli about of supreme harmony after the dream and Jung's reflections to it together with the number '137' of the footnote constitute reliable basis to the serious consideration of our analysis. Not to speak of the frivolous fact that Pauli died, as it is familiar in the room '137' of the Red Cross Hospital of Zurich.

Here we can assume also a unifying fifth meta-language based on number archetype, and geometric and algebraic system theories. Jung often refers to the old familiar saying: *"in the Olympus numbers reign"*. Following this pattern, we regard the supreme ruler of number archetypes the specific formula introduced for fine-structure constant.

6 First Manifestations of Pauli's Preoccupation

Coming back to the original thought, already in his school years, Pauli was fascinated by RYDBERG's famous Formulae about the length of the periods in the periodical system of elements. In his famous book *Atombau und Spektrallinien* [40], Sommerfeld called this 'two-p-square' equation a Cabbalistic formula. This is exactly the book which has the atomic model structure identifiable from the measured spectral lines in its focus. Further the introduction of the fine structure constant and the emphasizing of its central importance in the construction of the atomic structure. This means that the purely synchronistic coherence of the word Cabbala[9], cabbalistic sequence of numbers, and the number '137' appear in the book at simultaneously. All these fascinated Pauli reading and listening to Sommerfeld. As Pauli said in a lecture appreciating Rydberg [33]:

9 The Hebrew letters of 'Kabbalah' sums up to 137.

"After the lecture of Professor Bohr I only briefly mention Rydberg's idea of a relation between the parity of multiplicity of spectral lines and the parity of the chemical valence. If the one is even the other is odd and vice versa. However, Rydberg was not sure of this idea being unable to give a general proof of it. The reason for it was first that the order and the resolution of more complicated spectra (as for instance copper) was not sufficiently advanced at that time to determine correctly this multiplicity. Secondly the difference between spark spectra and arc spectra made complications, as it was not known at that time that the first are emitted by ions. Only much later it became clear that the rule holds without exception if the chemical valence is replaced by the number of electrons in the emitting atom. This exact rule of alternation (»Wechselsatz«) was called after Rydberg by Sommerfeld."

The Rydberg formula is explained in detail later in the same text [33]:

"A further progress in the order of the periodic system was made by Rydberg in his paper »Elektron, der erste Grundstoff« of 1906. Here (p. 11) he said for the first time that the 3 numbers 2, 8, 18 for the periods in the system of the elements are represented by $2 \cdot 1^2$, $2 \cdot 2^2$, $2 \cdot 3^2$. There was still some uncertainty about the number of the rare earths, which Rydberg assumed to be 36 instead of 32. His atomic numbers were still too high but not as high as in his earlier papers.

In a big paper »Untersuchungen über das System der Grundsoffe« of 1913 he goes one step further. After the quotation of the earlier formulas $2 = 2 \cdot 1^2$, $8 = 2 \cdot 2^2$ and $18 = 2 \cdot 3^2$ he goes on (§3): »the continuation would be $2 \cdot 4^2 = 32$, $2 \cdot 5^2 = 50$ etc.« This is the famous formula $2 \cdot p^2$ (p integer) which Sommerfeld called »cabbalistic« in his book »Atombau und Spectrallinien« and which impressed me very much as student. Definitely he says now about »the group G4« (p =4, rare earths) that it consists of 32, not of 36 elements."

The formula yields to numbers 2, 8, 18, 32, 50, 72, 98, 128, …, and Rydberg dealt especially with 32 of them. The even numbers 8, 32, 72, and 128 have a particular connection, and indeed, play a central role in the number system of Cabbala. We remark that 8, 128 and the unity sums up to 137, moreover, 32 and 72 sums up to 104, and these two numbers play important role in *Bahir* ("Book of the Brightness"). Here 72 is the number of *Tamar* (cf. Exodus 15:27) who bore the Messiah twins *Zarah* and *Pharez* (Genesis 38:29–30) – they can be suited to paired spectral lines, the so-called *doublets* which appeared frequently in Pauli's dreams.

It is also interesting to mention in connection with Rydberg that he expected the period-length to be 36 instead of 32. Therefore, the curious young Pauli became very early acquainted with the ambivalence, and at the same time with the identity of 32 and 36 (this is perceptible e.g. in the dream of World Clock).

The ambivalence of pair 32–36 together with unity is always a reference to '137', forasmuch, according to Pauli: *"The initial stage is a dyadic archetype whose proton corresponds to the same one and whose electron correspond to the other one. Through reflection of the unconscious a quaternity is produced."* [27]

Numbers 36 and 32 play decisive role in Bahir, namely, the 36 *Dekan* (the 36 hidden men) walk on the 32 ways of Wisdom. This is in Hebrew written as 'LVLB' and pronounced as 'Lulav' (לולב = 36 – 32) that is the shoot on the Date Palm 'Tamar'. Therefore we get again a reference to Tamar, to the twins, or to the doublets that were primordial – as Pauli himself stated in his Nobel Lecture [30] – in the formulation of the *Exclusion Principle*.

At the same time, in the introductory part of the Lecture he reported that the first impulse for his research work was given by Sommerfeld, who explained the structure of atoms using Rydberg's Formula and the numbers 2, 8, 18, and 32 (the natural period lengths of chemical elements). However, Sommerfeld emphasized number 8 among the former, and used the geometry of the *cube* to demonstrate it. With Pauli's own words:

"Sommerfeld to the structure of the atom – somewhat strange from the point of view of classical physics. I was not shared the shock which every physicist, accustomed to the classical way of thinking, experienced when he came to know of Bohr's 'Basic postulate of quantum theory' for the first time. At that time there were two approaches to the difficult problems connected with the quantum of action. One was an effort to bring abstract order to the new ideas by looking for a key to translate classical mechanics and electrodynamics into quantum language which would form a logical generalization of these. This was the direction which was taken by Bohr's Correspondence Principle. Sommerfeld, however, preferred, in view of the difficulties which blocked the use of the concepts of kinematical models, a direct interpretation as independent of models as possible, of the laws of spectra in terms of integral system, an inner feeling for harmony. Both methods, which did not appear to me irreconcilable, influenced me. The series of whole numbers 2, 8, 18, 32… giving the lengths of the periods in the natural system of chemical elements, was zealously discussed in Munich, including the remark of the if 'n' takes on all integer values. Sommerfeld tried especially to connect the number 8 and the number corners of a cube."

It is interesting, that the other inspiring lecture that led Pauli to the Exclusion Principle was given by Niels BOHR, and also was connected with the period-length of the periodic system. However, Bohr emphasized number 2 instead of number 8.

"A new phase of my scientific life began when I met Niels Bohr personally for the first time. This was in 1922, when he gave a series of guest lectures at Göttingen, in which reported on his theoretical investigations on the periodic system of elements. I shall recall only briefly that the essential progress spherical symmetric atomic model. […]

It made a strong impression on me that Bohr at that time and in later discussions was looking for a general explanation which should hold for the closing of every electron shell and in which, the number 2 was considered to be as essential as 8 in contrast to Sommerfeld's approach." [30]

Thus, the starting points of the Nobel Lecture were the interpretation of Sommerfeld and the cabbalistic numbers of Rydberg, and logically, the closing statements were emphasizing the significance of fine structure constant itself.

“From the view of logic my report on ‘Exclusion principle and quantum mechanics’ has no conclusion. I believe that it will only be possible to write the conclusion if a theory will be established which will determine the value of the fine structure constant and will thus explain the atomistic of electric fields actually occurring in nature.” [30]

In the preceding paragraphs we have shown that the properly given and explained formulae of fine structure constant ‘137’ (see Eq. 1) can lead far beyond physics in the narrow sense. Therefore the young (19–20 years old) Pauli – without any previous knowledge (!) – was enormously effected by Cabbala, especially by the *number archetypes* of the proto-cabbalistic Bahir. These synchronicities concerning number ‘137’ were in close correlation with his conscious physical aims. (This is certified also by his dreams relating to Sommerfeld.)

These associations does not decrease the spontaneous character of Pauli’s dreams that are analyzed by Jung in *Psychologie und Alchemy* [14], moreover, they emphasize the presence of a strong synchronistic influence –a prematurely constellated archetype using the common term of Pauli and Jung.

Therefore, both at the beginning and at the termination[10] of Pauli’s career there emerging the Cabbala, the Jewish traditions, and the number ‘137’. *“According Quispel, Pauli, searching for a meaning to his life while confronting his death, came to reassert his Jewish tradition.”* [27]

7 Understanding Synchronicity

(1) The Dream of World Clock in the final (both English and German) version of Jung’s *Collected Works* [15] is associated with a footnote numbered by ‘137’. This is a replacement of *asterisk* (character ‘*’, in Latin ‘asteriscus’), which – beyond emphasizing some meaning – symbolically denotes the crown. (Both the celestial star as a physical body, the linguistic symbol, the liturgical object of Eucharist, and the Holy Crown of Hungary are equally corresponding to the asterisk.)

10 The “final meeting” of Pauli with the number ‘137’ can be considered as an astonishing “subjective case” and example of synchronicity. “It was a mystery how Pauli was taken to his death, for on being carried into hospital, the physicist was told that he would be put into room 137. According to one version of story, learning his room number, Pauli said: ‘I will never get out of here’, and it happened so really (he died shortly after)” [36].

The content of footnote No. 137 is about augmentation, and the expanded version can be found in Jung's book *Psychology and Religion* [13]. Therefore the opus of *alchemy* expands, and – through religion – is connected naturally to the meaning of *redemption* and *eternal life*.

(2) The death of Pauli in the room No. 137 of Red Cross's Hospital in Zurich represents the same idea: number '137' mediates between physical existence, the psychic experience of death and resurrection, and spiritual eternity.

(3) Jung commenting the Dream of World Clock two times mentions that the dream has a possibly *cabbalist* character, besides its obvious connections with alchemy. Then in *Psychology and Alchemy* [14] there are numerous examples from (mostly Christian) cabbalist writings.

On the other hand, Jung emphasizes that the dreamer (Pauli) has a highly biased intellectual modality, and does not possess any (deeper) knowledge of mythology, mystics, or religions. It is important for Jung to emphasize the spontaneous formation of those dreams that are connected to alchemic and cabbalist motifs.

However, as it was demonstrated earlier in this paper, the *number archetype* had a decisive effect on Pauli in his student years, and exactly, in connection with the co-occurrence of word 'Cabbala' and number '137'.

Moreover, the well-known conversation[11] between Victor WEISSKOPF, a leading physicist and a former assistant of Pauli and Gershom SCHOLEM, one of the most eminent scholars of Jewish mysticism is also clearly imply the direct connection between number '137' and Cabbala. It is also an interesting synchronicity that after World War II Jung was the one who invited Scholem from his Palestinian isolation to the *Eranos Lectures*, and therefore introduced him into the European scientific community. Pauli wrote in a letter to Scholem the following:

"I read your book »Major Trends in Jewish Mysticism« already several years ago, and since then I have taken the opportunity to cite it, especially concerning the ideas of Isaac Luria, in my paper on Kepler (see »Naturerklarungen in Psyche« especially p. 149 – Rascher Verlag, Zurich 1952).

As you can see from this work, among others, I sit on the ground between the two stools of Orthodoxy and rationalism, but stronger still, I consider this to be the only fair and rational position. I find that the Rationalist authors, (and should not be surprised if I was attacked by this group because of my Kepler paper), also misunderstand the mental processes totally, because, in part they judge everything from the point of view of pure psychology consciousness and partly they hold on to

11 When Scholem met Weisskopf, he asked about the prominent unsolved problems in physics. Weisskopf said, "Well, there's this number, 137…" And Scholem's eyes lit up! He said, "Did you know that one hundred thirty-seven is the number associated with the Cabala?" [21] (We want to clarify that this means that the *Hebrew letters of 'Kabbalah' sums up to* 137.)

a world view which has been superseded by modern physics long ago – on the other hand, I think that I am sensitive to where a psychological mental phenomenon begins for the Orthodox, which I would call »wavering«. (It seems that in Jewish and Christian orthodoxy, this is not essentially different.) I think I also detect this wavering in Luria quite clearly (see the cited passage in my article). Also, I have some doubts whether Jewish mysticism is really fundamentally different from the non-Jewish, – I am interested in mysticism in general."

Semantic quternio is determined by the connection of the four persons Jung, Scholem, Pauli and Weisskopf who know well each other. One can often meet such semantic quaternion with the dream interpretation of Pauli and Jung (for example Pauli, Bohr, Einstein and Jung).

(4) An interesting synchronistic phenomenon (properly a precognition) happened in October 1949. In his important dream we can read: *"I am with colleagues on one of the upper floors of a house where a local conference on mathematics and physics is being held. I see that under my name a course of cookery is announced: 'Start: December 15.' Surprised, I ask a young man near me why the course begins so late in the year. Ha answers: 'Because the Nobel prize will be granted.' Now I notice that a fire has started in the adjacent room."* Without any detailed interpretation of the dream we all can feel the deep dark atmosphere of the precognition for the death of dreamer on 15th December, cf. [41]. (The Nobel Lecture in English was delivered at Stockholm, on December 13, 1946.)

It is part of the synchronicity of the dream from 1949 that the sum of the preindicated date in dream (15/12/1949) is '32', just like the sum of the numbers in the date of Pauli's death (15/12/1958). It is interesting that the opposition day of 15th of December is 16th of June to a large extend (according to 182 + 5/8 calculation). These two days make a contrary just like the twins *Castor* and *Pollux,* or the day and the night, or *Olympos* and *Hades*, the past and the future etc. These two twin points of time referred to '32' and '36' in the years of 1949 and 1958, respectively. Taking twice 32 and 36 extending with the unique entity we obtain 137. The background physical conception of Castor and Pollux[12] explained by Pauli is revealed to his letter to Jung dated 23/10/1956. Bear in mind that 16/06/1958 is the day of the execution of the Hungarian Prime Minister Imre NAGY, while the former date is the date of the outburst of the Hungarian revolution. Synchronicity is supported further by Pauli's letter to Jung dated 16/06/1936 where the sum of

[12] It is almost certain that Pauli did not know the next original version (or its tranlation) from *Nostradamus* concerning the "fate of Hungary". It seems to be an astonishing example for the conjunction (the most important purpose of Pauli's intellectuel and spiritual endevour) of the precognition and retrocognition in the synchronistical phenomena.
"Par vie et mort changé regne d'Ongrie La loy fera plus aspre que feruice En grand cité urlemens plains et crys Castor et Polux ennemis dans la lyce." (Nostradamus, Centuries, II. 90.)

the numbers in the date gives 32. In this letter he deals with the very important dream on 'Ace Club' and the card '7' (it is the 16th initial dream in Jung's *Psychology and Alchemy*).

Pauli's letter to Jung dated 16/06/1948 is also worth mentioning because a so called "Pauli effect" took place during the foundation ceremony of the Jung Institute. Synchronicity and Hungarian references are demonstrated in the last (13th) dream written to Jung in 23/10/1956. Pauli has dreamed that on the day of St. Stephen (26/12/1955) a king visited him, a great authority who praises him for the ability to see the important phenomena of the world in a "twin-like" manner. St. Stephens's day, the king and the mirror symmetry, together with October 23, 1956 make the strong Hungarian reference highly probable. The fundamental Hungarian references of *Bloomsday* (June 16 in the Ulysses of James JOYCE) are familiar just to refer to Bloom's coronation with the crown of St. Stephen of Hungary in the 15th Chapter. The Hungarian references of number '137' has been mentioned in our previous paper [44].

8 New Ways of Cognition

In a late work in 1957 Jung supposed that the era traditional art closed because the modern art completely diverged from the archetypal source of the great art.

"The pleasingness of the artistic product is replaced by chill abstractions of the most subjective nature which brusquely slam the door on the naive and romantic delight in the senses and on the obligatory love for the object. This tells us, in plain and universal language, that the prophetic spirit of art has turned away from the old object-relationship towards the -for the time being -dark chaos of subjectivisms. Certainly art, so far as we can judge of it, has not yet discovered in this darkness what it is that could hold all men together and give expression to their psychic wholeness. Since reflection seems to be needed for this purpose, it may be that such discoveries are reserved for other fields of endeavour.

Great art till now has always derived its fruitfulness from myth, from the unconscious process of symbolization which continues through the ages and, as the primordial manifestation of the human spirit, will continue to be the root of all creation in the future. The development of modern art with its seemingly nihilistic trend towards disintegration must be understood as the symptom and symbol of a mood of universal destruction and renewal that has set its mark on our age." [16].

Similarly, 1956 Pauli had the same impression on the contemporary philosophy:

"For my impression is that the philosophy of the contemporary philosopher specialist is not really produced with and for the intellect but comes across as a complex and involved emotional attitude […] I regard it is a regression into the undifferentiated." [27]

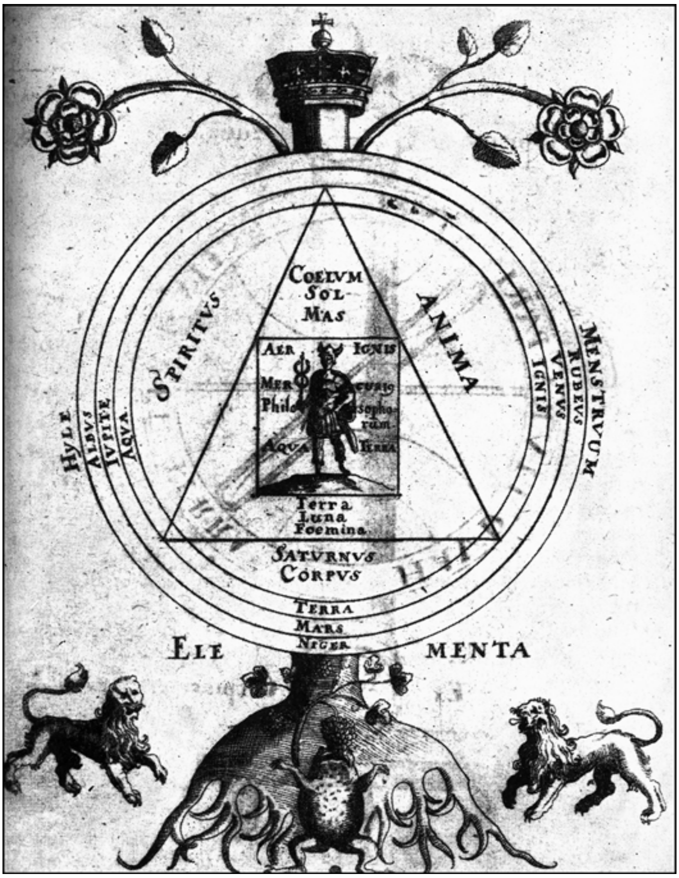

Figure 7
Symbol of Hermetic transformation: the *homo philosophicus* Mercurius (Samuel Norton, *Mercurius redivivus*, 1630)

Without concerning the truth of their judgments the question arises for us what the substitutive and constructive new endeavors are, which have the attribute of reflection. Beside his correspondence with Jung, Fierz and von Franz, Pauli introduces the idea of *background processes* in his famous Kepler studies in a joint book with Jung (*Naturerklärung und Psyche* [19]). The essence of this idea for the alchemists is that the creative (partly cognitive, partly emotional) background processes together with the structure of magico-animistic knowledge form the entirely pattern of the science.

"Furthermore, my feeling is that the purely psychological interpretation only apprehends half of the matter. The other half is the revealing of the archetypal basis of the terms actually applied in modern physics. What the final method of observation must see in the production of 'background physics' [Hintergrundsphysik]

through the unconscious of modern man is a directing of objective toward a future description of nature that uniformly comprises physics and psyche, a from of description that at the moment we are experiencing only in a pr-scientific phase. To achieve such a uniform description of nature, it appears to be essential to have recourse to the archetypal background of the scientific terms and concepts." [27]

It was shown by him that this cognitive and emotional character of the background processes was observable in the main work of Kepler creating really new forms of scientific thinking. At the end of his study Pauli emphasizes the significance of this feature of the scientific creation. *"Just because in our times the possibility of such symbolism has become an alien idea, it may be considered especially interesting to examine another age to which the concepts of what is now called classical scientific mechanics were foreign but which permits us to prove the existence of a symbol that had, simultaneously, a religious and a scientific function."* [35]

In our times the background processes are also associated with the development of scientific knowledge, causing a decisive influence on them. However, from the 18th Century the above background processes cannot be observed directly, i.e. confined to the unconsciousness. In his correspondence Pauli also deals with these problems hoping that in the future a new kind of the cognition will demand both the scientific approaches and symbolic nature of the so called creative background processes. This could be a precognition of a new form of knowledge when the subject of the creative cognition is a scientific mediator and medium at the same time who is confronted with the directions and autonomy of the creative background processes. Furthermore as a scholar applying the prospective productive form and efficient methods of a future artificial intelligence can survey, and then analyze the possible largest fields of the contemporary human knowledge. Supposedly, he will also relate to the above factual knowledge with the symbolic potential of unconsciousness using the direction and dynamics of the creative background processes.

Therefore, in Karl Popper's so-called *First World* the acting scholar mediator/medium perceives intuitively the actual unconscious patterns of the extended *Second World*; at the same time, applies the methods of the artificial intelligence. Thus, he will be able to achieve the suitable data and appropriate pattern of the knowledge of the *Third World* [37]. The conscious intellect of the new scholar can be obtained help from the future artificial intelligence. Similarly his 'mediumship' of the new scholar artist (Cabbalo-Alchemistic artist) can gain a today still unknown support from the unconscious background processes. Naturally, it is true only in case of active cooperation of unconsciousness ('deo concendente'). In the case of Pauli whom can be considered as a prototype of this new scholar-artist (with his intellect and mediumship) the artificial intelligence was replaced by the extended knowledge of the professors of his 'loved Polytechnic' (ETH) together with the professor of University of Zurich.

"What is particularly praiseworthy is the caution with which Pauli applies Jung's 'amplification method'. One cannot help admiring his courage, honestly, and attention to detail. When it came to work of this nature, Pauli was in a very fortunate position in Zurich, for at the university and the ETH he could call on a number of highly qualified advisers from all branches of science. This was a major factor in Pauli feeling so much at home there; he loved 'his Polytechnic' dearly and was happy to return after the war." [27]

The amplifying application of the potential knowledge of the unconsciousness could be realized by the hypothetical principles of spiritualistic communication according to the ideas William JAMES , James HYSLOP and others (see the concept of physical and psychical 'fields' of W. James). The best example/allusion for the possible future development is Pauli himself who could make a strong influence unconsciously under physical measurement equipments. This is so-called *Pauli-effect* that is not a legend, but a pure fact, the relatively frequently occurred events. At the same time this is a good example for wholeness, and for at least a close relationship between the psychical and physical processes. Because even the split of conscious-unconscious psychic state of Pauli which 'destroys synchronistically' the laboratory measurement systems. In the future these phenomena could be an allusion for the artificial amplification of the psychical and physical relationship. Thus, the symmetry structure of the three worlds and their interfaces could be considered as a complete one at least according to our disputable speculations.

Conclusions

Let us finish with the words of Professor Charles Enz [35]:

"An important detail in this thinking about duality is the meaning of the numerical value of the electric charge which in the form of Sommerfeld's fine-structure constant is approximately 1/137. Pauli repeatedly stressed that progress in quantum field theory was linked to an understanding of this number […]. But the number 137 also had an irrational, magic meaning for Pauli; it was the room number 137 in the Red Cross Hospital in Zurich where he died on 15 December 1958 […]

The enigmatic conjecture »that the observer in present-day physics is still too completely detached« also has a meaning beyond physics. Indeed, in his article for Jung's 80th birthday […], Pauli compares the observational situation in physics with that in psychology: »Since the unconscious is not quantitatively measurable, and therefore not capable of mathematical description and since every extension of consciousness ('bringing into consciousness') must be reaction alter the unconscious, we may expect a 'problem of observation' in relation to the unconscious, which, while it present analogies with that in atomic physics, nevertheless involves considerably greater difficulties.«

For Pauli this analogy had implications in both directions: On the one hand, in the concluding remarks of the birthday article for Jung […] he expresses the expectation that in the future the idea of the unconscious should emerge from the *purely therapeutical realm and become more a problem of objective research. On the other hand, he thought that in physics the remedy for the too complete detachment of the observer may lie in the integration of the subjective, psychic. Indeed, in Science and Western Thought […] Pauli asks the question: »Shall we able to realize, on a higher plane, alchemy's old dream of psycho-physical unity, by the creation of a unified conceptual foundation for the scientific comprehension of the physical as well as the psychical?«*

This quest for a unity of physics and psyche is a recurrent theme in the exchange between Pauli and Jung and is the main concern in Pauli's Background Physics in which he was guided by his dream motives (Hintergrundsphysik […])."

References

[1] A. Ádám, L. Jánossy, P. Varga, *Acta Phys. Hung.*, Vol. 4, p. 301, 1955

[2] C. Anderson, "The Positive Electron," *Phys. Rev.*, Vol. 43, p. 491, 1933

[3] A. Antoulas, Ed., *Influence of R. Kalman et al., Mathematical System Theory*. Springer Verlag, 1991

[4] A. Aspect, J. Dalibard, G. Roger, "Experimental Test of Bell's Inequalities Using Time-Varying Analyzers," *Phys. Rev. Lett.*, Vol. 49, No. 25, p. 1804, 1982

[5] W. Barker, F. Glover, "Reduction of Relativistic Two-Particle Wave Equations to Approximate Forms," *Phys. Rev.*, Vol. 99, No. 1, pp. 317-324, 1955

[6] J. Bokor, L. Keviczky, *Topics in Stochastic Systems: Modelling, Estimation and Adaptive Control*, ser. Lecture Notes in Control and Information Sciences Series 161, Springer Verlag, 1991, ch. Parametrizations of Linear Stochastic Systems, pp. 47-65

[7] J. Bokor, L. Nádai, "Controllability of Quantum Bits – from the von Neumann Architecture to Quantum Computing," in *3rd Intl. Conference on Computational Intelligence and Intelligent Informatics*, Agadir, Morocco, March 28-30 2007, on CD-ROM

[8] J. Dan, "Three Types of Ancient Jewish Mysticism," in *7th Rabbi Feinberg Memorial Lec. in Judaic Studies*, University of Cincinnati, 1984

[9] R. Kalman et al., *Mathematical System Theory*. Springer Verlag, 1971, ch. On Invariants, Canonical Forms, Moduli for Linear Constant Final Dimensional Dynamical Systems

[10] H. Grotch, D. Yennie, "Effective Potential Model for Calculating Nuclear Corrections to the Energy Levels of Hydrogen," *Rev. Mod. Phys.*, Vol. 41, No. 2, pp. 350-374, 1969

[11] I. Hargittai, Ed., *International Conferences on Symmetry Research*, Vols. 1-2, SpringerVerlag, 1987-89

[12] I. Joó, P. Várlaki, "Stabilization of Dirac Expansions by Riesz and Other Means," *Annales Univ. Sci. Budapest*, Vol. 39, pp.113-123, 1996

[13] C. Jung, *Psychology and Religion*, ser. Terry Lectures. Yale University Press, 1938

[14] ——, *Psychologie und Alchemy*. Olten: Walter Verlag, 1972

[15] ——, *Collected Works*, ser. Bolingen. Princeton University Press, 1973, Vols. 8, 9, 11, 12, 14

[16] ——, *Collected Works*. Princeton University Press, 1973, Vol. 10. Civilization in Transition, ch. The Undiscovered Self (Present and Future)

[17] ——, *Synchronicity*, ser. Bolingen. Princeton University Press, 1973

[18] ——, *Geheimnisvolles am Horizont : von Ufos und Ausserirdischen*. Olten: Walter Verlag, 1992

[19] C. Jung, W. Pauli, *Naturerklärung und Psyche*. Zürich: Rascher Verlag, 1952, in English *The Interpretation of Nature and the Psyche*, New York, 1955

[20] R. Kalman, P. Falb, M. Arbib, *Topics in Mathematical System Theory*. McGrow-Hill, 1969

[21] E. Klopfer, T. Raj, *The Number 137*, Budapest: Makkabi, 2006, in Hungarian

[22] A. Koestler, *The Roots of Coincidence*. New York: Random House

[23] D. Lindorff, *Pauli and Jung: The Meeting of Two Great Minds*. Quest Books, 2004

[24] A. Lindquist, G. Picci, "On the Stochastic Realization Problem," *SIAM J. Control and Optimization*, Vol. 17, pp. 361-389, 1979

[25] M. Maimonides, *The Guide of the Perplexed*. Chicago and London: The University of Chicago Press, 1963

[26] G. Marx, *Quantum-Electrodynamics*. Budapest: Tankönyvkiadó, 1972, in Hungarian

[27] C. Meier, Ed., *Atom and Archetype: The Pauli/Jung Letters*, 1932-1958. London: Routledge, 2002

[28] G. Michaletzky, J. Bokor, P. Várlaki, *Representability of Stochastic Systems*. Budapest, Hungary: Akadémiai Kiadó, 1998

[29] W. Pauli, "Space, Time and Causality in Modern Physics," *Scientia*, Vol. 59, pp. 65-76, 1936, expanded version of a lecture to the Philosophical Society in Zurich in November 1934

[30] ——, "Exclusion Principle and Quantum Mechanics," Nobel Lecture in English, delivered at Stockholm, December 13 1946

[31] ——, "Sommerfeld's Contributions to Quantum Theory," *Die Naturwissenschaften*, Vol. 35, p. 129, 1948, dedicated to A. Sommerfeld on his 80th birthday on 5th December, 1948

[32] ——, "Einstein's Contribution to Quantum Theory," in *Albert Einstein: Philosopher Scientist. The Library of Living Philosophers*, P. Schilpp, Ed. Evanston, 1949, Vol. 7, pp. 149-160

[33] ——, "Rydberg and the Periodic System of Elements," in *Rydberg Centennial Conference on Atomic Spectroscopy*, Vol. 50, No. 21. Lund, Sweden: Universitetes Årsskrift, 1954

[34] ——, "Phenomenon and Physical Reality," *Dialectica*, Vol. 11, pp. 35-48, March 15 1957, introduction to a Symposium on the occasion of the International Congress of Philosophers in Zürich, 1954

[35] W. Pauli, (Eds. C. Enz, K. V. Meyenn), *Writings on Physics and Philosophy*. Springer, 1994

[36] D. Peat, *Synchronicity: A Bridge Between Matter and Mind.* Toronto, New York: Bantam Books, 1987

[37] K. Popper, J. Eccles, *The Self and its Brain. An Argument for Intractionism*. London, New York: Routledge and Kegan Paul, 1986

[38] Committee on Data for Science and Technology, http://www.codata.org/

[39] P. Ricoeur, "Structure et herméneutique," in *Le Conflit des Interprétations*. Seuil, 1969, pp. 31-63

[40] A. Sommerfeld, *Atombau und Spektrallinien*. Braunschweig: Friedrich Vieweg und Sohn, 1919, in English see *Atomic Structure and Spectral Lines*, translated from the third German edition by H. L. Brose, Methuen, 1923

[41] H. van Erkelens, "Wolfgang Pauli's dialog with the spirit of matter," *Psychological Perspectives*, Vol. 24, pp. 34-53, 1991

[42] P. Várlaki, J. Bokor, "Cognitive Symmetry-Structures in Stochastic Control Theory and Early Quantum Physics," in *MAKOG I. – Cognitive Systems*, Visegrád, Hungary, 1993, in Hungarian

[43] ——, "On Synchronistic Random System's Theory: Cognitive Symmetry-Structures, Natural and Geometric Structure-Numbers in Stochastic Control and Quantumphysics," MTA SZTAKI–BME," *Research Report*, 1993

[44] P. Várlaki, J. Bokor, L. Nádai, "Historical Background and Coincidences of Kalman System Realization Theory," in *5th IEEE Int. Conference on Computational Cybernetics*, Gammarth, Tunisia, 2007, on CD-ROM

[45] P. Várlaki, L. Kóczy, "A Comparative Study of Pictures From Pala d'oro in St. Mark Catherdral of Venice and from the Holy Crown of Hungary," in

Intl. Conference on Genealogy and Heraldry, Nagykanizsa, Hungary, 2006, in Hungarian

[46] P. Várlaki, L. Nádai, J. Bokor, "Numbers and System Representations in Perspective of the Pauli-Jung Correspondence," in CD–ROM of the *6th Slovakian-Hungarian Joint Symposium on Applied Machine Intelligence and Informatics (SAMI 208)*, Herlany, Slovakia, 2008

[47] P. Várlaki, B. Vasvári, personal consultations, Physical Institute of BUTE, 1989

[48] Laurikainen, K., *Beyond the Atom: The Philosophical Thought of Wolfgang Pauli*, Springer Verlag, Berlin, 1988

[49] Keve, T., *TRIAD*, Rosenberger & Krausz, London, 2000 (n. 215)

[50] A. S. Eddington, *Fundamental Theory*. Cambridge University Press, 1946

Control an Electromechanical System in the Real-Time Linux Environment

Matej Dobšovič*, Martin Kratmüller**

* Ajilon Consulting, 33 Regent Street, London SW1Y 4NB
dobsovic@hp.com, tel: 00421-903-259-181

** Siemens PSE SK, Dúbravská cesta 4, Bratislava
martin.kratmueller@siemens.com, tel: 00421-2-59686358

Abstract: This article is concerned with possibilities of control in real-time using Real-Time Linux (RTL). It describes the control abilities of a Personal Computer (PC) with RTL operating system, cooperation with industrial control cards, the architecture of control software and possibilities of use in real applications. We are putting forward experience with a simple interconnection between the RTL kernel and a control card using the Comedi software package.

1 Preface

Along with the massive expansion of robotics the experts keep facing the objective of control in accurate time intervals. In the high-frequency field the standard operating systems working on PC platform (Personal Computer) are not able to work with sufficient time precision. The answer to the increasing demand of time precision is real-time control.

The control via PC has its advantages. It is mostly the ability of easy and fast modification of control software, the possibility of further data processing for vizualisation or interaction with other control parts. There are several kinds of real-time add-ons for Linux in the field of real-time automation by PCs. One of the most widely known is RT-Linux & RTAI. There are also some alternatives for the Microsoft Windows platform – for example RTX (by VenturCom).

The Real-Time Linux has developed in two ways by now. One of them is a well-supported commercial version called RT-Linux PRO. Later the non-commercial version called RT-Linux Free was created, based on RTL PRO. The Free version of RTL may be considered as an independent product because the programming environment is quite different.

We decided to use the non-commercial version RT-Linux Free for the purposes of real-time control. As I/O interface we used the Advantech PCI1711 multifunction card. It is a powerful but low-cost card for the PCI bus [1]. The operating system was based on the Debian Linux distribution and the installation of RT-Linux patch required kernel recompilation.

The physical model consists of a direct-current electromotor physically interconnected with a speed-voltage generator. The speed-voltage generator is used as a rotation speed sensor. The rate between rpms and voltage is approximately 1000 mV ~ 1000 rpm.

The electromotor was powered by a voltage changer with output up to 5 V (defined as maximal DC motor input voltage). The analog input of the card range was set on ± 5 V and output on 0 to 5 V. Digital representation of the range was 4096 units. It means that the input sensitivity was about 2,5 rpms and about 1,2 for output.

2 The Concept of Connection

We used a PC with the Intel Celeron 600 MHz processor and 256 M RAM, an Advantech 1711 PCI card. Measured time for reading of data from an analog input followed by writing random data to analog output was stable, about 19 µs. The control software described below then allowed changing the modes and improving the access performance.

The control card PCI1711 has several functions including 16 single-ended analog inputs, a 12-bit A/D converter & 16 digital inputs and outputs. But in our concept, you can use a few hundreds control cards from different vendors without any radical change of the control software.

It is possible to create a device driver and apply it as the connection between RT-Linux and a hardware device according to manufacturers' specifications. A simpler way is to use an already existing software package, Comedi [2].

Comedi consists of three parts:

- Drivers
- Comedilib (API for standard Linux environment)
- KComedilib (API for real-time environment)

The software package contains drivers for more than 300 control cards manufactured by different vendors, (e.g. ADLink, Advantech, Amplicon, Analog Devices, ComputerBoards, DataTranslation, ICP, ITL, Keithley Metrabyte, National Instruments, Measurement Computing, RealTime Devices, Winsystems).

These drivers are represented by kernel modules and are able to work in real-time mode under RTL. The real-time tasks are handled using the KComedilib library. This library does not contain the whole scale of standard Comedilib functions, but it is not a big obstacle with regard to our needs. The communication between the control module and the driver module is shown in the following picture (Figure 1).

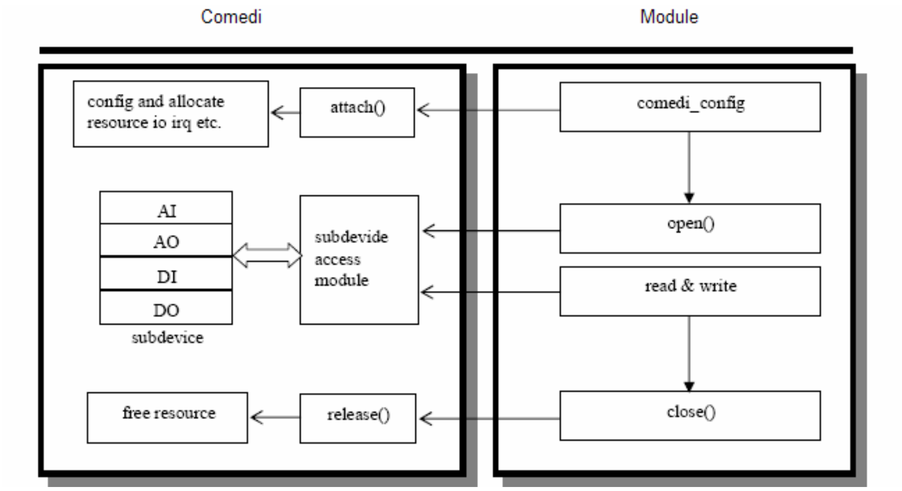

Figure 1
Communication between driver and control module

The control is done directly in the real-time running module. Because of that it's quite important to eliminate and remove all bugs from the control code. Every simple error makes the kernel going down on kernel level.

The kernel of RT-Linux does not use the preemptible kernel concept. It means that it cannot temporarily terminate one task and assign the processor time to another task with a higher priority. Due to this architecture it is advisable to make the control module as simple as possible and thus effective. RT-Linux does not allow the use of additional Linux libraries. The restricted set of functions ensures a low time tasks consumption and this restriction also makes up for the absence of preemptible concept.

The operation part of software (user interface) is a standalone application running in user-space. Communication between the user interface and the operation part is realized through the so called FIFO pipe. Unlike the standard FIFO, RT-Linux supports the real-time FIFOs. It allows you to work with FIFO immediately, for example write the data immediately after measuring. This data are written to the pipe directly and the operation application may change the control method or may be used for visualization.

KComedilib as the basic API for communication with cards contains a relatively sophisticated and very scalable functions. For example it is the subdevice selection (analog I/O, digital I/O, timer, …), the I/O range settings, timer options or the data collection mode. All these may be set through simple and intuitive functions included in this library. The library provides four access modes which specify the way of data-collecting. The simplest way is the 'Simple R/W mode'. The functionality of this mode is shown in the following picture (Figure 2).

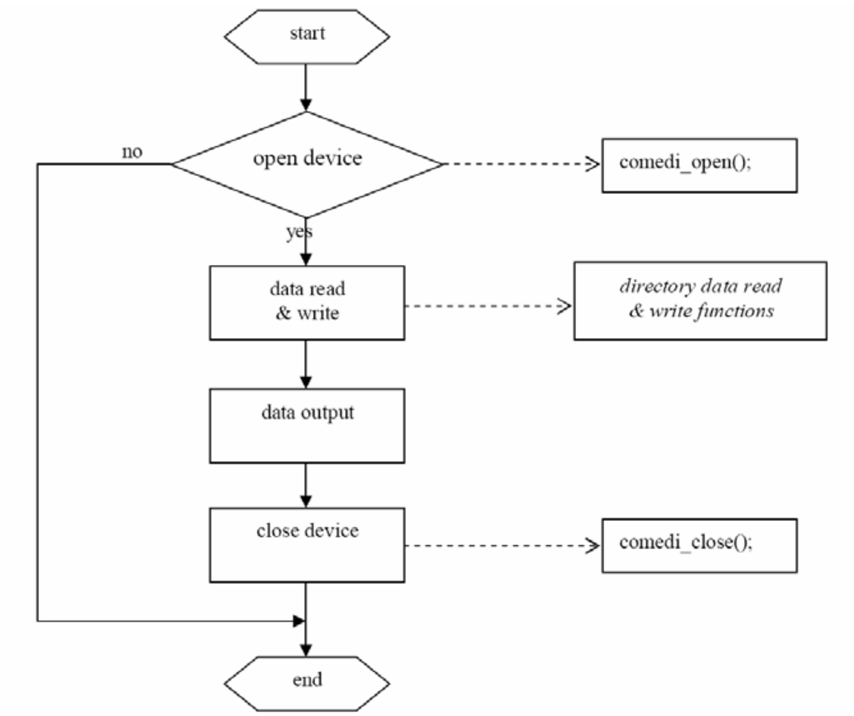

Figure 2
Process diagram of R/W mode

3 Control

The software is divided into two parts. The kernel module contains hardware initialization, creation of real-time pipes and the control algorithms. The user interface is an independent application running in the standard user-space. One named pipe is used as the control pipe (in the application->module direction). All the others are operating in the opposite direction. The architecture of the module is designed as a multi-threaded whereby any thread may read from the generic control pipe.

The responsibilities of the user interface application and the real-time module are as follows (Figure 3).

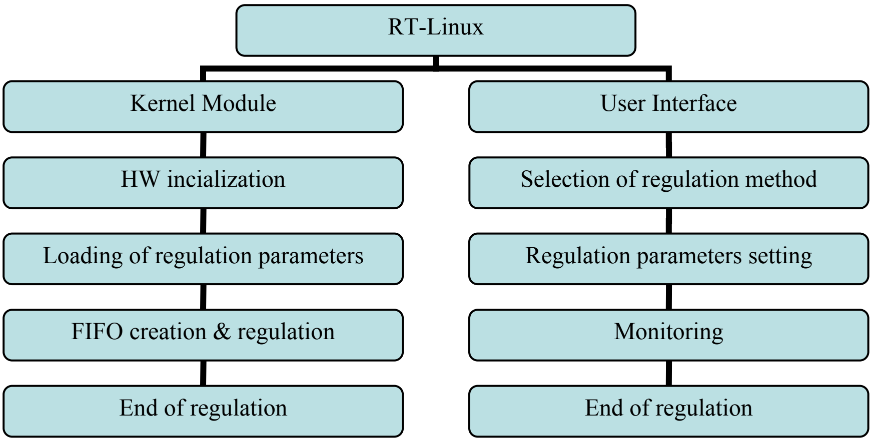

Figure 3
The responsibilities of the real-time and user-space parts of application

Our application contains three threads which, due to the module architecture, can run at the same time and communicate through the same control FIFO. The communication in the opposite direction passes through three real-time FIFOs, one for each thread. The measured data are transferred this way.

The whole concept of the control module is based on a similar architecture described in the module named 'Frank'. This module is supplied as an example in RT-Linux and it is easily alterable. The basic attributes of this concept are multi-threaded architecture, the usage of only one control pipe generic for all the threads and independence of the threads.

Because of that you can use each thread independently. For example one thread for adaptive control, one for control using the PSD algorithm and so on.

As the control card has several inputs and outputs, we can easily automate several processes by one computer with one I/O card in real-time mode.

The architecture allows you to change the control methods. It means that it can choose the best control method according to the current status of measured data. We think it would be really interesting to explore this feature more precisely and extend the possibilities of this method to further real systems.

We've tested three different algorithms of control. The first one is a simple adaptive regulator called 'Lambda Tracking' published in [3, 4]. It was used for simulation of chemical rectifying column control. In our case it was not usable as our system is more dynamic. It causes frequent excitation of the DC motor up to the limit value. Practically it causes lifetime degradation of the actuating unit. Therefore we've abandoned the use of this method and we've tried to regulate the process using standard PSD algorithms.

The most applicable is a speed PSD algorithm with action intervention limitation (AntiWind-Up) [5].

$$\Delta u(k) = P.\big(e(k) - e(k-1)\big) + P_I e(k) + P_D\big(e(k) - 2e(k-1) + e(k-2)\big) -$$

$$-\frac{1}{P + P_I + P_D}\left[(P_I - P_D)\big(u(k-1) - u_M(k-1)\big) + P_D\big(u(k-2) - u_M(k-2)\big)\right] \qquad (1)$$

Where u is intervention, e is control aberration, P, P_I a P_D are the constants of the PSD regulator and u_M is the maximal value of u (the limitation).

According to this control method we eliminated the initial overshoot and reduced the oscillation of required value to minimum. As start control method we used modified Ziegler-Nichols one [6]. We have reached very good results. The control aberration after settlement was about 2% of required value.

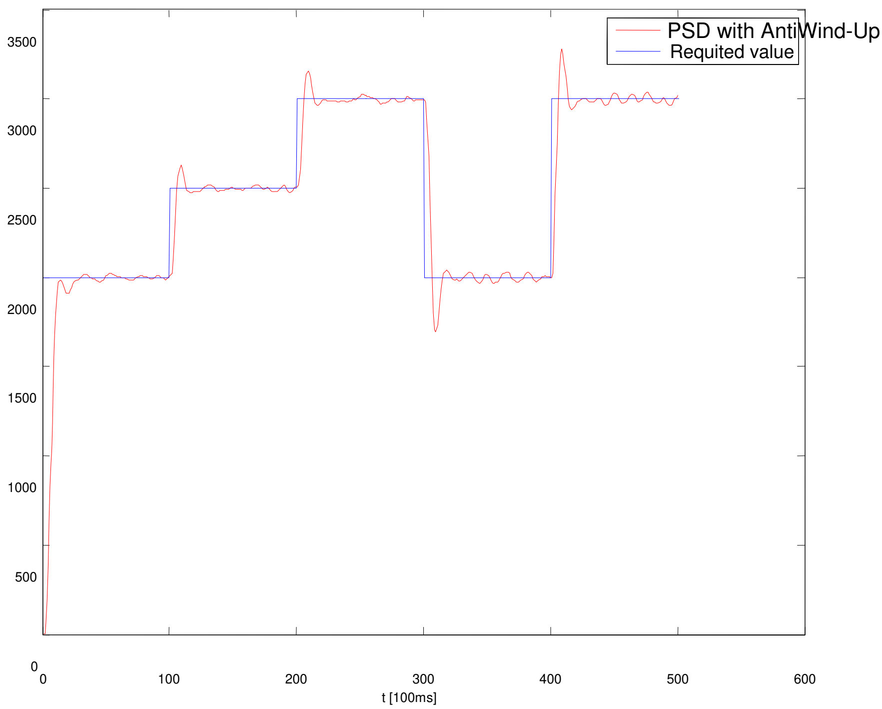

Figure 4
Time response of required value and of output from the system using the PSD algorithm with AntiWind-Up technology

Figure 4 illustrates the abilities of real-time PSD control with AntiWind-Up technology to minimize the control aberration in the closed control circuit. During 50 seconds we made five step changes of required value and we monitored the response of the closed control loop. The constants of PSD control for electro-mechanical system were used as follows:

P=0,8

$P_I = 0{,}1$

$P_D = 0{,}2$

After comparing the time responses of a standard PSD and a PSD with AntiWind-Up it becomes clear that AntiWind-Up is able to regulate the system much better during the start-up and step changes (Figure 5).

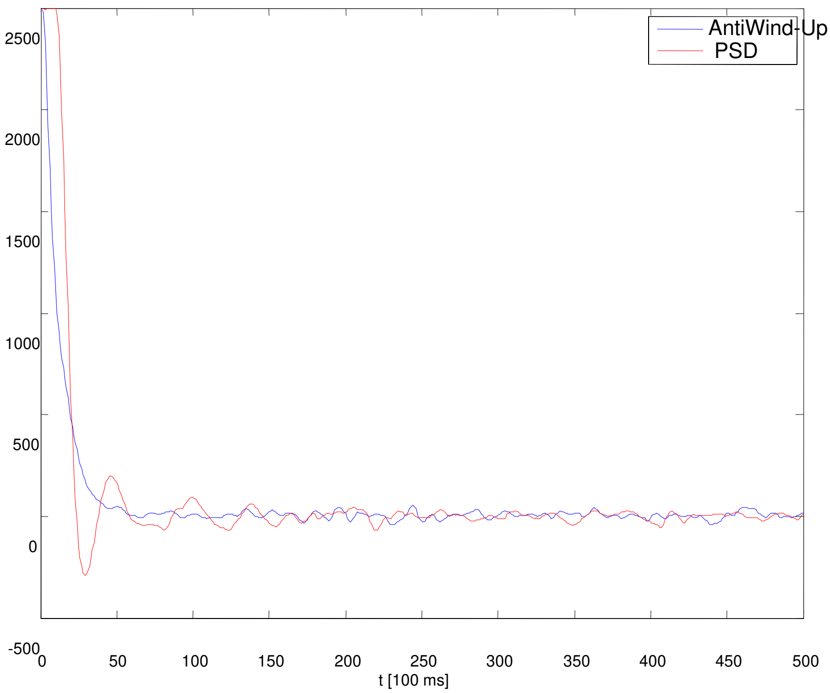

Figure 5

The comparison of time responses for standard PSD and PSD with AntiWind-Up technology

Conclusions

In our thesis we tested different control methods on a real system. We focused mainly on control abilities using the specific interconnection of RT-Linux and software package Comedi.

Because of the successful cooperation of Comedi and Real-Time Linux we showed that the real-time workable solution may be created very simply and inexpensively. All the software used in this thesis was developed under the GNU GPL (General Public License, free software license). Therefore the modification is legal, free and simple thanks to the open-source.

Acknowledgement

This article was written based on diploma project of Matej Dobšovič at Institute of Control and Industrial Informatics in 2007 year.

References

[1] RT-Linux manual, www.fmslabs.com

[2] Comedi Project Refencel, The Control and Measurement Device Interface handbook, http://www.comedi.org/doc/index.html

[3] Ilchmann, A., Ryan, E. P.: Universal λ-Tracking for Nonlinearly-Perturbed Systems in the Presence of Noise, Automatica, Vol. 30, No. 2, pp. 337-346, 1994

[4] Allgöwer, F., Ashman, J. Ilchmann, A.: High-Gain Adaptive λ-Tracking for Nonlinear Systems, Automatica, Vol. 33, No. 5, pp. 881-888, 1997

[5] B. C. Kuo, Automatic Control Systems, 6^{th} ed. Englewood Cliffs, NJ: Prentice-Hall, 1991

[6] C. C. Hang, K. J. Aström, W. K. Ho, "Refinements of the Ziegler-Nichols Tuning Formula," Proc. IEE, Pt. D., Vol. 138, pp. 111-118, 1991

Creating an Active Awareness System for Humans in Robotic Workcell

Olesya Ogorodnikova

Department of Mechatronics, Optics and Technological Information,
Budapest University of Technology and Economics
Műegyetem rkp. 3-9, H-1111 Budapest, Hungary
E-mail: olessia@git.bme.hu

Abstract: This article is devoted to the problem of human's security and safety at the time of any interaction with robots. We already discussed robot system reliability in the previous paper [1], now we emphasize the role of attention and human awareness with respect to the robot's performance in vicinity. We analyzed human's cognitive and physical abilities in ambient environment perception with the aim at effective warning system implementation. We introduced warning system interface on the basis of vibrotactile cuing, proposed an algorithm for robot controller and interface that impart tactile and visual information to human basing on data acquired from external sensory unit.

Keywords: Human-Robot Interaction,awareness, vibrotactile stimuli

1 Introduction

More than 40 years have passed, since robots have been introduced for industrial applications. Robotics technology, which has been matured enough to be applied to several systems in our daily life, is expected to be one of the key technologies for the aging society.

In industrial applications, robots usually isolated from human to ensure his/her safety, however, some advanced tasks, where robots cannot perform alone without superior human's capabilities, there is a necessity in the work spaces intersection, where conventional safeguarded systems are not sufficient anymore.

According to safety standards robot must be fully isolated from the other machines and any interactions with human are allowed only if the robot power is cut off, and the performance is interrupted, what usually brings to losses in productivity. Robotic cell usually occupies significant amount of space, hard for access and visual monitoring, often safety systems and controls are complicated in installation and usage, and at the same time don't provide absolute security and safety. There is a definite need to develop advanced dynamic active safety system where human's presence and work would be fully protected and convenience provided. Nowadays robots can be either slowed down or stopped if hazardous situation has been identified [2], moved to evade contact [3], or impact force can be minimized if contact occurred [4]. There are also examples of robot's physical redesign using visco-elastic covering [5], spherical and compliant joints [6], or distributed actuation [7]. However, to redesign already existed robots or change the control is not always possible solution in some applications, it requires time, additional costs and instrumentations.

In spite of the training programs, warnings and experience people incline to commit mistakes (errors), whether due to inattention, poorly designed workplace or accidentally because of the faulty cognition, perception or unawareness about ambient environments: robot's current state can be misperceived, speed and range of movements underestimated, warning sights unnoticed, etc.

In our approach we propose to minimize the volume of safeguards around the robot and consider more lightened robotic cell where human could enter invisible working zone, but where not only the robot would be aware about this presence, modifying its state, but also the human would 'feel' the level of danger depending on the current distance against the manipulator. Human safety will be provided by means of advanced warning system that consists of multisensory safety system and human active interface on the base of vibrotactile stimuli.

2 Human's Awareness and Modalities

Study of cognitive science would be essential to make better understanding human's mechanism of attention and utilize to reduce their errors. People have several cognitive abilities to perceive environment and process information about it. We receive the information visually or via an auditory organ (perception), understand and interpret the meaning of the perceived information and make decisions interacting with the knowledge stored in the memory system (cognition). (See Fig. 1) There are various levels of perception that depend on the stimulus and the task confronting a person. The most basic form of perception is a simple detection, the most complicated are identification and recognition. We perceive an environment via internal sensors, then this visual, auditory, tactile, etc. information is synchronized, processed in the brain and only then we evaluate the response according to the chosen behavior.

The act of perception involves prior experiences and learned associations. However, on abilities in perception nature of stimuli, attention and awareness have significant influence. Even the act of the simple detection depends on the quantity and quality of the stimuli and an attitude toward them.

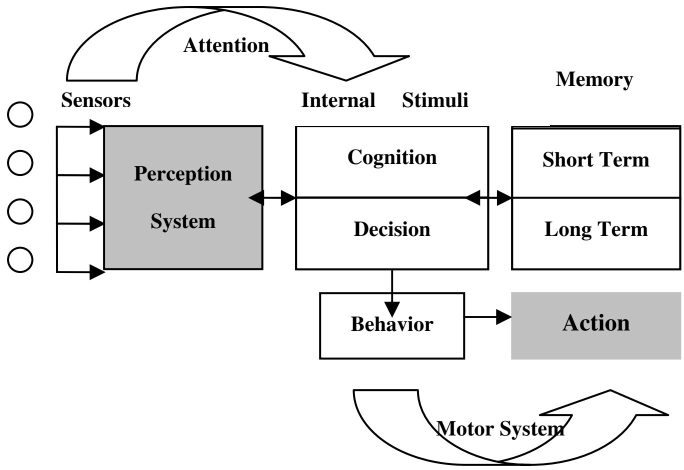

Figure 1
Human cognitive model

2.1 Awareness

Usually awareness involves matching information received from the human senses to the prelearned situational templates that allow us to understand the situation and predict what is likely happen, unfortunately we have some internal cognitive limitations and very sensitive to the external factors.

In fact, the person is aware that the stimulus will occur within a short time and within a specific spatial area. However, most of the time, several different stimuli may appear requiring different responses and the more natural the relationship between stimuli and responses, the faster reaction time (RT) and more accurate action will be. Our expectations have profound impact on the performance, they can be related whether to a robot state, range of motions, or to warnings and displayed information. Important concept in this issue is compatibility of stimuli and responses to human expectations. [8]

There are also significant limits on people's ability to divide attention across multiple aspects of the environment, particular within simple modality, such as vision or sound. Humans tend to lock in on certain aspects or features of the environment, they are trying to process and drop their scanning behavior. However, the human perceptual system can be more sensitive to certain features, so-called salient factors: loud noises, large shapes, things that physically have the advantage to catching a person's attention. By understanding and studying these natural responses more closely we could develop a concept of workplace and safety system design where human performance, attention and awareness would be enhanced.

2.2 Modalities

When we design human working space or warning system with the aim to optimize his/her work or effectively attract attention to particular case we need to take into consideration human's natural capabilities and constrains. Among the human sensing capabilities we highlighted 3 main forms:

2.2.1 Visual Sensing

The visual sensing system is very well developed. The structure of visual stimuli is well understood, and display technology can produce extremely expressive stimuli. However, one aspect that limits its general usage is the need for the user to attend to the visual stimulus. If the user is looking in a different direction, or is preoccupied with another visual task, stimuli can be missed. Visual stimuli are generated by combinations of varying hue, saturation, and intensity. It is often difficult to direct human attention rapidly toward appropriate areas of space. For instance, vision as a sensory input channel may become overloaded by the numerous parallel sources of information. Human cognitive system has a limited processing ability. Therefore, in general, human directs attention to only one of the objects presented. Asa result, for instance, during warning system design we need to take into account that the background color and ambient illumination can interact to influence the ability of people to detect and respond to lights of different colors. In the case of flashing lights, the rates of about 3-10 per second (with duration at least 0,05 s) have been recommended for attracting attention. The spurce should be situated within 30° of the human normal line of sight and subtend at least 1° of visual angle, stimulus presented in the peripheral field of view (45° from the fovea) are responded to about 15-30 ms slower than the centrally presented. [8]

There is potential competition among visual attention and other cognitive tasks for limited working memory capacity and additional sensory cues may reduce the demands of visual attention on working memory. So as a consequence aiming at improve human awareness and augment human sensitivity to stimuli we need to compensate visual cues with complementary ones (tactile, auditory) and without any signal suppression.

2.2.2 Auditory Sensing

The sense of hearing is very developed as well. Humans are sensitive to temporal, spatial, and waveform characteristics of audio signals. Audio cues are omni-directional in the sense that the listener does not need to be facing in a certain direction to attend to the sound. A sound signal is made up of waves of varying frequency and amplitude. This makes the general use of sound attractive for alerts, as well as for information display. It was discovered that the most noticeable audio signals are: beep with frequency 425 Hz and yeow (descending change in frequency from 800 to 100 Hz every 1.4 s). Reaction time to these signals decrease with increased signal intensity. However, the high-intensity signals elicit a startle reflex, which may be helpful only if the same response is demanded. We also should take into consideration the fact that the frequencies between 500-3000 Hz are most sensitive for the human ear and that the smallest frequency detectable is 28 Hz. [9]

2.2.3 Tactile Sensing

The sense of touch is arguably the most complex of the three modalities. This is partially due to several types of sensations all being attributed to this single 'sense'. There are several kinds of receptors, each allowing us to sense a different type of stimulus, such as thermal properties, vibration of varying frequencies, pressure, and pain. The sense of touch is the only one where the entire system conducts both sensing and actuation. [10]

The tactile sensory threshold is defined as the minimum stimulus intensity that is barely perceivable by a human. It is one of the most basic measures of human perception. The sensitivity vary depending on experimental conditions such as contact area, contact force, contact location, temperature of the skin, use of a rigid surround, stimulus duration, the participant's age, etc.

Importantly, when we design warning interface with multiple channels of auditory, visual and tactile signals we should take into account that stimulus requiring individual responses should be separated temporally (more than 0,25 s) to avoid overlapping. In addition, their number should be reduced, actual important signals be more intense and uncertainty of unnecessary signals reduced.

3 Active Interface. Vibrotactile Stimuli

Our aim was to design attention-aware system that would activate human's tactile and visual sensing, attracting attention and augmenting awareness about possible hazard in vicinity during work performance inside the robotic workcell. To do this we decided to implement vibrotactile stimuli, attached to the wearable device. Vibration with different intensity and flashes provides human with complementary information about possible hazard in proximity. This knowledge enables to take quick actions and avoid unwilling consecuances.

We have chosen this method of tactile cuing because this sense is the most reliable among the others modalities. The risk of signal overlapping is very low, and the skin sensitivity for the local signal exposure is very high. Moreover, it was studied that human's reaction time on tactile stimuli is much smaller with comparison to the audio and visual signals; don't have neither any spatial constrains nor dependance on human's current visual and audio attention.

3.1 Related Studies

A number of research groups have been exploring the use of additional stimuli modulities transmitted through different divices for improving human's capabilities in performance.

The effectiveness of tactile cues for spatial orientation, navigation, and situational awareness was demonstrated from the studies on pilots and astronaut [11]. In experiment conducted by Ho, et al. [12] vibrotactile display with two tactors attached to a belt fastened around the participant's waist was used to provide additional stimuli to drivers. The tactors were driven by sinusoidal signal at intensity sufficient to deliver clearly perceive vibrotactile stimuli. The results revealed that participants responded significantly more rapidly in the cued condition than in the uncued; results also highlighted a significantly bigger safety margin with the vibrotactile cuing. In another research Tacta ArmBand system was deployed. [13] To support the delivery of vibrotactile stimuli, this group designed the TactaBoard system and looked at determining the limits of perception in terms of vibration intensity, location discrimination and wearable system application for information transmission. There was also experiment where back of a person was chosen to be interfaced with a haptic display consisted of 3 by 3 tactor array to impart tactile information to user and to form multiverbal interfaces with other existing visual and auditory interfaces [14]. Studies were related to attentional and directional cuing. During research it was measured to what extent haptic cueing can affect an observer's visual spatial attention and was found that reaction times decrease with the valid haptic cues and increase with the invalid ones.

Thus, it was showed and experimentally proved that additional stimulating cuing is effective method to enhance human's attention and accordingly performance. Depending on the application and signal nature we may attract (or distract) people, make them more attentive or vice versa confused about ambient environment. Therefore it is very important to design and allocate signalizing elements properly considering all possible effects which they might have on human's perception.

4 Design

In proposed interface we have chosen the wrist of a person to be stimulated with tactile stimuli that impart non-verbal information to its user about close danger, i.e. robot in vicinity. For a tactile attention cueing, we propose to use one or two tactors attached inside of the band (see Figure 2).

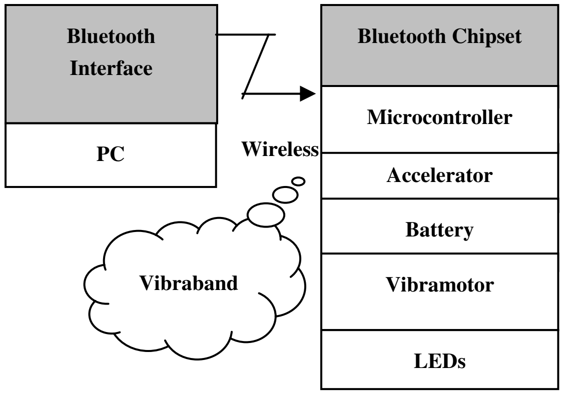

Figure 2
Vibraband structure and connection

The entire system consists of a host computer (PC) to send control commands via wireless interface, a microcontroller that transfers converted signals with predefined output voltages to vibromotor, accelerator, battery, and LEDs with 3 colors for visual signal indication. The system runs from battery power, and uses a Bluetooth wireless serial bridge connection to provide control from the host computer. The frequency of the vibration can be changed with voltages, thus we can vary the intensity of the signal, color of LEDs and control strength. For human skin (hand) the most sensitive frequency band was found between 200 Hz and 300 Hz [15]. Figure 3 displays this assumption with tactile perception for the hand, detection threshold and sensation level data for continuous sinusoidal vibration

However, for the moving person this frequency won't be enough and we should increase this magnitude to 500-550 Hz. Pulse interval was defined as 500 msec (for users can easily recognize it) [16]. This device, placing on the human's wrist, activates the vibration that imparts so-called precollision or warning information when a human is detected in the robot's (scanner) work zone with a high probability of impact. Figure 4 shows employed elements and connections between them.

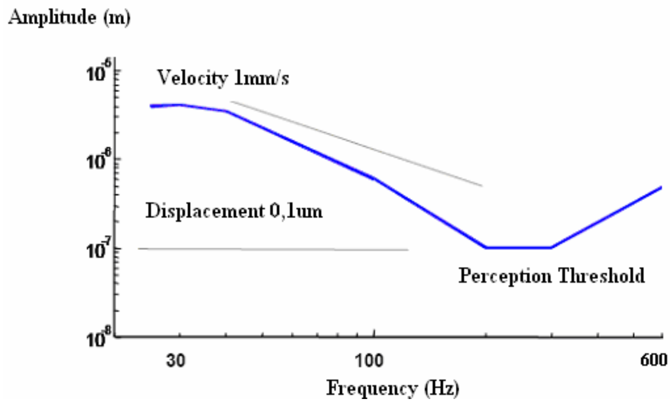

Figure 3
Human hand sensation threshold

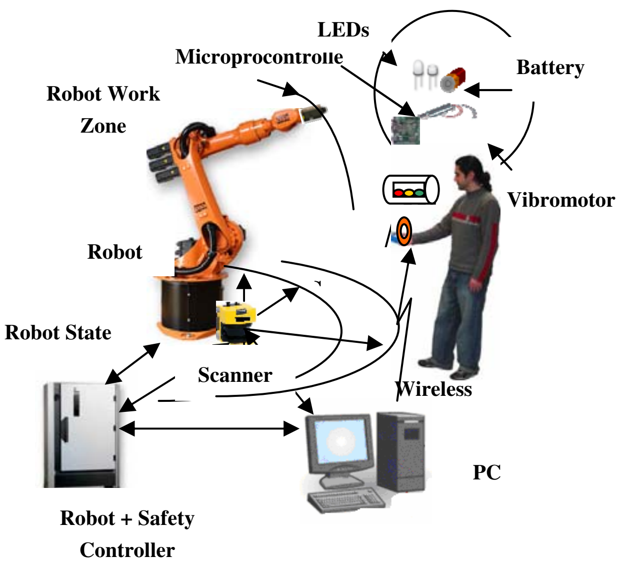

Figure 4
Elements and Connections

4.1 Robot Safety System

When robots are assigned to work in proximity to humans they should meet safety requirements related to the application. To comply with the safety regulations we use the robot with certified safety system and sensory feedback to monitor its normal operation and working space with human detection in it. Our robot comprises its own embedded safe technology that monitors velocity and acceleration of the axes and enables a safe operational stop of the robot.

The working space of the linear unit is limited by adjustable software limit switches for all axes backed up by mechanical limit stops in case these limit switches are overrun. The motion of the robot in space always corresponds to that calculated by the controller. The current axis positions are then compared with the saved reference positions, and loss of mastering in the event of a fault is detected immediately. The module responsible for monitoring the safety functions is designed using failsafe technology with built-in redundancy. Based on the reliably determined position values for each robot axes the limit values of the ranges for individual axes and axis groups are monitored on the safety module, along with their velocity and acceleration. The safety-relevant parameters and limit values are configured directly in the robot controller.

4.2 Sensory Safety System

For detecting humans in working zone we propose multisensory system consisting of ultrasonic sensory system mounted on the robot's wrist, that enable to detect objects very quickly from the close distance and to avoid collision with them by sending stopping signal on the robot controller. Ultrasonic sensors operate by using sound waves to detect targets. They generate a short, intense sound burst from a piezoelectric transducer, which is reflected back by the object. The sensor determines the distance to an object by measuring the time that elapses between the emission of an ultrasonic burst and the arrival of the echo reflected by the target. This sensing method ensures reliable operation regardless of the object's color or opacity. Other safety unit for human detection is scanning range finder that monitors the surface around the robot within its reachable range. Scanning Laser Range Finder (SLRF) is wired directly to the robot controller. We chose SLRF with wide scanning window high accuracy and its scanning speed. The scan area is defined as 240º semicircles with maximum radius 4000 mm. Scantime is 100 msec/scan. Pitch angle is 0.36º and sensor outputs the distance measured at every point (683 steps) [17]. The principle of distance measurement is based on calculation of the phase difference, due to which it is possible to obtain stable measurement with minimum influence from object's color and surface gloss. Scanner constantly monitors the surface, when the detection of object take place derived distance passes directly to the controller.

We also proposed to use stereo camera that would grid the surface similarly to the scanner, but with a larger zone. However, application of too many sensors can be resulted in signal overlapping and reduce the speed of signal processing in controller. The choice of safety equipment is not final, and requires more investigations and trial data to evaluate effectiveness and compatibility to application.

5 Distance Definition and Control Algorithm

The ranges of controlled distances to the robot were evaluated from the investigations provided by various researches in term of human's psychological attitude toward robot spatial motions, physiology, cognitive capabilities of the human (how much time we need to perceive and react on stimuli and anticipate hazard); robot system characteristics: maximum stopping time, speed, acceleration, range of movements, operational mode, etc.; safety sensors characteristics (range of operation, accuracy, response time.) For the criterion of the safe distance definition, we also examined the human spatial behavior during human-human interaction [18]. In this work the human-human interpersonal space was divided into four regions marked by the distance from the each person: 1) Intimate space (0–0.45 m): space reserved for interaction with close friend. Within this distance interaction with robots possible only if special operational mode is available or safety features to robot are embedded (compliance, force control, sensors, soft covering, etc.); 2) Personal space (0.45 m–1.2 m): distance for interaction with a friends. This zone we usually use for robot teaching and team work, where robots are assigned to assist humans or vise versa; 3) Social space (1.2 m–3.6 m): this distance is usually kept for a non-friend interaction (business, formal or casual acquaintance interaction). In robot application this area is kept for distant observation, teaching, monitoring tasks, etc.; 4) Public space (>3.6 m): in general this space is set for public speaking with where personnel contact is avoided. With respect to robotics distant monitoring and work observation can be carried out here. (See Fig. 5)

Drawing a parallel with a human-human interaction we defined that the distance people prefer to keep from the robot equal to those that they usually hold interacting with unknown person, not too close, distance enough for auditory conversation (personal space, 0,5-1,2 m). However, if there is a necessity in physical contact (peer to peer operation) this value can be diminished to 20-30 cm, but for operating within this area special safety rules, guidelines and additional guards should be considered.

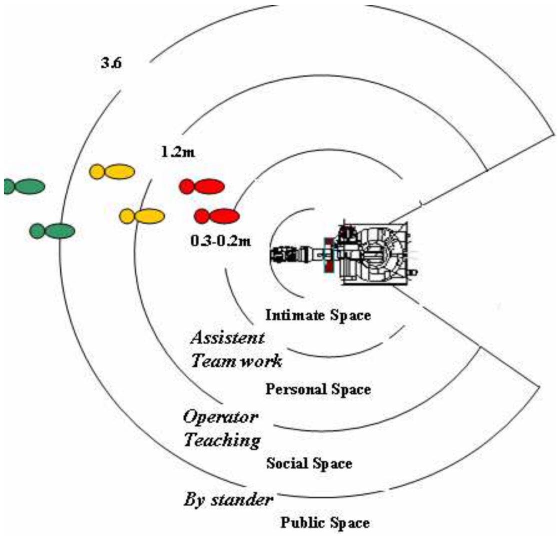

Figure 5
Interaction workspaces

5.1 Algorithm

The scanner range finder continuously monitors the robot working zone within the maximum range. Information about detected object i.e. measured distance directs to the robot controller and microcontroller of the vibraband. According to this value and other characteristics related to the robot state at the moment on the base of algorithm below robot and human interfaces (VB) change their functional conditions following the frame deployed further. When human approaches the robot within the distance smaller than 'public space' (225 cm), robot's speed should be decreased to the safe value (150 cm/s), and low frequency signal transmitted to the vibraband. LEDs are indicated with a yellow color. Robot's movements should be smooth and unlinear, working zone is restricted by: software, mechanical safety brakes. Robot's trajectory is predefined to each task and any changes in joints' angels mean failure and activate signal to stop any motions. When human approaches the robot within the 'social space' (110-70 cm) the intensity of the signal increasing thereby attracting attention to a hazardous area, LEDs are indicated with a red color. When human approaches the robot within a critical distance (>30 cm) emergency stop is activated.(See Fig. 6)

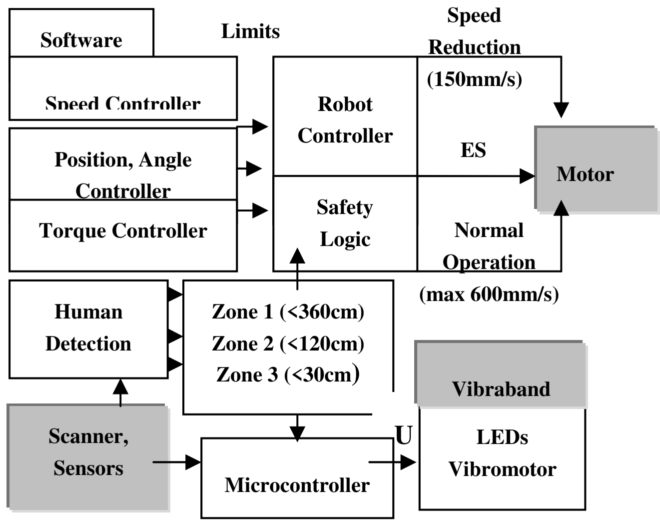

Figure 6
Control Algorithm

Robot controller constantly receives information from sensors about Robot current state from speed, acceleration, torque controller and environment (laser scanner finder, external sensors). Combination of this information with the safety logic generates corresponding signals to the motors and/or to the microcontroller of the activibroband system. Table 1 below represents a programmed functional logic for the robot controller and microprocessor by means of KRL and Delphi languages.

Table 1
Programmable logic for RC and VB Microcontroller

Frame Logic for Robot Controller (KRL)	**Frame Logic for Microcomputer (Delphi)**

If D<=360 AND D>120 Then
$VEL.CP = 450
Else If D<=120 AND D>30 Then
$VEL.CP = 150
Else If D<=30 Then
$VEL.CP = 0
Break ; Emergency Stop
EndIf
	Procedure CheckDistance(D : Integer);
	begin
	If (D>=120) And (D<360) Then
	Vibrotactile(200, 0.8, clGreen)
	Else If (D>=70) And (D<120) Then
	Vibrotactile(300, 1.2, clYellow)
	Else If (D<70) Then
	Vibrotactile(500, 1.6, clRed); end;

Conclusion and Future Work

The increasing provision of complex technologies means that human may become increasingly distracted. In guidelines for robotic safety it states: 'Audible and visible warning systems are not acceptable safeguarding methods but may be used to enhance the effectiveness of positive safeguards…' [19]. Our proposed warning system can be defined as visibly-tactile, and with compliance to the statement above cannot be considered as full right safeguard system, but we convinced that the idea of implementation this kind of instrumentation will significantly increase human's attentive qualities and help to respond on dangerous situations more accurately and quicker.

Following this rule we agree that warning a human by means of tactile and visual stimuli it is necessary but not sufficient method, therefore additional measures should be carried out where robot's hazardous characteristics would be taken into account. Also the whole workplace should be organized so that human's attentive capabilities, awareness, cognitive process would be enhanced, irrelevant information filtered and removed before it reaches the brain. In our future work we are planning to investigate the effect of directional cuing, by means of which humans would perceive incoming tactile information that indicate not only the close hazard but also its spatial location.

References

[1] Ogorodnikova O.: The Role of Risk Assessment in Human Centered Robotic Work Cell, in Proceedings of RAAD conference, Ljubljana 2007, p. 84

[2] Zurada J., Wright A. L., Graham J. H.: A Neuro-Fuzzy Approach for Robot System Safety, IEEE Transactions on Systems, Man and Cybernetics - Part C: Applications and Reviews, Vol. 31, pp. 49-64, 2001

[3] Traver V., Pobil A. P., Perez-Francisco M.: Making Service Robots Human-Safe, in Proceedings of International Conference on Intelligent Robots and Systems, 2000, pp. 696-701

[4] Lew J., Jou Y. T, Pasic H.: Interactive Control of Human/Robot Sharing Same Workspace, in Proceedings of International Conference on Intelligent Robots and Systems, 2000, pp. 535-539

[5] Yamada Y., Hirawawa Y, Huang S, Umetani Y, Suita: Human–Robot Contact in the Safeguarding Space. IEEE/ASME Transactions on Mechatronics, 2(4), 1997, pp. 230-236

[6] Bicchi A. et al.: Compliant Design for Intrinsic Safety: General Issues and Preliminary Design, in Proceedings of conference on Intelligent Robots, 2001, pp. 1864-1869

[7] Zinn M., Khatib O., Roth B.: A New Actuation Approach for Human Friendly Robot Design, in Proceedings of International conference on Robotics, USA, 2004, pp. 249-254

[8] McCormick E., M. Sanders: Human Factors in Engineering and Design, McCraw-Hill, 1992

[9] Gavriel S.: Human Factors and Ergonomics, 3d Ed., Handbook, Wiley, 2006, p. 1654

[10] Robert W., Lindeman, Yanagida Y, Noma H, Hosaka K.: Wearable Vibrotactile Systems for Virtual Contact and Information Display, Virtual Reality, 2006, pp. 203-213

[11] Gilson, R. D., Fenton, R. E.: Kinesthetic-Tactual Information Presentations In-Flight Studies, IEEE Transaction Systems. Man Cybernetics. SMC-4, 1974, pp. 531-535

[12] Ho, C., Tan, H. Z., Spence, C.: Using Spatial Vibrotactile Cues to Direct Visual Attention in Driving Scenes, Res. F: Traffic Psychological Behavour, 8, 2005, 397-412

[13] Lindeman RW, Sibert JL, Lathan CE, Vice JM.: The Design and Deployment of a Wearable Vibrotactile Feedback System, in Proceedings of the 8th IEEE International Sympoisum on Wearable Computers. Arlington, VA, 2004, pp. 56-59

[14] Tan H. Z., Pentland A.: Tactual Displays for Wearable Computing, in Proceedings of the International Symposium on Wearable Computers, Cambridge, Massachusetts, USA, 1997, pp. 225-230

[15] Linjama J., Puhakka M., Kaaresoja T.: User Studies on Tactile Perception of Vibrating Alert, in Proceedings of International conference on HCI, Crete, Greece, June 2003, pp. 22-27

[16] Tsukada K, Yasumura M.: ActiveBelt: Belt-type Wearable Tactile Display for Directional Navigation, http://mobiquitous.com/pub/ubicomp 2004-activebelt.pdf

[17] http://www.active-robots.com/products/sensors/urg-04lx.shtml

[18] Schmid A. J., Gorges N, Worn H: Towards Tightly-coupled Robot-Human Interaction, in Proceedings of International conference on Intelligent Robots and Systems, Beijing, China,October 2006

[19] STD 01-12-002 - PUB 8-1.3 - Guidelines For Robotics Safety

Investigation of Failure Systems

Alice Horváth

Faculty of Theachers Training, Eötvös Lóránd University
Kiss János u. 40, H-1126 Budapest, Hungary, hotvath.alice@chello.hu

Abstract: In an earlier paper A. Horváth and A. Prékopa [3] applied the Boole-Bonferroni lower and upper bounds to determine the expected time to failure of systems. The main goal of this paper is to show that the so called hypermultitree bounds developed by J. Bukszár [1] also can be applied for investigation of the expected time to failure systems.

Keywords: Boolean-Bonferroni bounds, hypermultitree bounds, time failor systems

1 Introduction

The reliability systems investigated in this paper belong to the field of serial and parallel interconnected systems. Serial system operates if and only if its all components operate. Parallel system operates if and only if at least one components of it operates.

Let us suppose that the components operate or do not operate independently from each other. Let p_i be the operational probability of component i, then the operational probability of serial and parallel systems can be given by the following formulas:

$$r = p_1 p_2 \cdots p_n \tag{1}$$

$$r = 1 - (1 - p_1)(1 - p_2)\cdots(1 - p_n) \tag{2}$$

In the practice one should investigate components with the random operating times, too. The investigation is simpler if they are independent random variables. Let X_i designate the lifespan of component i and $F_i(t)$ its distribution function:

$$F_i(t) = P(X_i \le t), \quad i = 1, \dots, n. \tag{3}$$

Then with the notations

$$p_i = 1 - F_i(t), \quad i = 1, \dots, n, \tag{4}$$

formulas (1) and (2) give the operational probability of the system at time t.

In practice besides determining the probability of reliable working state of the system at a certain time, we want to determine other data as well. One of the most important characteristics of this type is the expected value of elapsed time until failure. In the case of serial systems, the elapsed time till the failure of the system equals:

$$X = \min(X_1, X_2, \dots, X_n) \tag{5}$$

and in the case of parallel systems it equals:

$$Y = \max(X_1, X_2, \dots, X_n). \tag{6}$$

As it was pointed out in the paper [3], the expected value of these random variables can be determined in the following way.

If a nonnegative random variable Z has probability distribution function $G(z)$, then it is well known that

$$E(Z) = \int_0^\infty [1 - G(z)]\, dz. \tag{7}$$

Using this formula, the expected value of the elapsed time until failure for serial systems can be calculated as

$$\int_0^\infty [1 - F(t)]\, dt, \tag{8}$$

where

$$\begin{aligned} F(t) &= P(X \le t) = \\ &= P(\min(X_1, \dots, X_n) \le t) = \\ &= 1 - P(\min(X_1, \dots, X_n) > t) = \\ &= 1 - P(X_1 > t, \dots, X_n > t) = \\ &= 1 - P(X_1 > t) \cdots P(X_n > t) = \\ &= 1 - (1 - F_1(t)) \cdots (1 - F_n(t)). \end{aligned} \tag{9}$$

In the case of parallel systems the same value can be calculated as

$$\int_0^\infty [1 - G(t)]\, dt, \tag{10}$$

where

$$
\begin{aligned}
G(t) &= P(Y \le t) = \\
&= P(\max(X_1, \ldots, X_n) \le t) = \\
&= P(X_1 \le t, \ldots, X_n \le t) = \\
&= P(X_1 \le t) \cdots P(X_n \le t) = \\
&= F_1(t) \cdots F_n(t).
\end{aligned} \tag{11}
$$

The situation is more complicated when the random variables $X_1, X_2, \ldots, X_n$ are stochastically dependent. Paper [3] pointed out that in this case one can use the so called Boole-Bonferroni bounds (see for example book [2] by A. Prékopa) to give good lower and upper bounds on the expected value of the elapsed time until failure of systems. In the next section we shortly define the hypermultitree probability bounds introduced by J. Bukszár [1] and in the last section we will show how can be applied these bounds in this context. We remark, that the hypermultitree probability bounds need less calculations and usually are more accurate than the Boole-Bonferroni bounds, so their application in this context may become extremely useful.

2 The Hypermultitree Probability Bounds

In the paper [1] J. Bukszár introduced the concept of (h,m)-hypermultitrees and based on this concept he developed good lower and upper bounds on the probability of union (resp. intersection) of events. As a possible application he estimated the value of the multivariate normal probability distribution function by his newly introduced probability bounds. The definition of the $\Delta = (V,\ {}_h\varepsilon_2, \ldots,\ {}_h\varepsilon_{m+1})$, (h,m)-hypermultitree is given in Definition 3 of paper [1]. In the definition V is the set of vertices and ${}_h\varepsilon_i$'s are sets of hyperedges containing $h+i$ vertices. Definition 4 of paper [1] introduces the concept of the weight of (h,m)-hypermultitrees in the following way. Let $A_1, A_2, \ldots, A_n$ be arbitrary events and suppose we can calculate the probability of their intersections up to $h+m+1$ number of events involved in the intersection. Then the weight of the (h,m)-hypermultitree $\Delta = (V,\ {}_h\varepsilon_2, \ldots,\ {}_h\varepsilon_{m+1})$ is given by the formula

$$
\begin{aligned}
w(\Delta) \quad = \quad & \sum_{(i_1,\ldots,i_{h+2})\in {}_h\varepsilon_2} P\left(A_{i_1}\cdots A_{i_{h+2}}\right) \\
- \quad & \sum_{(i_1,\ldots,i_{h+3})\in {}_h\varepsilon_3} P\left(A_{i_1}\cdots A_{i_{h+3}}\right) \\
\vdots \quad & \\
+ \quad & (-1)^{m+1} P\left(A_{i_1}\cdots A_{i_{h+m+1}}\right).
\end{aligned} \tag{12}
$$

J. Bukszár proved the following inequalities in paper [1] (see Theorem 1):

If $A_1, A_2, \ldots, A_n$ are arbitrary events, and $\Delta = \left(V,\ {}_h\varepsilon_2, \ldots,\ {}_h\varepsilon_{m+1}\right)$ is an arbitrary (h, m)-hypermultitree then the following inequalities hold:

(i) if h is even, then

$$P(A_1 + \cdots + A_n) \le \sum_{k=1}^{h+1} (-1)^{k-1} S_k - w(\Delta), \tag{13}$$

(ii) if h is odd, then

$$P(A_1 + \cdots + A_n) \ge \sum_{k=1}^{h+1} (-1)^{k-1} S_k + w(\Delta), \tag{14}$$

where

$$S_k = \sum_{1\le i_1 < \cdots < i_k \le n} P\left(A_{i_1}\cdots A_{i_k}\right). \tag{15}$$

From these formulae one can see that the probability bounds given by the (h, m)-hypermultitrees are closer to the exact probability value when their weight is larger. The problem of finding the best possible hypermultitree is NP hard, however J. Bukszár in Section 3 of his paper [1] developed very efficient algorithms for finding hypermultitrees with heavy weight. In the next section we will use only the special cases $h = 0$ and $h = 1$ as it was proposed by J. Bukszár.

3 Application of the Hypermultitree Probability Bounds for Investigation of Failure Systems

In the case of serial systems we can apply the formulae (13) and (14) in a straightforward way for the events $A_i = \{X_i \le t\},\ i = 1, \ldots, n,$ as we have

$$F(t) = P(X \le t) = P(\min(X_1, \ldots, X_n) \le t) = P(A_1 + \cdots + A_n). \tag{16}$$

To do this first we introduce the notations:

$$F_{i_1,\ldots,i_k}(t) = P\left(X_{i_1} \le t, \ldots, X_{i_k} \le t\right), 1 \le i_1 < \cdots < i_k \le n, \tag{17}$$

$$S_k(t) = \sum_{1 \le i_1 < \cdots < i_k \le n} F_{i_1,\ldots,i_k}(t), \; k = 1, \ldots, n. \tag{18}$$

Now for $h = 0$ we get the upper bound:

$$\begin{aligned} F(t) \;\; \le \;\; & S_1(t) \\ - \;\; & \sum_{(i_1,i_2) \in_0 \varepsilon_2} F_{i_1,i_2}(t) \\ + \;\; & \sum_{(i_1,i_2,i_3) \in_0 \varepsilon_3} F_{i_1,i_2,i_3}(t) \\ \vdots \;\; & \\ + \;\; & (-1)^m \sum_{(i_1,\ldots,i_{m+1}) \in_0 \varepsilon_{m+1}} F_{i_1,\ldots,i_{m+1}}(t), \end{aligned} \tag{19}$$

and for $h = 1$ we get the lower bound:

$$\begin{aligned} F(t) \;\; \ge \;\; & S_1(t) - S_2(t) \\ + \;\; & \sum_{(i_1,i_2,i_3) \in_1 \varepsilon_2} F_{i_1,i_2,i_3}(t) \\ - \;\; & \\ \vdots \;\; & \\ + \;\; & (-1)^{m+1} \sum_{(i_1,\ldots,i_{m+2}) \in_1 \varepsilon_{m+1}} F_{i_1,\ldots,i_{m+2}}(t). \end{aligned} \tag{20}$$

In the case of parallel systems we have

$$1 - G(t) = P(Y > t) = P\left(\max\left(X_1, \ldots, X_n\right) > t\right) = P\left(\overline{A}_1 + \cdots + \overline{A}_n\right),. \tag{21}$$

so we have to introduce further notations:

$$\overline{F}_{i_1,\ldots,i_k}(t) = P\left(X_{i_1} > t, \ldots, X_{i_k} > t\right), 1 \le i_1 < \cdots < i_k \le n, \tag{22}$$

$$\overline{S}_k(t) = \sum_{1 \le i_1 < \cdots < i_k \le n} \overline{F}_{i_1,\ldots,i_k}(t), \; k = 1, \ldots, n. \tag{23}$$

Now for $h = 0$ we get the lower bound:

$$
\begin{aligned}
G(t) \quad &\geq \quad 1-\overline{S}_1(t) \\
&+ \quad \sum_{(i_1,i_2)\in_0 \varepsilon_2} \overline{F}_{i_1,i_2}(t) \\
&- \quad \sum_{(i_1,i_2,i_3)\in_0 \varepsilon_3} \overline{F}_{i_1,i_2,i_3}(t) \\
&\vdots \\
&+ \quad (-1)^{m+1} \sum_{(i_1,\ldots,i_{m+1})\in_0 \varepsilon_{m+1}} \overline{F}_{i_1,\ldots,i_{m+1}}(t),
\end{aligned} \tag{24}
$$

and for $h=1$ we get the upper bound:

$$
\begin{aligned}
G(t) \quad &\leq \quad 1-\overline{S}_1(t)+\overline{S}_2(t) \\
&- \quad \sum_{(i_1,i_2,i_3)\in_1 \varepsilon_2} \overline{F}_{i_1,i_2,i_3}(t) \\
&- \\
&\vdots \\
&+ \quad (-1)^{m} \sum_{(i_1,\ldots,i_{m+2})\in_1 \varepsilon_{m+1}} \overline{F}_{i_1,\ldots,i_{m+2}}(t).
\end{aligned} \tag{25}
$$

To get the expected value of the elapsed time until failure of the systems one can apply the general integration formula (7) as it was done before.

Conclusions

In this paper the (h,m)-hypermultitree bounds introduced by J. Bukszár were applied to determine lower and upper bounds on the expexted time to the failure of serial and parallel systems. As these bounds proved to be more efficient than the Boole-Bonferroni inequalities are, they are hoped to be useful tools in investigation of failure systems.

References

[1] Bukszár J.: Quickly Computable Bounds on the Probability of a Union, in Proceedings of the 4th GAMM/IFIP Workshop on “Stochastic Optimization: Numerical Methods and Technical Applications”, held at the Federal Armed Forces University Munich, Munich/Neubiberg, June 27-29, 2000, Lecture Notes in Economics and Mathematical Systems, 513, Springer-Verlag, Berlin, Heidelberg, New-York, 2002, pp. 79-89

[2] Prékopa A.: Stochastic Programming, Kluwer Academic Publishers, Norwell, 1995

[3] Horváth A., Prékopa A.: Lower and Upper Bounds for the Expected Time to Failure of Systems, Alkalmazott Matematikai Lapok, 21 (2004) pp. 131-149 (in Hungarian)

Ontologies used for Competence Management[1]

Victoria Iordan, Alexandru Cicortas

Computer Science Department
Faculty of Mathematics and Computer Science
University of the West Timisoara
e-mail: iordan@info.uvt.ro, cico@info.uvt.ro

Abstract: Developing multi-agent systems requires an adequate modeling of knowledge in order that the agents and the human person are able to understand and accept the concepts of domain area in the same way. Ontologies allow developing an coherent framework for a specified domain. Based on the concepts used ant the relations between them, the agents are able to understand, reason and act in the domain in order to accomplish their goals and finally the system functions. Based on appropriate ontologies defined for a particular domain, a multi-agent system that allows managing, searching and matching the user competences with the existent competences is presented.

Keywords: ontology, compentence modeling, multi-agent systems

1 Introduction

Complex systems from the real world are a challenge for designers due their complexity and their requirements. Designing such complex systems should include also some intelligent behavior due to their complexity. Knowledge management is one of the most important aspects that offer to the designers a real base for developing intelligent features of the system. Multi-agent systems had an expansively evolution in the last decades. For every domain the knowledge must be appropriately represented and understood by all the participants being humans or intelligent agents.

From [13] ontology is a formal theory within which not only definitions but also a supporting framework of axioms is included (perhaps the axioms themselves provide implicit definitions of the terms involved). Regarding the multi-agent systems, we can state that ontology is a description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents.

[1] The research for this paper has been partial supported by the project Ceex 05-D8-66/2005

For reducing the complexity we intend to define the ontology related necessary terms for:

- multi-agent communication;
- requirements from the user concerning the agents and requirements between agents;
- competence description, competence comparison and competence evaluation and related information that are given to the agents in order that they are able to accomplish the requirements.

Based on [14] the:

- entities will refer to the communication, requirements concerning the agents and competences;
- representations (their composite representation and representational units) concerning the competences will be given in the form of competence description or association of activities (sub competences).

The inclusion of a competence in the other one it is not similarly with the inheritance.

The paper is organized as follows. In the next section the ontology concepts are reviewed. The third section gives specific problems concerning the competences The fourth section introduces an example of multi-agent system for constructing and finding competences. The fifth section details schemas and gives some particular examples of necessary ontologies based on XML files. The last section suggests the contents of future works.

2 Ontology Concepts

For facilitating the sharing and the use the ontologies were developed in AI many domain models. From [13] ontology is a formal theory within which not only definitions but also a supporting framework of axioms is included (perhaps the axioms themselves provide implicit definitions of the terms involved). Concerning the multi-agent systems, we can state that ontology is a description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents [15]. But between the multitude of ontology definition the following is one of the most appropriated to our purposes: *An ontology is a formal, explicit specification of a shared conceptualization* (the terms are detailed in [7]). *Conceptualization* refers to an abstract model of some phenomenon in the world which identifies the relevant concepts of that phenomenon. *Explicit* means that the type of concepts used and the constraints on their use are explicitly defined. *Formal* reflects to the fact that the ontology should be machine readable. *Shared* reflects the notion that an ontology captures consensual knowledge that is accepted by a group.

The role of ontologies in AI is to facilitate the construction of a domain model. An ontology provides a vocabulary of terms and the relations between them in order to being able for modeling the domain. Due to the fact that ontologies aim at consensual domain knowledge their development is often a cooperative process involving different people or organizations. Concerning the people is said that they commit themselves to that ontology if they agree to accept that ontology.

Ontologies are introduced to facilitate the knowledge sharing and reuse between various agents, regardless of whether they are human or artificial in nature. They are supposed to offer this service by providing a consensual and formal conceptualization of a certain area. As a conclusion, ontologies are formal and consensual specifications of conceptualization that provide a shared understanding of a domain, an understanding that can be communicated across people and specification systems. From the fact that the ontologies define:

- formal semantics for information, allowing that it is processed by a computer and
- real-world semantics, that makes it possible to link machine-processable content with meaning for humans based on consensual terminologies;

these considerations allow to argue how can ontologies be used to communicate real-world semantics between human and artificial agents and also these implies that the ontologies have as important features their dynamicity and a network architecture.

Due to the fact that the ontologies are conceived in order to cover many domains from the simplest ones to the most complexes ones the designers must make a choice form the from the following [7]:

- Domain ontologies- capture the knowledge valid for a particular type of domain (e.g. teaching, IT);
- Meta data ontologies that provide a vocabulary for describing the content of on-line information sources;
- Generic or common sense ontologies that aim capturing general knowledge about the world, providing basic notions and concepts for things like: time, space, state event, etc.;
- Representational ontologies do not commit themselves to any particular domain (frames);
- Method and task ontologies. Task ontologies provide terms specific for particular tasks and method ontologies provide terms specific to particular problem-solving methods. These ontologies provide a reasoning point of view on domain knowledge.

Concerning our own purposes we will make a choice in order to satisfy the requirements of our model given in example.

3 Considerations on Competence Modeling

Concerning the competences there are many works that develop and details specific concepts [12], [10]. One of the major problem that occurs is that the comparing compentences [9]. The IEEE Reusable Competency Definitions [10] provide a model for the representation of competences, the objective being referencing and cataloging a competency but not classifying it. The model does not provide any means to specify the relationships between the competencies. The relationships must be taken into account that the competences are (is composed) from competency proficiency level and context. Different scales qualitative and quantitative are useful in order to represent proficiency levels.

One of the possibilities is that to represent as an ordered list the proficiency level scale. In such list the minimum value (subsumed by any other in the list) is given by the first element and the maximum is given by the last one. Therefore the order in the list represent presumption relationships, that is, the first element is subsumed by the second one and so on. In order to improve the interoperability and matching among scales, an optional field is included for mapping to the universal scale (e.g., *[0,1]*). The reason why this mapping field is optional is that even though it would be useful to include it, in some context it may not be possible to find a suitable mapping or it may not even be necessary.

Competence descriptions can refer to specific items of these scales in order to represent the proficiency level required/acquired. Algorithms could take relationships among proficiency levels into account in order to find out how much training/learning is required to reach a determined employee/learner proficiency level [3].

The context can be defined: interrelated conditions in which something exists or occurs or the circumstances and conditions which surround.

Regarding the competences, context may refer to different concepts like:

- the specific occupation in which a competence is required;
- a set of topics within a domain;
- even the personal settings related to the student.

These are contexts which may be part of a competence. Context descriptions cannot be defined in general, but these depend on the scope and the purpose of the competence descriptions to which they are attached. In addition, the context definitions may be reused.

Modeling contexts is a complex task; it may coincide with modeling the whole domain knowledge of an institution. Competences generally can be described [6] as reusable domain knowledge. Any model representing competences describes what a competence is and how it is composed of sub-competences. Due to the fact that the competences are referenced in different situations like: certifications, job descriptions and personalizing relevant competences for their business that are included in job offers projects descriptions.

Based on these the competence must be adequate represented and described in order to:

- how a competence may be achieved (ex: by acquiring some sub-competences);
- to which level each competence should be acquired;
- whether sub-competences must be all achieved or simply a subset of them;
- if the sub-competences must be acquired in a specific order.

Another significant problem is that the capability of the model to represent aggregate and alternative structures of the competence. The aggregation allows that the competence is composed from several sub-competences all of them required. Alternative competence can be viewed as a set of competences and there can be possible to specify by a numeric interval what the number of alternatives that must be acquired is. Due to the multiple usages of the model, it is also important that the equivalence relationships between the competences to be well defined understood and used by all users.

4 Management of Competences Using a MAS

The competences are frequently used in the relations between the universities and the future students, between the companies and the future employees. Our model intend to allow to the universities, students, employees and companies to construct and maintain their own competences; to evaluate their competences; to match their own competences with the other competences and to search the desired competences in appropriate domains. Comparing the competences for an efficient usage it is intended to offer a tool that make an exhaustive analysis concerning the competences. It will mainly based on the details that are given for a competence the components of the competence. Here the specific agent will compare the occurrence of the competence components scoring the matching between the two competences also the order of components will be taken into the account; the resources used for gaining a competence, the effort that must be fulfilled by the student in order to gain certain competence. The students that intend to obtain some qualification (and some competences) can make some suppositions concerning the financial effort and their own effort and time and it will be offered in an adequate manner. The details concerning the compentences in the domain, their description and their comparioson are given in [4], [5], [8].

The model based on a multi-agent system constituted from appropriate agents that will fulfill these objectives. The agents of our model are Competence Creation Agent (CCA), Evaluator Agent (EvA) and Broker Agent (BrA) and it is represented in Figure 1 as it was geiven in [4], [5], [8].

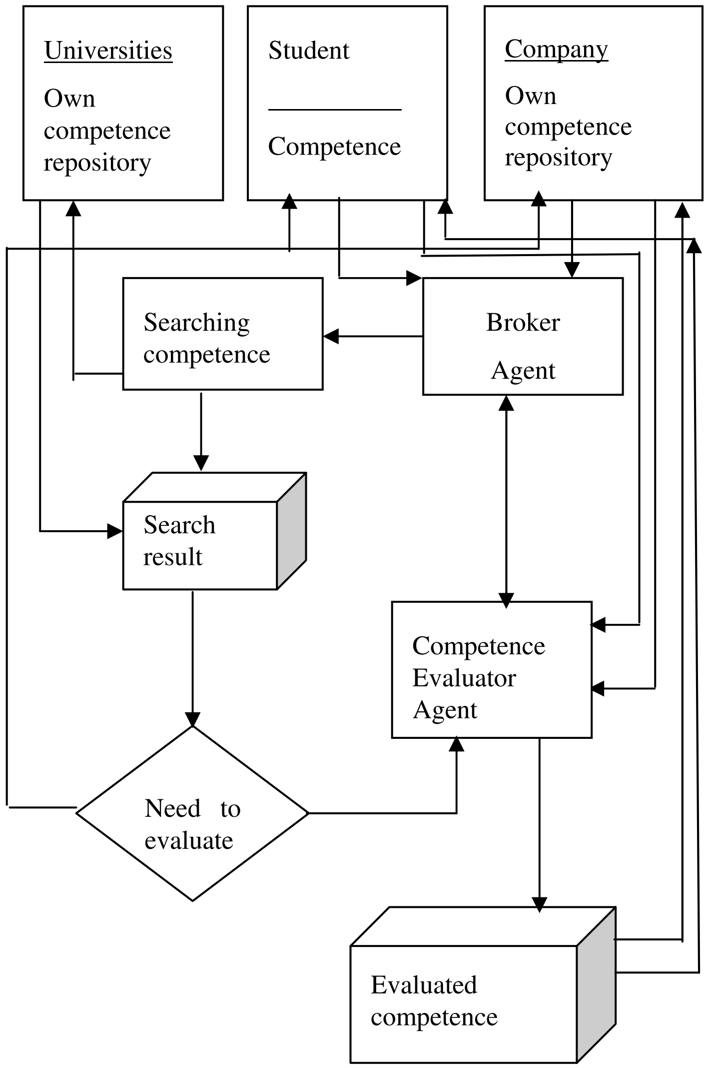

Figure 1
The multi-agent system for the proposed model

The user can be a university, a company, a student or an employee. The user can submit to the CCA requirements to create competences from basic sub-competences. The CCA creates and furnish the competences to the user that can place them into a Competence Repository or use immediately in new requirements. In Figure 2 is illustrated how the CCA interact with the users.

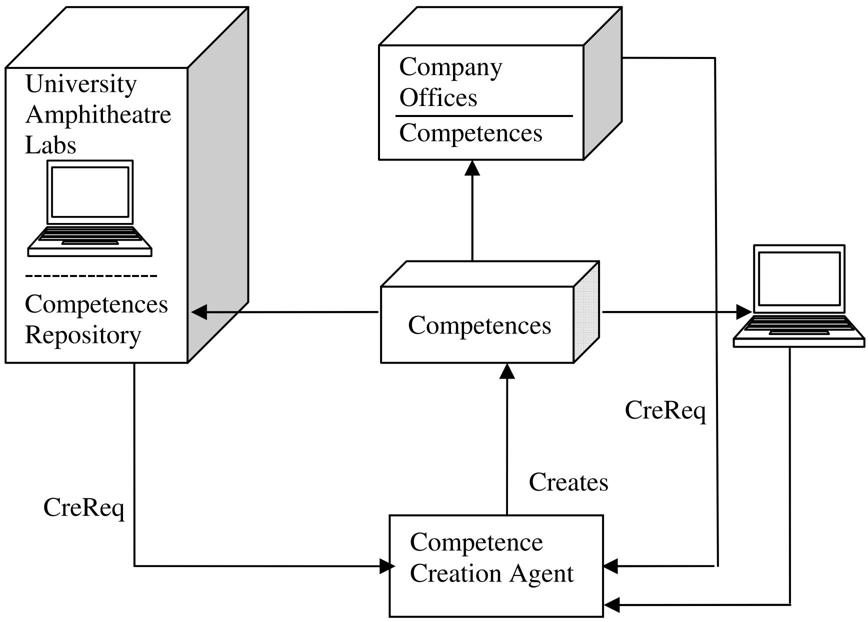

Figure 2
The interaction of the CCA with the users

Every user can construct its own Competence Repository. When a user (student, employee, company) want to find some competence that is placed in a university or company he must furnish the generic competence and eventually the sub-competences, the conditions that allow acquiring the competence and other own requirements concerning the competence.

All these requirements are taken by the BrA that search (usually on the Internet) and tries to match the user requirement with those that were found. The matching can be in a wide range as was presented above, starting from a simple matching (the competence name) through a complex one where a lot of actions are executed by the EvA: matching of competences with appropriate scores; matching the sub-competences and scoring the matching; matching of conditions and giving the scores in some order (preferred by the user or a predefined system order).

The interdependence between the agents and user is shown in Figure 3.

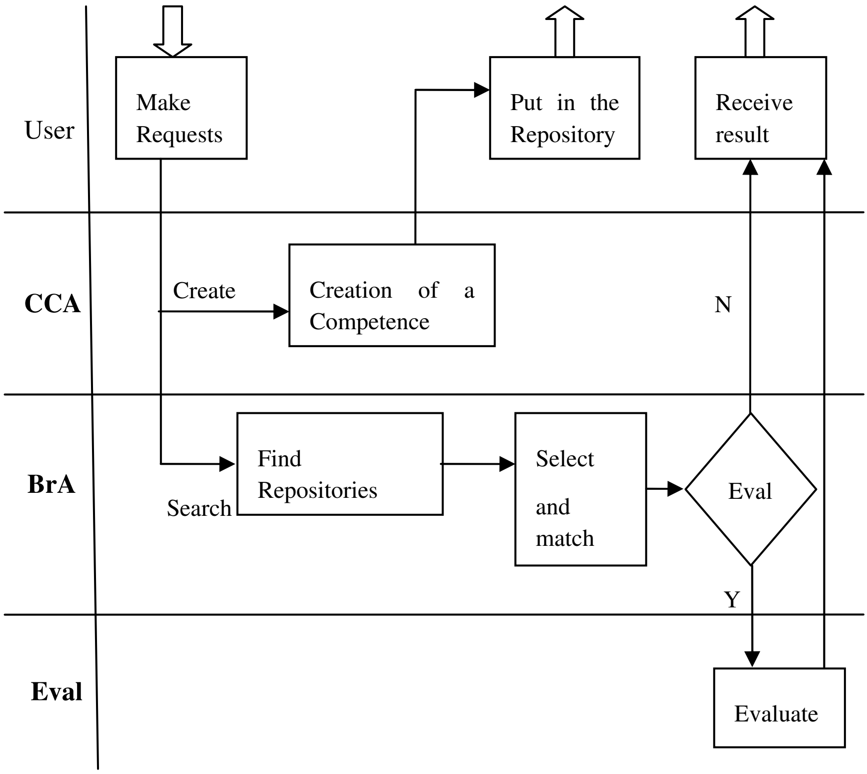

Figure 3
The collaboration between the users and the agents

The system work as follows. The companies and universities have their Competence Repositories that are posted as web pages. The users: students, employees, applicants, the universities and companies can define the requirements that are addressed to the system. The system agents try to satisfy the requirements in different levels of details and complexity, as was stated above.

5 Detailed Examples for Used Ontologeies

The necessary ontologies for the model of the system proposed in previous section are based on the following schemas and are illustrated with some examples. The following schema illustrates the specific elements used in course description and the competences and their components (subcompetences).

```
<?xml version="1.0" encoding="UTF-8"?>
<!--W3C scheme generated by XMLSpy v2007 sp2 (http://www.altova.com)-->
<xs:scheme xmlns:xs="http://www.w3.org/2001/XMLscheme">
 <xs:element name="subcomponent" type="xs:string"/>
 <xs:element name="subCompetences">
            <xs:complexType>
           <xs:sequence>
           <xs:element ref="subcomponent" maxOccurs="unbounded"/>
           </xs:sequence>
           </xs:complexType>
           </xs:element>
<xs:element name="objectives">
<xs:complexType>
           <xs:sequence>
           <xs:element ref="objective" maxOccurs="unbounded"/>
           </xs:sequence>
           </xs:complexType>
           </xs:element>
<xs:element name="objective" type="xs:string"/>
<xs:element name="name">
           <xs:simpleType>
           <xs:restriction base="xs:string"/>
           </xs:simpleType>
           </xs:element>
<xs:element name="courses">
           <xs:complexType>
           <xs:sequence>
           <xs:element ref="course" maxOccurs="unbounded"/>
           </xs:sequence>
           </xs:complexType>
           </xs:element>
<xs:element name="course">
           <xs:complexType>
           <xs:sequence>
           <xs:element ref="name"/>
           <xs:element ref="basic_competence"/>
           <xs:element ref="description"/>
           <xs:element ref="objectives"/>
```

```
<xs:element ref="subCompetences"/>
</xs:sequence>
</xs:complexType>
</xs:element>
<xs:element name="description" type="xs:string"/>
<xs:element name="basic_competence" type="xs:string"/>
</xs:scheme>
```

As an concrete example the Computer Network course is detailed based on previous schema.

```
<?xml version="1.0" encoding="UTF-8"?>
<courses xmlns:xsi="http://www.w3.org/2001/XMLScheme-instance" xsi:noNamespaceSchemeLocation="scheme_A.xsd">
<course>
    <name> Computer Networks </name>
    <basic_competence>Using and administration of computer networks</basic_competence>
    <description>administrating, configuring and maintaining computer networks</description>
    <objectives>
    <objective>Course gives a general view of concepts in computer networks, giving a synthesis the basic concepts concerning the architectures, protocols, administration, interconnexion</objective>
    </objectives>
    <subCompetences>
        <subcomponent>computer network administration </subcomponent>
        <subcomponent>computer network configuring </subcomponent>
        <subcomponent>software installing </subcomponent>
        <subcomponent>user and resource management </subcomponent>
    </subCompetences>
  </course>
</courses>
```

As can it be seen in the example all the users should accept the proposed ontological elements. It is obviously to remark that between the subcompetences we can find some inappropriate features that here are presented (given) as subcompetneces for many of possible users. As an example of such possible inadvertences *can be the capability to communicate with other people* or *capability to work in a team* that can not have a general acceptance between all the possible users.

6 Future Work

Based on the real model that was describy and refinning the capabilities, some important features will be analyzed and developed. The relationships between the components of an ontology and the relationships between the related ontologies should be clearly stated and therefore formalized. Concerning the relationships between the ontology components, these must be refind due to the following:

- subsuming are incluzion i.e. one component subsumes another one (i.e. excellent subsumes advances in speaking english);
- part_of;
- is_a.

As it can be seen the inheritance from object oriented does not satisfy in the description of above kind of relationships.

On the future works will concentrate in order to allow to give all above features of relationships. Another direction of our future work will refine the quantifications of ontology component. These will be very useful in mathing of requirements with offers. The long term research will focus to the ability of model in order to match two defferent ontologies for the same domain. More explicitely, for example an ontology contains the knowledges of an specialist and another one ontology contains the capabilities of a specialist in the same domain. The model must offer the inference capabilities in order to express the quantified matching between the two ontologies. A such feature will use an intelligent system, able to make such kind of interences.

XML schemas for ontology representations will be extehded with new details which allow to extend capabilities of ontology description and usage. Using the XML representations of ontologies, will be improved a better tool for their processing.

References

[1] R. H. Bordini, M. Dastani, J. Dix, A. El. F. Seghrouchni: Multi-Agent Programming, Languages, Platforms and Applications, Springer 2005, ISBN-10:0-387-24568-5 (HB)

[2] R. H. Bordini, J. F. Hubner: Jason: A Java-based AgentSpeak, manual, version 0.9.5 edition, Feb., 2007. http://jason.sourceforge.net/

[3] G. Cheetam, G. Chivers: Professions Competence and Informal Learning, Edgard Elgar Publishing Limited, Journal of Interprofessional Care, Volume 20, Number 5, October 2006, pp. 569-570

[4] Alexandru Cicortas, Victoria Iordan: Multi-Agent System for Competence Modeling, Proceedings of the 1st Conference on Knowledge Egeneering Principles and Techiques KEPT 2007, pp. 274-281, ISBN: 978-973-610-556-2, pp. 274-281

[5] Alexandru Cicortas, Victoria Iordan: Representing and Comparing Competences Using Agents, Proceedings SACI 2007, 4th International Symposium on Applied Computational Intelligence and Informatics, Timisoara, Romania, May 17-18, 2007, IEEE Catalog Number: 07EX1788, ISBN:1-4244-1234-X, pp. 225-230

[6] Juri L. De Coi, E. Herder, A. Koesling, C. Lofi, D. Olmedilla, O. Papapetrou, W. Siberski: A Model for Competence GAP Analysis, In WEBIST 2007, Proceedings of the Third International Conference on Web Information Systems and Technologies: Internet Technology/Web Interface and Applications, Barcelona, Spain, Mar 2007

[7] Fensel, D., Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce, Springer-Verlag Heidelberg, ISBN3-540-00302-9 2004

[8] Victoria Iordan, Alexandru Cicortas: Multi-Agent Systems Used for Managing Competences, The Proceedings of the 8th International Conference on Informatics in Economy, May 2007, ASE Publishing House, 2007, Informatics in Knowledge Society, ISBN 978-973-594-921-1, pp. 633-638

[9] INSTICC Press. ***, HR-XML Measurable Competencies, 2004. http://www.hr-xml.org

[10] ***, IEEE 1484.20.1/draft - draft standard for Reusable Competency Definitions (RCD), 2005

[11] http://ieeeltsc.org/wg20Comp/Public/IEEE_1484.20.1.D3.pdf

[12] ***, “IMS Reusable Definition of Competency or Educational Objective (RDCEO)”, 2002, http://www.imaglobal.org/competencies

[13] B. Smith: Ontology. An Introduction, Luciano Floridi (ed.), Blackwell Guide to the Philosophy of Computing and Information, Oxford: Blackwell, 2003, pp. 155-166

[14] B. Smith, W. Kusnierczyk, D. Schober, W. Ceusters: Towards a Reference Terminology for Ontology Research and Development in the Biomedical Domain, Proceedings of KR-MED 2006

[15] http://www.w3.org/TR/2004/REC-owl-guide-20040210/