Encoding Named Channels Communication by Behavioral Schemes

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: Our new approach to the calculus of mobile ambients is suitable for expressing the dynamic properties of mobile code applications, where the main goal is to avoid the ambiguities and possible maliciousness of some constructions. We define a behavioral scheme assigned to process types that statically specifies and checks access rights for authorization of ambients and threads to communicate and move. As an expressiveness test, we showed that the well-known π-calculus of concurrency and mobility can be encoded in our calculus in a natural way.

Keywords: process calculi; mobile code; type system

1 Introduction

The calculus of mobile ambients [1] is based on a concurrency paradigm represented by the π-calculus [2]. It introduces the notion of an ambient as a bounded place where concurrent computation takes place, which can contain nested subambients in a hierarchical structure, and which can move in and out of other ambients, i.e., up and down the hierarchy that rearranges the structure of ambients. Communication can only occur locally within each ambient through a common anonymous channel. Communication between different ambients has to be performed by movement and by dissolution of ambient boundaries.

Mobile ambients model several computational entities: mobile agents, mobile processes, messages, packets or frames, physical or virtual locations, administrative and security domains in a distributed system and also mobile devices. This variety means in principle there are no differences among various kinds of software components when expressing by mobile ambients. In mobile ambients there are implicitly two main forms of entities which we will respectively call *threads* and *ambients*. Threads are unnamed sequences of

primitive actions to be executed sequentially, generally in concurrence with other threads. They can perform communication and drive their containers through the spatial hierarchy but cannot individually go from one ambient to another. Ambients are named containers of concurrent threads. They can enter and exit other ambients, driven by their internal processes, but cannot directly perform communication. It is very important to ensure indivisibility and the autonomous behavior of ambients (this is also important e.g. for objects).

Communication between ambients is represented by the movement of other ambient of usually shorter life, which have their boundaries dissolved by an *open* action to expose their internal threads performing local communication operations. Such capability of opening an ambient is potentially dangerous [3, 4, 5]. It could be used inadvertently to open and thus destroy the individuality of an object or mobile agent. Remote communication is usually emulated as a movement of such ambients (communication packages) in the hierarchy structure.

Table 1
Abstract syntax

$M ::=$	**mobility operations**
$\mid\ n$	name
$\mid\ in\ M$	move ambient into M
$\mid\ out\ M$	move ambient out of M
$\mid\ move\ M$	move thread into M
$\mid\ M.M'$	path
$P ::=$	**processes**
$\mid\ \mathbf{0}$	inactive process
$\mid\ P \mid P'$	parallel composition
$\mid\ !P$	replication
$\mid\ M[P]$	ambient
$\mid\ (\nu n : \mathbf{P}[\mathcal{B}])P$	name restriction
$\mid\ M.P$	action of the operation
$\mid\ \langle M \rangle.P$	synchronous output
$\mid\ (n : \mu).P$	synchronous input

We explore a different approach, where we intend to keep the purely local character of communication so that no hidden costs are present in the communication primitives, but without *open* operation. This solves the problem of dissolving boundaries of ambients but disables interactions of threads from separate ambients. We must introduce a new operation, *move,* for moving threads

between ambients. The idea comes from mobile code programming paradigms [6], where moving threads can express a strong mobility mechanism, by which the procedure can (through *move* operation) suspend its execution on one machine and resume it exactly from the same point on another (remote) machine. This solves the problem of threads mobility and by moving threads between ambients we can emulate communication between the ambients.

2 Overview of the Calculus

1.1 Syntax

The abstract syntax of the terms of our calculus in Table 1 is the same as that of mobile ambients, except for the absence of *open* and the presence of the new operation *move* for moving threads between ambients.

Table 2
Free (a) and bound (b) names

(a)	(b)
$fn(n) = \{n\}$	$bn(n) = \varnothing$
$fn(in\ M) = fn(M)$	$bn(in\ M) = bn(M)$
$fn(out\ M) = fn(M)$	$bn(out\ M) = bn(M)$
$fn(move\ M) = fn(M)$	$bn(move\ M) = bn(M)$
$fn(M.M') = fn(M) \cup fn(M')$	$bn(M.M') = bn(M) \cup bn(M')$
$fn(\mathbf{0}) = \varnothing$	$bn(\mathbf{0}) = \varnothing$
$fn(P \mid P') = fn(P) \cup fn(P')$	$bn(P \mid P') = bn(P) \cup bn(P')$
$fn(!P) = fn(P)$	$bn(!P) = bn(P)$
$fn(M[P]) = fn(M) \cup fn(P)$	$bn(M[P]) = bn(M) \cup bn(P)$
$fn((\nu n : \mathbf{P}[\mathcal{B}])P) = fn(P) - \{n\}$	$bn((\nu n : \mathbf{P}[\mathcal{B}])P) = bn(P) \cup \{n\}$
$fn(M.P) = fn(M) \cup fn(P)$	$bn(M.P) = bn(M) \cup bn(P)$
$fn(\langle M \rangle.P) = fn(M) \cup fn(P)$	$bn(\langle M \rangle.P) = bn(M) \cup bn(P)$
$fn((n : \mu).P) = fn(P) - \{n\}$	$bn((n : \mu).P) = bn(P) \cup \{n\}$

1.2 Operational Semantics

The operational semantics is given by reduction relation along with a structural congruence, in the same way as those for mobile ambients.

Table 3
Structural congruence

equivalence:	
$P \equiv P$	(SRefl)
$P \equiv Q \Rightarrow Q \equiv P$	(SSymm)
$P \equiv Q, Q \equiv R \Rightarrow P \equiv R$	(STrans)
congruence:	
$P \equiv Q \Rightarrow P \mid R \equiv Q \mid R$	(SPar)
$P \equiv Q \Rightarrow !P \equiv !Q$	(SRepl)
$P \equiv Q \Rightarrow M[P] \equiv M[Q]$	(SAmb)
$P \equiv Q \Rightarrow (\nu n : \mathbf{P}[\mathcal{B}])P \equiv (\nu n : \mathbf{P}[\mathcal{B}])Q$	(SRes)
$P \equiv Q \Rightarrow M.P \equiv M.Q$	(SAct)
$P \equiv Q \Rightarrow \langle M \rangle.P \equiv \langle M \rangle.Q$	(SCommOut)
$P \equiv Q \Rightarrow (n : \mu).P \equiv (n : \mu).Q$	(SCommIn)
sequential composition (associativity):	
$(M.M').P \equiv M.M'.P$	(SPath)
parallel composition:	
$P \mid Q \equiv Q \mid P$	(SParComm)
$(P \mid Q) \mid R \equiv P \mid (Q \mid R)$	(SParAssoc)
$P \mid \mathbf{0} \equiv P$	(SParNull)
replication:	
$!P \equiv P \mid !P$	(SReplPar)
$!\mathbf{0} \equiv \mathbf{0}$	(SReplNull)
restriction and scope extrusion:	
$n \neq m \Rightarrow (\nu n : \mathbf{P}[\mathcal{B}])(\nu m : \mathbf{P}[\mathcal{B}'])P \equiv (\nu m : \mathbf{P}[\mathcal{B}'])(\nu n : \mathbf{P}[\mathcal{B}])P$	(SResRes)
$n \notin fn(Q) \Rightarrow (\nu n : \mathbf{P}[\mathcal{B}])P \mid Q \equiv (\nu n : \mathbf{P}[\mathcal{B}])(P \mid Q)$	(SResPar)
$n \neq m \Rightarrow (\nu n : \mathbf{P}[\mathcal{B}])m[P] \equiv m[(\nu n : \mathbf{P}[\mathcal{B}])P]$	(SResAmb)
$(\nu n : \mathbf{P}[\mathcal{B}])\mathbf{0} \equiv \mathbf{0}$	(SResNull)
garbage collection:	
$(\nu n : \mathbf{P}[\mathcal{B}])n[\mathbf{0}] \equiv \mathbf{0}$	(SAmbNull)

Each name of the process term can figure either as free (Table 2a) or bound (Table 2b).

We write $P\{n \leftarrow M\}$ for a substitution of the capability M for each free occurrences of the name n in the term P. The similarly for $M\{n \leftarrow M\}$.

Structural congruence is shown in Table 3 and it is standard for mobile ambients.

In addition, we identify processes up to renaming of bound names (α-conversion) as shown in Table 4.

Table 4
α-conversion

$(\nu n : \mathbf{P}[\mathcal{B}])P = (\nu m : \mathbf{P}[\mathcal{B}])P\{n \leftarrow m\} \quad m \notin fn(P)$	(SAlphaRes)
$(n : \mu)P = (m : \mu)P\{n \leftarrow m\} \quad m \notin fn(P)$	(SAlphaCommIn)

The reduction rules in Table 5 are those for mobile ambients, with the obvious difference consisting in the synchronous output and the missing *open* operation, and with the new rule for the *move* operation similar to the "migrate" instructions for strong code mobility in software agents.

Table 5
Reduction rules

basic reductions:	
$n[in\ m.P \mid Q] \mid m[R] \rightarrow m[n[P \mid Q] \mid R]$	(RIn)
$m[n[out\ m.P \mid Q] \mid R] \rightarrow n[P \mid Q] \mid m[R]$	(ROut)
$n[move\ m.P \mid Q] \mid m[R] \rightarrow n[Q] \mid m[P \mid R]$	(RMove)
$(n : \mu).P \mid \langle M \rangle.Q \rightarrow P\{n \leftarrow M\} \mid Q$	(RComm)
structural reductions:	
$P \rightarrow Q \Rightarrow P \mid R \rightarrow Q \mid R$	(RPar)
$P \rightarrow Q \Rightarrow n[P] \rightarrow n[Q]$	(RAmb)
$P \rightarrow Q \Rightarrow (\nu n : \mathbf{P}[\mathcal{B}])P \rightarrow (\nu n : \mathbf{P}[\mathcal{B}])Q$	(RRes)
$P' \equiv P, P \rightarrow Q, Q \equiv Q' \Rightarrow P' \rightarrow Q'$	(RStruct)

3 Overview of the Type System

The restriction of the mobility operations is defined by types applying a *behavioral scheme*. The scheme allows setting up the access rights for traveling of threads and ambients in the ambient hierarchy space of the system.

3.1 Types and Behavioral Scheme

Types are defined in Table 6 where we present communication types and message types.

Table 6
Types with behavioral schemes

$\kappa ::=$	**communication type**
$\mid \perp$	no communication
$\mid \mu$	communication of messages of type μ
$\mu ::=$	**message type**
$\mid$ **P**$[\mathcal{B}]$	process with behavioral scheme $\mathcal{B}$
$\mid$ **O**$[\mathcal{B} \mapsto \mathcal{B}']$	operation which changes behavioral scheme $\mathcal{B}$ to $\mathcal{B}'$

The behavioral scheme is the structure $\mathcal{B}=(\kappa, Reside, Pass, Move)$ which contains four components:

- κ is the communication type of the ambient's threads
- *Reside* is the set of behavioral schemes of other ambients where the ambient can stay
- *Pass* is the set of behavioral schemes of other ambients that ambient can go through, it must be $Pass \subseteq Reside$
- *Move* is the set of behavioral schemes of other ambients where ambient can move its containing thread

3.2 Typing Rules

The type environment is defined as a set $\Gamma = \{n_1 : \mu_1, \ldots, n_l : \mu_l\}$ where each $n_i : \mu_i$ assigns a unique type μ_i to a name n_i.

The domain of the type environment is defined by:

1 $Dom(\varnothing) = \varnothing$

2 $Dom(\Gamma, n : \mu) = Dom(\Gamma) \cup \{n\}$

We define two type formulas for our ambient calculus:

1 $\Gamma \vdash M : \mu$

2 $\Gamma \vdash P : \mathbf{P}[\mathcal{B}]$

Typing rules are shown in Table 7 and they are used to derive type formulas of ambient processes. We say the process is *well-typed* when we are able to derive a type formula for it using our typing rules. Well-typed processes respect the communication and mobility restrictions defined in all behavioral schemes of the system.

Table 7
Typing rules

$\dfrac{n:\mu \in \Gamma}{\Gamma \vdash n:\mu}$	(TName)
$\dfrac{\Gamma \vdash M:\mathbf{P}[\mathcal{B}] \quad \mathcal{B} \in Pass(\mathcal{B}')}{\Gamma \vdash in\ M:\mathbf{O}[\mathcal{B}' \mapsto \mathcal{B}']}$	(TIn)
$\dfrac{\Gamma \vdash M:\mathbf{P}[\mathcal{B}] \quad \mathcal{B} \in Pass(\mathcal{B}') \quad Reside(\mathcal{B}) \subseteq Reside(\mathcal{B}')}{\Gamma \vdash out\ M:\mathbf{O}[\mathcal{B}' \mapsto \mathcal{B}']}$	(TOut)
$\dfrac{\Gamma \vdash M:\mathbf{P}[\mathcal{B}] \quad \mathcal{B} \in Move(\mathcal{B}')}{\Gamma \vdash move\ M:\mathbf{O}[\mathcal{B} \mapsto \mathcal{B}']}$	(TMove)
$\dfrac{\Gamma \vdash M:\mathbf{O}[\mathcal{B}'' \mapsto \mathcal{B}'] \quad \Gamma \vdash M':\mathbf{O}[\mathcal{B} \mapsto \mathcal{B}'']}{\Gamma \vdash M.M':\mathbf{O}[\mathcal{B} \mapsto \mathcal{B}']}$	(TPath)
$\dfrac{}{\Gamma \vdash \mathbf{0}:\mathbf{P}[\mathcal{B}]}$	(TNull)
$\dfrac{\Gamma \vdash P:\mathbf{P}[\mathcal{B}] \quad \Gamma \vdash P':\mathbf{P}[\mathcal{B}]}{\Gamma \vdash P \mid P':\mathbf{P}[\mathcal{B}]}$	(TPar)
$\dfrac{\Gamma \vdash P:\mathbf{P}[\mathcal{B}]}{\Gamma \vdash !P:\mathbf{P}[\mathcal{B}]}$	(TRepl)
$\dfrac{\Gamma \vdash P:\mathbf{P}[\mathcal{B}] \quad \Gamma \vdash M:\mathbf{P}[\mathcal{B}] \quad \mathcal{B}' \in Reside(\mathcal{B})}{\Gamma \vdash M[P]:\mathbf{P}[\mathcal{B}']}$	(TAmb)
$\dfrac{\Gamma, n:\mathbf{P}[\mathcal{B}'] \vdash P:\mathbf{P}[\mathcal{B}]}{\Gamma \vdash (\nu n:\mathbf{P}[\mathcal{B}'])P:\mathbf{P}[\mathcal{B}]}$	(TRes)
$\dfrac{\Gamma \vdash M:\mathbf{O}[\mathcal{B} \mapsto \mathcal{B}'] \quad \Gamma \vdash P:\mathbf{P}[\mathcal{B}]}{\Gamma \vdash M.P:\mathbf{P}[\mathcal{B}']}$	(TAct)
$\dfrac{\Gamma \vdash P:\mathbf{P}[\mathcal{B}] \quad \Gamma \vdash M:\mu \quad \kappa(\mathcal{B}) = \mu}{\Gamma \vdash \langle M \rangle.P:\mathbf{P}[\mathcal{B}]}$	(TCommOut)
$\dfrac{\Gamma, n:\mu \vdash P:\mathbf{P}[\mathcal{B}] \quad \kappa(\mathcal{B}) = \mu}{\Gamma \vdash (n:\mu).P:\mathbf{P}[\mathcal{B}]}$	(TCommIn)

4 Encoding Named Channels

A standard expressiveness test for the ambient calculus and our variant is the encoding of communication on named channels via local anonymous communication within ambients. We consider a core fragment of the typed monadic synchronous π-calculus, given by the following grammar:

$$P ::= x(y:\sigma).P \mid \overline{x}\langle z\rangle.P \mid \mathsf{new}\ a:\sigma\ P \mid P_1 \mid P_2 \mid !P$$

where P, P_1,P_2 denotes processes and x, y, z, a are named channels from the set of all names $\mathcal{N}$ and lets k be their number.

Table 8
Free and bound names in π-calculus

$fn_\pi(x(y:\sigma).P) = \{x\} \cup fn_\pi(P)$	$bn_\pi(x(y:\sigma).P) = \{y\} \cup bn_\pi(P)$
$fn_\pi(\overline{x}\langle z\rangle.P) = \{x,z\} \cup fn_\pi(P)$	$bn_\pi(\overline{x}\langle z\rangle.P) = bn_\pi(P)$
$fn_\pi(\mathsf{new}\ a:\sigma\ P) = fn_\pi(P) - \{a\}$	$bn_\pi(\mathsf{new}\ a:\sigma\ P) = bn_\pi(P) \cup \{a\}$
$fn_\pi(P_1 \mid P_2) = fn_\pi(P_1) \cup fn_\pi(P_2)$	$bn_\pi(P_1 \mid P_2) = bn_\pi(P_1) \cup bn_\pi(P_2)$
$fn_\pi(!P) = fn_\pi(P)$	$bn_\pi(!P) = bn_\pi(P)$

Table 9
Structural congruence in π-calculus

$P \equiv_\pi P$	(πSRefl)
$P \equiv_\pi Q \Rightarrow Q \equiv_\pi P$	(πSSymm)
$P \equiv_\pi Q, Q \equiv_\pi R \Rightarrow P \equiv_\pi R$	(πSTrans)
$P \equiv_\pi Q \Rightarrow x(z:\sigma).P \equiv_\pi x(z:\sigma).Q$	(πSCommIn)
$P \equiv_\pi Q \Rightarrow \overline{x}\langle z\rangle.P \equiv_\pi \overline{x}\langle z\rangle.Q$	(πSCommOut)
$P \equiv_\pi Q \Rightarrow \mathsf{new}\ a:\sigma\ P \equiv_\pi \mathsf{new}\ a:\sigma\ Q$	(πSRes)
$P \equiv_\pi Q \Rightarrow P \mid R \equiv_\pi Q \mid R$	(πSPar)
$P \equiv_\pi Q \Rightarrow !P \equiv_\pi !Q$	(πSRepl)
$P \mid Q \equiv_\pi Q \mid P$	(πSParComm)
$(P \mid Q) \mid R \equiv_\pi P \mid (Q \mid R)$	(πSParAssoc)
$!P \equiv_\pi P \mid !P$	(πSReplPar)
$a \neq b \Rightarrow \mathsf{new}\ a:\sigma\ \mathsf{new}\ b:\sigma'\ P \equiv_\pi \mathsf{new}\ b:\sigma'\ \mathsf{new}\ a:\sigma\ P$	(πSResRes)
$a \notin fn_\pi(P) \Rightarrow \mathsf{new}\ a:\sigma\ (P \mid Q) \equiv_\pi \mathsf{new}\ a:\sigma\ P \mid Q$	(πSResPar)
$a \notin fn_\pi(P) \Rightarrow \mathsf{new}\ a:\sigma\ P \equiv_\pi P$	(πSResSkip)

The set of free and bound names of π-calculus terms are in Table 8.

Table 10
Free and bound names in π-calculus

$P \to_\pi Q \Rightarrow \mathsf{new}\ a:\sigma\ P \to_\pi \mathsf{new}\ a:\sigma\ Q$	(πRRes)
$P \to_\pi Q \Rightarrow P \mid R \to_\pi Q \mid R$	(πRPar)
$x(y:\sigma).P \mid \overline{x}\langle z\rangle.Q \to_\pi P\{y \leftarrow z\} \mid Q$	(πRComm)
$P' \equiv_\pi P, P \to_\pi Q, Q \equiv_\pi Q' \Rightarrow P' \to_\pi Q'$	(πRStruct)

Structural operational semantics of π-calculus is given by the structural congruence $\equiv_\pi$ (Table 9) and reduction relation $\to_\pi$ (Table 10).

Table 11
Typing rules in π-calculus

$\dfrac{a:\sigma \in \Gamma}{\Gamma \vdash a:\sigma}$	(πTName)
$\dfrac{\Gamma \vdash x:\mathrm{CH}(\sigma) \quad \Gamma, y:\sigma \vdash P}{\Gamma \vdash x(y:\sigma).P}$	(πTCommIn)
$\dfrac{\Gamma \vdash x:\mathrm{CH}(\sigma) \quad \Gamma \vdash z:\sigma \quad \Gamma \vdash P}{\Gamma \vdash \overline{x}\langle z\rangle.P}$	(πTCommOut)
$\dfrac{\Gamma, a:\sigma \vdash P}{\Gamma \vdash \mathsf{new}\ a:\sigma\ P}$	(πTRes)
$\dfrac{\Gamma \vdash P_1 \quad \Gamma \vdash P_2}{\Gamma \vdash P_1 \mid P_2}$	(πTPar)
$\dfrac{\Gamma \vdash P}{\Gamma \vdash !P}$	(πTRepl)

The type of communication channel is denoted by σ and is given by the grammar:

$$\sigma ::= \mathrm{CH}() \mid \mathrm{CH}(\sigma)$$

The type environment of the term P and the type formula in the type system of the π-calculus are defined in the following way:

$$\Gamma = \{\forall a:\sigma \mid a \in fn_\pi(P)\}$$

$$\Gamma \vdash a:\sigma \quad \Gamma \vdash P$$

where $a:\sigma$ is pair of communication name *a* and its communication channel type σ. Formula $\Gamma \vdash a:\sigma$ denotes that $a:\sigma \in \Gamma$ (where σ is the unique type of name *a*) and formula $\Gamma \vdash P$ denotes that term *P* is correctly typed in type environment Γ.

The type formulas in the type system of π-calculus are derived using the typing rules shown in Table 11.

The main idea of communication channels encoding in the system of mobile ambients is the representation of the channel by an ambient. The process term whose prefix is a communication action on channel *x* is expressed by a thread which is at first moved to an ambient $x[\ldots]$ and then it is moved back to the original ambient. While the move operation allows moving only to the neighbor (concurrent) ambient, we must define ambient *p* concurrent to ambient of channel *x*. The ambient *p* will encode the following term *P* from the π-calculus i.e. $p[[\![P]\!]] \mid x[\ldots]$.

Table 12
Encoding of π-calculus terms

$[\![P]\!]_{\mathcal{N}} = p[[\![P]\!]] \mid [\![\mathcal{N}]\!]$
$[\![\{a_1,\ldots,a_k\}]\!] = a_1[\ldots] \mid \ldots \mid a_k[\ldots]$, ´where *k* the number of names in $\mathcal{N}$
$[\![x(y:\sigma).P]\!] = move\ x.(y:[\![\sigma]\!]).move\ p.[\![P]\!]$
$[\![\overline{x}\langle z\rangle.P]\!] = move\ x.\langle z\rangle.move\ p.[\![P]\!]$
$[\![\text{new } a:\sigma\ P]\!] = (\nu a:[\![\sigma]\!])(a[out\ p] \mid [\![P]\!])$
$[\![P_1 \mid P_2]\!] = [\![P_1]\!] \mid [\![P_2]\!]$
$[\![!P]\!] = ![\![P]\!]$
$[\![\Gamma]\!] = \{p:\mathbf{P}[\mathcal{B}_p], \forall a:[\![\sigma]\!] \mid a:\sigma \in \Gamma\}$
$[\![\sigma]\!] = \mathbf{P}[\mathcal{B}_i]$, where *i* is the level of nested CH() in σ

The sequence of hierarchical types $\mathrm{CH}(), \mathrm{CH}(\mathrm{CH}()), \ldots, \mathrm{CH}^l()$ is expressed as the same sequence of behavioral schemes $\mathcal{B}_0, \mathcal{B}_1, \ldots, \mathcal{B}_l$, where *l* is the deepest level of nested CH(). The behavioral schemes have following structure:

$$\mathcal{B}_0 = (\bot, \{\mathcal{B}, \mathcal{B}_p\}, \{\mathcal{B}_p\}, \{\mathcal{B}_p\})$$
$$\mathcal{B}_1 = (\mathbf{P}[\mathcal{B}_0], \{\mathcal{B}, \mathcal{B}_p\}, \{\mathcal{B}_p\}, \{\mathcal{B}_p\})$$
$$\vdots$$
$$\mathcal{B}_i = (\mathbf{P}[\mathcal{B}_{i-1}], \{\mathcal{B}, \mathcal{B}_p\}, \{\mathcal{B}_p\}, \{\mathcal{B}_p\})$$
$$\vdots$$
$$\mathcal{B}_l = (\mathbf{P}[\mathcal{B}_{l-1}], \{\mathcal{B}, \mathcal{B}_p\}, \{\mathcal{B}_p\}, \{\mathcal{B}_p\})$$

where $\mathcal{B}_p$ is the behavioral scheme of ambient p and $\mathcal{B}$ is the behavioral scheme of the whole encoded process term, i.e. term $p[[\![P]\!]] \mid a_1[\ldots] \mid \ldots \mid a_k[\ldots]$. Behavioral scheme $\mathcal{B}_p$ of ambient p has following structure:

$$\mathcal{B}_p = (\bot, \{\mathcal{B}\}, \varnothing, \{\mathcal{B}_0, \mathcal{B}_1, \ldots, \mathcal{B}_l\})$$

The encoding of term P with the set of name $\mathcal{N}$ and type system Γ of π-calculus to mobile ambient is given by Table 12.

The correctness of π-calculus encoding is shown in following two theorems.

Theorem 1: (respecting types) Let P is the term of π-calculus with the set of names $\mathcal{N}$ and $\Gamma \vdash P$. Then $[[\![\Gamma]\!] \vdash [[\![P]\!]_{\mathcal{N}} : \mathbf{P}[\mathcal{B}]$ for some behavioral scheme $\mathcal{B}$.

Proof: By induction on the structure of the process.

Let $P = x(y:\sigma).P'$ and according (πTCommIn) there is $\Gamma \vdash x(y:\sigma).P'$ by assumption $\Gamma \vdash x : \mathrm{CH}(\sigma)$ and $\Gamma, y:\sigma \vdash P'$. After encoding we get $[[\![P]\!]_{\mathcal{N}} = p[move\ x.(y:[[\![\sigma]\!]).move\ p.[[\![P']\!]] \mid [[\![\mathcal{N}]\!]$, $[[\![\Gamma]\!] = \{p : \mathbf{P}[\mathcal{B}_p], \forall a : [[\![\sigma]\!] \mid a : \sigma \in \Gamma\}$, $[[\![\sigma]\!] = \mathbf{P}[\mathcal{B}_i]$, and $[[\![\mathrm{CH}(\sigma)]\!] = \mathbf{P}[\mathcal{B}_{i-1}]$. Then according (TPar), (TAmb), (TAct), (TMove), and (TCommIn) there is $[[\![\Gamma]\!] \vdash [[\![P]\!]_{\mathcal{N}} : \mathbf{P}[\mathcal{B}]$ for some $\mathcal{B}$.

Let $P = \overline{x}\langle z\rangle.P'$ and according (πTCommOut) there is $\Gamma \vdash \overline{x}\langle z\rangle.P'$ by assumption $\Gamma \vdash x : \mathrm{CH}(\sigma)$, $\Gamma \vdash z : \sigma$, and $\Gamma \vdash P'$. After encoding we get $[[\![P]\!]_{\mathcal{N}} = p[move\ x.\langle z\rangle.move\ p.[[\![P']\!]] \mid [[\![\mathcal{N}]\!]$, $[[\![\Gamma]\!] = \{p : \mathbf{P}[\mathcal{B}_p], \forall a : [[\![\sigma]\!] \mid a : \sigma \in \Gamma\}$, $[[\![\sigma]\!] = \mathbf{P}[\mathcal{B}_i]$, and $[[\![\mathrm{CH}(\sigma)]\!] = \mathbf{P}[\mathcal{B}_{i-1}]$. Then according (TPar), (TAmb), (TAct), (TMove), and (TCommOut) there is $[[\![\Gamma]\!] \vdash [[\![P]\!]_{\mathcal{N}} : \mathbf{P}[\mathcal{B}]$ for some $\mathcal{B}$.

Let $P = \mathsf{new}\ a:\sigma\ P'$ and according (πTRes) there is $\Gamma \vdash \mathsf{new}\ a:\sigma\ P'$ if $\Gamma, a:\sigma \vdash P'$. After encoding we get $[[\![P]\!]_{\mathcal{N}} = p[(\nu a : [[\![\sigma]\!])(a[out\ p] \mid [[\![P']\!])] \mid [[\![\mathcal{N}]\!]$, $[[\![\Gamma]\!] = \{p : \mathbf{P}[\mathcal{B}_p], \forall a : [[\![\sigma]\!] \mid a : \sigma \in \Gamma\}$, and $[[\![\sigma]\!] = \mathbf{P}[\mathcal{B}_i]$. Then according (TPar), (TAmb), (TRes), again according (TPar), (TAmb), and according (TOut) there is $[[\![\Gamma]\!] \vdash [[\![P]\!]_{\mathcal{N}} : \mathbf{P}[\mathcal{B}]$ for some $\mathcal{B}$.

Let $P = P' \mid P''$ and according (πTPar) there is $\Gamma \vdash P' \mid P''$ by assumption $\Gamma \vdash P'$ and $\Gamma \vdash P''$. After encoding we get $[[\![P]\!]_{\mathcal{N}} = p[[[\![P']\!] \mid [[\![P'']\!])] \mid [[\![\mathcal{N}]\!]$, $[[\![\Gamma]\!] = \{p : \mathbf{P}[\mathcal{B}_p], \forall a : [[\![\sigma]\!] \mid a : \sigma \in \Gamma\}$, and $[[\![\sigma]\!] = \mathbf{P}[\mathcal{B}_i]$. Then according (TPar), (TAmb), and again according (TPar) there is $[[\![\Gamma]\!] \vdash [[\![P]\!]_{\mathcal{N}} : \mathbf{P}[\mathcal{B}]$ for some $\mathcal{B}$.

Let $P = !P'$ and according (πTRepl) there is $\Gamma \vdash !P'$ if $\Gamma \vdash P'$. After encoding we get $[[\![P]\!]_{\mathcal{N}} = p[![[\![P']\!])] \mid [[\![\mathcal{N}]\!]$, $[[\![\Gamma]\!] = \{p : \mathbf{P}[\mathcal{B}_p], \forall a : [[\![\sigma]\!] \mid a : \sigma \in \Gamma\}$, and

$[\![\sigma]\!] = \mathbf{P}[\mathcal{B}_i]$. Then according (TPar), (TAmb), and (TRepl) there is $[\![\Gamma]\!] \vdash [\![P]\!]_{\mathcal{N}} : \mathbf{P}[\mathcal{B}]$ for some $\mathcal{B}$.

Theorem 2: (encoding correctness) Let P is the term of π-calculus with the set of name $\mathcal{N}$. If $P \equiv_\pi Q$ then $[\![P]\!]_{\mathcal{N}} \equiv [\![Q]\!]_{\mathcal{N}}$ and if $P \to_\pi Q$ then $[\![P]\!]_{\mathcal{N}} \to^* [\![Q]\!]_{\mathcal{N}}$.

Proof: By induction on the structure of the process.

1. Let P is the term of π-calculus with the set of names $\mathcal{N}$. If $P \equiv_\pi Q$ then $[\![P]\!]_{\mathcal{N}} \equiv [\![Q]\!]_{\mathcal{N}}$.

(πSRefl) Let $P \equiv_\pi P$ then $[\![P]\!]_{\mathcal{N}} \equiv [\![P]\!]_{\mathcal{N}}$.

(πSSymm) Let $Q \equiv_\pi P$ then $[\![Q]\!]_{N} \equiv [\![P]\!]_{N}$.

(πSTrans) Let $P \equiv_\pi R$ and $R \equiv_\pi Q$ for some R and let $[\![P]\!]_{\mathcal{N}} \equiv [\![R]\!]_{\mathcal{N}}$ and $[\![R]\!]_{\mathcal{N}} \equiv [\![Q]\!]_{\mathcal{N}}$ then $[\![P]\!]_{\mathcal{N}} \equiv [\![Q]\!]_{\mathcal{N}}$.

(πSCommIn) Let $P = x(y:\sigma).P'$, $Q = x(y:\sigma).Q'$, $P' \equiv_\pi Q'$ end let $[\![P']\!]_{\mathcal{N}} \equiv [\![Q']\!]_{\mathcal{N}}$. According definition of the structural congruence $\equiv$ there is $p[move\ x.(y:[\![\sigma]\!]).move\ p.[\![P']\!]] \mid [\![\mathcal{N}]\!] \equiv p[move\ x.(y:[\![\sigma]\!]).move\ p.[\![Q']\!]] \mid [\![\mathcal{N}]\!]$, what is $[\![x(y:\sigma).P']\!]_{\mathcal{N}} \equiv [\![x(y:\sigma).Q']\!]_{\mathcal{N}}$.

(πSCommOut) Let $P = \overline{x}\langle z\rangle.P'$, $Q = \overline{x}\langle z\rangle.Q'$, $P' \equiv_\pi Q'$ and let $[\![P']\!]_{\mathcal{N}} \equiv [\![Q']\!]_{\mathcal{N}}$. According definition of the structural congruence $\equiv$ there is $p[move\ x.\langle z\rangle.move\ p.[\![P']\!]] \mid [\![\mathcal{N}]\!] \equiv p[move\ x.\langle z\rangle.move\ p.[\![Q']\!]] \mid [\![\mathcal{N}]\!]$, what is $[\![\overline{x}\langle z\rangle.P']\!]_{\mathcal{N}} \equiv [\![\overline{x}\langle z\rangle.Q']\!]_{\mathcal{N}}$.

(πSRes) Let $P = \mathsf{new}\ a:\sigma\ P'$, $Q = \mathsf{new}\ a:\sigma\ Q'$, $P' \equiv_\pi Q'$ and let $[\![P']\!]_{\mathcal{N}} \equiv [\![Q']\!]_{\mathcal{N}}$. According definition of the structural congruence $\equiv$ there is $p[(\nu a:[\![\sigma]\!])(a[out\ p] \mid [\![P']\!])] \mid [\![\mathcal{N}]\!] \equiv p[(\nu a:[\![\sigma]\!])(a[out\ p] \mid [\![Q']\!])] \mid [\![\mathcal{N}]\!]$, what is $[\![\mathsf{new}\ a:\sigma\ P']\!]_{\mathcal{N}} \equiv [\![\mathsf{new}\ a:\sigma\ Q']\!]_{\mathcal{N}}$.

(πSPar) Let $P = P' \mid R$, $Q = Q' \mid R$, $P' \equiv_\pi Q'$ and let $[\![P']\!]_{\mathcal{N}} \equiv [\![Q']\!]_{\mathcal{N}}$. According definition of the structural congruence $\equiv$ there is $p[[\![P']\!] \mid [\![R]\!]] \mid [\![\mathcal{N}]\!] \equiv p[[\![Q']\!] \mid [\![R]\!]] \mid [\![\mathcal{N}]\!]$, what is $[\![P' \mid R]\!]_{\mathcal{N}} \equiv [\![Q' \mid R]\!]_{\mathcal{N}}$.

(πSRepl) Let $P = !P'$, $Q = !Q'$, $P' \equiv_\pi Q'$ and let $[\![P']\!]_{\mathcal{N}} \equiv [\![Q']\!]_{\mathcal{N}}$. According definition of the structural congruence $\equiv$ there is $p[![\![P']\!]] \mid [\![\mathcal{N}]\!] \equiv p[![\![Q']\!]] \mid [\![\mathcal{N}]\!]$, what is $[\![!P']\!]_{\mathcal{N}} \equiv [\![!Q']\!]_{\mathcal{N}}$.

(πSParComm) Let $P = P' \mid P''$ a $Q = P'' \mid P'$. According definition of the structural congruence $\equiv$ there is $p[[\![P']\!] \mid [\![P'']\!]] \mid [\![\mathcal{N}]\!] \equiv p[[\![P'']\!] \mid [\![P']\!]] \mid [\![\mathcal{N}]\!]$, what is $[\![P' \mid P'']\!]_{\mathcal{N}} \equiv [\![P'' \mid P']\!]_{\mathcal{N}}$.

(πSParAssoc) Let $P = (P' \mid P'') \mid P'''$ and $Q = P' \mid (P'' \mid P''')$. According definition of the structural congruence $\equiv$ there is $p[[\![P']\!] \mid ([\![P'']\!] \mid [\![P''']\!])] \mid [\![\mathcal{N}]\!] \equiv p[([\![P']\!] \mid [\![P'']\!]) \mid [\![P''']\!]] \mid [\![\mathcal{N}]\!]$, what is $[\![(P' \mid P'') \mid P''']\!]_{\mathcal{N}} \equiv [\![P' \mid (P'' \mid P''')]\!]_{\mathcal{N}}$.

(πSReplPar) Let $P = !P'$ and $Q = P' \mid !P'$. According definition of the structural congruence $\equiv$ there is $p[![\![P']\!]] \mid [\![\mathcal{N}]\!] \equiv p[[\![P']\!] \mid ![\![P']\!]] \mid [\![\mathcal{N}]\!]$, what is $[\![!P']\!]_{\mathcal{N}} \equiv [\![P' \mid !P']\!]_{\mathcal{N}}$.

(πSResRes) Let $P = \mathsf{new}\ a : \sigma\ \mathsf{new}\ b : \sigma'\ P'$, $Q = \mathsf{new}\ b : \sigma' \mathsf{new}\ a : \sigma\ P'$ and $a \neq b$. According definition of the structural congruence $\equiv$ there is $p[(\nu a : [\![\sigma]\!])(a[out\ p] \mid (\nu b : [\![\sigma']\!])(b[out\ p] \mid [\![P']\!]))] \mid [\![\mathcal{N}]\!] \equiv p[(\nu b : [\![\sigma']\!])(b[out\ p] \mid (\nu a : [\![\sigma]\!])(a[out\ p] \mid [\![P']\!]))] \mid [\![\mathcal{N}]\!]$, what is $[\![\mathsf{new}\ a : \sigma\ \mathsf{new}\ b : \sigma'\ P']\!]_{\mathcal{N}} \equiv [\![\mathsf{new}\ b : \sigma'\ \mathsf{new}\ a : \sigma\ P']\!]_{\mathcal{N}}$.

(πSResPar) Let $P = \mathsf{new}\ a : \sigma\ (P' \mid P'')$, $Q = \mathsf{new}\ a : \sigma\ P' \mid P''$ a $a \notin fn_{\pi}(P')$. There is $fn([\![P']\!]) = fn_{\pi}(P') \cup \{p\}$ and $a \neq p$. If $a \notin fn_{\pi}(P')$, then $a \notin fn([\![P']\!])$. According definition of the structural congruence $\equiv$ there is $p[(\nu a : [\![\sigma]\!])([\![P']\!] \mid [\![P'']\!])] \mid [\![\mathcal{N}]\!] \equiv p[(\nu a : [\![\sigma]\!])[\![P']\!] \mid [\![P'']\!]] \mid [\![\mathcal{N}]\!]$, what is $[\![\mathsf{new}\ a : \sigma\ (P' \mid P'')]\!]_{\mathcal{N}} \equiv [\![\mathsf{new}\ a : \sigma\ P' \mid P'']\!]_{\mathcal{N}}$.

(πSResSkip) Let $P = \mathsf{new}\ a : \sigma\ P'$, $Q = P'$ and $a \notin fn_{\pi}(P')$. There is $fn([\![P']\!]) = fn_{\pi}(P') \cup \{p\}$ and $a \neq p$. If $a \notin fn_{\pi}(P')$, then $a \notin fn([\![P']\!])$. According definition of the structural congruence $\equiv$ there is $p[(\nu a : [\![\sigma]\!])(a[out\ p] \mid [\![P']\!])] \mid [\![\mathcal{N}]\!] \equiv p[(\nu a : [\![\sigma]\!])a[out\ p] \mid [\![P']\!] \mid] \mid [\![\mathcal{N}]\!]$ and $a \notin fn([\![P']\!])$, what is $[\![\mathsf{new}\ a : \sigma\ P']\!]_{\mathcal{N}} \equiv [\![P']\!]_{\mathcal{N}}$.

2. Let P is the term of π-calculus with the set of names $\mathcal{N}$. If $P \rightarrow_{\pi} Q$ then $[\![P]\!]_{\mathcal{N}} \rightarrow^* [\![Q]\!]_{\mathcal{N}}$.

(πRRes) Let $P = \mathsf{new}\ a : \sigma\ P'$, $Q = \mathsf{new}\ a : \sigma\ Q'$ and $P' \rightarrow_{\pi} Q'$. We need to show, if $fn_{\pi}(\mathsf{new}\ a : \sigma\ P') \subseteq \mathcal{N}$, then $[\![\mathsf{new}\ a : \sigma\ P']\!]_{\mathcal{N}} \rightarrow^* [\![\mathsf{new}\ a : \sigma\ Q']\!]_{\mathcal{N}}$. If $fn_{\pi}(\mathsf{new}\ a : \sigma\ P') \subseteq \mathcal{N}$, then $fn_{\pi}(P') \subseteq \mathcal{N} \cup \{a\}$, what means that $a \notin \mathcal{N}$. According $[\![P']\!]_{\mathcal{N} \cup \{a\}} \rightarrow^* [\![Q']\!]_{\mathcal{N} \cup \{a\}}$ and by repeat usage of (RRes) and structural congruence $\equiv$ we get $(\nu a : [\![\sigma]\!])[\![P']\!]_{\mathcal{N} \cup \{a\}} \rightarrow^* (\nu a : [\![\sigma]\!])[\![Q']\!]_{\mathcal{N} \cup \{a\}}$. Then $(\nu a : [\![\sigma]\!])[\![P']\!]_{\mathcal{N} \cup \{a\}} = (\nu a : [\![\sigma]\!])(p[[\![P']\!] \mid [\![\mathcal{N} \cup \{a\}]\!]) \equiv (\nu : [\![\sigma]\!])(p[[\![P']\!] \mid [\![\mathcal{N}]\!] \mid a[\ldots]) \equiv p[[\![\mathsf{new}\ a : \sigma\ P']\!]] \mid [\![\mathcal{N}]\!] = [\![\mathsf{new}\ a : \sigma\ P']\!]_{\mathcal{N}}$ and the same way we get

$(\nu a:[\![\sigma]\!])[\![Q']\!]_{\mathcal{N}\cup\{a\}} \equiv [\![\text{new } a:\sigma\ Q']\!]_{\mathcal{N}}$, what is $[\![\text{new } a:\sigma\ P']\!]_{\mathcal{N}\cup\{a\}} \rightarrow^* [\![\text{new } a:\sigma\ Q']\!]_{\mathcal{N}}$.

(πRPar) Let $P = P' \mid R$, $Q = Q' \mid R$, $P' \rightarrow Q'$ and let $[\![P']\!]_{\mathcal{N}} \rightarrow^* [\![Q']\!]_{\mathcal{N}}$ and by repeat usage of (RPar) and structural congruence $\equiv$ we get $[\![P']\!]_{\mathcal{N}} \mid [\![R]\!] \rightarrow^* [\![Q']\!]_{\mathcal{N}} \mid [\![R]\!]$, what is $[\![P' \mid R]\!]_{\mathcal{N}} \rightarrow^* [\![Q' \mid R]\!]_{\mathcal{N}}$.

(πRComm) Let $P = x(y:\sigma).P' \mid \overline{x}\langle z\rangle.P''$ and $Q = P'\{y \leftarrow z\} \mid P''$. We need to show, if $fn_\pi(x(y:\sigma).P' \mid \overline{x}\langle z\rangle.P'') \subseteq \mathcal{N}$, then $[\![x(y:\sigma).P' \mid \overline{x}\langle z\rangle.P'']\!]_{\mathcal{N}} \rightarrow^* [\![P'\{y \leftarrow z\} \mid P'']\!]_{\mathcal{N}}$. Let $[\![x(y:\sigma).P' \mid \overline{x}\langle z\rangle.P'']\!]_{\mathcal{N}} = p[move\ x.(y:[\![\sigma]\!]).move\ p.[\![P']\!] \mid move\ x.\langle z\rangle.move\ x.[\![P'']\!]] \mid [\![\mathcal{N}]\!]$. By assumption $x \in \mathcal{N}$, $[\![\mathcal{N}]\!]$ must contain $x[\ldots]$. After reduction we get $[\![P']\!]\{y \leftarrow z\} \mid [\![P'']\!] \mid [\![\mathcal{N}]\!]$, where $[\![P']\!]\{y \leftarrow z\}$ is equivalent to $[\![P'\{y \leftarrow z\}]\!]$. That gives $[\![x(y:\sigma).P' \mid \overline{x}\langle z\rangle.P'']\!]_{\mathcal{N}} \rightarrow^* [\![P'\{y \leftarrow z\} \mid P'']\!]_{\mathcal{N}}$.

(πRStruct) Let $P \equiv_\pi P'$, $Q \equiv_\pi Q'$, $P' \rightarrow_\pi Q'$ and let $[\![P']\!]_{\mathcal{N}} \equiv [\![P]\!]_{\mathcal{N}}, [\![P]\!]_{\mathcal{N}} \rightarrow^* [\![Q]\!]_{\mathcal{N}}, [\![Q]\!]_{\mathcal{N}} \equiv [\![Q']\!]_{\mathcal{N}}$. According transitivity of structural congruence $\equiv$ we get $[\![P']\!]_{\mathcal{N}} \rightarrow^* [\![Q']\!]_{\mathcal{N}}$.

Conclusions

The main choice in designing a calculus with mobile (lightweight) processes is one of the mobility primitives for them. We have chosen to introduce, for the moment, only one primitive *move* since it is already present, though in a context of immobile locations, in well established concurrent calculus, such as Dπ [7]. Also, this primitive might be argued to naturally model the elementary instruction by which an agent moves from one location to another at the same level. A natural alternative, or a natural extension, would be a thread mobility analogous to that for ambients, i.e., capabilities to go one step up or down the tree hierarchy, by exiting or entering an ambient.

We used this approach to encode standard π-calculus which expresses the communication of named channels by our approach in a mobile ambient system. The encoding was presented as an expressiveness test of our ambient calculus with behavioral schemes [8].

References

[1] Cardelli, L., Gordon, A. D.: Mobile Ambients. Theoretical Computer Science, Vol. 240, No. 1, 2000, pp. 177-213

[2] Milner, R., Parrow, J., Walker, D.: A Calculus of Mobile Processes, Part 1 – 2. Information and Computation, Vol. 100, No. 1, 1992, pp. 1-77

[3] Levi, F., Sangiorgi, D.: Controlling Interference in Ambients. Proceedings of POPL'00, ACM Press, New York, 2000, pp. 352-364

[4] Bugliesi, M., Castagna, G.: Secure Safe Ambients. Proceedings of POPL'01, ACM Press, New York, 2001, pp. 222-235

[5] Bugliesi, M., Castagna, G., Crafa, S.: Boxed Ambients. In B. Pierce (ed.): TACS'01, LNCS 2215, Springer Verlag, 2001, pp. 38-63

[6] Fuggeta, A., Picco, G. P., Vigna, G.: Understanding Code Mobility. IEEE Transactions on Software Engineering, Vol. 24, No. 5, May 1998, pp. 342-361

[7] Hennessey, M., Riely, J.: Resource Access Control in Systems of Mobile Agents. Technical Report 2/98, Computer Science Department, University of Sussex, 1998

[8] Tomasek, M.: Expressing Dynamics of Mobile Programs. PhD thesis, Technical university of Kosice, 2004

Surface Roughness Prediction in Machining Castamide Material Using ANN

Şeref Aykut

Department of Mechanical Engineering, Faculty of Engineering Architecture
Bitlis Eren University, 13000 Bitlis, Turkey
e-mail: saykut@bitliseren.edu.tr

Abstract: Castamide, a kind of casting polyamide, is widely used in industry because of its light weight and high corrosion resistance, and because of its impact-resistant, oil-free and silent operation. The scope of its usage has been increasing. It is used in the packaging, textiles, chemicals, leather, construction and heavy machinery manufacturing sectors. Particularly, in the manufacture of machine parts like gears surface roughness of which is crucial, it has superseded many metals because it is important to be able to predict the surface roughness to get more qualified materials. The aim of this study is to predict the surface roughness of Castamide material after machining process using ANN (artificial neural network). In this study, experiments on Castamide were done in CNC milling using high speed steel and hard metal carbide tools. The cutting parameters (cutting speed, feed rate and depth of cut) were changed and the average surface roughness (Ra, μm) values were obtained. In the experiments, the effects of cutting tools with the same diameters, but with different cutting edges and tool materials on average surface roughness were also investigated. The data were used to train and test a dynamic ANN model. It is quite clear from the model results that the surface roughness predicted by the ANN model matches well with the training data as well as the test data. The developed model has managed to is to the surface roughness with correlation rate of 83.6% and minimum error rate of 0.02.

Keywords: Surface roughness; Castamide; ANN; End milling; Cutting edge

1 Introduction

End milling is one of the most fundamental and commonly encountered chip removal operations occurring in a real manufacturing environment. In this machining process, the surface finish is a key factor in evaluating and determining the quality of a part. In practice, a desired surface roughness value is usually designated, and the appropriate cutting parameters are selected to achieve the desired quality of a specified part. Typically, surface inspection is carried out through manually inspecting the machined surfaces and using surface profilometers with a contact stylus at fixed intervals. Being a post-process

operation, this procedure is both time-consuming and-labor intensive. In addition, a number of defective parts can be produced during the period of surface inspection, thus leading to additional production cost. Another disadvantage is that the accuracy of surface measurements might be significantly influenced by serious interference or vibration from the surroundings [1].

The usage of engineering plastics has increased in today's designs due to its light weight, low cost and strength. It is used in almost all fields of industry [2]. One of the most commonly used engineering plastics is polyamide. The type which is obtained by casting, using mechanical techniques and improved with specific additives, is called cast-polyamide or Castamide, to use its industrial special name. Castamide takes the place of many metals due to its being a cheap, easily-processed, lightweight, high-resistance, abrasion-resistant and quiet working engineering material. It is preferred because of being cheaper than metals such as aluminum, copper and brass.

Many studies have been carried out since the 1960s to discover the different characteristics of polyamides, during the time they have been used as an engineering material. Some of these studies are based on their friction condition [3]. Friction forces of dry Castamides which do not contain any lubricant are lower than other metals. In order to decrease this friction force even more, different lubricants are added in Castamide materials [4-7].

With the inclusion of lubricant in the Castamide materials, the operating life of machine elements such as frictional beds, shafts, slides and cams are extended. Castamide materials are processed with metal cutting. The workability of different polyamide types, the cutting force and surface roughness observations are other areas of the experiment. A lot of parameters, such as cutting types, cutting speed, depth of cut, material used, etc. can be effective on the cutting force and surface quality [2, 8].

There are various studies aimed at determining the relationship between surface roughness and cutting conditions. Wang has studied the effect of special cutting conditions for micro cutters on the formation of surface roughness by using a miniature bench and composing a mathematical model of these effects [9]. Davim has studied the difference between processing conditions of turn bench and surface roughness formation of glass-fiber-reinforced and non-reinforced PA66 polyamide materials [2]. In some studies, the processing of rough metals, such as cobalt alloy, and change in their cutting parameters are analyzed or optimized by using techniques such as the Taguchi method on the experimental results [10, 11]. The observation of surface roughness values by using optimization methods and predictions based on artificial mind techniques can be made. Ozcelik and Bayramoglu have designed a model based on the predictability of surface roughness value with statistical methods [12]. While most formulas are developed by studying the relationship between the controlled cutting conditions that are created and surface roughness, in some conditions the shape of chip waste formed

during treatment is observed as well [13]. Generally in the metal cutting processes, cutting conditions, cutting tool geometry, cutting tool type, the usage or non-usage of coolant, the rigidity of work bench used, the cutting method used and the material type used all have effects on average surface roughness. Cutting parameters, i.e. feed rate, depth of cut, cutting speed, cutting edge and the number of cutting tool, also have effects on cutting [14, 15].

Several modeling techniques of input–output and in-process parameter relationship using ANN sets offer a distribution-free alternative and have attracted the attention of manufacturing practitioners and researchers alike when they run into difficulties in building empirical models in metal cutting process control. These techniques can offer a cost effective alternative in the field of machine tool design and manufacturing approaches, and have thus received wide attention in recent years [16].

A few applications of ANN-based input-output relationship modeling for metal cutting processes are reported in literature. The literature is rich with relevant investigations on choosing the best machining parameters for low surface roughness during different machining processes. Lee et al [17] used abductive network modeling for the drilling process to predict surface roughness. Chien and Chou [18] presented an ANN approach to predict the surface roughness of AISI 304 stainless steel, the cutting forces and the tool life. Then the genetic algorithm was introduced to find the optimum cutting conditions for the maximum material removal rate under the constraints of the expected surface roughness. Risbood et al [19] utilized a neural network to predict surface roughness and dimensional deviation based on the cutting forces and vibrations in turning of rolled steel bars containing about 0.35% carbon.

Nabil and Ridha [20] developed an approach that combined the design of experiments (DOE) and the ANN methods to establish accurate models for ground surface roughness parameter prediction. Erzurumlu and Oktem [21] have developed an ANN and surface response model to predict surface roughness in milling mould parts. A statistical design consisting of 243 experiments was adopted to collect the Ra measurement data. An effort has been made to predict surface roughness in the end milling process by using an ANN model based on the design experiments of Oktem et al [22].

Topal [23] discovered the role of the step-over ratio in surface roughness prediction studies in flat-end milling operations. Machining experiments were performed under various cutting conditions by using sample specimens. The surface roughness of these specimens was measured. Two ANN structures were constructed. The Zain et al. [24] model for surface roughness in the milling process could be improved by modifying the number of layers and nodes in the hidden layers of the ANN network structure, particularly in order to predict the value of the surface roughness performance measure. As a result of the prediction, the recommended combination of cutting conditions to obtain the best surface roughness value is a high speed with a low feed rate and a radial rake angle.

The ANN model of Zain et al. [25] predicted the surface roughness performance measured in the machining process by considering the Artificial Neural Network as the essential technique for measuring surface roughness. The ANN technique predicted the value for surface roughness; a real machining experiment is referred to.

This study investigates the application of dynamic ANN to predict the surface roughness values. Feed speed, cutting speed, depth of cut and tool type are selected as the inputs of the model while the surface roughness is the output variable. The obtained results show that the used ANN model has given similar values to experimental data with sufficient accuracy.

2 The Importance of the Study

In this study, the cutting parameters of Castamide materials on a CNC vertical machining bench are changed under control, and the average surface roughness values are experimentally observed. In the experimental studies, Castamide samples and cutters in the same diameters and of different types are processed. In order to detect the average surface roughness (Ra) value, experiments were carried out by changing the cutting speed (V_c), the feed speed (f) and the depth of cut (a_p). The experiments on Castamide were done in CNC milling using high speed steel and hard metal carbide tools. In the experiments, the effects on the average surface roughness of cutting tools with the same diameters but with different cutting edges and tool materials were also investigated. The measured data were used to train and test a dynamic ANN model. It is quite clear from the model results that the surface roughness predicted by the ANN model matches well with the training data, as well as with the test data.

3 Experimental Design

In the study, a TAKSAN trademark TMC700VC CNC vertical machining centre was used. The bench is composed of a system that can be programmed in ISO format with 15 kW of power, in metrical and inch units, and can do linear and circular interpolation in three cycles. The control unit is FANUC serial O-M. The bench table moves automatically in the directions of X, Y, Z. The tools are set to CNC milling machine. Then, the part is mounted on the table of the milling machine. Later, NC codes are transmitted to the CNC and the parts are measured after the machining process. These processes are shown in the figure below. The used CNC vertical machining centre and schematic picture of experiment setting can be shown in Fig. 1.

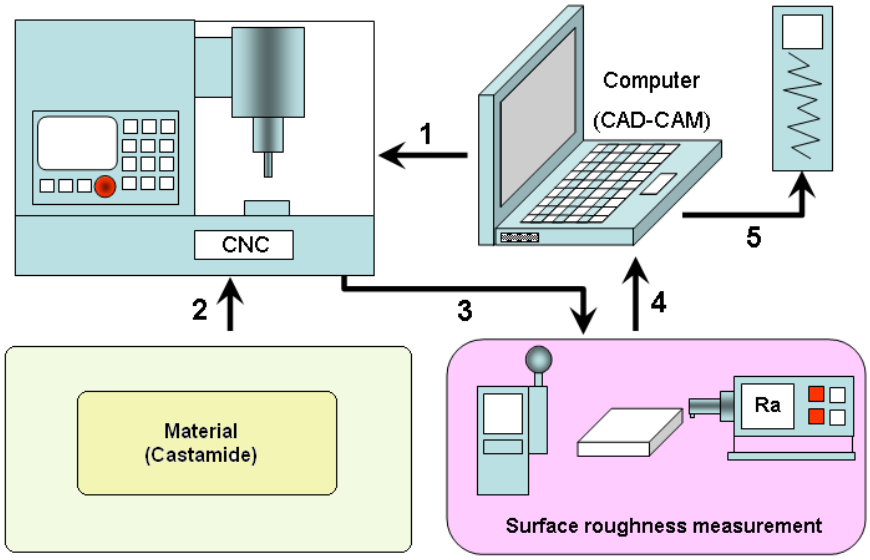


Figure 1
Schematic picture of experiment setting

For the measurement of roughness, a MarSurf PS1 portable surface roughness measurement unit was used. The measurement needle has a diameter of 2 µm and an average pressure force of 0.7 mN. The measurement scanning length was adjusted to 5.6 mm.

3.1 The Cutting Tool and Conditions

3.1.1 The Cutting Tool

In the cutting process, 14 mm end milling cutter tools made of high speed steel (HSS) and hard metal carbide were used. An HSS cutting tool, a cutting tool with four cutter edges in DIN standards, can easily do machining in profile processing, and in the process of high strength steel and other hard processed steels. It has four cutter edges and a 30° helix angle. Hard metal carbide cutter tools are produced in the DIN 81800 standard, with four cutter edges, a 30° helix angle, WC (Tungsten Carbide) rate of 87.7%, a cobalt rate of 12.3%, a TRS (Transverse rupture strength) MPa, a 92.5 HRA hardness, a 0.5 µm grain size and high corrosion and effect ability. The cutting tools are bound to a spindle with the help of pincers. In the milling process, the average surface roughness is obtained with the help of cutting parameters: cutting speed, feed speed, depth of cut, type of cutting equipment and Castamide material.

3.1.2 The Machining Conditions

The experimental work was done on a CNC milling machine. The surface roughness was investigated by the effect of the cutting rate, the feed rate and the cutting depth. Cutting speeds were 100, 120, 140 m/min; the feed rates were 75, 100, 125 mm/min; and the cutting depths were selected as 1.0; 1.5; 2.0 mm. Prefeasibility parameter values were selected as recommended for polyamide material [26, 27]. A sample taken out of a dry container was brought to the experiment in desired process parameters after serially connected to the bench. The cutting conditions can be seen in Table 1.

Table 1
Cutting conditions for end milling

Terms of cutting condition	Unit	Value
Cutting speed	(m/min)	100, 120 and 140
Feed rate	(mm/min)	75, 100 and 125
Depth of cut	(mm)	1,0; 1,5 and 2,0
Dimension of cutting tool	(mm, ¬)	14
Tool and the number of cutting edge	-----	CARBIDE (4), HSS (4) and HSS (6)
Material	-----	Castamide
Coolant	-----	Dry

3.2 Workpiece Material

The Castamide material in 46 mm plates used in the experiments was supplied by Polimersan. It is called POLIKES® (PA6 G) in the firms product catalogue. The plates were cut in the dimensions of 112*82*46 mm. The mechanical and physical features of the PA6G used in the expiraments are shown in Table 2.

Table 2
Mechanical and physical features of tested material

Properties of Castamide	Unit	Value
Specific gravity	gr/cm^3	1,15
Thermal elongation	$1/K*10^5$	8-9
Pulling resistance	N/mm^2	55-85
Breaking resistance	N/mm^2	88-90
Breaking elongation	%	10-40
Elastic module	N/mm^2	3900-4200
Water absorption	%	6-7
Resistance as per volume	Ω*cm	$>10^{15}$
Resistance as per surface	Ω	$>10^{12}$
Dielectric resistance	KV/mm	80-100
Rockwell	HRC	M88
Ball notch 358/30	N/Mm^2	110-160

4 The Artificial Neural Network and the Prediction of Surface Roughness

There are various simple surface roughness amplitude parameters used in industries, such as roughness average (Ra), root-mean-square (rms) roughness (Rq), and maximum peak-to-valley roughness (Ry or Rmax), etc. [28]. The parameter Ra was used in this study. The average roughness (Ra) is the area between the roughness profile and its mean line, or the integral of the absolute value of the roughness profile height over the evaluation length (Fig. 2). Therefore, the Ra is specified by the following equation;

$$Ra = \frac{1}{L}\int_0^L |Y(x)|dx \tag{1}$$

where Ra is the arithmetic average deviation from the mean line, L is the sampling length and Y the ordinate of the profile curve. The average surface roughness Ra was measured within the sampling length of 5.6 mm.

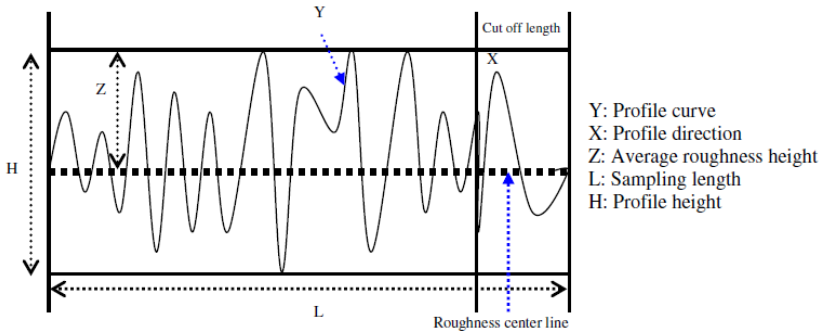


Figure 2
Surface roughness profile [29]

The obtained database (Table 3) by the surface roughness of milled specimens was used to train and test the neural network model. A four-layer network having one input layer, two hidden layers and one output layer were formed for the present study. The input layer had 6 neurons and the output layer was determined by the output parameter (surface roughness, Ra). The number of neurons in the hidden layer is optimized to achieve optimum output accuracy. The number of hidden neurons depended on both input vector size and the number of input classifications (first hidden layer had 2 neurons and second hidden layer had 2 neurons). Too few neurons could lead to under-fitting whereas too many neurons could result in over-fitting.

Fig. 3 shows the schematic layout of the neurons within the network, each arrow representing a link between neurons, or a synapse. Each of these synapses has a weight attached to it which governs the output of the neuron. By adjusting the

values of these synaptic weights, the outputs of the neural network can be altered. Training involves adjusting the synaptic weights within the network until the mean of absolute error between the predicted and experimental output values has been minimized.

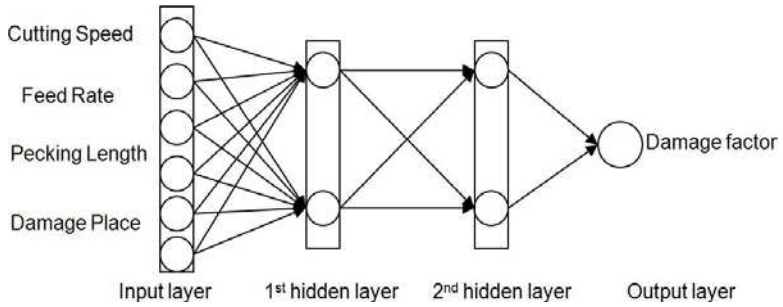


Figure 3
ANN model

Table 3
Sample training and testing data set

a_p	V_c	f	Tool	Ra	Train	a_p	V_c	f	Tool	Ra	Train
2	100	100	CARBIDE–4	0,588	Training	2	125	200	CARBIDE–4	0,644	Testing
1	100	100	HSS–4	0,811	Training	2	125	100	HSS–4	0,945	Testing
2	125	100	CARBIDE–4	0,915	Training	1,5	75	150	HSS–4	1,293	Testing
1,5	125	150	CARBIDE–4	1,003	Training	1,5	100	200	CARBIDE–4	1,369	Testing
1	100	200	HSS–4	1,05	Training	1,5	125	100	CARBIDE–4	1,374	Testing
1	75	100	HSS–6	1,058	Training	1	75	200	HSS–4	1,449	Testing
1	75	150	HSS–4	1,204	Training	1	75	150	CARBIDE–4	1,473	Testing
2	100	200	CARBIDE–4	1,239	Training	1,5	75	150	CARBIDE–4	1,512	Testing
1,5	125	200	CARBIDE–4	1,25	Training	1	125	150	HSS–4	1,617	Testing
1,5	125	200	HSS–4	1,256	Training	2	75	150	CARBIDE–4	1,619	Testing
1,5	100	150	HSS–6	1,264	Training	1	75	200	CARBIDE–4	1,646	Testing
1,5	100	100	HSS–4	1,277	Training	1,5	75	100	HSS–4	1,836	Testing
1	100	200	CARBIDE–4	1,307	Training	1	125	100	CARBIDE–4	1,874	Testing
1,5	75	100	HSS–6	1,313	Training	2	100	150	CARBIDE–4	2,056	Testing
2	100	200	HSS–4	1,313	Training	2	75	200	HSS–4	2,086	Testing

For the present study, the input variables were discrete and set values while the values of output variables were in a range. Since three different tools were used, the names of the tools were accepted as the value of the variable: hard metal CARBIDE-4, HSS-4 and HSS-6. The cutting speed and feed rate have discrete values (75, 100, 125 m/min and 100, 150, 200 mm/min). 81 data sets were used. 56 data sets (approximately 70%) were used to train the model and 25 data sets were used to test the model.

5 Results and Discussion

For each input pattern, the predicted value of surface roughness was compared with the respective experimental Ra (μm) value. It was found that the predicted average surface roughness were close to the experimental values. Fig. 4 shows the graph of the predicted and the experimental values of the surface roughness for training patterns. The maximum error was 1.59 and the minimum error was 0.015. The absolute error was found to be less than 20% for most cases.

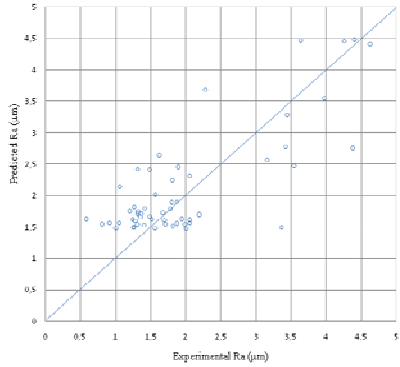


Figure 4
Predicted Ra values for training data sets

The network was then tested with the 25 test data points which were not used for the training purpose. The comparison of the predicted and the experimental values of surface roughness for the test data sets are presented in Fig. 5. From the test results, it can be observed that the predicted values are close and follow almost the same trend as the experimental values. While the maximum error was 1.18, the minimum error was 0.024. The absolute error was found to be less than 20% for most cases. In conclusion, our model has managed to cover experimental data with a 83.6% correlation factor.

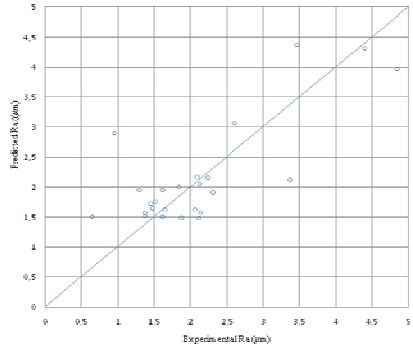


Figure 5
Predicted Ra values for testing data sets

It is quite obvious from the results of the ANN predictive model that the predicted accuracy was good and the predicted results matched well with the experimental values. The prediction of the surface roughness of Castamide is important for applications. It was earlier established that the cutting parameters (feed rate, depth of cut, cutting speed, cutting edge number of cutting tool) have an impact on cutting. The obtained results hereby are in compliance with the earlier findings. The difference may be due to the type of investigated materials. Obtained surface roughness quality varies between 0.5 and 5. This change makes it hard to find a good correlation among cutting parameters, tool type and surface roughness. In addition, considering that tool type also affected the results, there is a high correlation for CARBIDE-4 and HSS-6 datasets. 83.6% correlation becomes 90-99% when only one tool type is considered. As the correlation between the machining and the surface roughness is strongly dependent on the material being machined, there is an imminent need to develop a generic predictive platform to predict surface roughness. The present investigation is a step in this regard. The proposed model is helpful in the judicious selection of the various machining parameters to minimize surface roughness.

Conclusions

An ANN-based predictive model for surface roughness in the process of milling of Castamide was developed. The model served as a tool to calculate the surface roughness for various sets of input parameters. Prior knowledge of the potential roughness could help engineers to select the optimum process parameters as well as the tool type for smooth surfaces on Castamide.

- The cutting speed, feed rate, depth of cut, and two types of tools were used as the input parameters, and the surface roughness was used as the output.
- The predicted values using ANN were found to be in close agreement with the training as well as the testing data. For both training and testing data, the obtained error was less than 20 percent in most cases.
- A more generic ANN model incorporating input variables at a greater number of levels can be developed and validated. The experimental data available in literature can be incorporated in the model to increase its applicability. An ANN model for predicting surface roughness can also be used in conjunction with an optimization tool to find the best drilling parameters, to reduce the damage prior to milling and at the start of any milling process of Castamide.

References

[1] Tsai, YH., Chen, JC., Lou, SJ., An In-Process Surface Recognition System Based on Neural Networks in End Milling Cutting Operations, International Journal of Machine Tools & Manufacture 39 (1999) 583-605

[2] Davim, J. P., Silva, L. R., Festas, A., Abrão, A. M., Machinability Study on Precision Turning of PA66 Polyamide with and without Glass Fiber Reinforcing, Materials and Design 30 (2009) 228-234

[3] Adams, N., Friction and Deformation of Nylons, J. Appl. Polym. Sci. 7 (1963) 2075-2103

[4] Samyna, P., Baets, P., Schoukens, G., Van Driessche, I., Friction, Wear and Transfer of Pure and Internally Lubricated Cast Polyamides at Various Testing Scales, Wear 262 (2007) 1433-1449

[5] Palabiyik, M., Bahadur, S., Mechanical and Tribological Properties of Polyamide 6 and High Density Polyethylene Polyblends with and without Compatibilizer, Wear 246 (2000) 149-158

[6] Samyn, P., Tuzolana, T. M., Effect of Test Scale on the Friction Properties of Pure and Internal-lubricated Cast Polyamides at Running-In, Polymer Testing 26 (2007) 660-675

[7] Liu, C. Z., Wu, J. Q., Li, J. Q., Ren, L. Q., Tong, J., Arnell, R. D., Tribological Behaviours of PA/UHMWPE Blend under Dry and Lubricated Condition, Wear 260 (2006) 109-115

[8] Mata, F., Reis, P., Davim, J. P., Physical Cutting Model of Polyamide Composites (PA66 GF30), Mater Sci Forum (2006) 514-516:643-7

[9] Wang, W., Kweon, S. H., Yang, S. H., A Study on Roughness of the Micro-End-milled Surface Produced by a Miniatured Machine Tool, The International Journal of Advanced Manufacturing Technology (2005);162-13:702-708

[10] Bağcı, E., Aykut, Ş., A Study of Taguchi Optimization Method for Identifying Optimum Surface Roughness in CNC Face Milling of Cobalt-based Alloy (stellite 6), Int J Adv Manuf Technol (2006) 29:940-947

[11] Aykut, Ş., Demetgül, M., Tansel, İ. N., Selection of Optimum Cutting Condition of Cobalt-based Super Alloy with GONN The International Journal of Advanced Manufacturing Technology (2010) 46:957-967

[12] Ozcelik, B., Bayramoglu, M., The Statistical Modeling of Surface Roughness in High-Speed Flat End Milling, Int J Mach Tools Manuf (2006) 46:1395-1402

[13] Kishawy, H. A., Dumitrescu, M., Ng, E. G., Elbestawi, M. A., Effect of Coolant Strategy on Tool Performance, Chip Morphology and Surface Quality during High-Speed Machining of A356 Aluminum Alloy, Int J Mach Tools Manuf (2005) 45:219-227

[14] Ertakin, Y. M., Kwon, Y., Tseng, T. L., Identification of Common Sensory Features for the Control of CNC Milling Operations under Varying Cutting Conditions, Int J Mach Tools Manuf (2003) 43.897-904

[15] Dabade, U. A., Joshi, S. S., Ramakrishnan, N., Analysis of Surface Roughness and Cross-Sectional Area while Machining with Self-propelled Round Inserts Milling Cutter, J Mater Process Technol (2003) 132:305-312

[16] Mukherjee, I. Ray, P. K., A Review of Optimization Techniques in Metal Cutting Processes, Computers & Industrial Engineering 50 (2006) 15-34

[17] Lee, B. Y., Liu, H. S., Tarng, Y. S., 1998. Modeling and Optimization of Drilling Process, J. Mater. Process. Technol. 74, 149-157

[18] Chien, W. T., Chou, C. Y., 2001. The Predictive Model for Machinability of 304 Stainless Steel. J. Mater. Process. Technol. 118, 442-447

[19] Risbood, K. A., Dixit, U. S., Sahasrabudle, A. D., 2003. Prediction of Surface Roughness and Dimensional Deviations by Measuring Cutting Forces and Vibrations in Turning Process, J. Mater. Process. Technol. 132, 203-214

[20] Nabil, B. F., Ridha, A., Ground Surface Roughness Prediction Based upon Experimental Design and Neural Network Models. Int. J. Adv. Manuf. Technol., 2006 31, 24-36

[21] Erzurumlu, T., Oktem, H., Comparison of Response Surface Model with Neural Network in Determining the Surface Quality of Moulded Parts, Materials & Design 200728, 459-465

[22] Oktem, H., Erzurumlu, T., Erzincanlı, F. Prediction of Minimum Surface Roughness in end Milling Mold Parts Using Neural Network and Genetic Algorithm, Materials & Design Volume 27, Issue 9, 2006, pp. 735-744

[23] Topal, EY, The Role of Step over Ratio in Predictionof Surface Roughnessin Flat Endmilling International Journal of Mechanical Sciences 51 (2009) 782-789

[24] Zain, AM., Haron, H., Sharif, S., Prediction of Surface Roughness in the End Milling Machining Using Artificial Neural Network, Expert Systems with Applications 37 (2010) 1755-1768

[25] Zain, AM., Haron, H., Sharif, S., Prediction of Surface Roughness in the End Milling Machining Using Artificial Neural Network, Expert Systems with Applications 37 (2010) 1755-1768

[26] Atakök, G., Kurt, M., Measurement and Evaluation of Force, Vibration, Thermal Changes and Roughness of Polyamide Materials' Machining at CNC Machines, 10. Denizli Malzeme Sempozyumu ve Sergisi, 2004, 895-902

[27] Xiao, K. Q., Zhang, L. C., The Role of Viscous Deformation in The Machining of Polymers, International Journal of Mechanical Sciences, 2002, pp. 44-52

[28] Benardos, P. G., & Vosnaikos, G. C. (2003). Predicting Surface Roughness in Machining: a Review. International Journal of Machine Tools and Manufacture, 43, 833-844

[29] Yang, J. L., & Chen, J. C. (2001) A Systematic Approach for Identifying Optimum Surface Roughness Performance in End-Milling Operation, Journal of Industrial Technology, 45, 110-120

The Treatment of the Surfaces of Mg-Al-Zn-Mn and Ti-Al-Zr-Nb Alloys by Shot Peening

Peter Mrva

Department of Aviation Engineering
Faculty of Aeronautics
Technical University of Košice
Rampova 7, 041 21 Košice, Slovakia
e-mail: Peter.Mrva@tuke.sk

Daniel Kottfer

Department of Technologies and Materials
Faculty of Mechanical Engineering
Technical University of Košice
Mäsiarska 74, 040 01 Košice, Slovakia
e-mail: Daniel.Kottfer@tuke.sk

Abstract: The influence of shot peening on the surfaces of Mg-Al-Zn-Mn and Ti-Al-Zr-Nb alloys is described by the authors. There are the estimates focusing on roughness of the shot peened surfaces and on a size of selected shot peened material - the corundum. Based on the results of measurements, the evaluation was oriented to the curve of roughness, the functionality of surface roughness R_a and the necessary quantity of the shot peening material q_{nR} of estimated material depending on the grain size d_z.

Keywords: shot peening; functional surface; surface roughness

1 Introduction

Surface strengthening by shot peening can make use of shot peening to increase the fatigue resistance of engineering components. Some alloys (on the basis of the magnesium) are inclined to the initiation and propagation of fatigue cracks. Defects like grains and systemless structures begin and accelerate these cracks – the stress concentration.

With high frequency cyclic loading, for a smaller number of cycles to failure the locations of cracks start appearing on the surface [1, 2]. For a higher number of cycles to failure, the location of cracks begins to appear in the thermal area of the experimental sample [3]. Some alloys after thermal treatment obtain structures with compound of balanced polyedric grains with concrete phases. It involves mechanical properties growth and resistance to fatigue as well [4, 13] and resistance to oxidation by high temperature [14].

Nowadays, there are enough developed imaginations of deposition process of thermal spraying coatings on the surfaces of magnesium [5] and steel engineering accessories [14].

There are a lot of methods for the research of the microgeometry of shot peened surfaces [15]. However, the following experimental method is adequate for the determination of the necessary quantity of abrasive material q_{nR}.

The determination of the necessary quantity of abrasive material q_{nR} is important. If the necessary quantity is less than q_{nR} (Fig. 2), the coverage of the shot peened surface and the adhesion of the deposited thermal layer will be inadequate. If the necessary quantity is equal q_{nR} (Fig. 2), the surface of the shot peened surface and the adhesion of deposited thermal layer will be appropriate. If the necessary quantity is more than q_{nR} (Fig. 2), the roughness, R_a, of the shot peened surface and the adhesion of deposited thermal layer will be sufficient as well. There are the cracks on the surface which create the stress concentration. This causes the decrease of limit of the material as well as the lifetime shrinkage of the machine [17].

The lifetime of coatings depends on the ideal adhesion of functional coatings [6, 7, 8]. Adhesion is conditioned by ideal pretreatment of the functional surface. Shot peening is one of the most frequently used technologies of mechanical pretreatment of surfaces under thermal spraying coatings. The surface is cleaned and the necessary microgeometry of the surface is created by shot peening.

It is known that the activation energy of surfaces made from the deformation of surface layers during shot peening definitely affects the coating adhesion [9]. The value of the activation energy is reduced exponentially by the material surface and background influence. Therefore the coating must be deposited till 1-3 hours after shot peening. The coating adhesion to the basic material can be evaluated by the mechanism of adhesion. The mechanical adhesion of coating on the surface relief of sample occupies between 50-80%. The Van der Walls forces make about 5% and power of chemical compounds make up 15-45%. In comparison with surface pretreatment by cauterization, shot peening is more convenient than cauterization. The technology of surface shot peening can be used for surface strengthening and the formation of good surface roughness. The experiment was focused on the influence of shot peening on the presented Mg alloy surface. The influence of the particular parameter on the formation of good, strong coating and on the substance was investigated.

The utilization of thermal spraying is an up-to-date technology for the renovation of aviation components made from light alloys on the base Al, Mg, Ti. The experiment was focused on the research impact of the sorts and dimensions of the shot peening material grain and the shot peening parameters on the necessary surface quality under thermal spraying coatings.

2 Experiment Methodology

This experimental research was directed towards the determination of the curve of roughness, the functionality of the surface roughness, R_a, and the necessary quantity of the shot peening material, q_{nR} , of the evaluated material depending on the grain size, d_z.

2.1 Evaluation of Shot Peened Surface Roughness R_a

Shot peening is a specific form of surface pretreatment of components. The character of the shot peened surface is typical for this technology. In the shot peening process the component surface is consequently roughened. The roughness is evaluated by a touch-profimeter. These appliances have bigger scale of measured parameters of surface roughness, for example R_a=30 μm. The average roughness, R_a, was selected for surface evaluation. A Profimeter HOMMEL Tester T2000 was used for the measured values and for the profile formation of the shot peened surfaces.

To obtain relevant results, the following conditions have been used [8, 9]:

- measured length L= 6,3 mm,
- terminal undulation (cut-off) l = 1,25 mm
- number of measurements n = 10.

Profiles were taken out following next conditions:

- measured length L= 6,3 mm,
- terminal undulation pinch l=∞.

The mean arithmetic value as a statistic value has been calculated from the measured values of roughness. Each surface was evaluated by two profilegrams.

2.2 Experimental Samples Preparation

Mg-Al-Zn-Mn and Ti-Al-Zr-Nb alloy samples [10] were used in the experiment. Sample dimensions were chosen in such a way as to eliminate the unwanted influence of the shot peening device on the shot peening process (e.g. the heterogeneous consistency of the grain touches in the entire field of shot peening beam). The samples' dimensions enabled us to realize the adhesion test after thermal spraying coatings on surfaces of the shot peened samples. The samples were made by turning into the form of a roll with a diameter of 30 mm and a thickness of 3 mm (Fig. 1). The functional surface of the samples before shot peening had the roughness of R_a=0,6 μm.

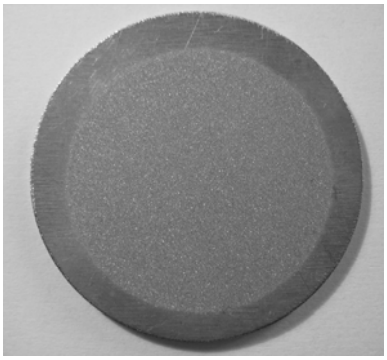


Figure 1
The sample of the Mg-Al-Zn-Mn alloy after shot peening

2.3 The Material Used for Shot Peening

There are not uniform selected criteria for shot peening material by now. Selection of the shot peening material was based on the basic material properties.

For the surfaces' shot peening process, corundum granular was used (STN EN 22 40 12). This material is produced in all granularities and the shot peening material is a polydisperse. To explain the influence of the grain size on the surface roughness, the shot peening material was selected via a wire screen with a specific grain diameter. The chosen gain diameter was according to STN 15 3105.

2.4 Shot Peening Process

Laboratory equipment [7] was used for the shot peening of the evaluated sample surfaces. The influence of the grain size of the shot peening material on the roughness of the surface of Mg-Al-Zn-Mn and Ti-Al-Zr-Nb alloys was tested during the experiment at speed v=78.1 ms^{-1}. The angle of incidence of the shot

peening material grain on the sample surface was α=75°. The sample distance to the shot peening wheel was L=200 mm. According to Matling and Steaffens, one cannot prevent hobbing of the shot peening material onto the shot peened surface [11]. It causes galvanic cells formation [12]. Impresses (by hobbing) can be made by shot peening of the surface with harder materials, too. Consequently, corundum with a minimum grain speed of v_1=78.1 ms^{-1} was used for shot peening of the Mg-Al-Zn-Mn and Ti-Al-Zr-Nb alloys. Corundum with medium grain diameters 0.1; 0.2; 0.315; 0.4; 0.5; 0.63; 0.71; 0.8; 1.0; 1.25 mm was used for the Mg-Al-Zn-Mn alloy. Corundum with medium grain diameters 0.36; 0.56; 0.71; 0.9; and 1.12 mm was used for Ti-Al-Zr-Nb alloy.

3 Methods for Determining the Roughening Curve of the Shot Peened Surface

The experiment was aimed at determining the necessary quantity of abrasive material, q_{nR}, which is needed to cover the shot peened surface completely. The initial ration was determined from the characteristic roughening curves. The roughening curve technique specifies the functional dependancy of the shot peened surface roughness on the quantity of the abrasive material which shapes the measured surface (Fig. 2). The area from the first to the second section is important for the determination of the covering grade of the shot peened surface by roughening curves (Fig. 2) [8].

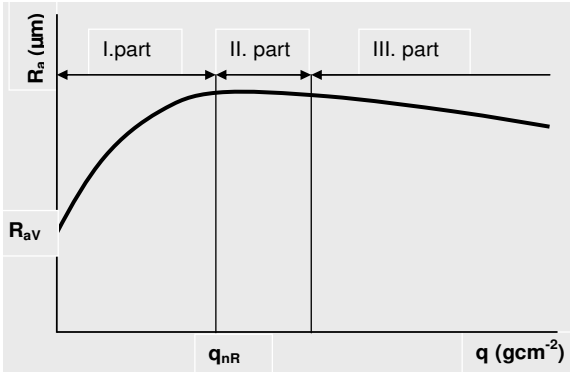


Figure 2
The roughening curve

3.1 Determining the Roughening Curves of the Mg-Al-Zn-Mn Alloy

Samples from the Mg-Al-Zn-Mn alloy were gradually shot peened with a ration of 1000 g of corundum, with covering grade q=0.5 gcm^{-2}, and with the number of rations 10. Next samples were shot peened by 1000 g of corundum with the fraction diameter d_z=0.1; 0.2; 0.315; 0.4; 0.5; 0.63; 0.71; 0.8; 1.0 and 1.25 mm. After each shot peening the roughness R_a was measured. Each R_a is the arithmetic average of 10 measured tests. After each shot peening the surface was evaluated by means of optical microscope. After each shot peening another material was selected. The number of rations in the experiment was selected so that the roughening curve could capture the first and second part, and partially the third one (Fig. 2). The roughening curves are in Fig. 3.

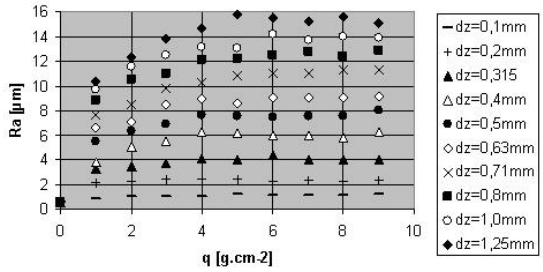


Figure 3
Roughening curves of the Mg-Al-Zn-Mn alloy with corundum fractions of the diameters d_z

3.2 Determining the Roughening Curves of the Ti-Al-Zr-Nb Alloy

Samples from the Ti-Al-Zr-Nb alloy were shot peened gradually with a ration of 2000 g of corundum, with covering grade q=1.0 gcm^{-2}, and the number of rations 10. Next samples were shot peened with six rations of 2000 g, fraction diameter d_z=0.36 mm; by seven rations of 2000 g, fraction diameter d_z=0.56 mm; by three rations of 2000 g, fraction diameter d_z=0.71 mm; by three rations of 4000 g, fraction diameter d_z=0.9 mm and by six rations of 4000 g, fraction diameter d_z=1.12 mm. After each shot peening the roughness R_a was measured. Each of R_a is the arithmetic average of 10 measured accounts. After each shot peening the surface was evaluated by means of optical microscope. After each shot peening another material was selected. The number of rations in the experiment was selected so that the roughening curve could capture the first and second part and partially the third one (Fig. 2). The roughening curves are in Fig. 4.

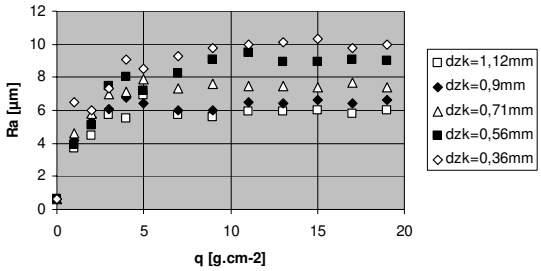


Figure 4

Roughening curves of the Ti-Al-Zr-Nb alloy with corundum fractions of the diameters d_{zK}

3.3 Determining the Necessary Quantity of Abrasive Material

To cover the shot peened surface with touches by shot peening, it is necessary to know the microgeometry of the shot peened surface. The grade of covering of this surface is to be n=1. Now, the necessary quantity of abrasive material for surface covering is on the roughening curve q_{nR}. It is expected that the linear and planar covering grade is 1 (Fig. 2).

The dependency of the necessary quantity of the abrasive material on the grain dimension can be solved from the roughening curves (Figs. 3, 4). The correlation of grain diameter d_z on the necessary quantity of shot peening material q_{nR} for Mg-Al-Zn-Mn alloy and Ti-Al-Zr-Nb alloy was determined (Figs. 5, 6). This way is valid for shot peening.

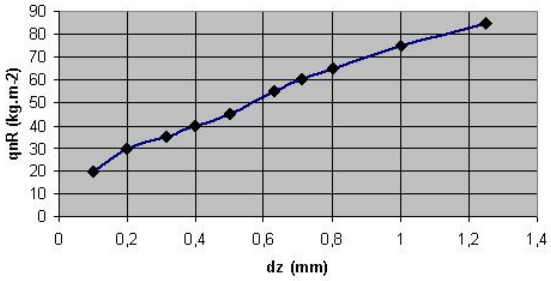


Figure 5

The correlation of grain diameter d_z on the necessary quantity of shot peening material q_{nR} of the Mg-Al-Zn-Mn alloy

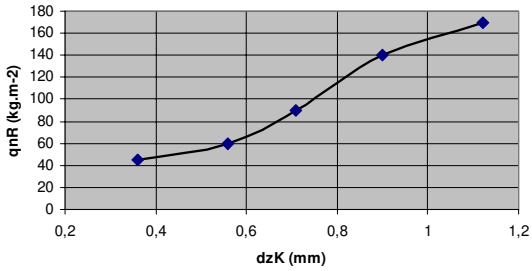


Figure 6

The correlation of grain diameter d_{zK} on the quantity of shot peening material q_{nR} of the Ti-Al-Zr-Nb alloy

3.4 The Influence of Grain Diameter of the Shot Peened Material on the Roughness R_a

The roughness value R_a of the shot peened surface of Mg-Al-Zn-Mn and Ti-Al-Zr-Nb alloys was determined as the arithmetic average from 10 measurements. The final values for particular grain diameters of the shot peening material by speed v_1=78.1 ms^{-1} are in diagram (Figs. 7, 8). From function dependency it is obvious that if the grain diameter grows, the roughness of the shot peened material grows, too. It is related directly to the touch size after the grains of the shot peening material fall on the surface.

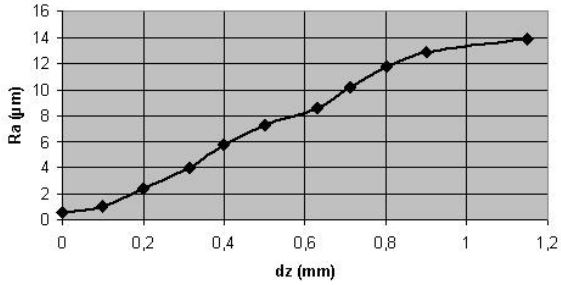


Figure 7

The correlation of grain diameter d_z on the surface roughness R_a of the Mg-Al-Zn-Mn alloy

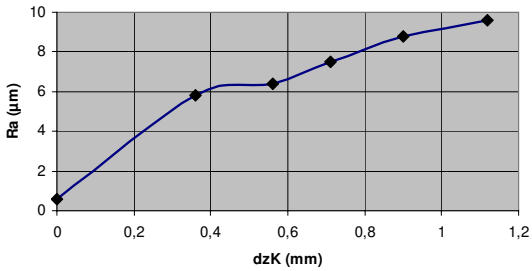


Figure 8
The correlation of grain diameter d_{zK} on the surface roughness R_a of the Ti-Al-Zr-Nb alloy

4 Discussion of the Experimental Results

The surface of the experiment samples was gradually roughened in the shot peening process. As a consequence, roughness R_a is different in measured values of tested grain diameter d_{zK} (Figs. 7, 8). The roughness R_a of the Mg-Al-Zn-Mn alloy for grain diameter d_z is 14 µm. The roughness R_a of the Ti-Al-Zr-Nb alloy for grain diameter d_{zK} is 9.6 µm. These values are material properties of the experimental samples.

The quantity of abrasive material q_{nR} for tested dimension of the grain d_z and for the Mg-Al-Zn-Mn alloy was determined from interval 20 to 85 kgm^{-2}. The quantity of abrasive material q_{nR} for tested grain diameter d_{zK} and for Ti-Al-Zr-Nb alloy was determined from interval 42 to 170 kgm^{-2}. These values are material properties of the experimental samples, as well.

The necessary quantity q of tested grain diameters for Mg-Al-Zn-Mn alloy is from the interval 4 to 6 gcm^{-2} (Fig. 3) and the necessary quantity of tested grain diameters in the case of Ti-Al-Zr-Nb alloy is from the interval 10 to 15 gcm^{-2} (Fig. 4). The value of the necessary quantity q_{nR} can be exactly assigned for the material (alloy) and for the chosen dimension of the grain (Figs. 3, 4).

Conclusions

According to the analyses, study and realized experiments, the following conclusions can be made:

- The determination of the necessary quantity of abrasive material q_{nR} is important. Too much abrasive material can cause crack formation on the shot peened surface. Subsequently, there is higher value of stress concentration (for necessary quantity > q_{nR} - Fig. 2, with constant speed). It causes a decrease in the fatigue limit of the material [17] as well as a decrease in the lifetime of the machine.

- The measured accounts were the basis for the drawing of the roughening curves – the change depends on the roughness R_a as well as on the quantity of the shot peening material q_{nR} which falls on the measured surface.
- The roughness of the shot peened surface is affected by the grain size of the shot peening material. If the diameter grows, the roughness of the shot peened surface grows, too.
- The necessary quantity of the shot peening material q_{nR} (corundum) influences the grain diameter d_z. The value of the necessary quantity q_{nR} can be exactly assigned for a material (tested alloy) and for a chosen dimension of the grain.
- The roughness R_a grows with the increasing quantity of the abrasive material.
- The necessary quantity of the shot peening material q_{nR} for the whole covering of shot peened surface can be assigned from the roughening curves [16]. The necessary quantity for the Mg-Al-Zn-Mn alloy is from the interval 4 to 6 gcm^{-2} and for the Ti-Al-Zr-Nb alloy it is from the interval 10 to 15 gcm^{-2} (Figs. 3, 4).
- The roughness R_a is constant for reached covering grade q. It is valid for q=5 $g.cm^{-2}$ of diameters of shot peening material – for corundum.
- Corundum is the most suitable substance for the shot peened Mg-Al-Zn-Mn alloy and Ti-Al-Zr-Nb alloy. Using corundum, we achieved cleanness of the shot peened surface and less necessary quantity of abrasive material q_{nR} with higher roughness.

Acknowledgement

This work was financially supported by the Slovak Grant Agency under the grant VEGA No. 1/0279/11.

References

[1] Ján Piľa, Aurel Sloboda, Aurel Sloboda jr.: Some Opportunities of Diagnostic Parameters Utilization in the Aircraft Proactive Maintenance Management, in Proceedings of 7th International Scientific Conference on Theory and Aplication of Techniques of Technical Diagnostic 2004, Košice, October 13-14, 2004, pp. 76-82

[2] Dušan Neštrák, Ján Piľa: Helicopter Aerodynamics, Structures and Systems (in slovak): Textbook: Akademické nakladatelství CERM (2006) p. 454

[3] Mariana Kuffová, Vladimír Bella, S. Wolny: Fatigue Resistance of Mg–Alloy AZ 63HP under High-Frequency Cyclic Loading. In Mechanika Kwartalnik Akademii Górniczo-Hutniczej imienia Stanislawa Stašika w Krakowie, Poland, 23, 3, 2004

[4] Mariana Kuffová: Microstructure of Magnesium Alloys after Heat Treatment. in Proceedings of Wear and Dependability, Diagnostic 2006, Brno, Czech Republik, p. 139

[5] Peter Mrva, Daniel Kottfer: Influence of Thermal Spray Coatings on the Thermal Endurance of Magnesium Alloy ML-5, Acta Polytechnica Hungarica, Vol. 6, No. 2, 2009, pp. 71-75

[6] Peter Mrva, Stanislav Kaliský: Mathematical Model Evaluation of the Shot Peening Technological Process of Ti Alloys onto Adhesion of Plasma Sprayed Coatings with Thermal Insulating Properties (in slovak), in Proceedings of 4th International Conference on Corosion and Corrosion Protection of Matrials, Trenčín, April 12-13, 2000, pp. 124-128

[7] Dušan Kniewald, P. Pivoda: Mechanical Pretreatment of the Surface of Parabolic Springs under Protective Al and Zn Thermal Sprayed Coatings (in slovak), In: Zborník vedeckých prác VŠT v Košiciach, 1978, pp. 309-318

[8] Mrva Peter: Research of the Influence of the Surface Pretreatment of Titanium Alloys (in czech), Research report VU 070 Brno, 1986

[9] Kniewald Dušan, Šefara Michal: Vorbehandlung der Metalloberfläche durch Strahlen als Vorbereirung fűr Schutzűberzűge aus Pulverkunststoffen, Zborník vedeckých prác VŠT v Košiciach, 1980, pp. 263-274

[10] M. Štěpánek, Dušan Neštrák: Handbook of the Aviation Technician (in slovak), Translation, Naše vojsko, 1989, Praha

[11] A. Matting, H. D. Steafens: Metalurgy, No. 10, 1968, pp. 13-17 (in german)

[12] Vladimír Sedláček: Metall Surfaces and Coatings (in czech), ČVUT Praha 1992, ISBN 1335-2393

[13] Karel Slámečka, Jaroslav Pokluda, Marta Kianicová, Štěpán Major, Ivan Dvořák: Quantitative Fractography of Fish-eye Crack Formation under Bending–Torsion Fatigue, International Journal of Fatigue, Volume 32, Issue 6, June 2010, pp. 921-928

[14] Dongming Zhu, Robert A. Miller: Development of Advanced Low Conductivity Thermal Barrier Coatings, Int. J. Appl. Ceram. Technol., 1 [1] 2004, pp. 86-94

[15] Omar Hatamleh, James Smith, Donald Cohen, Robert Bradley: Surface Roughness and Friction Coefficient in Peened Friction Stir Welded 2195 Aluminum Alloy, Applied Surface Science, Volume 255, Issue 16, 30 May 2009, pp. 7414-7426

[16] Daniel Kottfer, Peter Mrva: Mechanical Pretreatment of Surface of Aluminum Alloy D16-T by Shot Peening, Acta Polytechnica Hungarica, Vol. 6, No. 4, 2009, pp. 75-82

[17] Peter Mrva, Daniel Kottfer: Effect of Shot Peening and NiAl Coating on Fatigue Limit of Mg-Al-Zn-Mn Alloy, Archives of Metallurgy and Materials, Vol. 56, No. 3, 2011 (in print)

A Legal Business Information System: Implementation Process Context

Alok Mishra, Deepti Mishra

Department of Computer Engineering, Atilim University
Incek 06836, Ankara, Turkey
alok@atilim.edu.tr, deepti@atilim.edu.tr

Abstract: Information Technology (IT) is fast becoming useful in implementing time, case, manpower and cost management strategies within judicial services. The legal system environment has adopted IT not just to save costs and time but also to give organizations a competitive edge and to ensure security as well. The Legal Business Information System is a fully operational and integrated system for a legal department. The mission of the department is to provide innovative and quality services in insolvency and trustee matters. Very few legal business information system implementations are documented in literature. Therefore this paper will facilitate understanding of system implementation in this sector.

Keywords: Legal business information system; Information Technology; Implementation; Process

1 Introduction

The significance of the information era is growing on legal practitioners and, to some extent, successful legal practice depends on it [1], [2]. This is a growing tendency and is compounded by the fast-developing nature of information technology, which stresses the need for sound legal information and knowledge management practices [3]. Court decisions are legal documents with a high potential for later retrieval; many interested parties can benefit from consulting for instance older decisions on cases similar to the ones they are involved in [4]. Legal interpretation is extremely difficult to automate. Years of hard work aimed at creating efficient systems supporting legal interpretation have barely brought this domain out of research laboratories into the common legal practice [5]. Zurek, and Kruk [5] further argue that the reasons why this proves so difficult are varied, including a great deal of common sense coming into play, a lack of precision and the ambiguity of legal provisions, the necessity to take the context of a given situation and the aim of the legislator into account, the choice of the line of interpretation, and so forth.

Law is a knowledge-based profession and its core 'legal practice' is about providing specialized knowledge and services to a variety of clients. A knowledge-based system is most effective in the management of semi-structured problems [6]. This knowledge or intellectual capital, i.e. the law firm's aggregated experience or collective wisdom, applied to delivering knowledge-based services, is one of the most important assets of a law firm [3]. Currently, activities related to law and legal issues involve a large group of professionals, and amount to a multi-billion overall value. People involved in such activities must deal with a challenging amount of text, information and knowledge contained in thousands of documents. Thousands of civil servants and legal professionals need every day to access and analyse the existing statutes in order to propose updates and changes. For them, dealing with information and knowledge is an essential part of their daily work. They are "knowledge workers" and vulnerable to suffering from information overload. Legal professionals and civil servants need to reduce the significant amount of their time spent finding, reading, analysing and synthesizing information in order to take decisions and prepare advice and trials [7]. In the field of legal knowledge management and representation, the problem of representing legal knowledge in the form of a variety of knowledge bases or ontologies has widely been recognized [8].

Legal information constitutes primary sources of information, i.e. statutes and cases, and secondary sources, such as legal reference works, digests, law reviews, legal periodicals, commentaries, and books and articles from specialised law publications. Legal knowledge, that is knowledge on law and its application, is used to procure, produce and manage legal work [3]. Several studies have shown that advances in information and communication technologies (ICTs) are increasingly transforming the methods that lawyers use to access, retrieve and process information in order to solve legal problems and deliver legal services to clients [9], [10]. Nowadays, the Internet can also be seen as an additional 'library' or 'information channel' from which statutes, regulations and cases reported by countries worldwide, as well as secondary sources, may be retrieved either freely or on a subscription basis [11], [12]. With the rapid increase in the amount of electronic and print information, as well as with the increasing availability of information technologies, information literacy has become one of the most vital skill sets of knowledge for workers in the information era [13]. In the context of the legal profession, Carroll, Johnston, and Thompson [14], view information literacy as: the ability to locate primary and secondary legal materials, which implies sufficient knowledge of retrieval tools and techniques; the evaluation of source authority and content relevance, applicability and value of the materials to the task at hand; management of information, e.g. sorting, categorising, annotating and ranking the information; and lastly the use of the information for the task at hand. An important aspect is the knowledge of how and where to seek information [3]. Although legal research entails much more than just information seeking, the vast majority of lawyers often engage in some form of information seeking and synthesising [15], [16], [17]. Plessis and Toit [3] argue that many legal researchers

nowadays do not ignore the change that was brought forward by the advent of the Internet, computerised legal databases, CD-ROMs and other electronic media channels. Rather, they emphasise that legal research encompasses using and mastering both print and electronic resources.

Technological advancements have greatly affected our lives today. The use of computer technology and communication via the Internet has connected people around the world in a way never before envisaged [18]. The business world has indeed surpassed the legal world when it comes to using technology. Often, the reason why businesses have moved to using technology is that it is more cost effective to share and store information digitally. Clients will expect their lawyers, and the courts, to do the same [19]. It is common to find many law firms, ranging from the large multinational or leading partnerships to the simple sole practitioner, investing in this new technology, not just to save time and costs but also to deliver professional legal advice in an increasingly competitive market [20]. The Internet now provides a wide range of legal information, and the benefit of information being provided in this way is that it can be kept up-to-date as the law changes. The Internet can also assist in court processes generally, that is, in trial preparation and also, in the court room throughout the hearing. The reason courts should embrace technology lies in the constantly increasing caseloads, the complexity of cases, the jurisdiction, resource constraints, and pressure to improve access to justice, the expectations for improvement in performance and the pressure to improve the efficiency and effectiveness of court administration and the delivery of justice [19]. Data should be entered only once into a system, and this accords with the notion of integrated justice [21]. The areas to be considered in addressing how courts should deal with these problems are [19]:

- Electronic appeals and electronic filing
- Legal research
- Medium neutral citations
- Inside the courtroom
- Virtual courtrooms
- Legislative changes
- Public examination of court information
- Privacy
- Free access

One of the problems with utilising e-filing is that forms are often required to be signed personally by the party who is filing the document [22]. One way to deal with this may be the use of digital signatures. It is recognised that legislation may require amendment so that information can be tendered in evidence and documents can be “signed” electronically. The advantages of the Internet are that the information is generally current and up-to-date, easily accessible and freely available. One disadvantage is that the information is not secure, and downloading

information can be slow depending on traffic. There is also a concern that reports published on the Internet can be sabotaged, although this can be overcome by ensuring that a duplicate copy of a court's database and that the Internet server is regularly updated via a one-way modem link, from the duplicate database [19]. A virtual courtroom is one which need not exist anywhere but electronically [19]. In the United States, Courtroom 21, "The Courtroom of the 21st Century Today", is located at the College of William-Mary Law School and is arguably the world's most technologically advanced trial and appellate courtroom [23]. The Singapore Supreme Court has successfully setup a virtual courtroom known as "The Technology Court" [24], which has a local area network (LAN) with an Internet connection and which allows the use of imaging, multimedia and video conferencing. It also has a Litigation Support System for Presentation (LSSP), a Computer Based Recording Transcription System (CBRT), a sophisticated Audio Visual System (AVS), which allows various types of audio and a video conferencing to allow foreign witnesses to give evidence in any proceedings.

Further, to the best of our knowledge, very few real life legal business information system implementations are documented in literature. Therefore, this paper will facilitate understanding of implementation in this sector. This paper is organized as follows: first, the case study with the background, the scope, and the different components and sub components of the project are described. Later, the implementation, including testing, is discussed. Section 3 presents the conclusion.

Generally, the case study method is a preferred strategy when "how" and "why" questions are posed and the researcher has little control over events [25]. The case study method, a qualitative and descriptive research method, looks intensely at an individual or small number of participants, drawing conclusions only about the participants or the group and only in the specific context [25]. The case study method is an ideal methodology when a holistic, in-depth investigation is required [26]. Case studies are often conducted in order to gain a rich understanding of a phenomenon and, in information systems research, the intensive nature, the richness of the description of a case study and the complexity of the phenomenon are frequently stressed in case study reports [27].

2 Case Study

2.1 Background

The legal business information system is responsible for ongoing programmes and services under relevant acts such as official Assignee, official receiver and public trustee. Systems that were used prior to the fully integrated system were the accounting system, asset realization system and liquidation management system. The proposed system was aimed at meeting the increasing business requirements

of management of corporate insolvency cases, case accounting and asset realization processes. It also has interfaces with relevant applications of the existing application systems to ensure the seamless flow of data and business processes. The key motivation behind the system was to adopt an integrated approach to streamline processes, computerize operations and enhance information flow and communications. The department had invited prospective software developers for a tender for the proposed system and a local IT company was awarded the contract for the supply, delivery, development, customization, installation, testing, data migration, data conversion, training, commissioning, implementation and support of the proposed system.

2.2 Scope

Several meetings were conducted between developers and users to gather the requirements. Afterwards functional specifications covering all the requirements were documented and signed by key users. It was found that the proposed system would consist of a Corporate Insolvency Management System (CIMS), a Case Accounting System (CAS), a Support Services Management System (SSMS), a File Management System (FMS), and a Workflow (Document/Content) Management System (WDMS).

2.2.1 A Corporate Insolvency Management System (CIMS)

A CIMS provides users with the facility to manage the administration of corporate insolvency cases. The system supports users in their business processes with the following modules:

- Case creation, search and summary: This module facilitates the creation of cases for users. It also provides functions for user to search and retrieve a specific case and then to view pertinent summary information regarding the case.
- Pre-winding up activities: This module helps users in the generation of letters to consent to act as liquidator or to suggest that a private liquidator be appointed as liquidator.
- Pursuance of Statement of Affairs (SA)/Statement of Assets and Liabilities (SAL) Filing: This module supports users in the process of pursuing officers of the wound-up entities to file SA or SAL respectively.
- Statement of affairs/statement of assets and liabilities: This module allows users to capture the details of the SA or SAL submitted. It also provides facilities for users to consolidate the details in the event of multiple submissions.
- Preliminary investigation: This module allows users to generate the summary of company affairs for each SA submitted.
- Asset management: This module supports users in the administration of the assets of the wound-up entity.

- Administration, adjudication and payment of claims: This module allows users to capture the details of the proof of debt submitted by claimants/creditors of the wound-up entity. It also supports users in the verification of the proof of debt submitted electronically. This module also provides users with the functions to manage the proof of debt filed; to manage claims that are both disclosed and not disclosed in SA; to manage tax on interest on investment; and to manage petitioner's costs.
- Return of capital to contributories: This module supports users in the process of returning capital to the contributories of the wound-up entity. It facilitates the generation of the various statutory forms and affidavit to be used in the process of returning capital.
- Regulation of private liquidators: This module supports users in the process of regulating private liquidators. It facilitates the generation of letters to the private liquidators and liquidator's report. It provides functions for users to manage the statutory forms submitted by private liquidators and to manage the refund of money to private liquidators.
- Preparation for release of liquidator and dissolution of company: This module supports users in the process of preparing for the release of the wound-up entity and the discharge as liquidator upon completion of administration of the wound-up entity.

2.2.2 Case Accounting System (CAS)

CAS provides users with the facility to create the chart of accounts manually or through respective systems of the business units. Once the chart of accounts has been created for the respective business unit, users can issue receipts for payments made by the public through the receipt function. For all cases, payments received from the public are to be paid to a case account, which already exists in the CAS.

The system also captures the details of payments such as distribution of dividend, distribution of motor accident compensation, etc. to be made to the public, thus expediting the payment process through electronic capturing and processing. The maintenance payment for the trust cases is fully automated. Business units through workflow can initiate payment process. The individual business units should approve of all the payment processes before they are routed to the CAS to be processed. The CAS user can also reject and route back the payment process to the individual business units due to insufficient information.

Bank reconciliation of transactions made between the legal department and banks can be done via the bank reconciliation functions, which match the legal department's bank transaction records against the bank statement.

Investment processes, such as the identification of cases for a new investment and the renewal and withdrawal of an investment are performed via investment functions.

Ad-hoc processing, enquiry and reports functions are made available to users for the overall completeness of the accounting system.

Every user of the system is authenticated before the use of any function. Only the authorized functions are displayed for the users. Audit trails for the critical functions are also provided. The old transaction records are archived or deleted after a period of time set by the users.

2.2.3 Support Services Management System (SSMS)

This system, SSMS, is established to provide an effective and efficient system to support the users in the management of court, asset management and prosecution matters.

2.2.3.1 Court

Petition details for both individual and corporate insolvency are created by the system through data exchange with the supreme court.

Besides the automated creation and updating of case records, the system provides functions for the user to maintain case details manually as well. As there are documents to be provided for each petition, the system provides facilities for the user to track the availability of documents.

As petitioning creditor's deposit needs to be paid for each petition, the system provides functions for tracking of petitioning creditor's deposit payment and sending of reminders to solicitors for missing payments.

Functions are provided for users to manage the hearings of petitions, release, discharges as well as other types of hearings. As multiple cases are heard on a day, a function is designed to enable users to update multiple records together for faster data capturing.

The system provides functions for printing of letters, notices and reports. A facility is provided for users to keep the documents in document management system.

Every user of the system is authenticated before the use of any function. Only the authorized functions are displayed for the users. An audit trail for the critical functions is also provided.

2.2.3.2 Asset Management

The system provides the ability to maintain the agent and contract details for asset seizure and realization work.

For those visitations, seizure and realization work carried out by agents, asset management officers need to keep track of the status of their realization work. Details pertaining to the realization of the assets are required to be captured into the system as well for information and tracking purposes.

The system provides the asset management officers with functions for maintaining the visitation, seizure, assets, lot allocation and details of the disposal assets.

In addition to capturing information, the system provides functions for printing of letters, notices and reports. A facility is provided for the user to keep important documents in Electronic Data Management System (EDMS) as well.

As the sale of assets involves payment of the sales amount to the legal department by the agent, a facility is provided for users to initiate such payment to the cashier via the system.

2.2.3.3 Prosecution

All cases referred to prosecution are consolidated under one function, whether the referrals are manually initiated or system initiated based on specific business requirements. There are 3 main areas covered under prosecution: investigation, summons, and high court warrant of arrest.

2.2.4 File Management System (FMS)

The FMS consists of the following modules:

- Maintain file: Maintain file module allows registry users to create or update the details of a physical main/sub file to the FMS.
- Maintain storage media (e.g. compact disc): Maintain storage media allows registry users to track the movement of storage media e.g. compact disc.
- Maintain multiple files: Maintain multiple files module provides an alternative for registry users to update movement and archival of records in bulk. Registry users can also perform bulk update of disposal and archival information via this module.
- Enquiry: The enquiry module consists of commonly used enquiries needed by the registry users to find out the information they required.
- Report: The report module allows registry users to generate out simple reports needed for daily processing, etc.

This system also integrates with the existing handheld device via batch job to upload the movements tracked by the device.

2.2.5 Workflow (Document/Content) Management System (WDMS)

The features included in WDMS include:

- Check-in and check-out documents
- Version controlling the documents
- Audit trails of document changes
- Indexing and keyword management to the documents
- Archiving
- Content/document search utilities based on document attributes, i.e. index fields and keywords

- Multiple index search at the same time
- Document life cycle
- Review and annotation capabilities
- Storage of all types of documents
- Web-based access to authorized users

Work/task shall be assigned using rule-based routing and ad-hoc routing with options for round- robin or load balancing capabilities. These routings policies are well defined to various sets of rules and conditions according to the business process. Various alerts like email, fax can be provided to the users with deadlines and schedules.

2.3 Design

The salient features of their proposed solution are as follows:

- Employs the best of breed and industry proven products.
- An open, generic and flexible solution, which allows for future expansion.
- Reusable system components and application modules to provide seamless system integration, thus minimizing administrative effort. Lower total-cost-of-ownership in long-term maintenance of the system can also be achieved. Effective integration is the key because if one of these links fail, the organization's performance may suffer and may not meet the expectations of its customers or the service level of its competitors [28].
- Highly integrated and workflow-driven intelligence, together with middleware to deliver data and images from one business process to another.

After a careful analysis of the requirements, the vendor used a combination of commercial off-the-shelf packages, application framework and custom development to deliver the solution for the legal department's system. This enabled the system to leverage an advanced and proven application framework and available solutions in the market wherever possible, that are able to satisfy the business and operational needs of the system so that a cost effective and quality system can be delivered within the expected timeframe.

The solution had been designed keeping in view the following major principles:

- **Open architecture:** The solution is based upon open standards to allow for ease of future expansion and services to be built upon the infrastructure proposed. Proven technology and standards such as J2EE, XML and relational database standards are deployed.
- **Scalability:** The solution has been designed so that servers can be added easily to accommodate increased users and volume. The design also takes care of the ease of potential extension of the usage of the system beyond the defined user environment, i.e. the Internet.

- **Ease of use:** The solution has been designed to allow the end user to access from the common web browser. The deployment of web solution also enables the users to perform their tasks without being constrained in their office environment.

The overall application architecture for the custom built application was based on the N-Tier web-based architecture. The architecture of the system adopted a layered service-based approach with collections of co-operating and communicating services.

Service driven architectures refer to architectures that provide a high quality of service delivery. This can be achieved by listing services in well-defined public interfaces with discretely packaged functionality. In the context of this system, the supporting application and security services encapsulate the business services in the application layer. The business components of the system like CIMS, SSMS, CAS, Insolvency e-Services, etc. can then ask for services rather than interacting with them directly. This loose coupling means that service-based architectures are less brittle and respond easily to changes in business demands.

This service-based concept allows the system to be further tuned to provide additional functionality/services. Subsequently, applications are able to share these added functionality/services, hence promoting reusability of complicated services and making the application more extensible and easier to manage.

In addition, the layered approach allows the design of this system to be modular by using the "building-blocks" concept. Components can quickly be developed and deployed using lower layer application services. By using this approach, the developed application modules can easily be reused. This has the added advantages of saving cost and manpower efforts as well as ensuring standardization in the implementation of proposed system application services.

The J2EE platform provides a reusable component model, using Enterprise JavaBeans and JavaServer Page technologies to build and deploy multi-tier applications that are platform and vendor-independent. The system leverages heavily on state-of-the-art technologies such as:

- Java and Object Oriented Design
- Design Patterns and Sun's Core J2EE Design Patterns
- The Internet and Internet based standards/protocols (eg. HTTP / HTML / RMI-IIOP / XML)
- J2EE standards like JSP (JavaServer Pages), Java Servlets, Enterprise JavaBeans (EJB) Component Architecture, JavaMail and JNDI (Java Naming and Directory Interface).

The N-Tier J2EE based architecture consists of the following tiers as shown in Figure 1:

- Client tier – Mainly front-end client browser

- Web tier – Hosts the web servers and other web services
- Business tier – Hosts the application servers and line of business servers
- Data and Interface tier – Stores and delivers data to the application. This tier possesses RDBMS and few interface systems

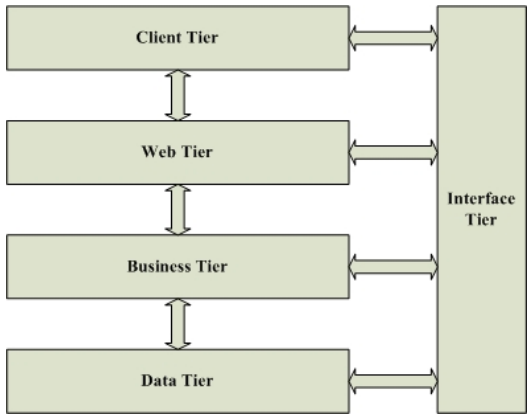


Figure 1
Implementation Architecture

2.4 Implementation Process

Prior to the implementation, a system load performance test was carried out on the production environment. To facilitate this testing, data converted in the user acceptance testing environment was ported over to the production environment. Users made use of these functions for verification and update of code and parameter lists, in preparation for the final phase implementation.

The data conversion process was carried out over a weekend before the implementation. Users were involved in the meetings where data mapping had to be established between the old system's data and new system's data structure. For some special scenarios, users had to decide on the conversion strategy. After data was converted, users tested the new system by accessing old converted records to verify that there are no issues with the conversion. After the data was successfully converted, users verified the data and performed wellness testing.

The system was implemented in two phases:

- Phase 1 – Web portal and FMS
- Phase 2 – Rest of the system

There was also a contingency plan in the event of implementation failure. In the event of the implementation failure, users would switch back to the existing systems. Before the commissioning of the system, all end users and technical support staff were trained by the vendor company.

2.5 Testing

The testing aimed at achieving the following objectives:

- Verifying that the system meets business requirements.
- Confirming that the system performs to acceptable levels.
- Confirming that the system operates correctly with interfacing systems.
- Verifying the usability of the system functions.

The scope of the testing included testing for the online and batch functions (including converted data) covering the following aspects:

- Functional - To ensure that the system satisfies the requirements spelt out in the functional specifications.
- Error handling - To ensure that the system detects incorrectly entered data and displays appropriate error messages in accordance with the validation rules.
- User interface - To check the page layout of all the functions and to ensure that the system complies with the user interface standard as spelt out in the proposed system's user interface design standards.
- Performance - To ensure that the system meets the performance standard as specified in the functional specifications.
- Control - To ensure that proper audit trail is created for critical functions as specified in the functional specifications.
- Inter-system - To ensure successful communication between all sub-systems of the proposed system.
- Compliance - To check that the functions have been developed in accordance with proposed system's project standards and procedures.

The testing approach adopted was based on each logical functional group/module identified in the various systems. Test specifications were created according to each logical functional group/module. For each incident encountered, there was proper follow-up to correct and retest the issue before closing it. When all problem reports were resolved and closed, the system was considered ready for user acceptance testing.

2.6 User Acceptance Testing

This was the final test and, as suggested by its name, stakeholders, IT managers and users played the main roles in this stage [29]. Test specifications were released to users for review before the actual start date. Each release was accompanied by an increase of the version for the user acceptance testing plan. The test specifications were reviewed and signed off by the users. A user acceptance testing briefing (of around 1 hour) was conducted during the 1st session of the user acceptance testing for each group of modules released. The user

proceeded with the user acceptance testing after the briefing. Users were actively involved in testing all functions of the system. They would report errors and after the development team resolved them, users would retest the functions. Some enhancements were also made during this stage. There were weekly review meetings to review the user acceptance testing progress (updated by users) and status of the problems/changes reported (updated by the developers). When all reported problems were resolved and closed, the system was considered ready for operation.

Conclusion

This study provides valuable insights towards understanding the implementation process, different components (modules), sub modules, design and testing issues to provide quality services in insolvency and trustee matters of legal business information system.

End-user satisfaction includes five components: content, accuracy, format, ease of use, and timeliness. User's feedback showed they were satisfied with the system operation. The system provided precise, concise information and the reports content met stakeholders' needs. Users found that the report format was clear and useful and they also appreciated the system's user-friendliness and up-to-date information content, which was delivered on time. It is also significant to mention the following issues related to this implementation:

- End users are satisfied. No major problems have been encountered since the system was implemented.
- Developers are also satisfied with work. The same team that developed the system is also maintaining the system. Hence users have no problem working with the IT team for enhancements or problems encountered with the system. Problems encountered have all been solved within the service level agreement period.
- There have also been frequent enhancements requested, which indicates that users have actually been using the system and would like to improve it further.

References

[1] Kennedy, D. (2001) The Connected Law Firm: Preparing your Firm to Thrive in the Internet Era [Online] Available: http://www.denniskennedy.com/connectedbook.htmS; Accessed Nov. 2001

[2] Maier, R. (2002) State-of-Practice of Knowledge Management Systems: Results of an Empirical Study. Upgrade, 3(1), 15-23

[3] Plessis, T.du, Toit, A. S. A. du (2006) Knowledge Management and Legal Practice, International Journal of Information Management, 26(2006), 360-371

[4] Stede, M., Kuhn, F. (2009) Identifying the Content Zones of German Court Decisions, W. Abramowicz and D. Flejter (Eds.): BIS 2009 workshops, LNBIP 37, 310-315

[5] Zurek, T., Kruk, E. (2009) Legal Advisory System for the Agricultural Tax Law, W. Abramowicz, D. Flejter (Eds.): BIS 2009 workshops, LNBIP 37, 304-309

[6] Mishra, A., Mishra, D. (2009) Customer Relationship Management – Implementation Process Perspective in *Acta Polytechnica Hungarica,* Vol. 6, Issue 4, 83-99

[7] Bianchi, M, Draoli, M., Gambosi, G., Stilo, G. (2009) A Support System for the Analysis and the Management of Complex Ruling Documents, W. Abramowicz and D. Flejter (Eds.): BIS 2009 workshops, LNBIP 37, 292-303

[8] Despres, S., Szulman, S. (2004) Construction of a Legal Ontology from a European Community Legislative Text, IOS Press Amsterdam

[9] Susskind, R. (2003). Transforming the Law: Essays on Technology, Justice and the Legal Marketplace. Oxford: University Press

[10] Van der Merwe, D. (2000) Computers and the Law (2nd ed.) Kenwyn: Juta

[11] Barratt, A., Snyman, P. (2002) Researching South African law. [Online]. Available: http://www.llrx.com/features/southafrica.htmS; Accessed October 2002

[12] Bast, C. M., Pyle, R. C. (2001) Legal Research in the Computer Age: A Paradigm Shift? Law Library Journal, 93(2), 285-302

[13] Bruce, C., Candy, P. (1995) Developing Information Literate Graduates: Prompts for Good Practice [Online] Available: http:// www.fit.qut.edu.au/InfoSys/bruce/inflit/prompts.htmlS; Accessed Sept. 2002

[14] Carroll, R., Johnston, S., Thompson, E., (2001) Information Literacy and Legal Research Skills Education in the UWA Bachelor of Laws Degree. In A. Herrmann, M. M. Kulski (Eds.) (2001) Expanding Horizons in Teaching and Learning. Proceedings of the 10th Annual Teaching Learning Forum. February 7-9, 2001, Curtin University of Technology, Perth, [Online] Available: http://cea.curtin.edu.au/tlf/tlf2001/carroll.htmlS; Accessed August 2002

[15] Dempsey, D. J. et al. (2000) Design and Empirical Evaluation of Search Software for Legal Professionals on the WWW. Information Processing and Management, 36(2), 253-273

[16] Leckie, G. J., Pettigrew, K. E., Sylvain, C. (1996) Modeling the Information Seeking of Professionals: A General Model Derived from

Research on Engineers, Health Care Professionals and Lawyers. Library Quarterly, 66(2), 161-193

[17] Wilkinson, M. A. (2001) Information Sources used by Lawyers in Problem Solving: An Empirical Exploration. Library and information science research, 23(3), 257-276

[18] Bhatt, J. K. (2005) Role of Information Technology in the Malaysian Judicial System: Issues and Current Trends, International Review of Law Computers and Technology, Vol. 19, No. 2, 199-208

[19] Stanfield, A. (1998) Cyber Courts: Using the Internet to Assist Court Processes, Computer Networks and ISDN Systems 30(1998), 559-566

[20] Bileta (2008) A Good Discussion of the Various Uses of IT to Practitioners in the Common Law System can be Found in the Various Bileta Conference avaialble at htttp://www.bileta.ac.uk

[21] Justice in the Balance 2020, Report of the Commission of the Future of the California Courts, available at www.courtinfo.ca.gov/reference/documents/2020.pdf

[22] Criminal Practice Rules 1900, available at www.legislation.qld.gov.au/LEGISLTN/.../C/CriminalCPracRu_01A_.pdf

[23] http://www.courtroom21.net

[24] http://www.gov.sg./judiciary/supermeet/computerisation/index.html

[25] Yin, R. K. (2003) Case Study Research: Design and Methods, 3rd Edition, Sage Publications, Thousands Oaks, CA

[26] Feagin, J., Orum, A., Sjoberg, G. (Eds.) (1991) A Case for Case Study, University of North Carolina Press, Chapel Hill, NC

[27] Van Der Blonk, H. (2003) Writing Case Studies in Information Systems Research, Journal of Information Technology, Volume 18, Number 1, March 2003, pp. 45-52

[28] Mishra, A, Mishra D. (2010) ERP System Implementation in FMCG Sector, *Technical Gazette* 17, 1(2010), 115-120

[29] Mishra, D, Mishra A. (2010) Improving Baggage Tracking, Security and Customer Services with RFID in the Airline Industry, *Acta Polytechnica Hungarica,* Volume 7 (2), 139-154

A WSMO-based Framework Enabling Semantic Interoperability in e-Government Solutions

Karol Furdík[*], Martin Tomášek[*], Ján Hreňo[**]

[*] Faculty of Electrical Engineering and Informatics, Technical University of Košice, Letná 9, 042 00 Košice, Slovakia, e-mail: karol.furdik@tuke.sk, martin.tomasek@tuke.sk

[**] Faculty of Economics, Technical University of Košice, Letná 9, 042 00 Košice, Slovakia, e-mail: jan.hreno@tuke.sk

Abstract: The paper presents a software framework that was designed and implemented within the FP6 IST EU project Access-eGov to integrate governmental services of various types, i.e. on-line (electronic and web services), as well as off-line (i.e. traditional, face-to-face) services by means of enhancing the service description with semantic information. The user-centric approach of life events is employed for presenting complex workflow sequences to the end users. The system architecture, functionality, and structure of underlying ontologies is described together with a mechanism for the orchestration and choreography of semantic web services in WSMO. The framework includes tools for the maintenance of semantically enriched services and for presenting the services to the citizens via customisable web interfaces. The paper concludes with an outline of the results obtained from the testing and evaluation of the implemented Access-eGov platform in real settings within public administrations in Slovakia, Poland, and Germany.

Keywords: semantic interoperability; e-Govenrment; web services; ontologies; WSMO

1 Introduction

Interoperability in the field of information software systems stands for an ability of the seamless interoperation of the possibly heterogeneous services which may be provided and consumed by various independent actors in a networked environment. The increasing demand on interoperable frameworks and solutions in the last five years is invoked by adopting the advancements of service-oriented architectures (SOA) and web services. It is particularly notable in the areas of e-Business [20], e-Health [13], or e-Government [18], where there is pressure on data and information exchange between the services, data resources, and applications distributed among a wide community of stakeholders.

The aspects of interoperability as a general concept or approach cover technical, syntactic, semantic, and organisational issues, usually referenced as interoperability layers [11]. These layers, which are related and mutually interconnected, deal with the following objects:

- *Technical interoperability level*: signals, low-level services and data transfer protocols;
- *Syntactic interoperability level*: data in standardised exchange formats, mostly based on XML or similar formalisms;
- *Semantic interoperability level*: information in various shared knowledge representation structures such as taxonomies, ontologies, or topic maps;
- *Organisational interoperability level*: processes, defined as workflow sequences of tasks, integrated in a service-oriented environment.

The main focus of this paper is on semantic interoperability; however, other levels are addressed as well. The framework presented in the next sections is built on established and widely accepted standards for data transfer and exchange (TCP/IP, XML), web services (WSDL, SA-WSDL) and process models (BPMN, BPEL). The integration platform is based on the WSMO framework (Web Services Modelling Ontology[1]), which provides an environment for the creation and development of underlying semantic knowledge structures - ontologies and semantically annotated web services that may be organised into a dynamic process workflow [17].

The outlined approach was adopted in the IST FP6 project Access-eGov (Access to e-Government Services Employing Semantic Technologies[2]) and will be described in more detail in the next sections of this paper. The consortium of this project consisted of 11 partners from five countries (Slovakia, Poland, Germany, Greece, Egypt) and was coordinated by the Technical University of Košice, which was responsible for the majority of design and implementation works as well. The main objective of the project was to develop a software platform that will be capable of providing support for citizens and businesses in their life event situations and business episodes related to various governmental services. The solution combines the user-centric paradigm of life events (on the side of user interface) with semantically interoperable service-oriented architecture (on the side of back-office) [15].

1 http://www.wsmo.org

2 http://www.access-egov.org

1.1 Related Research

In the field of e-Government, interoperability was recognised as a precondition for the implementation of European e-Government services already in the eEurope Action Plan [4] and was explicitly addressed as one of the four main challenges in the i2010 EU strategy [7]. This is important especially for the integration and co-operation of existing services - employing solutions based on existing standards, open specifications and open interfaces [8], [11].

One of the most promising approaches to interoperability is the employment of semantic technologies [18], [1]. Semantics provides a capability to model and represent knowledge within a domain by means of explicit formalisation of key domain concepts, their attributes and relations, as well as workflow sequences and structures. Considering the heterogeneous and distributed nature of the e-Government domain, semantics can be very effectively used as a common background platform for describing the processes and services provided by governmental institutions on various levels. The common platform then allows for integrating the services, making them interoperable and transparent for the end users, citizens and businesses.

Intensive research in the application of semantics in the e-Government field is going on, mostly focused on the integration of back-offices, employing SOA and web services enriched by a semantic description [5]. This research can be documented, for example, by projects supported by the European Commission within its 6th and 7th Framework Programme. Most of the solutions apply semantic technologies to ease the system design by modelling the citizen's behaviour, to enable or enhance interoperability of services, to provide a platform for creation of semantically described web services, etc. The provision of better and more integrated public services to citizens and businesses can be recognised as a common goal of all the research efforts. In the following paragraphs, we will briefly mention some of the R&D projects which can be considered as examples of existing solutions and approaches.

The *Terregov* project[3] is focused on the semantic requirements of governments at local and regional levels for building flexible and interoperable tools to support the change towards e-Government services. The Terregov solution provides a specialised ontology as well as a platform for enhancing existing government web services with a semantic description. Such semantically enhanced web services can then be detected, accessed, and orchestrated in an interoperable way. However, the Terregov solution only operates on a regional level of administration and, as such, it lacks a more global point of view. In addition, the Terregov solution requires a suite of already existing web services on the side of public administrations. The support for transforming other types of services (such

[3] http://www.terregov.eupm.net

as traditional face-to-face services, or electronic services provided by web forms) into required web services is rather limited.

The *SemanticGov* project[4] is aimed at supporting the provision of pan-European services to resolve semantic incompatibilities amongst public administration systems. The focus is put on the discovery, composition, mediation, and execution of services within complex scenarios, and the global ontology of semantic components needed for web service description is provided. Again, this approach requires an existence of web services on the side of public administrations. Contrary to the Terregov project, the global level of government services is covered, but the application of the solution on the level of local public authorities is not directly supported.

The *OntoGov* project[5] provides a semantics-based platform for the consistent composition, reconfiguration, and evolution of e-Government services. The solution includes a set of ontologies to describe and support the lifecycle of eGovernment services. The OntoGov approach mainly focuses on the software engineering side rather than on detection and orchestration of the government services; as a consequence, the interpretation on how the ontologies can be used in practical scenarios can be rather vague. In addition, the maintenance and usage of the OntoGov solution requires expert knowledge and lacks a certain degree of transparency for public servants when using the system.

In addition to the outlined projects, we can also mention some e-Government interoperability frameworks such as e-GIF in the UK[6], SAGA in Germany[7], European EIF IDABC[8] or SEMIC.EU[9]. These frameworks provide detailed information and guidelines about central government systems (on the national or European level). However, they fail to introduce specific information and rules for building eGovernment solutions for local administration [10].

To sum it up, there exist quite a wide range of approaches, proposals, frameworks, and projects in the area of semantic interoperability in e-Government domain, especially in creation and maintenance of semantic web services. However, the practical outcomes of the current research in this area (see e.g. [18]) are lagging behind expectations. The lack of supporting methodology, specialised tools, and guidelines describing how to create and maintain formal semantic descriptions of the services in practice may be one of the reasons. Another reason may be weak support for existing types of governmental services, and necessity to change

4 http://www.semantic-gov.org

5 http://www.ontogov.com

6 http://www.govtalk.gov.uk

7 http://www.cio.bund.de/DE/Standards/SAGA/saga_node.html

8 European Interoperability Framework for pan-European eGovernment services, http://ec.europa.eu/idabc/en/document/2319/5938.html

9 The Semantic Interoperability Centre Europe, http://www.semic.eu/semic/

(reengineer) dramatically how governmental services are provided, e.g. by implementing them as semantically described web services.

One of the main advantages of the semantic enhancement of government services is the capability of formally describing the meaning and context of government services, both traditional (i.e. face-to-face, "paper-based") as well as electronic ones (provided as electronic forms or web services), without the necessity of modifying the services themselves. In Access-eGov, this issue was targetted by developing tools as well as a methodology enabling the semantic interoperability of government services in practical applications [9], [2].

2 Approach, Methodology, and Technology

Following the main objective of the Access-eGov project, the access to government services (either traditional or electronic) needs to be provided in an "integrated" manner, in accordance with the pre-defined life event situations and business episodes of system users, citizens and businesses. From this perspective, the central position of life events (business episodes) as expressions of a user's needs is in correspondence with the life event approach [12], an effective and frequently used method in the user-oriented e-Government solutions. The life event is a situation in the life of the citizen (similarly, a business episode is a situation in the life cycle of the business organisation), which requires the provision of government services and should be semantically described within the system. Life events are usually complex and can be decomposed into several mutually dependent sub-goals. The fulfilment of the sub-goals leads to the solution of the given situation. Each sub-goal can be resolved to (i.e. fulfilled by) a set of government services that are provided either in a traditional way (requiring face-to-face communication and mostly based on some paper forms) or in an electronic way (available on-line via web service interfaces or web forms).

Sub-goals can be conditioned, organised in workflow structures using if-then-else constructs, cycles, and dependencies on outputs of other services – according to the specific case of the citizen or the organisation. During execution, the list of sub-goals for a life event is customised (e.g. by information provided by the user to specify his/her case) and then dynamically evaluated [14]. Services, which resolve sub-goals, may require some additional inputs provided by other services, so sub-goals can be further decomposed to sub-sub-goals and so on. During the service resolution process, the system may dynamically create a user scenario of the life event by evaluating the conditions of sub-goals. The user is then navigated to the proper services that are capable of fulfilling the goals and solving the life event situation.

Ontologies, as powerful knowledge representation formalism for modelling real-world concepts, were chosen as a basic mechanism for semantic modelling and the annotation of life events, goals, sub-goals, services, and other specific concepts from the public administration (PA) domain. This approach allows for the integration of existing (and future) systems and government services, as well as their functional interconnection on the technical, syntactic, semantic, as well as organisational level. To design a concrete ontology structure and content according to the purposes of the Access-eGov project, three basic resources were identified, namely:

- a *conceptual model* provided by selected semantic framework,
- existing and available *ontology resources*,
- formalised *requirements* that were provided by user partners of the project and were systematically organised into an ontology-like structure.

2.1 Conceptual Framework

After a detailed survey and analysis of existing and most used approaches (RDF-S, WSDL-S, WSMO, and OWL-S for ontologies; BPEL4WS for modelling web services in a business process interaction, etc.), we decided to apply the WSMO as a basic conceptual framework and implementation platform. The WSMO framework provides a consistent conceptual model for the semantic description of web services, with the inclusion of mediators and the distinction between goals and capabilities. It also provides the WSMX execution environment, WSML language specification for ontology formalisation, as well as the WSMO Studio visual development environment [3]. Based on this technological framework and the specified functional description, the architecture of the Access-eGov system was proposed [14].

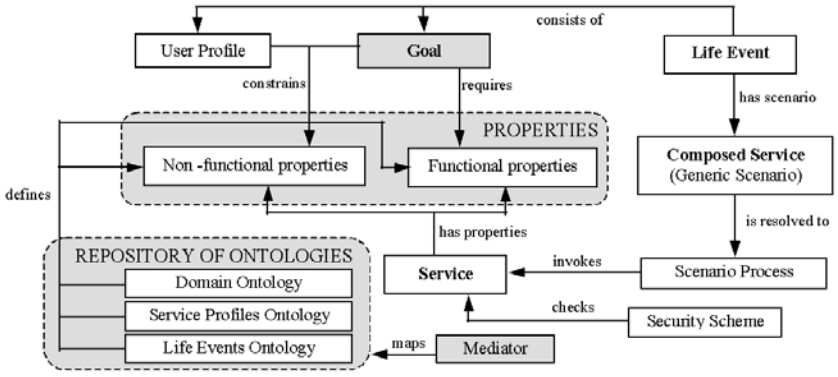


Figure 1
The WSMO-based conceptual model adapted for the life-event approach

In accordance with the life event approach, the conceptual framework of WSMO was extended, as it is depicted in Figure 1. The parts reused from the original WSMO model, such as Goal, Mediator, and Ontology top-level elements and the descriptive properties, are marked with a gray background. Two new top-level WSMO elements were added:

- *Life Event* element as a formal model of user's needs, consisting from multiple goals and services organised into a generic scenario and expressed by orchestration construction consisting from workflow, control-flow and data-flow sequences.
- *Service* element as a generalisation of the web service concept, already provided by WSMO. This extension enables the description of both electronic and traditional government services by means of a service profile, containing functional and non-functional properties, capabilities, and interfaces.

In addition, the WSMO conceptual model was enriched by the workflow extensions that are capable of representing a process model of the interactions with human actors in the e-Government domain [19]. The current WSMO specification provides the process model based on abstract state machines and is not structured in a way suitable for interaction with human actors, which is required for e-Government applications. The extended Access-eGov process model is based on the workflow CASheW-s model [16]. The state signature is reused from the WSMO specification and replaces the transition rules with the workflow constructs. The shared ontology state signature allows for reusing the grounding of input and output concepts to the communication protocols via WSDL for the invocation of web services. The workflow model consists of activity nodes connected with the control-flow or data-flow links. The nodes can be divided into atomic nodes (*Send*, *Receive*, *AchieveGoal* and *InvokeService*) and control nodes (*Decision*, *Fork* and *Join*).

An example of WSML statements for the orchestration interface that represents the high level process connected with the life event “Establish an enterprise” is presented in the following listing:

```
interface EstablishEnterpriseLifeEventInterface
orchestration
   workflow
     perform n1_1 receive ?x memberOf Q1.
     perform n1_2 achieveGoal RegisterInLocalGovernmentGoal
     perform n1_3 achieveGoal RegisterInStatisticalOfficeGoal
     perform n1_4 achieveGoal RegisterInTaxOfficeGoal
     perform n1_5 achieveGoal RegisterInSocialInsuranceAgencyGoal
   controlFlow
     source n1_1 target n1_2
     source n1_2 target n1_3
     source n1_3 target n1_4
     source n1_4 target n1_5
   dataFlow /**/
```

By interpreting this formal description, first the batch of answers to the pre-defined questions (*Q1*, invoked by the *Receive* node) needs to be received from the user by the process. Then other sub-goals need to be achieved in the right order. As can be seen, one of these goals is *RegisterInStatisticalOfficeGoal*. Transitions in the *controlFlow* part express that all nodes are executed in a sequence. The *dataFlow* part is empty in this case, since there is no direct use of some variable between these workflow nodes.

An example of the choreography interface, which composes a set of external services from a perspective of upper service, can be defined as follows:

```
interface RegisterInStatisticalOfficeInterface
choreography
   workflow
        perform n2_1 receive ?x memberOf Q3.
        perform n2_3 receive ?x memberOf FormRG_1.
        perform n2_4 decision
        perform n2_5 receive ?x memberOf FormRG_RD.
        perform n2_6 send ?x memberOf REGON.
   controlFlow
        source n2_1 target n2_3
        source n2_3 target n2_4
        source n2_4 target n2_5 guard ?x[q1 hasValue moreThanThree].
        source n2_5 target n2_6
        source n2_4 target n2_6
   dataFlow
        source n2_1{?x} target n2_4{?x}
```

This formal description can be interpreted as follows. At the beginning, the batch of answers to the pre-defined questions (*Q3*) needs to be received from the user. Then the process needs to receive a given form, which is represented by the *FormRG_1* variable. The *Decision* node in the next statement means that some of the following nodes are optional – in this case, only the node *n2_6* is optional, which is determined by the controlFlow part. Next, the process might need to receive another form *FormRG_RD*. Finally, the process sends the *REGON* number. The controlFlow part contains one conditional transition. The transition between the *Decision* workflow node and the following *Receive* workflow node depends on the answer to the question about the number of business activity types (i.e. if q1 is more that 3). The process can thus reach the final node right after this *Decision* node or from the last *Receive* node depending on the decision result. The dataFlow part specifies that the variable from the first node (*n2_1* - the batch of question) is equivalent with the variable from the *Decision* node (*n2_4*).

2.2 Ontology Construction

Semantic structures of the Access-eGov platform were created upon the specified and extended conceptual model of WSMO. The second resource used for the design of resource ontologies resulted from our survey of the ontology resources available worldwide. Using already existing ontologies assures consistency with

the widely accepted standards and avoids unnecessary double work. After a detailed analysis of about 25 ontology resources and standards, we finally reused the following ontologies:

- WSMO ontologies[10] for description of date, time, and location;
- vCard ontology[11] for addresses and personal data;
- Dublin Core[12] for metadata and document types;
- Terregov, DIP, DAML, GEA, GOVML, AGLS metadata set, and IPSV ontologies for description of specific e-Government concepts.

Existing ontology resources were used to produce some fragments of the whole ontology structure, mostly the definitions of non-functional properties for services. The example below presents an implementation of vCard ontology for WSML representation of the ontology concept *Organization*:

```
namespace{ _"http://www.accessegov.org/ontologies/core/",
  dc _"http://purl.org/dc/elements/1.1/",
  v _"http://www.w3.org/2006/vcard/ns#" }
concept Organization
  v#relation ofType Link
  v#organizationName ofType _string
  v#organizationUnit ofType _string
  v#addr ofType (1 1) v#Address
```

Finally, as the third resource of the ontology design procedure, requirements from the Access-eGov user partners were collected in a systematic way to produce ontology models of life events, sub-goals, and provided services for the pilot applications to be carried out within the project (cf. Section 4). The so-called *requirement-driven approach* [9], a method originally designed and developed within the Access-eGov project by one of the project partners (the German University of Cairo), was used as the main resource for ontology creation. This 7-step procedure starts with the identification of users' information needs for a particular case, which are provided by public administrations in a free-text format, e.g. as user scenarios. The information needs are then analysed with respect to the required properties, such as scope, relevance, etc. A list of proposed services together with related laws and regulations, documents needed to negotiate between users and public administrations, and other requirements concerning information quality are provided.

The descriptions and background materials are processed and a glossary of topics and terms is generated. The terms are organised into the controlled vocabulary, which contains a hierarchy of categories and subcategories created from the glossary by grouping the terms into hierarchical subgroups. This means that the

10 http://www.wsmo.org/WSMO_ontologies.html
11 http://www.w3.org/2006/vcard/
12 http://dublincore.org

terms in the controlled vocabulary are connected by *is_a* relations. In the next step, a set of other relations and mutual dependencies is identified between the terms. New categories (terms, concepts) can also be defined here if it is needed for the consistence of the whole structure. An ontology-like structure is provided as the output of this step. The core fragment of such a structure, which served in Access-eGov as a "seed" of Life events ontology, is presented in Figure 2. The grey background identifies so-called boundary concepts that will be annotated as non-functional properties of the services.

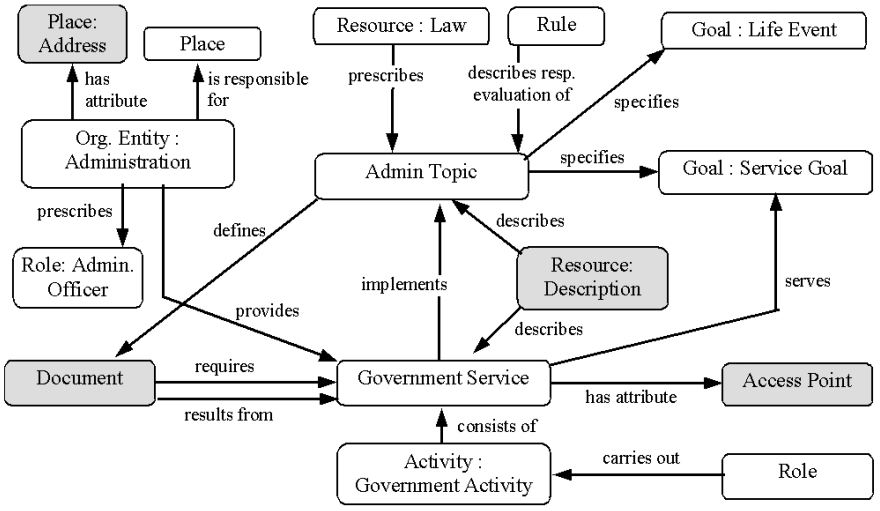


Figure 2
The ontology-like structure of identified concepts and relations

The ontology-like structure is then formalised and expressed by WSML statements. It requires fixing the meaning of the terms and relations defined in the controlled vocabulary, as well as verifying that formal meaning reflects informal description in the glossary. For example, a hierarchy of certificates can be expressed in WSML notation as follows:

```
concept certificate
  subConceptOf document
  concept birth_certificate
    subConceptOf certificate
```

The produced ontology is still rather static, consisting of declarative statements that express the concepts, their attributes, and mutual relations. In many cases, the conceptualisation needs to be enriched by "business rules" that can be, for example, conditional if-then-else expressions, loops, and workflow sequences. The enhanced process model of WSMO, described in Section 2.1 above, provides the means to semantically describe the life events, goals, and services in a dynamic manner. The example below presents the WSML formalisation of the life

event for marriage (expressed as complex goal) by means of the orchestration interface:

```
namespace {_"http://www.access-egov.org/ontologies/shg/",
  dc _"http://purl.org/dc/elements/1.1#",
  aeg _"http://www.access-egov.org/ontologies/core/"}
goal MarriageLifeEvent
  nfp dc#title hasValue "Marriage" endnfp
  interface MarriageLifeEventInterface
   orchestration
    workflow
      perform n1_1 receive ?x memberOf Q1.
      perform n1_2 achieveGoal ApplyForMarriageGoal
      perform n1_3 achieveGoal WeddingPlaceReservationGoal
      perform n1_4 achieveGoal WeddingCeremonyGoal
    controlFlow
      source n1_1 target n1_2
      source n1_2 target n1_3
      source n1_3 target n1_4
    dataFlow
      source n1_1{?x} target n1_2{?x}
```

By interpreting this formal description, first a batch of answers to the pre-defined questions (*Q1*) needs to be provided by the user. Then other sub-goals (*ApplyForMarriageGoal*, etc.) need to be achieved in a proper order. Transitions in the controlFlow part express that all the nodes are executed in a sequence. The dataFlow part specifies that the variable from the first node (*n1_1*, the batch of questions) is equivalent to the variable from the decision node (*n1_2*).

As a result of the 7-step procedure applied in Access-eGov, the following ontologies were created and formalised by WSML language:

- The *Core ontology* containing definitions of basic elements (concepts, attributes, relations) that are shared among the pilot applications and used for the annotation of the atomic services.
- The *Life Events ontology* containing conceptual descriptions of life events, complex goals (also referenced as generic scenarios), and elementary sub-goals for the pilots. Separate Life-Events ontologies were produced for each of the Access-eGov pilot applications.
- *Domain ontologies*, providing domain-specific information for the pilots. The ontologies are fully localized (concepts have labels in several languages – in this case the labels are in the English, German, Polish, and Slovak languages) and include concepts for description of forms, documents, certificates, location constraints, fees, questions, notification messages, etc. that are necessary to model the inputs and outputs of the provided government services. Separate domain ontologies were produced for each of the pilots.

The produced ontologies were uploaded as an asset on the SEMIC.EU portal[13] and are, after the required registration, freely available for further reuse, exploitation, or customisation.

3 System Architecture and Functionality

The Access-eGov system architecture, schematically depicted in Figure 3, consists of four main functional modules:

- The *AeG resource ontology*, a persistent data repository and a knowledge base that contains WSML representations of the life events and goals. In addition, it contains generic service concepts and service templates that enable the service annotation, as well as the instances of already annotated services.
- The *AeG core components* module, which includes the inner business logic of the system. The components are responsible for the decomposition of a given life event or goal into sub-goals, for the orchestration, composition, and mediation of the sub-goals within a workflow thread, for the semantic matching and discovery of the services for a given goal, as well as for the execution of the retrieved and resolved services.
- The *Annotation tool* (AT) for the semantic description (i.e. annotation) of the services that are to be integrated by the Access-eGov system. The web-based interface allows information providers to specify the non-functional properties for various service types, including traditional face-to-face services (in this case, the service is described by an explanatory HTML text that is presented to citizens), electronic, and web services. Capability interfaces, required inputs and provided outputs, and related workflow sequences are determined by a service template used during the annotation. The resulting WSML representations of the annotated service instances are stored in the resource ontology.
- The *Personal Assistant client* (PAC), a tool that enables the citizens to browse and navigate through the life event and corresponding sub-goals. This web-based tool is implemented as a kind of wizard that enables the personalisation and customisation of the thread of sub-goals by the answering of a set of customisation questions, which can be defined in a process model of the semantic representation of corresponding life events and sub-goals.

[13] http://www.semic.eu/semic/view/Asset/Asset.SingleView.xhtml?id=270

Operations performed during the design time on the side of information provider (i.e. a public administration) are represented in Figure 3 by thin arrows and are labelled by capital letters. An ontology designer uses the WSMO Studio tool [3] to create the resource ontology and customise it according to a given application case (A). The steps of the requirement-driven approach can be employed to specify the life events and goals, as well as the services and service types (templates). The structure of life events and goals is then automatically populated to the PAC to be presented for information consumers (B). However, the services that should correspond to particular goals need to be created separately, by means of semantic annotation (steps C-F).

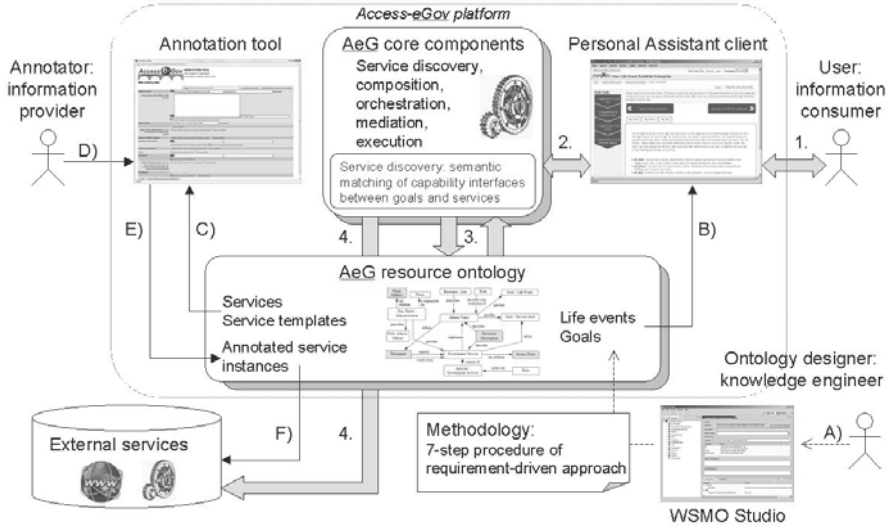


Figure 3
The architecture and control flow within the Access-eGov platform

The structure of generic services and service templates, created in the resource ontology, is automatically populated to the AT (C). An annotator then uses the AT to semantically describe the services, i.e. to specify concrete values for particular non-functional properties, defined by the employed service template (D). A WSML representation of the annotated services is created automatically and is uploaded into the resource ontology as a set of service instances (E). The service instances may contain a reference to an external web service or to an existing web content (i.e. a portion of a web page). This reference is specified as a non-functional property during the step D. After uploading the service instance to the ontology, the reference is evaluated, the external resource is validated, invoked, and the returning data are set as default value for the service instance (F).

Operations performed by the information consumer during the run time are numbered and represented in Figure 3 by thick arrows. A citizen uses the PAC to

browse the life events and goals (1). Some of the goals may require an additional input that concretises the citizen's needs – then the citizen provides answers to the customisation questions. The core system evaluates the responses obtained from the citizen for a given goal and dynamically creates a new thread of sub-goals, which is then returned back to the PAC (2). The process model of the goal is modified by the provided answers and its evaluation includes the procedures as service discovery, composition, orchestration, mediation, and execution. The core system communicates with the ontology to decompose a complex goal to sub-goals, to orchestrate, mediate, and compose the sub-goals into a workflow thread (3). For atomic goals that cannot be decomposed to sub-goals, the semantic matching procedure is used to discover and dynamically resolve a set of proper services. The core system then transforms the resolved services, according to their type, to an executable form and invokes the referenced external services (4). The input values for the external services are populated from the input provided by users and/or calculated during the evaluation of the goal's process model (step 2). The output values provided by the invoked external services are returned to the process model, which is then modified accordingly and is presented to the citizen in the PAC (step1).

The Access-eGov system is built on the WSMO framework, using the WSMX execution environment for discovery, selection, mediation, and invocation of semantic web services. It is implemented in Java; the WSMO4j API[14] is employed for parsing the WSML ontologies and obtaining the respective WSMO elements. The architecture of the system combines principles of SOA-based web services and service-oriented peer-to-peer architecture, which brings the advancements of modularity, possibility of local or remote accessibility from any platform, fault tolerance, scalability, and ease of deployment.

3.1 The Annotation Tool

The Annotation Tool (Figure 4) was implemented as a standard web application, using the extended WSMO object model and JSF technology. The tool provides public administration personnel a set of forms for the specification of preconditions and non-functional properties as parameters of the government services. A template mechanism was implemented to ease the maintenance of predefined workflow sequences for the annotated services. A simple user access control and multilingual support, on both the interface and data level, is also included in the AT. In addition, a simple "content grabber" functionality enables to link particular a field in the form (i.e. the value of a service parameter, e.g. service hours of an office) with an element on an existing web site of the public administration. This solution enables the annotation of the external web pages and the semantic integration of their content into a unified e-Government application.

[14] http://wsmo4j.sourceforge.net.

Figure 4
The user interface of the Annotation tool

3.2 The Personal Assistant Client

On the side of citizens, the Personal Assistant client (Figure 5) was developed as a tool that provides browsing, discovery, and execution capabilities of proper services for citizens and businesses according to a specified life event or goal.

Again, PAC was implemented as a web application using the JSF technology. The layout, structure, and ordering of tabs in the interface are dynamically created from the annotated services and are customised by the answers provided by the given user. After selecting a life event, a corresponding navigation structure of sub-goals and services is displayed for users in a form of textual information, hyperlink, a field for inserting a specified input value, or an interface for invocation of a web service. Users can browse sub-goals and provide their answers when customisation input is requested. Then the system automatically resolves the sub-goals and navigates the user to a new set of sub-goals and services inferred from the conceptual model. The Access-eGov system can also directly invoke electronic services provided via a standardised web service interface. Finally, the user obtains all available information on the life event customised to his/her case, and has also the possibility to execute the actions required for particular services needed for the accomplishing of the life event.

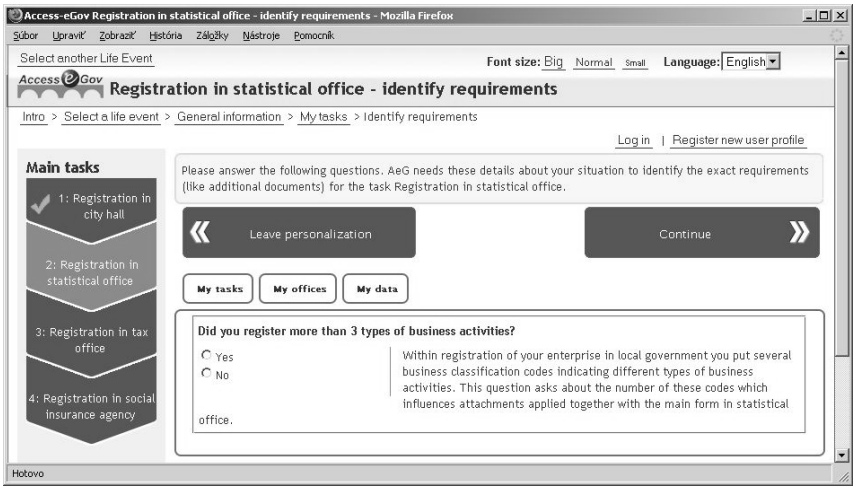


Figure 5
The user interface of the Personal Assistant client

4 Platform Testing and Evaluation

The Access-eGov platform in the scope of the described functionality was tested on three pilot applications in Germany, Poland, and Slovakia.

The *German pilot application* took place in the federal state of Schleswig-Holstein and dealt with the scenario "Getting married". This application case was chosen as a prototype example [2], but the goal was to integrate the services of the different administrations of 1,120 municipalities located in the federal state.

The *Polish pilot application*, focusing on the "Establishing a new enterprise" scenario, ran in the Silesian region around the city of Gliwice. The main aim of this pilot application was to provide a portal-like interface that would integrate all the relevant information (provided in Poland mostly in a form of traditional services) of a rather complex process in one place, and would guide the citizens and businesses through the life event and related sub-goals.

The *pilot application in Slovakia* covered the area of Kosice Self-governing Region and the municipality of Michalovce as its part. The goal of this pilot was to provide a personalised guidance for citizens during the process of obtaining permits for building a house, including services related to the land-use planning and final approval proceedings.

The testing was carried out in two trials within each of three pilot applications. The first trial resulted in a set of requirements that were taken by system

developers as a basis for system enhancements. The evaluation of the achieved results [6] demonstrated the feasibility of the proposed solution as a platform for front-office service integration. However, several issues were identified and requested as important for further improvement, specifically related to the graphical user interface of the PAC, navigation and browsing between the goals (tasks). The enhancements of the PAC invoked the necessity of adapting the core components accordingly, though the data structures and general functionality were not fundamentally changed. In addition, invocations of semantically annotated web services and related XSLT data transformations were fully integrated into the Access-eGov platform, namely into the PAC for presentation of the service results on the web user interface.

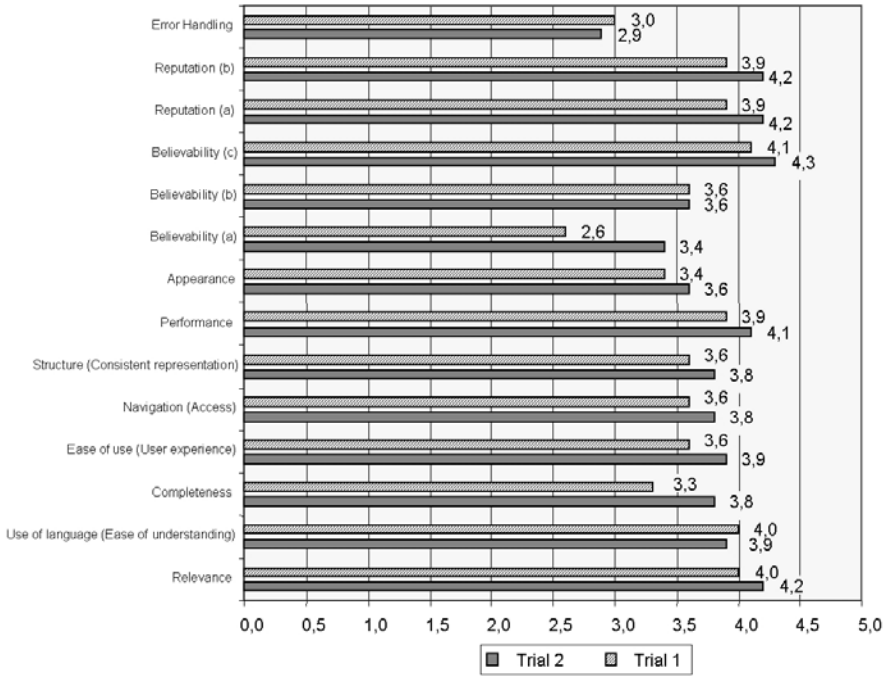


Figure 6
Overall results of the user evaluation after the second trial

The updated system was tested in the second trial on the same pilot applications as for the first trial. The testing was aimed at the ability of the system to integrate the external web services and data resources, and globally at proving the functionality and behaviour in real-world settings, especially focusing on the interoperability of heterogeneous distributed services and information resources. Technical testing, which was focused on the correctness, speed, and overall performance of the platform, resulted in the identification of significant improvements, in the service discovery and the overall speed of system response.

The method of online questionnaire was adopted for collecting the feedback from involved public. The chart presenting the responses obtained on the usability of the updated PAC is depicted in Figure 6. The levels of believability include a) ability to identify a provider (author) of the presented information, b) ability to determine which of the presented links lead to an external web site, and c) overall correctness of the provided information. The reputation covers a) a conviction to take the Access-eGov as a good (relevant, significant) source of information, and b) trustfulness of the information provided by the PAC. The usability evaluation of the second trial was quite successful, since average results of all the investigated aspects were better than those achieved in the first trial. The largest improvement was achieved in believability (+15%) and in completeness of presented information (+10%). Moreover, the people involved in the testing expressed rather positive feedback, stating that the solution is easy-to-use and provides very useful information for the modelled life events.

Acknowledgement

The work on this paper is the result of the project implementation: Development of the Center of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120030) supported by the Research & Development Operational Program funded by the ERDF.

References

[1] Abecker, A., Sheth, A., Mentzas, G., Stojanovich, L. (eds.): Proceedings of AAAI Spring Symposium "Semantic Web Meets eGovernment", Technical Report SS-06-06, AAAI Press, Menlo Park, CA, 2006

[2] Bednar, P. et al: Semantic Integration of eGovernment Services in Schleswig-Holstein. In: Electronic Government, 7th International EGOV Conference, Springer, LNCS 5184, 2008, pp. 315-327

[3] Dimitrov, M. et al: WSMO Studio Users Guide, v. 1.27. July 25, 2007. Available at http://www.wsmostudio.org/doc/wsmo-studio-ug.pdf

[4] eEurope 2005: An information society for all. COM (2002) 263, 2002

[5] eGovernment RTD 2020 project. Visions and Conceptions of European Citizens. Deliverable D 4.1: Final Roadmapping Workshop Report, 2007

[6] Furdik, K., Mach, M., Sabol, T.: Practical Experiences with Enhancing Semantic Interoperability in eGovernment using WSMO. In: Proceedings of MeTTeG 2008, Halley Editrice, 2008, pp. 23-35

[7] i2010 - A European Information Society for growth and employment, COM (2005) 229, 1 June 2005

[8] Interoperability for Pan-European eGovernment Services. COM (2006) 45 final, Brussels, 13.2.2006

[9] Klischewski, R., Ukena, S.: Designing Semantic e-Government Services Driven by User Requirements. In: Electronic Government, 6th International EGOV Conference, Proceedings of ongoing research, project contributions and workshops, Trauner Verlag, Linz, 2007, pp. 133-140

[10] Koussouris, S. et al: Building a Local Administration Services Portal for Citizens and Businesses: Service Composition, Architecture and Back-Office Interoperability Issues. In: Electronic Government, Springer LNCS 4656, 2007, pp. 80-91

[11] Kubicek, H., Cimander, R.: Three Dimensions of Organizational Interoperability. Insights from Recent Studies for Improving Interoperability Frame-Works. In: European Journal of ePractice, www.epracticejournal.eu, No. 6, January 2009

[12] Leben, A., Vintar, M.: Life-Event Approach: Comparison between Countries. In: Electronic Government, LNCS 2739, 2003, pp. 434-437

[13] Lopez, D. M., Blobel, B.: A Development Framework for Semantically Interoperable Health Information Systems. In: International Journal of Medical Informatics 78(2), 2009, pp. 83-103

[14] Mach, M., Bednár, P., Hreno, J.: Execution and Composition of Government Services. In: Proceedings of MeTTeG 2007, Halley Editrice, 2007, pp. 139-153

[15] Mach, M., Sabol, T., Paralic, J.: Integration of eGov Services: Back-Office versus Front-Office Integration. In: Proceedings of ESCW 2006, Budva, Serbia - Monte Negro, June 2006, pp. 48-52

[16] Norton, B., Pedrinaci, C.: 3-Level Service Composition and Cashew: A Model for Orchestration and Choreography in Semantic Web Services. In: Springer LNCS 4277, 2006, pp. 58-67

[17] Roman, D., Scicluna, J., Fensel, D., Polleres, A., de Bruijn, J.: D14: Ontology-based Choreography of WSMO Services. WSMO working draft, DERI, 2006. Available at http://www.wsmo.org/TR/d14/v0.4/

[18] Scholl, J., Klischewski, R: E-Government Integration and Interoperability: Framing the Research Agenda. In: International Journal of Public Administration (IJPA), Vol. 30, Issue 8-9, 2007, pp. 889-920

[19] Skokan, M., Bednar, P.: Semantic Orchestration of Services in eGovernment. In: Proceedings of Znalosti (Knowledge) 2008, STU Bratislava, Slovakia, 2008, pp. 215-223

[20] Vernadat, F. B.: Technical, Semantic and Organizational Issues of Enterprise Interoperability and Networking. In: Annual Reviews in Control, Volume 34, Issue 1, April 2010, pp. 139-144

On the Support of the Resultant Vector of d'Alembert's Fictitious Forces for Bars and Plates in Rotation Motion

Mihail Boiangiu

Department of Mechanics, "Politehnica" University of Bucharest
Splaiul Independentei 313, sector 6, cod 060042, Bucharest, Romania
E-mail: mboiangiu@gmail.com

Adina Boiangiu

Technical College "Edmond Nicolau", Bucharest, Romania

Abstract: This paper presents the determination of the central axis of d'Alembert's fictitious forces system for plane bars and plates in uniform rotation motion around an axis in their plane. General cases of bars and plates are studied. A general law for the determination of the support of the resultant vector of d'Alembert's fictitious forces is established. This law is based on the position of the center of mass of the rotation surface and body generated by the bar and the plate in rotation, respectively.

Keywords: plate; bar; rotation; d'Alembert's fictitious forces; support of the resultant vector

1 Introduction

A number of problems on the dynamics of rigid bodies [1] are solved by the application of d'Alembert's principle [2], [3], [4], [5].

In order to solve the problems on plates and bars having a uniform rotation motion by using d'Alembert's principle, it is necessary to know the position of the support of the resultant vector of d'Alembert's fictitious forces. Finding the central axis can sometimes be difficult. In the technical literature [4], [5] the support of the resultant vector of d'Alembert's fictitious forces is obtained by the integration of the elementary moment of the d'Alembert's elementary fictitious force and by stating the condition that the resultant moment be zero.

In this paper the authors propose a general method for the determination of the support of the resultant vector of d'Alembert's fictitious forces for bars and plates having a uniform rotation motion. The law proposed is original and is based on the position of the center of mass of the rotation surface and body generated by bar and plate in rotation, respectively.

2 Bars in Rotation Motion

Let us consider a homogeneous plane curve bar (Figure 1a), articulated in point A. The bar rotates with constant angular velocity ω, around the vertical axis which passes through point A and is contained in the plane of the curve.

The system of d'Alembert's fictitious forces (parallel forces) is equivalent to a resultant force on the central axis ($\overrightarrow{M^{in}} = 0$) [5]. In these conditions we want to find the support of the resultant vector of d'Alembert's fictitious forces.

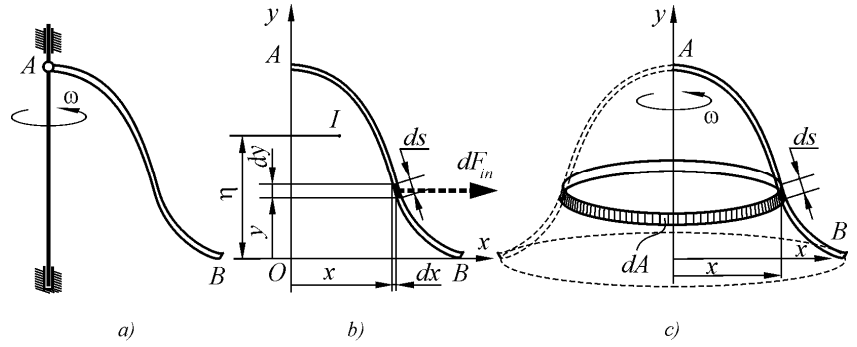


Figure 1

Determination of the support of the resultant vector of d'Alembert's fictitious forces for a bar having a rotation motion: a) bar in rotation motion; b) application of d'Alembert's principle; c) rotation surface generated by the bar in rotation

We relate the curve to a Cartesian reference system (Figure 1b). We isolate an element of infinite little length ds. The elementary d'Alembert's fictitious force dF^{in} is written as follows:

$$dF^{in} = \omega^2 x\, dm = \omega^2 x \rho\, ds\,, \tag{1}$$

where ρ represents the linear density.

If η denotes the *y*-coordinate of a point I on the support of the resultant vector, then the moment of the d'Alembert's elementary fictitious force with respect to this point will be:

$$dM_I^{in} = dF^{in}(y-\eta) = \omega^2 \rho x(y-\eta)ds\,. \tag{2}$$

Finally, the resultant moment of d'Alembert's fictitious forces is:

$$M_I^{in} = \int_{(D)} dM_I^{in} = \int_{(D)} \omega^2 \rho x(y-\eta)ds = \rho\omega^2 \int_{(D)} xy\,ds - \eta\rho\omega^2 \int_{(D)} x\,ds\,. \tag{3}$$

As point I is situated on the support of the resultant vector of d'Alembert's fictitious forces, the resultant moment of d'Alembert's fictitious forces is zero. Taking this condition into account, it follows that:

$$\eta = \frac{\rho\omega^2 \int_{(D)} xy\,ds}{\rho\omega^2 \int_{(D)} x\,ds}\,. \tag{4}$$

If we amplify the expression of η by 2π, we obtain the product $2\pi x\,ds$, which is equal to the elementary area dA generated by the rotation of the element of infinite little length ds (Figure 1c):

$$\eta = \frac{\rho\omega^2 \int_{(D)} xy\,ds}{\rho\omega^2 \int_{(D)} x\,ds} = \frac{\int_{(D)} xy\,ds}{\int_{(D)} x\,ds} = \frac{\int_{(D)} y\,2\pi x\,ds}{\int_{(D)} 2\pi x\,ds} = \frac{\int_{(D)} y\,dA}{\int_{(D)} dA} = y_C\,. \tag{5}$$

It follows that, for a plane bar in rotation motion, the support of the resultant vector of d'Alembert's fictitious forces passes through the center of mass of the rotation surface generated by the curve in rotation.

Let us consider the example of a homogeneous bar ABO (Figure 2a), consisting of two parts: a circle arch AB of radius R and a straight part $OB = R$. The linear density of the material is ρ (kg/m). The circle arch is articulated in A. The straight part *OB* is supported in point O. The plane which contains the bar rotates with constant angular velocity ω around the vertical axis that passes through points O and A. We want to find the angular velocity ω so that the bar will not touch the rotation axis in O.

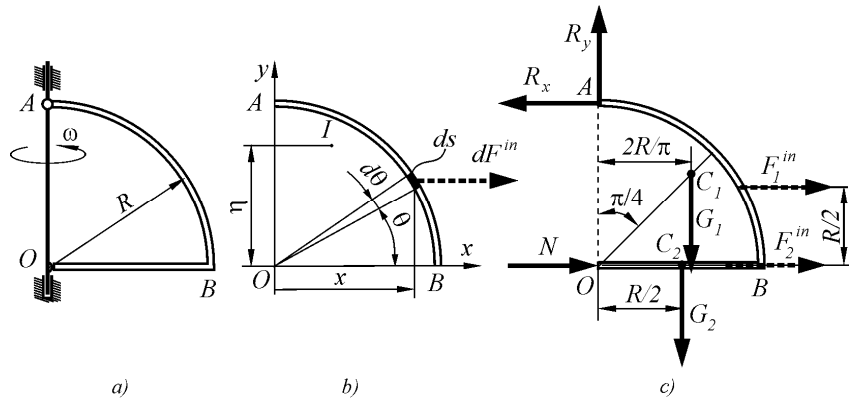


Figure 2

Application for a bar in rotation motion: a) bar in rotation motion; b) determination of the support of the resultant vector of d'Alembert's fictitious forces by the "classic" method; c) application of relation(8)

The support of the resultant vector of d'Alembert's fictitious forces for the circle arch AB can be determined by using relation (5), or in the "classic" way.

First we will use the "classic" method [4]. Let us consider an element of the bar of length ds which corresponds to an angle dθ (Figure 2b). For this element the d'Alembert's elementary fictitious force is:

$$dF^{in} = \omega^2 x\, dm = \omega^2 x \rho\, ds = \omega^2 \cdot R\cos\theta \cdot \rho \cdot R\, d\theta = \rho\omega^2 R^2 \cos\theta\, d\theta\ . \tag{6}$$

We consider that the support of the resultant vector of d'Alembert's fictitious forces crosses the rotation axis at distance η from the center of the circle arch AB.

The moment of the d'Alembert's elementary fictitious force with respect to a point I, on the support of the resultant vector, is:

$$dM_I^{in} = dF^{in}(\eta - R\sin\theta) = \rho\omega^2 R^2 \cos\theta(\eta - R\sin\theta)\, d\theta\ . \tag{7}$$

The resultant moment will be:

$$M_I^{in} = \int_{(D)} dM_I^{in} = \int_0^{\pi/2} \rho\omega^2 R^2 \cos\theta(\eta - R\sin\theta)\, d\theta = \rho\omega^2 R^2 \eta \int_0^{\pi/2} \cos\theta\, d\theta -$$

$$- \rho\omega^2 R^3 \int_0^{\pi/2} \sin\theta\cos\theta\, d\theta = \rho\omega^2 R^2 \eta \int_0^{\pi/2} \cos\theta\, d\theta - \frac{1}{2}\rho\omega^2 R^3 \int_0^{\pi/2} \sin 2\theta\, d\theta\ . \tag{8}$$

With the change of variable $u = 2\theta$, from relation (8) we obtain:

$$M_I^{in} = \rho\omega^2 R^2 \eta + \frac{1}{4}\rho\omega^2 R^3 \int_0^{\pi} (-\sin u)\, du = \rho\omega^2 R^2 \eta - \frac{1}{2}\rho\omega^2 R^3\ . \tag{9}$$

The resultant moment M_I^{in} is zero on the support of the resultant vector of d'Alembert's fictitious forces. From relation (9) it results that $\eta = R/2$.

The same result can be quickly obtained by applying relation (5). By rotation, the curve bar AB generates a hemisphere surface. This is a special case of a spherical zone. The center of mass is on the rotation axis at the distance *R/2* from the center of the sphere. The support of the resultant vector of d'Alembert's fictitious forces crosses the rotation axis in this point.

The resultant vector of d'Alembert's fictitious forces is:

$$F_I^{in} = m_{OA} a_{C_1}^{\nu} = \rho \frac{\pi R}{2} \frac{R \sin \frac{\pi}{4}}{\frac{\pi}{4}} \sin \frac{\pi}{4} \omega^2 = \rho R^2 \omega^2 . \tag{10}$$

The force of gravity is:

$$G_1 = \rho g l_{OA} = \frac{\rho \pi R g}{2} . \tag{11}$$

For the straight bar OB the support of the resultant vector of d'Alembert's fictitious forces is in the direction of the line OB and crosses the rotation axis in point O.

The resultant vector of d'Alembert's fictitious forces for the straight bar OB is:

$$F_2^{in} = m_{AB} a_{C_2}^{\nu} = \frac{\rho R^2 \omega^2}{2} . \tag{12}$$

The force of gravity corresponding to the straight bar OB is:

$$G_2 = \rho g l_{AB} = \rho R g . \tag{13}$$

Now we will apply d'Alembert's principle. We isolate the bar ABO and we obtain the system of forces in Figure 2c. We write the moment's equation with respect to point A:

$$N \cdot R + \rho R^2 \omega^2 \cdot \frac{R}{2} + \frac{\rho R^2 \omega^2}{2} \cdot R - \frac{\rho \pi R g}{2} \cdot \frac{2R}{\pi} - \rho R g \cdot \frac{R}{2} = 0 . \tag{14}$$

We obtain the expression for the reaction in point O:

$$N = \rho R \left(\frac{3}{2} g - R \omega^2 \right) . \tag{15}$$

From the condition $N \le 0$ we obtain the value for the angular velocity ω:

$$\omega \ge \sqrt{\frac{3g}{2R}} . \tag{16}$$

3 Plates in Rotation Motion

Let us consider a homogeneous plane plate (Figure 3a). This plate rotates with a constant angular velocity ω, around a vertical axis which is identical to a straight leg. The axis is contained in the plane of the plate.

The system of d'Alembert's fictitious forces (parallel forces) is reduced to a resultant vector on the central axis ($\overrightarrow{M^{in}} = 0$) [5]. In these conditions we want to find the support of the resultant vector of d'Alembert's fictitious forces.

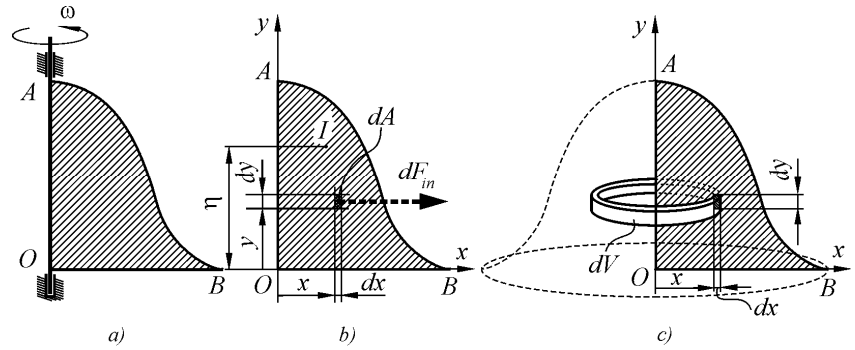


Figure 3

Determination of the support of the resultant vector of d'Alembert's fictitious forces for a plate in rotation motion: a) plate in rotation motion; b) application of d'Alembert's principle; c) rotation body generated by plate in rotation

We relate the plate to a Cartesian reference system (Figure 3b). We isolate an element of infinite little area dA=dxdy. The elementary d'Alembert's fictitious force dF^{in} is written:

$$dF^{in} = \omega^2 x\, dm = \omega^2 x \rho dA = \omega^2 x \rho dxdy\,, \tag{17}$$

where ρ represents the surface density.

If η denotes the *y*-coordinate of a point I on the support of the resultant vector, then the moment of the d'Alembert's elementary fictitious force with respect to this point will be:

$$dM_I^{in} = dF^{in}(y-\eta) = \omega^2 \rho x(y-\eta)dxdy\,. \tag{18}$$

Finally, the resultant moment of d'Alembert's fictitious forces is:

$$M_I^{in} = \int\limits_{(D)} dM_I^{in} = \int\limits_{(D)} \omega^2 \rho x(y-\eta)dxdy = \rho\omega^2 \int\limits_{(D)} xy\,dxdy - \eta\rho\omega^2 \int\limits_{(D)} x\,dxdy\,. \tag{19}$$

As point I is situated on the support of the resultant vector of d'Alembert's fictitious forces, the resultant moment of these forces is zero. Taking this condition into account it follows that:

$$\eta = \frac{\rho\omega^2 \int\limits_{(D)} xy\,dxdy}{\rho\omega^2 \int\limits_{(D)} x\,dxdy}. \tag{20}$$

If we amplify the expression of η by 2π, we obtain the product $2\pi x\,dxdy$, which is equal to the elementary volume dV generated in rotation by the element of infinite little area dA (Figure 3c):

$$\eta = \frac{\rho\omega^2 \int\limits_{(D)} xy\,dxdy}{\rho\omega^2 \int\limits_{(D)} x\,dxdy} = \frac{\int\limits_{(D)} xy\,dxdy}{\int\limits_{(D)} x\,dxdy} = \frac{\int\limits_{(D)} y\,2\pi x\,dxdy}{\int\limits_{(D)} 2\pi x\,dxdy} = \frac{\int\limits_{(D)} y\,dV}{\int\limits_{(D)} dV} = y_C. \tag{21}$$

It results that, for a plane plate in rotation motion, the support of the resultant vector of d'Alembert's fictitious forces passes through the center of mass of the rotation body generated by the plate in rotation.

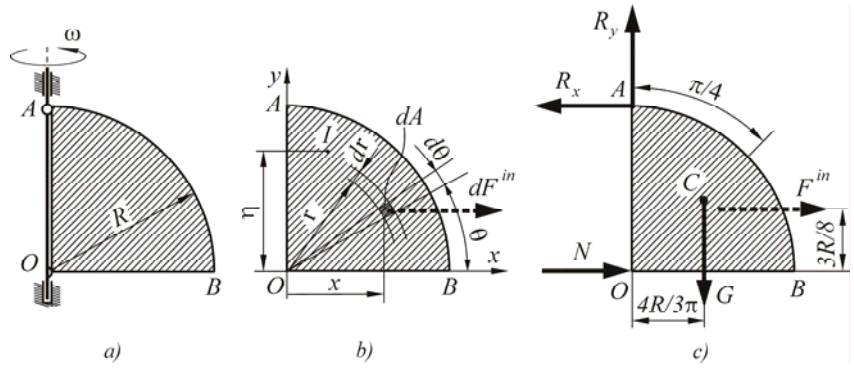


Figure 4

Application for a plate in rotation motion: a) plate in rotation motion; b) determination of the support of the resultant vector of d'Alembert's fictitious forces by "classic" method; c) application of relation(21)

Let us consider the example of a homogeneous plane plate OAB, a quarter of disc (Figure 4a) of radius *R*. The surface density of the material is ρ (kg/m^2). The plate rotates with constant angular velocity ω around the vertical axis that passes through points O and A. We want to find the angular velocity ω so that the plate will not touch the rotation axis in O.

The resultant vector support of d'Alembert's fictitious forces can be found using the "classic" method or relation (21).

First we will use the "classic" method [4]. Let us consider an element of the plate of area dA which corresponds to an angle dθ and a radius r (Figure 4b). For this element the d'Alembert's elementary fictitious force is:

$$dF^{in} = \omega^2 x\, dm = \omega^2 x \rho\, dA = \omega^2 \cdot r\cos\theta \cdot \rho \cdot r\, d\theta\, dr = \rho\omega^2 r^2 \cos\theta\, dr\, d\theta\,. \tag{22}$$

We consider that the support of the resultant vector of d'Alembert's fictitious forces crosses the rotation axis at the distance η from the center O. The moment of the d'Alembert's elementary fictitious force with respect to a point I, on the support of the resultant vector, is:

$$dM_I^{in} = dF^{in}(\eta - r\sin\theta) = \rho\omega^2 r^2 \cos\theta(\eta - r\sin\theta)\, dr d\theta\,. \tag{23}$$

The resultant moment will be:

$$M_I^{in} = \iint_{(D)} dM_I^{in} = \int_0^R \int_0^{\pi/2} \rho\omega^2 r^2 \cos\theta(\eta - r\sin\theta)\, dr d\theta = \rho\omega^2 \eta \int_0^R \int_0^{\pi/2} r^2 \cos\theta\, dr d\theta -$$

$$- \rho\omega^2 \int_0^R \int_0^{\pi/2} r^3 \sin\theta\cos\theta\, dr d\theta = \rho\omega^2 \left(\eta \int_0^R \int_0^{\pi/2} r^2 \cos\theta\, dr d\theta - \frac{1}{2} \int_0^R \int_0^{\pi/2} r^3 \sin 2\theta\, dr d\theta \right). \tag{24}$$

With the change of variable $u = 2\theta$, from relation (24) we obtain:

$$M_I^{in} = \rho\omega^2 \frac{r^3}{3}\eta + \frac{1}{4}\rho\omega^2 \int_0^R \int_0^{\pi} r^3(-\sin u)\, dr du = \rho\omega^2 \frac{r^3}{3}\eta - \frac{1}{8}\rho\omega^2 R^4\,. \tag{25}$$

The resultant moment M_I^{in} is zero on the support of the resultant vector of d'Alembert's fictitious forces. From relation (25) it follows that $\eta = \frac{3}{8}R$.

We obtain the same result by applying relation (21). By rotation, the plate generates a hemisphere. The centre of mass is on the rotation axis at the distance $\frac{3}{8}R$ from the center of the sphere. In this point the support of the resultant vector of d'Alembert's fictitious forces crosses the rotation axis.

The resultant vector of d'Alembert's fictitious forces is:

$$F^{in} = ma_C^v = \rho\frac{\pi R^2}{4} \cdot \frac{2}{3} \frac{R\sin\frac{\pi}{4}}{\frac{\pi}{4}} \sin\frac{\pi}{4}\omega^2 = \frac{1}{3}\rho R^3\omega^2\,. \tag{26}$$

The force of gravity is:

$$G = \rho A g = \frac{1}{4}\rho\pi R^2 g\,. \tag{27}$$

Now we will apply d'Alembert's principle. We isolate the plate OAB and we obtain the system of forces in Figure 4c. We write the moment's equation with respect to point A:

$$N \cdot R + \frac{1}{3}\rho R^3 \omega^2 \cdot \frac{5}{8}R - \frac{1}{4}\rho \pi R^2 g \cdot \frac{4R}{3\pi} = 0 . \tag{28}$$

We obtain the expression for the reaction in point O:

$$N = \rho R^2 \left(\frac{1}{3} g - \frac{5}{24} R\omega^2 \right). \tag{29}$$

From the condition $N \le 0$ we obtain the value for the angular velocity ω:

$$\omega \ge \sqrt{\frac{8g}{5R}} . \tag{30}$$

Conclusions

In this paper the authors have proposed a rule for the determination of the support of the resultant vector of d'Alembert's fictitious forces.

To sum up, for a plane bar having a rotation motion, the support of the resultant vector of d'Alembert's fictitious forces passes through the center of mass of the rotation surface generated by the curve in rotation. Also, for a plane plate in rotation motion, the support of the resultant vector of d'Alembert's fictitious forces passes through the center of mass of the rotation body generated by the plate in rotation.

If we consider the fact that the positions of the center of mass for a large number of bodies can be found in the technical literature, the method proposed here is accessible because it replaces the integral calculus.

References

[1] M., Vukobratović, B., Borovac, Why should Robots in Unstructured Environments Perform a Dynamically Balanced Regular Gait? Acta Polytechnica Hungarica, Vol. 6, No. 1, 2009, pp. 39-62

[2] V., I., Arnold, Mathematical Method of Classical Mechanics, second edition, Springer science, 1989

[3] G., R., Fowles, G., L., Cassiday, Analytical Mechanics, Harcourt College Publishers, 1998

[4] M., Radoi, E., Deciu, Mecanica, Editura Didactica si Pedagogica, Bucharest, 1993, p. 542

[5] R., Voinea, D., Voiculescu, V., Ceausu, Mecanica, Editura Didactica si Pedagogica, Bucharest, 1975, p. 85

Modeling and the Use of Simulation Methods for the Design of Lighting Systems

Miroslav Badida, Ružena Králiková, Ervin Lumnitzer

Department of Environmental Studies, Faculty of Mechanical Engineering, Technical University of Košice, Park Komenského 5, 040 01 Košice, Slovakia
<miroslav.badida, ruzena.kralikova, ervin.lumnitzer>@tuke.sk

Abstract: The article deals with designing internal artificial lighting as part of the work on the environment, which is subject to certain rules, derived from the nature of lighting. Good lighting exerts an impact on visual comfort – which contributes to overall psychological well-being, and indirectly also to the quality and productivity of performance, to reliability, and to visual performance – and which must be maintained, especially in long-term operations and in adverse conditions, to ensure quality of work and safety. One of the most frequently raised requirements for project design nowadays is speed; therefore a successful project cannot be done without good application software. The result is hundreds of drawings and various other outputs, which, without good software, cannot be handled within a short period of time. Modeling and simulation technologies are tools to streamline the presentation and assess the risks for the implementation.

Keywords: working environment; light; lighting microclimate; simulation

1 Introduction

The lighting of workplaces puts on light-technical solution the following requirements:

1) sufficient horizontal and vertical lighting value for a particular type of the work performed,
2) appropriate distribution of brightness in the area,
3) suppressing the creation of glare and protecting against it,
4) satisfactory psychological impact of the colour of the light and colour of the administration premises,
5) appropriate colour change in the environment,
6) stabile lighting,
7) reasonable uniformity,
8) suitable orientation of the impact of light on the desktop. [1]

In compliance with all the quantitative and qualitative parameters of illumination, we must design a lighting system based on the principles of maximum performance. By selecting a new generation of lamps, i.e. long life and high efficiency ones, we can economise on electricity. Lighting systems with streamlined operation, regulation and management of lighting may also significantly contribute to energy savings. [10]

2 Methodological Procedure of Light-Technical Design

The project of a lighting system is a complex and laborious task that requires not only technical knowledge, but also knowledge of architecture, production, and the physiology of vision. The role of the designer is not only to select the type of solution; this task is often complex and might be of a research character, leading to the development and manufacture of the lighting systems testing, analysis, and finding the optimum lighting conditions of the workplace and the area as a whole.

To develop a quality project of the lighting system, we should have construction in hand, technological and health technical drawings of the lighting the object, and we should also be familiar with the technology or the purpose of the premises. In addition to the quantitative and qualitative parameters of the workplace, the lighting area or the surrounding area should maintain well-observed and fault-free lighting system functions, the possibility of comfortable handling the luminaries and lighting efficiency. [12] The design of the lighting system is divided into light-technical, electric, and budget sections. The light-technical part of the interior lighting consists basically of two main parts: technical reports and the drawing section.

The technical report includes:

- description of the area to be lighted,
- demands on visual activity according to the category and work class,
- lighting values,
- qualitative lighting indicators (brightness distribution, direction of light, flare, lighting, durability, colour and colour submissions, etc.),
- draft operation and maintenance of the lighting system, choice of lamps, etc.,
- computational methods employed and specific calculations of lighting,
- colour adjustment of the immediate surroundings,
- assistant addressing, security, and replacement of emergency lighting,
- proposal for economic recovery.

The drawing section contains:

- footprints and cuts of lighting facilities,
- prescribed value of lighting on certain points and value quality parameters,
- electrical distribution, involvement and control of lighting systems,
- deployment of lamps, their specifications and type of the light resources,
- isoline diagrams and marking control points by which the agent glare was assessed.

In addition to the documents belonging to the base set of the project documentation, it is also necessary to produce drawings of the various elements of installation illumination, drawings of complete assembly nodes, drawings of connections and typical control components and drawings needed for the implementation of the proposed lighting.

3 Modelling of Light-Technical Parameters

In the past, there existed three basic types of light-technical models:

- calculation (without taking into account the actual dimensions, by means of tables),
- accurate (in models in the 1:1 scale),
- mock-ups that generate a display similar to visual perception of the lighting system designed.

Currently, a different approach is applied in the light-technical modeling, which is based on computer visualization of the spatial scenes of the lighting system designed. With computer visualization, whose goal is photo-realistic imagining, the propagation of light in space is often described in detail and simulated. Modern visualization programs can reproduce brightness, colour and surface structures of complex three-dimensional spaces in a quite realistic way, since the calculations include inter-reflection of light between various surfaces in space and quite a number of optical effects arising in daylight, in artificial or joint lighting. Simulation methods are based on classical optical, thermodynamic, or light-technical models of the spread of radiation [14].

3.1 Simulation Methods

There exist two basic methods employed in computer simulations of the light environment, namely the Monte Carlo method, which applies the technology of tracing the light rays (ray tracing is the name used for the follow-up of rays; one

also uses the term of "ray casting" - sending the light ray when a ray of light comes from the light source), and the radiation method (also radiosity). From a physical point of view, both of the methods are similar; the difference lies in algorithmization.

3.1.1 The Monte Carlo Simulation Method and the Calculation of Direct and Indirect Lighting

We have initially considered only spectacular reflections of light in a manner of subsequently applied probability calculations and other components of illumination. The stochastic (probability) method of light calculation, often referred to as the Monte Carlo method, is conveniently applied in furnished rooms with surfaces that have different optical properties. In general, this method is one of the operational methods of research used for the simulation of technical, economic, and social situations [8]. There exists a number of variants of this method.

In general, these methods employ a large number of randomly cast light rays or energy bearing particles. Their movement in the area is subject to physical laws and is monitored. A completely accurate calculation can only be made if the path of each photon can be followed, which, of course, is impractical for a number of reasons. However, if a sufficient number of rays (particles), e.g. 50 million, is accidentally sent out, the calculation of the lighting capacity will also correspond to high demands for accuracy. If the propagation of light is monitored from the source to the environment, one usually talks about the method of monitoring the particles (Fig. 1).

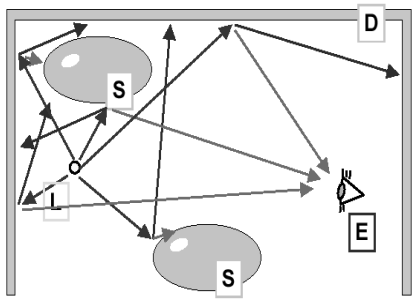


Figure 1
Behaviour of light rays in the Monte Carlo simulation method of ray casting
D - Wall space S - lighting operator, L - light source, E - Observer

In terms of computer graphics, ray tracing in the direction of the light source to the observer's eye or camera lens is onerous. Quantity rays are "lost" before the eye reaches the observer. [12] It is therefore a frequently used method of tracing rays (Fig. 2) when the monitor path of light rays is in the direction of the observer to the light source.

In this way, the algorithms take into account the particles that are mostly involved in the lighting of the scene as seen by observers. In this case, lighting of a place is proportionately dependent on the number of light particles which hit it, and on the density of luminous flux carried by each of these particles.

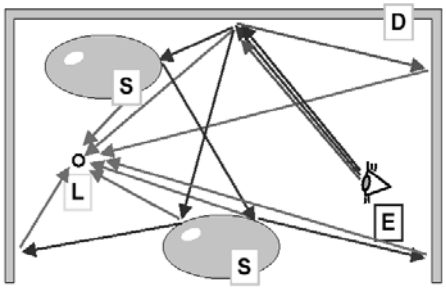


Figure 2
Behaviour of light rays in the Monte Carlo simulation method of ray tracing
D - space walls S – object being lighted, L - light source, E - observer

In the method of back tracing the rays, a virtual ray of light is cast in the direction of the observer through each of the imagining points on the display screen (pixels), and its intersection is tested along with all the objects in that space. Rays are cast in the direction of the light source to determine whether a visible place is overshadowed by an object. If the object surface is shiny, it mirrors the reflection of the primary ray. If the surface is transparent, rays are created, representing light reflection and refraction according to optical properties of the transparent material. If the surface is non-transparent, rays are generated (often more than 100) mimicking the light reflection from the surface concerned.

In the case that the location of the intersection of the primary ray with a certain object in space is illuminated by any of the light sources (or a mirror reflection of a certain material), its lighting or brightness is calculated. The term of direct lighting is employed in computer graphics for this lighting in contrast to the overall lighting containing the contribution of the reflected light, which in this field of science is called global lighting.

The nearest intersection is determined for each secondary ray, and the process is repeated until the ray leaves the space, or until the amount of light (or brightness) represented by the imaginary ray falls below the selected value. In some of the algorithms, the ray is monitored until it is returned in the eye of the virtual observer, or only a specified number of reflections is considered. In this way, the geometry of the space is modelled simultaneously with its synthetic (colour) imagining. Maps of direct and overall lighting are stored in the computer memory, which are further processed to achieve a smooth transition of shadows, in order to describe optical phenomena, among other things. In principle, the ray tracing technique solves the following integral equation (1) for the energy balance of each nearly the same surfaces in space [8].

$$L_r(\theta_r,\varphi_r) = L_e(\theta_r,\varphi_r) + \iint L_i(\theta_i,\varphi_i).\rho_{bd}(\theta_i,\varphi_i,\theta_r,\varphi_r).|\cos\theta_i|.\sin\theta_i.d\theta_i.d\varphi_i \qquad (1)$$

where: θ - polar angle as measured from the surface at normal levels,

φ - azimuthal angle of the surface at normal levels,

$L_e(\theta_r,\varphi_r)$ - its own radiation [W.sr^{-1}.m^{-2}]

$L_r(\theta_r,\varphi_r)$ - the total radiation [W.sr^{-1}.m^{-2}]

$L_i(\theta_i,\varphi_i)$ - incident radiation [W.sr^{-1}.m^{-2}]

$\rho_{bd}(\theta_i,\varphi_i,\theta_r,\varphi_r)$ - two-way function of the reflectivity distribution [sr^{1}].

3.1.2 Radiation Methods and Radiation Equation

Although the ray tracing algorithm produces perfect results in modeling the mirror reflectivity and undispersional refractional transparency, the algorithm has a shortcoming; specifically, it does not take into account the physical laws of some of the important visual effects, for example shade staining by the influence of the reflection of light from another object. It is due to the fact that ray tracing only monitors the final number of rays emanating from the observer's eye. The radiation method is attempts to remove this shortcoming. [2]

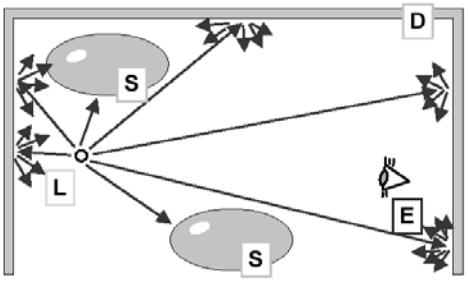


Figure 3
Behaviour of light rays in the radiation method
D - space walls S – object lighted, L - light source, E - observer

The radiation method may be seen as a certain generalization of the method of monitoring the ray. This method assumes that all the surfaces are ideal diffuse primary or secondary light sources (Fig. 3) or a combination of the given types of sources. The advantage of this method in terms of visualization and algorithm development is that the surfaces lighting is calculated independently from the direction of view on the simulated scene [9].

The radiation method is based on the principles of the spread of light energy and the energy balance. Unlike conventional rendering algorithms, this method first determines all the mutual light interactions in space from various independent

perspectives. Then one or more perspectives are calculated by defining a visible surface and interpolation shading.

In the algorithm of shading, the light sources have always been considered independently from the surfaces that are lighted. In contrast to the above, the radiation method allows any surface to emit light, i.e., all the light sources are modeled naturally as an active surface. Consider the distribution of the environment as a final number of *n* discrete surfaces (patches), each of which has its final respective size and emits and reflects light evenly across its surface. The scene then consists of surfaces acting both as light sources and reflective surfaces creating a closed system. If we consider each of the surfaces as an opaque Lambertian diffuse emitter and reflector, then the following equation applies for the surface due to energy conservation (2):

$$B_i = E_i + p_i \sum_{1 \le j \le n} B_j F_{j-i} \frac{A_j}{A_i} \tag{2}$$

where:

B_i, B_j - intensity of radiation areas *i* and *j* measured in units of energy per unit of surface ($W.m^{-2}$)

Ei - power of light radiated from the surface *i* and has the same dimension as radiation,

p_i - the reflection coefficient (reflectivity) of the surface *i* and is dimensionless,

F_{j-i} - dimensionless configuration factor (form-factor), which specifies the energy leaving the surface *i* and the energy incoming to the surface and taking into account the shape, relative orientation of both of the surfaces, as well as the presence of any areas that could create an obstacle. The configuration factor takes its values from the interval <0,1>, while for the fully covered surfaces it takes the value of 0,

A_i, A_j - surface levels *i* and *j*.

Equation (2) shows that the energy leaving the unit part of the surface is the sum total of both light emitted and reflected. The reflected light is calculated as a product of the reflection coefficient and the sum total of the incident light. On the contrary, the incident light is the sum total of the light leaving the whole surface changed in the part of the light which reaches the receiving unit content of the receiving surface. B_jF_{j-i} is the amount of light leaving the unit content of the surface *Ai* area and incident on the entire surface of A_i. It is therefore necessary to multiply the equation by the ratio of *Ai*/*Ai* for the determination of light leaving the entire surface *Ai* and incident on the entire surface A_i. [4]. A simple relationship is valid between the configuration factors in the diffuse medium:

$$A_i F_{i-j} = A_j F_{j-i} \tag{3}$$

By simplifying equation (2) using equation (3) we obtain the equation:

$$B_i = E_i + p_i \sum_{1 \le j \le n} B_j F_{i-j} \tag{4}$$

By subsequent treatment we get the equation in the form:

$$B_i - p_i \sum_{1 \le j \le n} B_j F_{i-j} = E_i \tag{5}$$

Interaction of light between the surfaces may be expressed in the matrix form [9]:

$$\begin{bmatrix} 1-p_1F_{1-1} & -p_1F_{1-2} & \dots & -p_1F_{1-n} \\ -p_2F_{2-1} & 1-p_2F_{2-2} & \dots & -p_2F_{2-n} \\ . & . & \dots & . \\ -p_nF_{n-1} & -p_nF_{n-2} & \dots & 1-p_nF_{n-n} \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ . \\ B_n \end{bmatrix} = \begin{bmatrix} E_1 \\ E_2 \\ . \\ E_n \end{bmatrix} \tag{6}$$

Note that the contribution of a part of the surface to its own reflected energy (which may be hollow, concave) must be taken into account. Thus, in general, each term on the diagonal need not necessarily equal to 1. Equation (6) must be solved for each group of wavelengths of light in the model, since *pi* and *Ei* depend on the wavelength. Form factors are independent of wavelength and are solely a function of geometry; therefore, they need not be recalculated, if the surface reflectivity or illumination changes. Equation (6) may be solved by employing the Gauss-Seidel method obtaining radiation for each area. In order for radiological methods to become partial, one had to start calculating the form factors for absorbed surfaces.

3.2 Form-Factor Calculation

To find the form factor, we must find the fractional contribution that a single patch makes upon another patch. This term is purely geometric, related only to the size, orientation, distance, and visibility between the two patches. The basic geometry for the form factor calculation is shown in Fig. 4.

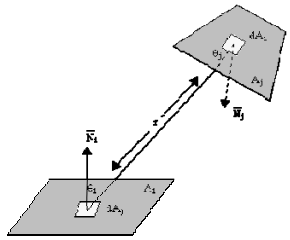


Figure 4
Form-factor geometry

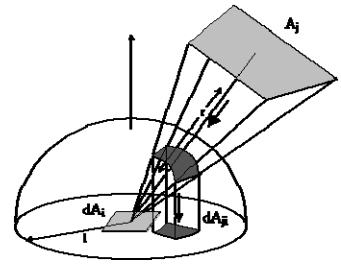


Figure 5
Projected area onto the hemisphere

If we look at Fig. 5, we see that the area A is related to the projected area, A_p, by $A_p=A.\cos_q$, and the contribution of the projected area A_p is related to the solid angle by (7)

$$\omega = \frac{A_p}{r^2} \tag{7}$$

The expression relating the contribution from one infinitesimal area to another is:

$$F_{dA_i - dA_j} = \frac{\cos\phi_i . \cos\phi_j . dA_j}{\pi . r^2} \tag{8}$$

The contribution from the infinitesimal area to the finite area is found by integrating over the receiving area:

$$F_{dA_i - dA_j} = \int_{A_j} \frac{\cos\phi_i . \cos\phi_j . dA_j}{\pi . r^2} \tag{9}$$

And from a finite patch to another finite patch, we take the area average of the previous equation:

$$F_{dA_i - dA_j} = \frac{1}{A_i} \iint_{A_i A_j} \frac{\cos\phi_i . \cos\phi_j . dA_j}{\pi . r^2} \tag{10}$$

There are several different methods for evaluating this integral. The contour integral is found by transforming the double integral by Stoke's Theorem [6]:

$$F_{dA_i - dA_j} = \frac{1}{A_i} \iint_{A_i A_j} \left(\ln(r) dx_i . dy_j + \ln(r) dy_i . dx_j + \ln(r) dz_i dz_j \right) \tag{11}$$

where ln(r) is the intensity for a particular wavelength. One limitation of this algorithm is that it does not take into account the visibility between one patch and another; another limitation is that it is extremely expensive computationally. Baum [1] also uses an analytical approach to find form factors. They integrate the outer integral numerically, while integrating the inner integral analytically by converting it into a contour integral. They then calculate the contour integral by piecewise summation.

$$F_{dA_j A_i} = \frac{1}{2.\pi} \sum_{g \in G_i} N_j . \Gamma_g \tag{12}$$

where:

G_i - is the set of edges in surface *i*,

N_j - is the surface normal for the differential surface *j*,

Γ_g - is a vector with magnitude equal to the angle gamma illustrated.

4 Outputs from the Proposal of Lighting System

Currently, the development of computer graphics software products exist to enable a comprehensive design and calculation of the parameters of lighting systems, which would reflect light effects that arise in artificial and day lighting. In consequence, there are on the market several light-technical programs with different purposes and uses. For the purposes of this paper, as to the possibilities utilisation simulations of light - technical parameters are presented the outputs created in the software DIALux 4.7. The above simulation programme offers the following options of the selected lighting system and various options for the presentation of results as chart values, isofotic lines (Fig. 6), light maps (colour scale) (Fig. 7), false colour rendering, summary tables of lighting or brightness, a three-dimensional model lighting, economic evaluation of brightness of the lighting project in terms of power consumption, visualization of sunshine, and so on. [11]

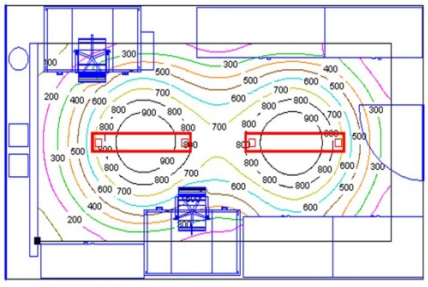


Figure 6
Isofotic lines

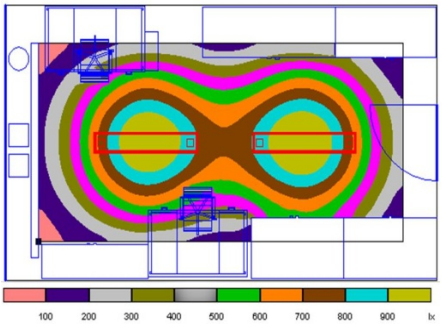


Figure 7
Light maps (colour scale)

Conclusion

In terms of the quantity of information, a person registers 80% to 95% of all the information visually in the work. The primary role in creating the work environment is to ensure optimal conditions of vision and ensure a safe working environment. Visibility must therefore be seen as a precondition for the implementation of high quality, safe, and reliable work operations. It is necessary to pay close attention to this issue. When dealing with light-technical projects, the visualization of lighting parameters is a useful tool by using programmes realistically displaying the lighting parameters.

Despite numerous possibilities that the current software tools offer, in some cases there is a difference between the modelled and actual light-technical parameters. One of the reasons affecting the result of the computer output may be the inadequate definition of certain inputs (the colour shades and quality of the room's

surfaces, the lightning effects on the scattering characteristics of light sources, etc.). However, these differences do not affect the overall relevance of computer outputs and may be virtually eliminated by qualified estimation.

Acknowledgement

This contribution was elaborated within the project KEGA No 3/7426/09 “Physical factors of environment - valuation and assessment" and KEGA No 3/7422/09 Creating of research conditions for preparation of modern university text book "Ecodesign in Mechanical engineering".

References

[1] Baum, D. R.; Winget, J. M.: Real Time Radiosity through Parallel Processing and Hardware Acceleration. Computer Graphics, Vol. 24/2, (1990) pp.67-75

[2] Budak, V. P.; Makarov, D., N.; Smirnov, P., A.: Přehled a porovnání počítačových programů pro navrhování osvětlovacích soustav, SVĚTLO 1/2006, Vol. 1, No. 1/2006, ISSN 1212-0812, pp. 50-54

[3] Chen, S. E.; Rushmeier, H.; Miller, G.; Turner, D.: A Progressive Multi-Pass Method for Global Illumination, Computer Graphics, Vol. 25/4, July 1991, pp. 165-174

[4] Cohen, M. F.; Greenberg D. P.: The Hemi Cube: A Radiosity Solution for Complex Environments. Symposium on Computational Geometry, 1985, pp. 31-40

[5] Fujiomoto, A; Tanaka, T.; Iwata, K.: ARTS Accelerated Ray Tracing System, IEEE C. G. & A., Vol. 6 (4), April 1986, pp. 16-26

[6] Goral, C. M.; Torrance, K. E.; Greenberg, D. P., Battaile, B.: Modeling the Interaction of Light between Diffuse Surfaces. Computer Graphics, 18(3), Vol. 18/3, 1984, pp. 213-222

[7] Klvač, P.: Logický postup při navrhování vnitřního umělého osvětlení, SVĚTLO 06/2008, ISSN 1212-0812

[8] Rybár, P. et al.: Denní osvětlení a oslunění budov, ERA group spol.s r.o., Brno, 2001, ISBN 80 – 86517 – 33 – 0, Brno, Czech Republic

[9] Sillion, F., Puech, C.: A General Two-Pass Method Integration Specular and Diffuse Reflection. Comp. Graphics, Vol. 23(3), 1989, Boston. p. 338

[10] Silion, F., Puech, C.: Radiosity and Global Illumination, Morgan Kaufmann, 1994

[11] Krupa, M.: Methods of Lightening for Workplaces In: Novus scientia 2005. Košice, Slovakia, ISBN 80-8073-354-6. pp. 215-220

[12] Smola, A.; Gašparovský, D.; Krasňan, F.: Navrhovanie vonkajšieho a vnútorného osvetlenia v nadväznosti na technické normy a právne predpisy, SAP, Bratislava, 2005, ISBN 80-89104-71-1

[13] Wallace, J. R.; Elmquist, K. A.; Haines, E. A.: A Raytracing Algorithm for Progressive Radiosity, Comp.Graphics, Vol. 23(3): pp. 315-324, 1989, Boston

[14] Tilinger, Á.; Madár, G.: Spectral Radiosity Rendering Application for Lighting Researches, Acta Polytechnica Hungarica Vol. 5, No. 3, 2008, pp. 141-145, ISSN 1785-8860

[15] Dénes, J.; Patkó, I.: Computation of Boundary Layers, Acta Polytechnica Hungarica Vol. 1, No. 2, 2008, pp. 79-87, ISSN 1785-8860

The Fight against Income Tax Evasion in Hungary

Zoltán Imre Nagy

Óbuda University
Népszínház u. 8, H-1081 Budapest, Hungary
E-mail: nagy.imre@kgk.uni-obuda.hu

Abstract: The evasion of income taxes causes the national budget of Hungary great damage. Just the partial prevention of it would greatly benefit the state balance and the realization of economic policy in Hungary. Repelling income tax evasion can only be achieved by various means. These are the following: To identify and abolish its reasons and causes; to post strict regulations consistent with EU legislation; to close legal gaps; not to concentrate on penalties; to reduce high taxes and the exorbitant income centralization of the state; to take the moral education for serious as well as the legal harmonization and effect analysis for the various decisions.

Keywords: evasion of income; national budget; reasons and causes; legal harmonization and effect analysis

1 Introduction

All over the world – even though not always with pleasure – taxes have to be paid. Tax evasion is causing problems today even in the most developed economies and in countries in which taxation systems have existed longer than is the case in Hungary. The tax evaders want to remain undetected by the tax offices by trying to make the conditions of their work and their earnings undetectable and invisible. This is why many experts use the terms of the invisible earnings and income, the invisible economy, the black economy as well as the underground economy. The term of grey economy is also used. This is understood to be a transitional form between the legal and the illegal economy. The higher the fraction of the invisible income is in comparison to the statistically proven earnings, the more often one describes the phenomenon as a black or grey underground economy. In my opinion, however, this terminology is not correct as it suggests an underlying independent structure, although a significant part of the black economy is very closely connected to the legal economy. Every owner of goods for which no invoice has been issued becomes both offender and victim at the same time, thus victim even in a dual sense. On the one hand, the buyer is not able to raise an

objection in the event of quality problems, if there are defects with the purchased goods or services. One the other hand, the person who makes use of illicit employment experiences disadvantages as a member of society, due to the fact that the reduced public revenue affects public services. That is problematic in particular when austerity measures are introduced as a consequence of reduced public revenue and increased national debt. As a result, the effectiveness and quality of public services should be increased with less financial means. Hence it is no coincidence that the national legislation in various countries ranks the group of people down who purchase without acquiring an invoice. This solution was also discussed in Hungary, but at present the population is encouraged to purchase and acquire an invoice by a lottery procedure. Each invoice for a service over 1000 HUF participates in a monthly lottery drawing. Excluded are citizens who work in municipal public utilities. This solution does not seem to be effective; more resolute measures of the government would be absolutely necessary.

The terminology of black economy can be looked at in terms of organisation, because the notion of black economy correctly indicates organisation. As examples, some of the biggest cases of tax evasion in Hungary will be analysed, such as the "invoice factory" in Vecsés, near Budapest, or the incidents which were colloquially referred to as the "oil blonding cases". In these cases diesel fuel for motor vehicles was produced by removal of the red colouring. Thereby mineral oil taxes as well as value added tax were avoided. This tax evasion normally ensured a high additional profit for the seller, very often as well for the buyer (In the end of the 1990s the capital of the oil blonding is said to have fled into the alcohol and drug industry due to the restrictive tax legislation). The term of a black economy (underground economy) – as also showed by the examples above – as well correctly points towards relations with crime. Income tax evasion very often leads to relationship with other criminal acts. It is well known that the notorious mafia boss Al Capone was finally arrested and convicted in the United States of America because of tax evasion. This illustrates how important it is, even for a criminal organisation such as the mafia, to respect the laws of taxation in a highly developed society of market economy. In the following the terms invisible income and tax evasion will be used instead of black economy, because these terms better express that no one is able to estimate well the societal losses due to the fact that the unwanted processes are indeed invisible.

2 The Size of the Invisible Income

According to the estimates of experts, the proportion of income tax evasion in developed industrial countries ranges between 5 and 20% in relation to the annual gross domestic product (GDP). In comparison to that, it is likely to be significantly higher in Hungary. Although there are some doubts regarding these approaches it is to notice that income tax evasion is particularly widespread in

Hungary compared to these countries. (Several experts even assume that the proportion of the invisible income in the developed industrial countries ranges between 20 and 50%.) In Hungary the GDP in 2007 was 25.374 billion HUF (the level of the GDP in 2008 was lower by 1%). Since the national income centralisation in Hungary amounts to more than 50% (in 2006 in Hungary the income centralisation was even 54,6%, which experts in the Central European region consider to be very high and harmful)[1], in Hungary one can calculate with an annual loss of national minimum income of 2.500 billion HUF (10 billion Euro). The tax centralisation in Hungary is about 38%, the overall national tax loss should be around at least 1.900 billion HUF each year (7 billion Euro). Of course the multiplier effect[2] alleviates this enormous sum, because it is not a matter of visible processes. This is a problem for Hungary and also it is made more difficult by the fact that 60% of the tax revenues and premium incomes originate from Budapest and the district of Pest (the surrounding area of Budapest). The reasons for the high level of invisible income in Hungary are traced back to various factors; among others, one fact that figures prominently is that Hungary's modern taxation system was only introduced in 1988 at the time of the change of regime.

Table 1

Tax centralisation in Hungary, the tax revenues and premium incomes of the Hungarian state budget in % of the GDP. Source: see Figure 1

Year	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
In% of GDP	39,1 1	39,2 2	38,6 3	38,34 4	37,9 5	37,7 6	37,5 7	37,2 8	38,2 9	39 10	39,8 11

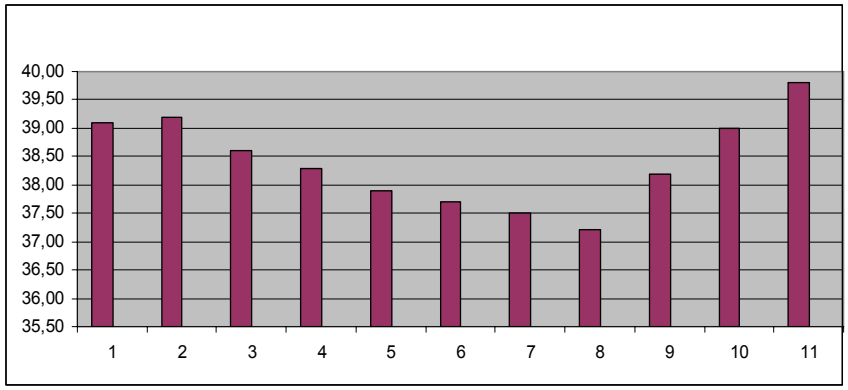


Figure 1

Tax centralisation in Hungary 1999-2008, 1 → 1999

[1] See lecture of Békesi 2006

[2] See paragraph 3

Sources: 1) Békesi, lecture, Mindentudás Egyeteme 2006: http://www.mindentudas.hu/bekesilaszlo/20060917bekesi1.html?pIdx=3

2) MTI June 26 2009: http://www.origo.hu/uzletinegyed/valsag/20090626-oszko-33-szazalekracsokkenhet-az-adocentralizacio.html

The extent of the Hungarian tax centralisation is slightly lower than the average tax centralisation of the EU countries (39-40%). Considering the rate of Hungary, however, in comparison with those of the Baltic countries and of Ireland, it is clear that the extent of tax centralisation in Hungary is essentially higher. Moreover, it is even higher than the level of the Visegrád states (Poland, Slovakia, Czech Republic and Hungary) (30-36%).

3 The Reasons for Income Tax Evasion

In general, it is correct that moral scruples against income tax evasion are decreasing globally. This fact is certainly related to the increasing egoism and decreasing collective spirit. In analysing the reasons, however, firstly the rational, economic, so called traditional models, must be mentioned, to explain income tax evasion as a risk charged decision problem, where the taxpayer wants to increase his income also by means of income tax evasion, while he probably takes the following aspects into consideration: the probability of detection, the extent of monetary fines and other penalties, the general tax burden, the level and the slope of the graduated tariff and the amount of the tax free income, as well as the amount of the real income.

The increased probability and the enhanced risk of detection decrease the amount of cases of income tax evasion. The successful inspections of fiscal authorities provide a very high degree of prevention. However, international researchers have not been able to prove statistically that the extent of penalties (surcharges and fines) had a significant effect on cases of income tax evasion.

The answer is quite simple. The determination of the penalties is naturally one of the competences of the legislature, while the inspection of the taxpayers remains a task of the tax offices. The parliament and the judiciary seem not to be capable of adjusting the penalties for income tax evasion with an appropriate speed. Thus no positive effects are achieved and the unwanted processes are not pushed back. Of course in the parliament the political aspects of the decisions are important. A penalty, on the other hand, for income tax evasion does not significantly have deterrent effects.

In researching the causes of income tax evasion, more and more psychological, sociological and other aspects are being brought into consideration. This is correct, as it is undisputed that the particular factors which influence income tax evasion are emphasised by psychological and sociological regularities of the behaviour of the taxpayers.

Worldwide the view is widespread that one of the main motives for income tax evasion is the high average tax burden. Moreover other elements added to high taxes lead to an even higher illegalisation of the income (health care and pension insurance contributions, employer's and employee's contribution to social insurance). As shown above, the Hungarian tax burden is particularly high in comparison to the neighbouring and competing eastern European countries. Experiences at the international level demonstrate that the taxpayers react to an increase of the average tax burden of 1% with an increase of tax income evasion of 8%. This stochastic interrelation naturally depends on different country-specific factors, but it illustrates that the extent of income tax evasion as a reaction to an increase in taxation is much greater than the extent of the original tax increase. Many taxation experts claim that this allegation is also true the other way around. This means: with a decrease in taxation income tax evasion would decrease as well and the state could in this case receive higher revenue. However, this cannot be proven, especially not in the short term. In my opinion there are highly complex connections which determine whether taxpayers are motivated towards income tax evasion.

Table 2

Hungarian graduated tariffs of the income tax for individuals 2005-10 in HUF

2005		
Tax base	Fixed tax	Tax %
0- 1 500 000		18
1 500 001-	270 000 +	38
2006		
Tax base	Fixed tax	Tax %
0- 1 550 000		18
1 550 001-	279 000 +	36
2007		
Tax base	Fixed tax	Tax %
0- 1 700 000		18
1 700 001-	306 000 +	36

2008		
Tax base	Fixed tax	Tax %
0- 1 700 000		18
1 700 001-	306 000 +	36
2009		
Tax base	Fixed tax	Tax %
0- 1 900 000		18
1 900 001-	342 000 +	36
2010		
Tax base	Fixed tax	Tax %
0- 5 000 000 *		17
5 000 000-	850 000+	32
*Super gross tax base		

In the years 2007 and 2008 a so-called special tax in the amount of 6.325.450 HUF (+ 4%) was raised. This is equivalent to a total tax burden of 40%. From 2002 to 2006 the minimum wage was tax-free by means of an income tax credit up to a monthly income of 62.500 HUF (minimum wage in 2006). While the minimum wage was increased to 65.500 HUF in 2007, to 69.000 HUF in 2008 and to 71.500 HUF in 2009, the amount of the tax credit was not changed. So on the minimum wage a higher and higher tax was imposed, even though it is only subject to a low tax burden of 1-2%.

Table 3
Inflation rate in the so called Visegrád countries and in Austria in %

Source: KSH (Central Statistical Office Hungary)

Country	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Austria	0,5	2,0	2,3	1,7	1,3	2,0	2,1	1,7	2,2	3,2
Poland	7,2	10,1	5,3	1,9	0,7	3,6	2,1	1,3	2,6	4,2
Slovakia	10,4	12,2	7,2	3,5	8,4	7,5	2,8	4,3	1,9	3,9
Czech Rep.	1,8	3,9	4,5	1,4	-0,1	2,6	1,6	2,1	3,0	6,3
Hungary	10,0	10,0	9,1	5,2	4,7	6,8	3,5	4,0	7,9	6,0

After in Hungary the tax burden – due to the missing adjustment of the graduated tariffs table – was increased in parallel with the inflation rate, this factor had increasing effects on income tax evasion.[3] In Hungary the minimum and the average wage are proportionally low[4] and even with a relatively low wage the income is taxed according to the higher tax zone. The presented tax tables show that the government has raised only the upper limit of the tax zones of 18% all in all three times to a small extent in the last five years. The largest zone broadening of the smaller tax rates was 200.000 HUF. This is an annual tax saving of 36.000 HUF (130 Euro annually). That is very little and it cannot be seen as an enhancing of value, especially not when one takes Hungary's inflation rate into consideration.[5]

Table 4
Monthly average wage and minimum wage in Hungary in HUF

Source: KSH (Central Statistical Office Hungary)

Year	2001	2002	2003	2004	2005	2006	2007	2008
Average wage	103 553	122 482	137 193	145 520	158 343	171 351	185 017	198 942
Minimum wage	40 000	50 000	50 000	53 000	57 000	62 500	65 500	69 000

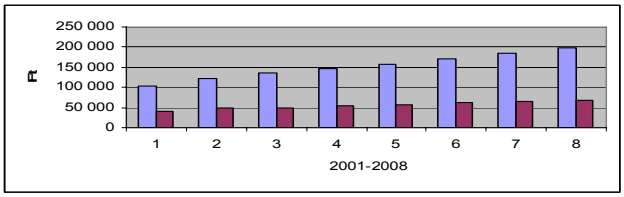


Figure 2
Monthly average wage and minimum wage in Hungary in HUF, Source: KSH (Central Statistical Office Hungary)

[3] See Table 2
[4] See Table 4
[5] See Table 3

Table 5
Monthly average and minimum wage in EU countries (January 2008), in Euro

Source: Pályacsúcs Magazine Hungary Feb 19 2008

Country	Minimum wage	Average wage	Country	Minimum wage	Average wage
Bulgaria	112	228	Spain	600	1500
Romania	138	431	Malta	617	1234
Lithuania	203	534	Greece	668	1712
Slovakia	224	658	Austria	1000	-
Latvia	227	687	France	1280	2723
Hungary	258	678	Belgium	1283	3207
Estonia	278	842	The Netherlands	1317	2863
Poland	282	854	Great Britain	1381	3732
Czech Republic	288	738	Ireland	1499	2882
Portugal	426	1065	Luxembourg	1570	3140
Slovenia	538	1169	**Average**	**676**	**1470**

The occurrence, the reinforcement and the continuance of income tax evasion is caused and supported by the impacts of a couple of factors. It is normal that income tax evasion often takes place in small organisations. There exists a direct relation between sellers, buyers, employers and employees. Over the past several years the number of small businesses has increased as well in highly developed countries. These small organisations "operate" within the network of the larger and provide country-specific "methods of cost cutting". In many cases the boundaries blur between the legal, half legal (grey) and illegal (black) activities. Very often the organisation itself is legal, but a part of its activities is illegal. The spread of income tax evasion is often brought into clear relation with the increase of small organisations compared to the total number of businesses. For example in Hungary large corporations are willing to employ people by means of a trade license for self employed entrepreneurship. That is how foreign big concerns very often bypass the payment of social insurance contributions. The legislation had to interfere, and the representatives interpreted these "business contracts" as fictitious contract. In cases in which the income of the small businesses was from just one client on a regular basis, the Hungarian occupational safety authority took actions against the suspects. (In addition, it is remarkable that the principal as well as one leading employee of the Hungarian occupational safety authority have been sitting in custody for months). There is an obvious relation between the increase of unemployment and poverty as well as the spread of income tax evasion. According to Hungary's Central Statistical Office (KSH) the number of unemployed persons sank in the third quarter of 2007 to 296.900. In the first quarter of 2009 the number of unemployed persons was 402.800. This is equivalent to an unemployment rate of 9.7%. In comparison to this the rate in the first quarter of 2007 was "only" 7.7%.

Table 6
Unemployment rate in the first quarter of 2009 in %, KSH (Central Statistical Office Hungary)

Austria	4,7	Czech Republic	5,8
Poland	8,4	Hungary	9,7
Slovakia	10,4	**EU-27**	**8,8**

It is very interesting that very often families with a high income take the risk of income tax evasion. This fact indicates inter alia that here the joint impact of several factors is involved. One of the factors is to cover the costs of living. However, this is by far not the most important factor. This means that the common standard of living considered by itself does not help much to disintegrate the black economy. Experiences on an international level show that privatisation increases the amount of cases and the extent of corruption and bribery. According to Transparency International, in 2009 only four of the tested 36 countries proved to be active in the fight against corruption and bribery. 21 countries – including Hungary – did little or nothing to realise the OECD Anti-Bribery Convention. In Hungary the situation of fighting corruption got worse due to a lack of an existing legal framework. Consequently Hungary fell back in the ranking of the international corruption index (CPI) by 5,1, from the 39th rank (2008) to the 47th rank (2009). In the Eastern and Central European region Hungary was overtaken by Slovenia and Estonia in 2008 and by the Czech Republic in 2009. As breeding grounds for the Hungarian corruption, Transparency International emphasized two reasons. One important factor is the opacity of the financing of the political parties and the campaigns. Hungary's parties make use of a multiple of the sources or financing which are defined by law. The other factor is the overcomplicated system of public procurement. The higher the proportion of the total income that is well-documented, the lower is the extent of income tax evasion. The increase of the part of the taxpayers above the employable age also has very positive effects. The explanation for this is, besides obvious psychological reasons, that the part of the population which does not work due to their high age mainly shows a well documented income.

As a result of the factors demonstrated so far, it can be shown that the probability of detection can be a factor which can be used to exert essential influence on the unwanted processes. Thorough international research points out that the doubling of amount of tax controls and inspections only causes a 15% reduction in income tax evasion. It is surprising what a small effect this instrument has. The fiscal authorities in numerous countries of the world are struggling with huge problems, such as, for example, the relatively bad remuneration in the tax offices, where there "is no perspective for the staff, just the eternal, boring treadmill". In addition the public in general judges the laborious work of the tax authorities as negative. Also the work is made more difficult by the complexity of the taxation law. German tax experts brought to my attention that the taxation law is a set of rules that drown in its own numerous exceptions. The same problem exists in Hungary, too. The consequence is a high fluctuation and a shortage of workforce in the tax

offices. The tax advisers take up the battle with the tax authorities in well-organised and well-equipped offices for high salaries. Often the best specialists just switch over to the other side. Of course it is worthwhile to increase the amount of inspections and to enhance the controls by the tax offices even under poor conditions. However, it is obvious that the remaining resources are insufficient therefore.

It is not sufficient and effective to try to solve the problem of the black economy only superficially by imposing penalties and sanctions. The fight against the black economy can only be successful if the reasons for the problem are detected and eliminated.

The restricted effectiveness of the financial inspections draws the attention towards psychological and sociological factors. A great number of researchers have analysed for instance the interrelation between the education of the taxpayers and the extent of income tax evasion. Taxpayers with lower qualifications are less knowledgeable in the area of taxation rules and do not know much about the penalties and sanctions, nor about the consequences of income tax evasion. This statement also holds true the other way round. Highly qualified taxpayers are in a position to better estimate the impact of their acts. On the other hand, it is also true that the well-qualified strata of the population better know their way around the jungle of taxation rules, so that they are in a better position to cheat. As a result of the mutual impact of the contradictory factors, it arises that cases of tax evasion are rarer among highly qualified taxpayers. This also means that an increase in tax inspections can only influence a small proportion of the taxpayers. This statistically proven piece of evidence is partially an explanation for the fact that no more than a decrease of 15% of income tax evasion can be achieved by doubling the amount of tax inspections. The frequent inspections produce active reactions of the taxpayers as they often experience during the examinations and hearings that the tax authorities can only detect a fraction of the total cases of income tax invasion. This experience can then later become an additional motivational factor. It is also wise to finish all inspections, where the tax authorities exceeded the usual control resources, possibly with detection and with a fact of evidence. The small level of information the population receives regarding penalties partly explains why the penalties are deterrent only to a limited extent. The international literature recommends punishing a particularly bad tax evasion with high monetary fines and imprisonment in order to force back income tax evasion. Until now international researchers were not able to provide statistically significant information about the correlation of particular population strata (single and married persons, women or men, foreign or domestic employees, casual or permanent labour, and groups with different religious backgrounds) and income tax evasion.

In a part of the highly developed industrial countries the tax lists, which are generated by self-administration, are being published. The possibility of insight by third parties reduces the extent of income tax evasion.

4 Specific Reasons for Income Tax Evasion in Hungary

In addition to the presented common international factors, there are also numerous other factors which contribute to an increase in income tax evasion in Hungary. While those reasons are also applicable to other countries, these are more specific and likely to apply to Hungary. The begin of the "second economy" in Hungary made the transformation to market economy essentially easier (From the beginning until the end of the 1990s Hungary and Slovenia were the driving forces in the extension of the market economy in Central and Eastern Europe). Through the "second economy" (economic joint ventures in industry and trade as well as individual home economies for agricultural production cooperatives and state farms) it was possible in Hungary to gain experience with market economy. In addition certain adaptability was developed as a reaction to the economy of scarcity of those times. The second economy was the only possibility to make autonomous decisions and was an important part of the reform process. However, it was not the best starting point for the development of the market economy, because the participants of the second economy could not plan with a long term perspective due to the frequent rule changes. Instead, they tried to maximise profit in the short term. This inevitable philosophy of profit maximisation was opposed to the philosophy of long-term orientated and sustainable management. Another problem was the increasing number of bribe payments and corruption, which became routine due to insecurities and problems of supply in the country. These negative elements have been solidly entrenched in the economic life since then. Nevertheless, it must be considered that before the change of system there existed no income tax in the Eastern European region. The incomes were determined in such a way that these neither had to cover several taxes nor contributions. In those times, the need of the main part of society could be satisfied from the income of main occupation. The purpose of an occupation in the second economy was "only" to secure additional income. However the introduction of the new taxation system in 1988 was not succeeded by a consequent change in the income system in Hungary. This means that the citizens needed to pay more and more taxes and contributions (social insurance contributions) with a continually diminishing quality of the services paid for. When the incomes were composed there was no inflationary adjustment. This resulted in a situation where the population spent considerably more money than they earned, so that the citizens had to tighten their belts and economise their way of living. One possible consequence of this economical behaviour is income tax evasion. Tax evasion and illicit work among employers is to some extent to be seen as a result of the chronic scarcity of capital and the consequential shortage of credit in Hungary. In this way the employers wanted (and could) accumulate capital owing to the liability to pay high taxes and contributions.

It is important to underline that the duties and taxes from the well-functioning private businesses at the time of the change of system continued to increase, caused by the bad performance of the public enterprises. The newly privatised enterprises reacted to that as well in Hungary with income tax evasion. In addition to the facts presented above, the very low tax morality can also be explained historically. During the time of the Turkish occupation, the avoiding of tax payment was considered to be brave and noble. An attitude similar to that can be found today as well as resistance against the authority of the state. In general the Hungarian public moral appraises this move as a positive one.

Furthermore, the distribution of the invisible income in Hungary is affected negatively by legal loopholes, which were particularly characteristic of the transitional phase. Additional factors are the difficult application of existing penalties and sanctions as well as the fact that trials take too long. One has to wait very long for a legally binding judgment (often more than 5 years). Obviously also in the case of income tax evasion the legal framework is crucial. One of the most important characteristics of income tax evasion is the bypassing of the laws and tax payments by the offender. Income tax evasion can often be looked at within one single country. The international market and the phenomenon of globalisation facilitate the bypassing of taxation of international capital. The capital can change its location and start and combine with different subsidiaries, branches and offshore companies in different countries. The term offshore has received a negative connotation in Europe – in particular because of the problematic nature of job relocations to Central and Eastern European as well as to Asian countries. Hungary is severely affected by that. The extent of income tax evasion by offshore companies cannot be assessed. An international union is in emerging against tax havens. As well, the Hungarian legislature is planning laws against the formation of offshore companies. Already in the year 2009, the so-called amnesty rules came into effect (originally the rules were in force until June 30, 2009, but this was extended until the end of the year). An amnesty rule refers to the loan of the shareholder. If the shareholder renounces for the benefit of the enterprise, no gift tax is raised. Thus he can legalise his own gift free of profit tax, paying just 10% withholding tax (instead of 25% dividend tax) and hence legally acquire his own present. This demonstrates the problem that the loans of shareholders often are only existent on paper. Another amnesty rule refers to such enterprises which repatriate their money from offshore companies and invest half of this money for 2 years in Hungarian government bonds. After this preferential treatment the enterprise has only to pay 25% of the usual taxes. The third amnesty rule refers to private persons who repatriate interest income, dividends or exchange profits from tax havens. The taxpayer pays a withholding tax in the amount of 10% and has to invest the half of the money for 2 years in Hungarian government bonds in order to be relieved from all consequences (also an "enrichment check" by the tax office). Tax authorities expect national tax revenues of 75 billion HUF from this amnesty law. A Hungarian citizen referred to the 2009 amnesty law in Hungary with the following statement: "I am asking for amnesty, too, because my income

was taxed for the last 38 years right down to the last cent. I did not buy myself a new vehicle with value added tax refunding, did not deduct operational costs from my tax base, did not account for my kitchen furniture as costs, did not live on minimum wage for years, while living in a luxury apartment and driving a luxury car... (as some businessmen do). After the ceased use of my missed amounts I am asking for a minimum compensation of 10%, immediate, tax-free cash on the nail." Nota bene: Following economic logic one could even demand a higher compensation. Further interesting ideas of the Hungarian government are: the taxation of the family allowance and property tax (real estate tax) under the symbol of justice, also to make profits from gambling tax-free for the winner and afterwards to raise a 5% tax according to the basis of the winnings (under the symbol of the "whitening" of the black economy". With effect July 1, 2009 the government raised the general rate of the value added tax from 20% to 25%, and the advantaged tax rate from 15% to 20%. At the same time taxes on mineral oil, tobacco and alcohol are increasing. On the one hand, these increases are not desirable from the perspective of increasing income tax evasion. And on the other hand these increases intensify the impacts of the financial crisis and consequently imply cost inflationary effects. A consequence of this can be several years of stagflation. The government regards the upcoming introduction of the property tax (real estate tax), according to a statement of the finance minister, as an "important tool in fighting the crisis".

While the state correctly was forced out of the market gradually, as a matter of course it has to secure the social services. For that reason high tax revenue is needed. However, it cannot be indifferent to its origin and nature – from whom and how it is achieved.

For the given reasons for income tax evasion and after analysing its nature, we conclude that income tax evasion in Hungary has increased in recent years and is very likely to exceed current estimates.

Conclusions

The particular areas of the invisible income can also be brought into being by maladjustment and mismanagement, and/or can be amplified. They are more or less necessary side effects of modern processes. The forcing back of the black economy – which is necessary everywhere – can only be achieved with various measures. Criminal actions such as drug trafficking or smuggling must be prevented by the organs of criminal prosecution. "According to the principal of the Hungarian office for national security, Sándor Laborc, the power of the specific criminal associations and organisations has increased in the last year ... and the frequency of occurrence of acts of violence will increase ... Increasingly strong is the local Hungarian Mafia. ... The criminal organisations are probably going to increase the veiled contact to the public administration..."[6] Strong

[6] See interview with Mr. Laborc

regulations in accordance with the EU legislation need to be created with the aim of closing the legal gaps. However, here it is necessary to treat the different situations appropriately and circumspectly. It must not be merely concentrated on penalties and sanctions. As an important basic principle for legislation, it must be emphasized that only such rules can be introduced which can be possibly complied with and be followed.

It is practical to reduce over-taxation and exaggerated national income centralisation. Obviously, not only fiscal measures are needed but also other regulating instruments. The black market needs to be forced back by all means. It is not sufficient and little effective to only tackle the surface of the problem by imposing sanctions. The fight against the black economy can only be successful if reasons and causes are detected and eliminated. In doing so, long-term-orientated methods also need to be considered, such as, for instance, moral education, the harmonisation of legislation, and impact analyses for the particular decisions. It should be realised that income tax evasion is a part of the modern market economy. We have to cope with it, while we of course keep on trying to proceed against the negative impacts of income tax evasion. But even income tax evasion has positive aspects. A means of fighting against it can be to seize the positive effects temporarily and not to remove them. This is not possible in every case. Positive impacts are that income tax evasion provides the poor, who are living under the subsistence minimum, an additional income. Thus they can ensure that they have means to live (illicit work). The trade of goods of bad quality and no guarantee reduces poverty, because these goods are available for all. The illicit work facilitates the survival of unemployment and creates new "jobs", which would otherwise not emerge due to the liabilities to pay high taxes and contributions and would therefore not be profitable. The illicit work also keeps alive such businesses as would have to go bankrupt because of the tax burden. The black economy secures very often higher incomes and greater independence than the legal economy. It is able to react quickly to shifts in demand and provide the desired goods and services. This is important especially in times of crisis and recession. It is also true that only a part of the population can benefit from these positive effects, while the greater part of the population meets more and more severe difficulties. The majority of the people are not capable to supplement their income in this way. Of course it would be better if this all would happen under legal circumstances. But until this all can be legalised, also these positive phenomena should be considered as successes in the fight against income tax evasion. In the reduction of income tax evasion we can and we must proceed determinedly but carefully.

References

[1] Amnesty law in Hungary, http://www.fntudosito.hu/riport/1639

[2] Békesi, László Dr.: Lecture, 2006, in Hungarian, http://www.mindentudas.hu/bekesilaszlo/20060917bekesi1.html?pIdx=3

[3] Beseitigung grenzübergreifender steuerlicher Hindernisse in der EU IWB 1/2011 S. 6 (NWB DokID: MAAAD-59245).in German Cowell, Franka A.:The economics of evasion, Cambridge,1990.ISBN 0-262-03153-1

[4] Engström, P., Holmlund, B. (2006): "Tax Evasion and Self-Employment in a High-Tax Country: Evidence from Sweden", CESifo working paper No. 1736

[5] Hardeck, Inga: Steuerhinterziehungsbekämpfungsgesetz - Regelungsinhalt und Implikationen für die Praxis, IWB Fach 3 Gruppe 1 Land Deutschland S. 2431; IWB 16/2009 S. 781 (NWB DokID: UAAAD-27142)

[6] Konz, Franz: 1000 ganz legale Steuertricks. Knauer Verlag 2010 ISBN 978-426-78330-6, in German

[7] Krekó, Judit, Kiss, P. Gábor: Tax evasion and tax changes in Hungary. MNB Bulletin April 2008, in English

[8] Krekó, Judit, Kiss, P. Gábor: Tax evasion and the Hungarian tax system), MNB Occasional Papers No. 65, in English

[9] Laborc, Sándor Dr.: The organised crime can integrate itself into the public administration. Radio statement of the principal of the Hungarian office for national security, April 20 2009 in Hungarian. http://www.mr1-kossuth.hu/index.php?option=com_content&task=view&id=79778

[10] Nagy, Imre Zoltán, Dr.: Die Gründe der Einkommenshinterziehung. MEB, 7th International Conference Budapest,2009. Proceedings p.185.in German

[11] Nagy, Imre Zoltán, Dr.: The psychological and sociological aspects of tax evasion. Adó (Tax), No. 1-2, Year X. 1996. January. In Hungarian

[12] Schneider, F. (1994): "Can the shadow economy be reduced through major tax reforms? An empirical investigation for Austria", supplement to Public Finance 49, pp. 137-152, in English

[13] Statistical data: www.ksh.hu

[] Vértes, András Dr.: Tax system and competitiveness, 2008, in Hungarian, http:wwww.magyarorszagholnap.hu/pdf/5_versenykepesseg.pdf

[14] Villalba, Miguel Sánchez: Tax evasion as a global game. Published by: Instituto Valenciano de Investigaciones Económicas, S.A. Printed in English, February 2010

Language for a Distributed System of Mobile Agents

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: Types with behavioral scheme for mobile ambients are suitable for expressing the dynamic properties of mobile code applications, where the main goal is to avoid the ambiguities and possible maliciousness of some standard ambient constructions. We can statically specify and check access rights for the authorization of ambients and threads to communicate and move. We define a language which expresses software agents migration in the space of distributed places. This allows us to understand various aspects of code mobility.

Keywords: code mobility; software agents; type system

1 Introduction

Communication between mobile ambients [1] based on a concurrency paradigm represented by π-calculus [2] is represented by the movement of other ambients of usually shorter life which have their boundaries dissolved by an open action to expose their internal threads performing local communication operations. Such capability of opening an ambient is potentially dangerous [3, 4, 5]. It could be used inadvertently to open and thus destroy the individuality of an object or mobile agent. Remote communication is usually emulated as a movement of such ambients (communication packages) in the hierarchy structure.

We intend to keep the purely local character of communication so that no hidden costs are present in the communication primitives, but without open operation. This solves the problem of the dissolving boundaries of ambients, but disables interactions of threads from separate ambients. We must introduce a new operation move for moving threads between ambients. The idea comes from mobile code programming paradigms [6] where moving threads can express strong mobility mechanism, by which the procedure can (through move operation) suspend its execution on one machine and resume it exactly from the same point

on another (remote) machine. This solves the problem of threads mobility and by moving threads between ambients we can emulate communication between the ambients.

The advantages of our approach are shown in the natural way of encoding the semantics of language for adistributed system of mobile agents. First, we discuss the code mobility for better understanding and then we show how to naturally express objective and subjective mobility implemented in various software applications. Respecting all aspects of code migration paradigms, we are able to propose the language for mobile agents distributed system specification.

2 Revised Calculus of Mobile Ambients

Abstract syntax and operational semantics of our calculus are based on abstract syntax and operational semantics of ambient calculus including our new constructions.

2.1 Abstract Syntax

The abstract syntax of the terms of our calculus is the same as that of mobile ambients except for the absence of open and the presence of the new operation *move* for moving threads between ambients. We allow synchronous output and the asynchronous version is its particular case. The abstract syntax consists of two domains:

M	::=	mobility operations
	n	Name
	$in\ M$	move ambient into M
	$out\ M$	move ambient out of M
	$move\ M$	move thread into M
	$M.M'$	Path
P	::=	Processes
	$\mathbf{0}$	inactive process
	$P \mid P'$	parallel composition
	$!P$	Replication
	$M[P]$	Ambient
	$(\nu n : \mathbf{P}[\mathcal{B}])P$	name restriction
	$M.P$	action of the operation

	$\langle M \rangle.P$	synchronous output
	$(n : \mu).P$	synchronous input

2.2 Operational Semantics

The operational semantics is given by reduction relation along with a structural congruence in the same way as those for mobile ambients.

Each name of the process term can figure either as free:

$$fn(n) = \{n\} \qquad fn(\mathbf{0}) = \varnothing$$
$$fn(in\ M) = fn(M) \qquad fn(P \mid P') = fn(P) \cup fn(P')$$
$$fn(out\ M) = fn(M) \qquad fn(!P) = fn(P)$$
$$fn(move\ M) = fn(M) \qquad fn(M[P]) = fn(M) \cup fn(P)$$
$$fn(M.M') = fn(M) \cup fn(M') \qquad fn((\nu n : \mathbf{P}[\mathcal{B}])P) = fn(P) - \{n\}$$
$$fn(M.P) = fn(M) \cup fn(P)$$
$$fn(\langle M \rangle.P) = fn(M) \cup fn(P)$$
$$fn((n : \mu).P) = fn(P) - \{n\}$$

or bound:

$$bn(n) = \varnothing \qquad bn(\mathbf{0}) = \varnothing$$
$$bn(in\ M) = bn(M) \qquad bn(P \mid P') = bn(P) \cup bn(P')$$
$$bn(out\ M) = bn(M) \qquad bn(!P) = bn(P)$$
$$bn(move\ M) = bn(M) \qquad bn(M[P]) = bn(M) \cup bn(P)$$
$$bn(M.M') = bn(M) \cup bn(M') \qquad bn((\nu n : \mathbf{P}[\mathcal{B}])P) = bn(P) \cup \{n\}$$
$$bn(M.P) = bn(M) \cup bn(P)$$
$$bn(\langle M \rangle.P) = bn(M) \cup bn(P)$$
$$bn((n : \mu).P) = bn(P) \cup \{n\}$$

We write $P\{n \leftarrow M\}$ for a substitution of the capability M for each free occurrences of the name n in the term P. Then similarly for $M\{n \leftarrow M\}$.

Structural congruence is standard for mobile ambients:

- equivalence

$$P \equiv P \qquad \text{(SRefl)}$$
$$P \equiv Q \Rightarrow Q \equiv P \qquad \text{(SSymm)}$$
$$P \equiv Q, Q \equiv R \Rightarrow P \equiv R \qquad \text{(STrans)}$$

- congruence

$P \equiv Q \Rightarrow P \mid R \equiv Q \mid R$ (SPar)

$P \equiv Q \Rightarrow !P \equiv !Q$ (SRepl)

$P \equiv Q \Rightarrow M[P] \equiv M[Q]$ (SAmb)

$P \equiv Q \Rightarrow (\nu n : \mathbf{P}[\mathcal{B}])P \equiv (\nu n : \mathbf{P}[\mathcal{B}])Q$ (SRes)

$P \equiv Q \Rightarrow M.P \equiv M.Q$ (SAct)

$P \equiv Q \Rightarrow \langle M \rangle.P \equiv \langle M \rangle.Q$ (SCommOut)

$P \equiv Q \Rightarrow (n : \mu).P \equiv (n : \mu).Q$ (SCommIn)

- sequential composition (associativity)

$(M.M').P \equiv M.M'.P$ (SPath)

- parallel composition (associativity, commutativity and inactivity)

$P \mid Q \equiv Q \mid P$ (SParComm)

$(P \mid Q) \mid R \equiv P \mid (Q \mid R)$ (SParAssoc)

$P \mid \mathbf{0} \equiv P$ (SParNull)

- replication

$!P \equiv P \mid !P$ (SReplPar)

$!\mathbf{0} \equiv \mathbf{0}$ (SReplNull)

- restriction and scope extrusion

$n \neq m \Rightarrow (\nu n : \mathbf{P}[\mathcal{B}])(\nu m : \mathbf{P}[\mathcal{B}'])P \equiv (\nu m : \mathbf{P}[\mathcal{B}'])(\nu n : \mathbf{P}[\mathcal{B}])P$ (SResRes)

$n \notin fn(Q) \Rightarrow (\nu n : \mathbf{P}[\mathcal{B}])P \mid Q \equiv (\nu n : \mathbf{P}[\mathcal{B}])(P \mid Q)$ (SResPar)

$n \neq m \Rightarrow (\nu n : \mathbf{P}[\mathcal{B}])m[P] \equiv m[(\nu n : \mathbf{P}[\mathcal{B}])P]$ (SResAmb)

$(\nu n : \mathbf{P}[\mathcal{B}])\mathbf{0} \equiv \mathbf{0}$ (SResNull)

- garbage collection

$(\nu n : \mathbf{P}[\mathcal{B}])n[\mathbf{0}] \equiv \mathbf{0}$ (SAmbNull)

In addition, we identify processes up to renaming of bound names (α-conversion):

$(\nu n : \mathbf{P}[\mathcal{B}])P = (\nu m : \mathbf{P}[\mathcal{B}])P\{n \leftarrow m\} \quad m \notin fn(P)$ (SAlphaRes)

$(n : \mu)P = (m : \mu)P\{n \leftarrow m\} \quad m \notin fn(P)$ (SAlphaCommIn)

The reduction rules are those for mobile ambients, with the obvious difference consisting in the synchronous output and the missing open operation, and with the new rule for the *move* operation similar to the "migrate" instructions for strong code mobility in software agents:

- basic reductions

$n[in\ m.P \mid Q] \mid m[R] \rightarrow m[n[P \mid Q] \mid R]$ (RIn)

$m[n[out\ m.P \mid Q] \mid R] \rightarrow n[P \mid Q] \mid m[R]$ (ROut)

$n[move\ m.P \mid Q] \mid m[R] \rightarrow n[Q] \mid m[P \mid R]$ (RMove)

$(n : \mu).P \mid \langle M \rangle.Q \rightarrow P\{n \leftarrow M\} \mid Q$ (RComm)

- structural reductions

$P \to Q \Rightarrow P \mid R \to Q \mid R$ (RPar)

$P \to Q \Rightarrow n[P] \to n[Q]$ (RAmb)

$P \to Q \Rightarrow (\nu n : \mathbf{P}[\mathcal{B}])P \to (\nu n : \mathbf{P}[\mathcal{B}])Q$ (RRes)

$P' \equiv P, P \to Q, Q \equiv Q' \Rightarrow P' \to Q'$ (RStruct)

3 Type System with Behavioral Scheme

The restriction of the mobility operations is defined by types applying a *behavioral scheme*. The scheme allows for setting up the access rights for traveling of threads and ambients in the ambient hierarchy space of the system.

We define types where we present communication types and message types:

$\kappa ::=$	communication type
$\mid \perp$	no communication
$\mid \mu$	communication of messages of type μ
$\mu ::=$	message type
$\mid \mathbf{P}[\mathcal{B}]$	process with behavioral scheme $\mathcal{B}$
$\mid \mathbf{O}[\mathcal{B} \mapsto \mathcal{B}']$	operation which changes behavioral scheme $\mathcal{B}$ to $\mathcal{B}'$

The behavioral scheme is the structure $\mathcal{B}= (\kappa, Reside, Pass, Move)$ which contains four components:

- κ is the communication type of the ambient's threads.
- $Reside$ is the set of behavioral schemes of other ambients where the ambient can stay.
- $Pass$ is the set of behavioral schemes of other ambients that the ambient can go through, it must be $Pass \subseteq Reside$.
- $Move$ is the set of behavioral schemes of other ambients where the ambient can move its containing thread.

3.1 Typing Rules

The type environment is defined as a set $\Gamma = \{n_1 : \mu_1, \ldots, n_l : \mu_l\}$ where each $n_i : \mu_i$ assigns a unique type μ_i to a name n_i.

The domain of the type environment is defined by:

$$Dom(\varnothing) = \varnothing \qquad Dom(\Gamma, n : \mu) = Dom(\Gamma) \cup \{n\}$$

We define two type formulas for our ambient calculus:

$$\Gamma \vdash M : \mu \qquad \Gamma \vdash P : \mathbf{P}[\mathcal{B}]$$

Typing rules are used to derive type formulas of ambient processes:

$$\frac{n : \mu \in \Gamma}{\Gamma \vdash n : \mu} \quad \text{(TName)}$$

$$\frac{\Gamma \vdash M : \mathbf{P}[\mathcal{B}] \quad \mathcal{B} \in Pass(\mathcal{B}')}{\Gamma \vdash in\ M : \mathbf{O}[\mathcal{B}' \mapsto \mathcal{B}']} \quad \text{(TIn)}$$

$$\frac{\Gamma \vdash M : \mathbf{P}[\mathcal{B}] \quad \mathcal{B} \in Pass(\mathcal{B}') \quad Reside(\mathcal{B}) \subseteq Reside(\mathcal{B}')}{\Gamma \vdash out\ M : \mathbf{O}[\mathcal{B}' \mapsto \mathcal{B}']} \quad \text{(TOut)}$$

$$\frac{\Gamma \vdash M : \mathbf{P}[\mathcal{B}] \quad \mathcal{B} \in Move(\mathcal{B}')}{\Gamma \vdash move\ M : \mathbf{O}[\mathcal{B} \mapsto \mathcal{B}']} \quad \text{(TMove)}$$

$$\frac{\Gamma \vdash M : \mathbf{O}[\mathcal{B}'' \mapsto \mathcal{B}'] \quad \Gamma \vdash M' : \mathbf{O}[\mathcal{B} \mapsto \mathcal{B}'']}{\Gamma \vdash M.M' : \mathbf{O}[\mathcal{B} \mapsto \mathcal{B}']} \quad \text{(TPath)}$$

$$\frac{}{\Gamma \vdash \mathbf{0} : \mathbf{P}[\mathcal{B}]} \quad \text{(TNull)}$$

$$\frac{\Gamma \vdash P : \mathbf{P}[\mathcal{B}] \quad \Gamma \vdash P' : \mathbf{P}[\mathcal{B}]}{\Gamma \vdash P \mid P' : \mathbf{P}[\mathcal{B}]} \quad \text{(TPar)}$$

$$\frac{\Gamma \vdash P : \mathbf{P}[\mathcal{B}]}{\Gamma \vdash !P : \mathbf{P}[\mathcal{B}]} \quad \text{(TRepl)}$$

$$\frac{\Gamma \vdash P : \mathbf{P}[\mathcal{B}] \quad \Gamma \vdash M : \mathbf{P}[\mathcal{B}] \quad \mathcal{B}' \in Reside(\mathcal{B})}{\Gamma \vdash M[P] : \mathbf{P}[\mathcal{B}']} \quad \text{(TAmb)}$$

$$\frac{\Gamma, n : \mathbf{P}[\mathcal{B}'] \vdash P : \mathbf{P}[\mathcal{B}]}{\Gamma \vdash (\nu n : \mathbf{P}[\mathcal{B}'])P : \mathbf{P}[\mathcal{B}]} \quad \text{(TRes)}$$

$$\frac{\Gamma \vdash M : \mathbf{O}[\mathcal{B} \mapsto \mathcal{B}'] \quad \Gamma \vdash P : \mathbf{P}[\mathcal{B}]}{\Gamma \vdash M.P : \mathbf{P}[\mathcal{B}']} \quad \text{(TAct)}$$

$$\frac{\Gamma \vdash P : \mathbf{P}[\mathcal{B}] \quad \Gamma \vdash M : \mu \quad \kappa(\mathcal{B}) = \mu}{\Gamma \vdash \langle M \rangle.P : \mathbf{P}[\mathcal{B}]} \quad \text{(TCommOut)}$$

$$\frac{\Gamma, n : \mu \vdash P : \mathbf{P}[\mathcal{B}] \quad \kappa(\mathcal{B}) = \mu}{\Gamma \vdash (n : \mu).P : \mathbf{P}[\mathcal{B}]} \quad \text{(TCommIn)}$$

We say the process is well-typed when we are able to derive a type formula for it using our typing rules. Well-typed processes respect the communication and mobility restrictions defined in all behavioral schemes of the system. It means such a process has the correct behavior.

4 Discussing Mobility of Software Agents

The new operation *move* with its semantic rule (RMove)

$$n[move\ m.P \mid Q] \mid m[R] \rightarrow n[Q] \mid m[P \mid R]$$

allows us to eliminate remote communication which is usually quite difficult to express. By moving threads among ambient we can move their communication part and return back the results of the communication. For example, the elimination of the remote communication between ambient helps us to encode π-calculus in mobile ambient.

Another interesting aspect of the *move* operation is the possibility to express objective mobility. Distinction between subjective mobility and objective mobility is very important. *Objective mobility* means the migration of the process term managed externally. When we want to move an ambient from one place to another, we can use the operation *move* independently of the inner ambient operations. On the other hand *subjective mobility* is the migration of process term which is managed itself. Using *in* and *out* primitives is the expression of subjective mobility of the ambient. In the theory of mobile ambients we sometimes define objective mobility [7] by primitive $go\,N.M[P]$ with its semantic rules

$$go\ (in\ m.N).n[P] \mid m[Q] \rightarrow m[go\ N.n[P] \mid Q]$$
$$m[go\ (out\ m.N).n[P] \mid Q] \rightarrow go\ N.n[P] \mid m[Q]$$

The *go* operation allows similar movement of the ambient as *in* and *out* where only one ambient boundary is crossed. The *move* operation moves process terms between neighbor ambients, which means crossing two ambient boundaries. This is a possible disadvantage, but it is in the interest of the dangerous *open* primitive avoidance. We decided to adopt this operation because of its importance in the context of software mobility and for its background in the Dπ [8] variant of π-calculus. Another argument is the simplicity and understandability of the type system.

The meaning of objective and subjective mobility we can show in the example of a server for software agents. The mobility of agents is the autonomous process and no external impact is needed. The migration is expressed by a travel plan as a sequence of *in* and *out* operations

$$s_1[...] \mid s_2[a[out\ s_2.in\ s_1.P] \mid ...] \rightarrow s_1[...] \mid s_2[...] \mid a[in\ s_1.P] \rightarrow s_1[a[P] \mid ...] \mid s_2[...]$$

where s_1 and s_2 represent two instances of the server and ambient *a* represents a mobile agent moving between them. In some cases the server can "banish" the agent for various reasons (abusing the system, lack of resources, system overload).

This aspect we can express by objective mobility where the server itself moves the agent to another place

$$s_1[...] \mid s_2[move\ s_1.a[P] \mid ...] \rightarrow s_1[a[P] \mid ...] \mid s_2[...]$$

Table 1
Abstract syntax of the language of mobile agents

τ ::=	**Agent type**
| **A**[]	Agent type without communication
| **A**[τ]	Agent type with communication type τ
System ::=	**Distributed system of places**
| *nothing*	Empty system
| *place p*[*Room*]	Place p with inner *Room*
| *System* | *System*	Composition of places in the system
Room ::=	**Inner of the place**
| *empty*	Empty place
| *agent a* : τ[*Body*]	Agent a of type τ with activity *Body*
| *Room* | *Room*	Compositions of agents in the place
Body ::=	**Agent activity**
| *null*	No activity
| *new a′* : τ[*Body′*].*Body*	Creation of new agent a' of type τ and activity *Body′* on the actual place
| *go p′*.*Body*	Moving agent to place p'
| *read* (m : τ).*Body*	Reading message m of type τ from input
| *write a′*⟨m⟩.*Body*	Writing message m to agent a' on the same place

5 Design of Language for Mobile Agents

Understanding the code mobility and mobility of software agents guide us to define the natural semantics of the mobile applications in the distributed computational environment. We define a language which expresses software agents migration in the space of distributed places. The only operation of agents we consider in this case is the agent communication.

The abstract syntax of the proposed language is in Table 1 together with the informal description of the language constructions. The language semantics is

defined by the encoding to mobile ambient and can be found in Table 2. We can see the encoding follows the dynamical hierarchy of agents and places, which is an advantage of the applied calculus.

Table 2
Denotation semantics of the language of mobile agents

$[\![\mathbf{A}[]]\!] = \mathbf{P}[\mathcal{B}_\perp]$ for $\mathcal{B}_\perp = (\perp, \{\mathcal{B}_{System}, \mathcal{B}_{Room}, \mathcal{B}_\perp\}, \{\mathcal{B}_{Room}, \mathcal{B}_\perp\}, \{\mathcal{B}_\perp\})$
$[\![\mathbf{A}[\tau]]\!] = \mathbf{P}[\mathcal{B}_\tau]$ for $\mathcal{B}_\tau = ([\![\tau]\!], \{\mathcal{B}_{System}, \mathcal{B}_{Room}, \mathcal{B}_\tau\}, \{\mathcal{B}_{Room}, \mathcal{B}_\tau\}, \{\mathcal{B}_\tau\})$
$[\![System]\!] : \mathbf{P}[\mathcal{B}_{System}]$ for $\mathcal{B}_{System} = (\perp, \varnothing, \varnothing, \varnothing)$
$[\![Room]\!]_p : \mathbf{P}[\mathcal{B}_{Room}]$ if $p : \mathbf{P}[\mathcal{B}_{Room}]$ for $\mathcal{B}_{Room} = (\perp, \{\mathcal{B}_{System}\}, \varnothing, \varnothing)$
$[\![Body]\!]_a : [\![\tau]\!]$ if $a : [\![\tau]\!]$
$[\![nothing]\!] = \mathbf{0}$
$[\![place\ p[Room]]\!] = p[[\![Room]\!]_p]$ for $p : \mathbf{P}[\mathcal{B}_{Room}]$
$[\![System \mid System]\!] = [\![System]\!] \mid [\![System]\!]$
$[\![nothing]\!]_p = \mathbf{0}$
$[\![agent\ a : \tau[Body]]\!]_p = (\nu a : [\![\tau]\!])a[[\![Body]\!]_a]$
$[\![Room \mid Room]\!]_p = [\![Room]\!]_p \mid [\![Room]\!]_p$
$[\![null]\!]_a = \mathbf{0}$
$[\![new\ a' : \tau[Body'].Body]\!]_a = (\nu a' : [\![\tau]\!])a'[out\ a.[\![Body']\!]_{a'}] \mid [\![Body]\!]_a$ for $a' \neq a$ and $a : [\![\tau]\!]$
$[\![go\ p'.Body]\!]_a = out\ p.in\ p'.[\![Body]\!]_a$
$[\![read\ (m : \tau)]\!]_a = (m : [\![\tau]\!]).[\![Body]\!]_a$ for $a : [\![\mathbf{A}[\tau]]\!]$
$[\![write\ a'\langle m\rangle.Body]\!]_a = move\ a'.\langle m\rangle \mid [\![Body]\!]_a$ for $m : [\![\tau]\!]$, $a : [\![\mathbf{A}[\tau]]\!]$ and $a' : [\![\mathbf{A}[\tau]]\!]$

Agents define communication type τ in the form of $\mathbf{A}[\tau]$, which expresses that the agent can communicate messages of type τ. A closer look shows us that the agent can communicate only to another agent of the same communication type no matter the direction of the communication. This is given by the possibility of the communication thread movement defined in the *Move* set of the agent's behavioral scheme. We can think of a more general solution, but for better understanding we use this limitation for one behavioral scheme. On the system level there is no communication, so its behavioral scheme defines no communication type. The same is for the distributed places.

Communication between agents takes place in the ambient of the agent accepting the message. We consider only communication of agents located on the same

place. Remote communication we can implement by e.g. complex information about communication place and moving agents there. The message exchange is asynchronous. Synchronous communication is more natural, but its expression is more complex.

Mobility rules are given very simply and statically assuming the places are immobile and agents are moved only through places. For simplicity and better understandability, we consider only one general behavioral scheme for all distributed places in the system. This does not allow us to restrict the movement through the places, but of course we can consider also more complex movement management. To keep the type system correct, we must allow moving agents through agents on the same place, which results from the command *new* from agent creation.

Conclusions

The usage of type system is limited by its very simplicity and it does not prevent more restrictive properties from being checked at runtime. In our related work [9] we proved the soundness theorem for the type system, we demonstrated the system by showing how to model some common mobile code paradigms, we demonstrated some typical mobile code applications and as an expressiveness test, and we showed that well-known π-calculus of concurrency and mobility can be encoded in our calculus in a natural way.

In this work we discussed mobility aspects of software agents and we identified the objective and subjective mobility. Understanding the code mobility was provided by our revised calculus of mobile ambient and types enhanced by behavioral scheme. We were able to propose a very simple language for distributed system of mobile agents. The agents' encoding respects the way of hierarchical distribution of ambients and naturally expresses mobility. The simplicity of the language does not allow us to show more complex constructions, e.g. remote communication and restriction of the movement.

References

[1] Cardelli, L., Gordon, A. D.: Mobile Ambients. Theoretical Computer Science, Vol. 240, No. 1, 2000, pp. 177-213

[2] Milner, R., Parrow, J., Walker, D.: A Calculus of Mobile Processes, Part 1 – 2. Information and Computation, Vol. 100, No. 1, 1992, pp. 1-77

[3] Levi, F., Sangiorgi, D.: Controlling Interference in Ambients. Proceedings of POPL'00, ACM Press, New York, 2000, pp. 352-364

[4] Bugliesi, M., Castagna, G.: Secure Safe Ambients. Proceedings of POPL'01, ACM Press, New York, 2001, pp. 222-235

[5] Bugliesi, M., Castagna, G., Crafa, S.: Boxed Ambients. In B. Pierce (ed.): TACS'01, LNCS 2215, Springer Verlag, 2001, pp. 38-63

[6] Fuggeta, A., Picco, G. P., Vigna, G.: Understanding Code Mobility. IEEE Transactions on Software Engineering, Vol. 24, No. 5, May 1998, pp. 342-361

[7] Cardelli, L., Ghelli, G., Gordon, A. D.: Mobility Types for Mobile Ambients. Proceedings of the ICALP'99, LNCS 1644, Springer Verlag, 1999, pp. 230-239

[8] Hennessey, M., Riely, J.: Resource Access Control in Systems of Mobile Agents. Technical Report 2/98, Computer Science Department, University of Sussex, 1998

[9] Tomasek, M.: Expressing Dynamics of Mobile Programs. PhD thesis, Technical university of Košice, 2004

Predicting the Seismic Performance of Cylindrical Steel Tanks Using Artificial Neural Networks (ANN)

Mehran S. Razzaghi, Alireza Mohebbi

Department of Civil Engineering, Islamic Azad University, Qazvin Branch, Iran
e-mail: mehran@qiau.ac.ir, ar.mohebi@gmail.com

Abstract: The main purpose of this study is to predict the seismic performance of liquid storage tanks using an Artificial Neural Network (ANN) model. In order to develop this model, 240 seismic data were collected from relevant literature. Fifty samples were randomly selected as a test set, while the remaining 190 samples were used to train the network. The data used in the ANN model were arranged in a format of six input parameters: peak ground acceleration (PGA), tank diameter (D), tank height (H), ratio of H/D, height of liquid during earthquake (HLIQ), and percent full (% Full). The output parameter, damage state (DS), was provided for measuring the seismic performance of the liquid storage tanks. The model outputs confirmed that an artificial neural network has acceptable potential for predicting the seismic performance of liquid storage tanks. The applicability of the developed technique was then validated by comparing the outputs to the actual damage states of the affected tanks according to HAZUS.

Keywords: Artificial Neural Network; cylindrical tank; seismic performance

1 Introductions

The performance of liquid storage tanks during past seismic events has shown that these structures are seismically vulnerable. Liquid storage tanks in oil refineries and petrochemical plants usually contain hazardous material. For this reason, damage to these structures may cause serious indirect impacts, such as explosions and environmental pollutions. Therefore predicting the seismic performance of existing liquid storage tanks is an important task in seismic risk analysis of industrial plants. The dynamic behavior of liquid storage tanks is very complex. The seismic performance of liquid storage tanks may be affected by several parameters such as H/D, % Full, etc. For this reason, it is very difficult to estimate the seismic performance of liquid storage tanks and to obtain a mathematical representation of uncertain and nonlinear dynamic processes [1]. Hence, the Artificial Neural Network (ANN) may be a useful tool for estimating the seismic performance of such a complex structure.

In conventional modeling methods, different statistical tools, such as regression analysis, are utilized for developing a model to predict the seismic performance of liquid storage tanks. Available fragility functions of liquid storage tanks are samples of conventional modeling methods. For the last two decades, various modeling methods based on Artificial Neural Network (ANN) have become popular and have been used by many researchers for a variety of engineering applications such as concrete engineering [2, 3], traffic engineering [4] and earthquake engineering [5, 6]. The ANN is able to solve very complex problems with the help of interconnected computing elements [3]. It is also a powerful data analysis tool that is able to capture and represent complex input/output relationships. The true power and advantage of neural networks lies in their ability to represent linear and non-linear relationships and in their ability to derive these relations directly from the data being modeled [7]. Traditional linear models are simply inadequate when it comes to modeling data that contains non-linear characteristics.

The objective of this study is to present a methodology designed by ANN for predicting the seismic performance of liquid storage tanks. The model is expected to determine the damage state of the tanks.

2 The Seismic Performance of Liquid Storage Tanks

Over the past few decades, many liquid storage tanks were damaged due to earthquakes. During an earthquake, the upper part of the contained liquid moves in a long-period motion. This part of the liquid may apply upward hydrodynamic pressure to the tank roof or may cause overflowing of the liquid. The other part moves rigidly with the tank [8]. Moreover, during an earthquake large amounts of hydrodynamic pressure can be applied to the tank shell. The hydrodynamic pressure may cause damage to the tank shell. Many of the on-grade tanks, even anchored ones, may experience shell uplift due to the strong ground motion. The shell uplift may cause ruptures of the shell-to-base-plate junction, rupturing of pipes and/or appurtenances. Elephant-foot buckling (Elastic-Plastic failure) may occur by large axial compressive stresses in the tank wall. Also, distortion of the tank roof or rupturing of the roof-to-wall junction may occur due to the strong ground motion.

There are various methods for classification of the damage states of cylindrical steel tanks. ATC 13 [9] and HAZUS [10] classifications are two common classifications of tank damage states. ATC 13 [9] considers seven different damage states for tanks which are: no damage, slight damage, light damage, moderate damage, heavy damage, major damage, and destroyed. HAZUS [10] considers five damage states which vary from no damage to collapsed tanks, based on the serviceability, loss of content, and the occurrence of shell buckling.

HAZUS damage states are described in Table 1. It should be mentioned that five of ATC13 damage states – none, light, moderate, heavy and destroyed – are equivalent to HAZUS DS1 to DS5 damage states respectively [11]. Herein HAZUS damage states are considered for classifications of damage in tanks.

Table 1
Description of damage states based on HAZUS

Damage State	Description
DS1	No damage.
DS2	Minor damage without loss of content or functionality. Damage to roof, localized wrinkles in steel.
DS3	Considerable damage with minor loss of content. Elephant-foot buckling without loss of content.
DS4	Severe damage. Tank going out of service. Elephant-foot buckling with loss of content.
DS5	Collapse. Losing all of content.

3 Architecture of the Artificial Neural Networks

The neural network-based modeling process involves five main aspects: (a) data acquisition, analysis and problem representation; (b) architecture determination; (c) learning process determination; (d) training of the networks; and (e) testing of the trained network for generalization evaluation [12]. There are different common architectures for artificial neural networks. The multi layer perceptron (MLP), radial basis function network (RBFN), the probabilistic neural network (PNN), the cascade correlation neural network (Cascor), the learning vector quantization (LVQ), and the self-organizing feature map (SOM) are some popular neural network architectures [13, 14]. They differ in aspects including the type of learning, the node connection mechanism, the training algorithm, etc. The most common neural network model is the multilayer perceptron (MLP). This type of neural network is known as a supervised network because it requires a desired output in order to learn. The goal of this type of network is to create a model that correctly maps the input to the output using historical data so that the model can then be used to produce the output when the desired output is unknown. A typical structure of an artificial neuron is shown in Fig. 1.

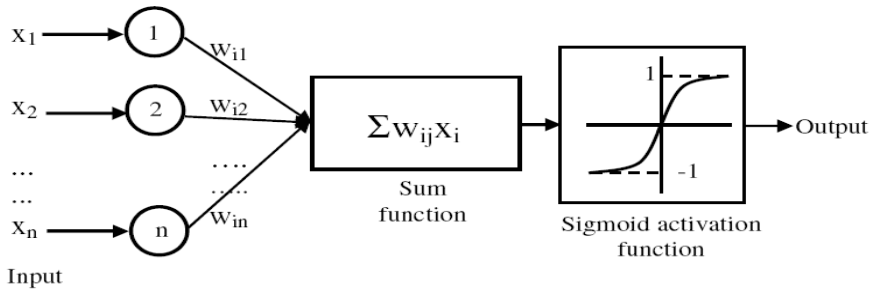


Figure1
A typical structure of an artificial neuron

An error incurred during the learning process can be expressed as a mean square error (MSE) or a root-mean-squared (RMS) as given in the following equation:

$$MSE = \frac{1}{p}\sum_j (o_j - t_j)^2 \tag{1}$$

$$RMS = \sqrt{(\frac{1}{p}) * \sum_j \left| t_j - o_j \right|^2} \tag{2}$$

In addition, the absolute fraction of variance (R^2) and sum of the squares error (SSE) can be calculated by utilizing Eqs. 3 and 4, respectively:

$$R^2 = 1 - \left(\frac{\sum_j (t_j - o_j)^2}{\sum_j (o_j)^2} \right) \tag{3}$$

$$SSE = \sum_j (o_j - t_j)^2 \tag{4}$$

where t is the target value, o is the output value and p is the pattern.

In this study, the back propagation (BP) algorithm is used to train and construct the present ANN model and the hyperbolic tangent function transfer function is adopted. The tangent function is nonlinear and, therefore, the original data before training the network are normalized. The overall flowchart of the procedure of this study is given in Fig. 2.

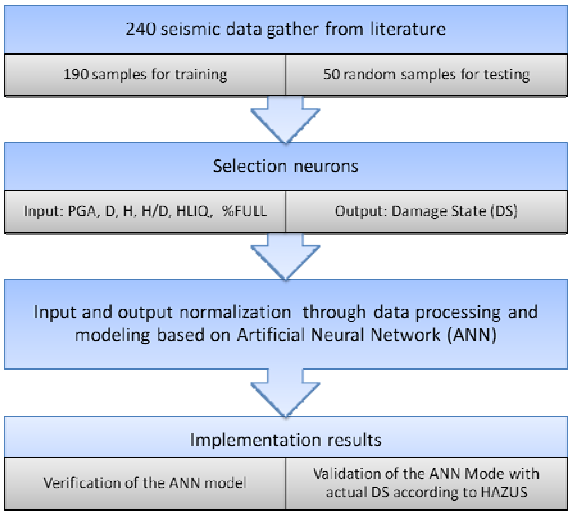


Figure 2
Flowchart of the methodology of this study

4 Proposed Neural Network Model

The ANN model developed in this study is used to predict the seismic performance of liquid storage tanks. In order to produce an effective ANN model, it is vital that the network be properly trained. Therefore, 240 tanks which experienced strong ground motion in past earthquakes were selected. The data of damaged tanks were adopted from [15] (see Table 2). The range of the six input variables, including peak ground acceleration (PGA), tank diameter (D), tank height (H), ration of H/D, height of liquid during earthquake (HLIQ), percent full (% full) and one output, damage state (DS), are given in Table 3.

Table 2
List of the selected triggered tanks

Seismic event	Year	PGA range (g)	Number of affected tanks	Reference
Long Beach	1933	0.17	37	15
Kern County	1952	0.19	23	15
Imperial Valley	1979	0.24-0.49	19	15
Coalinga	1983	0.71	11	15
Loma Prita	1989	0.13	86	15
Landers	1992	0.15-0.56	26	15
Northridge	1994	0.55-1.0	38	15

Table 3
Range of input-output parameters in databases

Input parameters	Minimum	Maximum
PGA (g)	0.13	1
D (m)	3.2	83.2
H (m)	4.9	19
H/D (%)	18	416
HLIQ (m)	0	15.2
% Full	0	100
Output parameter		
DS	1	5

In order to determine the best structure, five different architectures as [6-6-2-1 to 6-6-6-1] are considered and, based on MSE and R^2 criteria, the best one is selected. Table 4 presents the obtained results for each structure. As can be observed, the decrease in MSE causes the increase of R^2 and is approaching to 1. Therefore, the best structure is [6 6 6 1] (see Table 4). The architecture of the proposed ANN is also illustrated in Fig. 3.

Table 4
Evaluation of ANN architecture

Structure	Mean Square Error	Train Error	Test Error	R^2
[6 6 2 1]	0.170	0.681	0.313	0.5328
[6 6 3 1]	0.147	0.586	0.223	0.6200
[6 6 4 1]	0.147	0.588	0.238	0.6182
[6 6 5 1]	0.242	0.967	0.345	0.4575
[6 6 6 1]	0.101	0.416	0.154	0.7793

To test the reliability of the proposed ANN model, 50 samples are randomly selected as the test set, while the remaining 190 samples were used to train the network. Herein, the Matlab neural network toolbox was used to construct and train the supervised network. In training a supervised ANN, weights between the neurons are adjusted to minimize the error in the output. The values of parameters used in this research are as follows:

- Number of input layer units = 6
- Number of hidden layers = 2
- Number of output layer units = 1
- Learning rate = 0.75
- Learning cycle = 1000

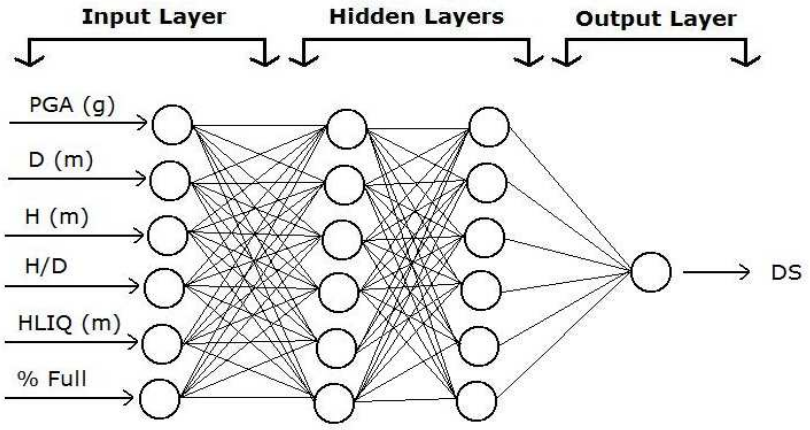


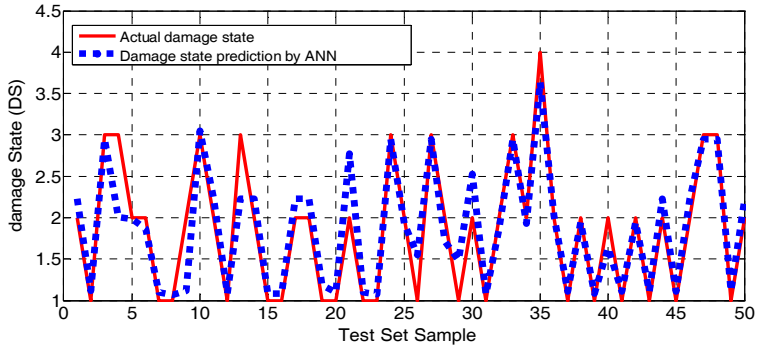
Figure 3
Proposed ANN architecture

5 Implementation Results

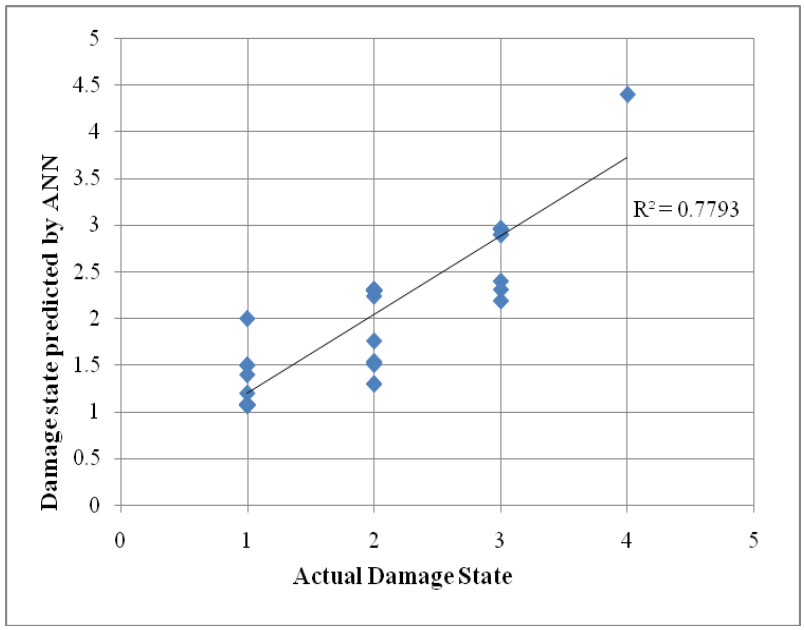
The developed ANN model in this research is utilized to predict the damage state for the seismic performance of liquid storage tanks. The error between the predicted and the target values for the damage state (DS) is plotted in Fig. 4, which includes row numbers up to 50. As indicated in Fig. 4, the neural network was capable of deriving the relationship of input variables and the output. The correlation factor is $R^2 = 0.7793$, which is acceptable for liquid storage tanks [11].

In order to indicate the accuracy of the ANN prediction, various earthquake-affected tanks of different H/D, %Full, and PGA were randomly selected. The actual damage states of the affected tanks (according to HAZUS) were compared to the ANN prediction. The comparison of actual performances and ANN predictions in different ranges of PGA are indicated in Figs. 5 to 8.

As can be observed in these figures, the ANN prediction is acceptable for various models – especially for PGAs less than 0.3 g (see Figs. 5 and 6). It is worth mentioning that the prediction of the ANN model in this study was not accurate enough for the higher PGAs (See Figs. 7, 8). The main reason for such unacceptable prediction is the lack of enough data for training the model in higher PGAs.



(a)



(b)

Figure 4
(a) Comparison of predicted data to test samples (b) Predicted data vs. actual damage state

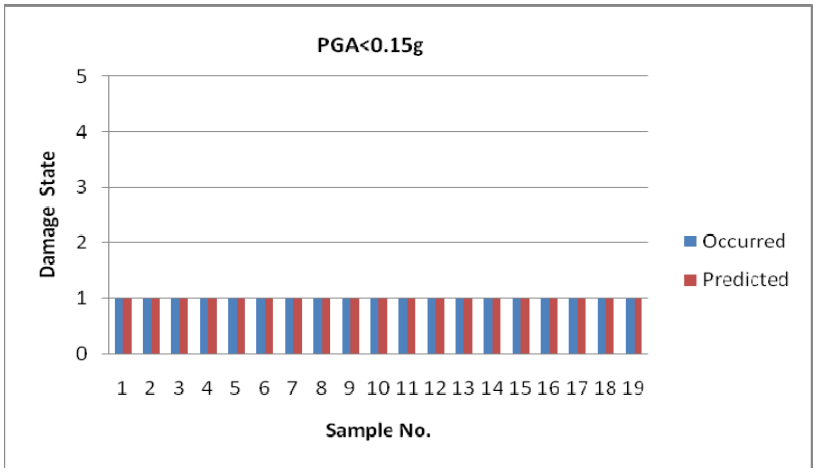


Figure 5
Comparison of ANN prediction to actual damage state for PGA<0.15 g

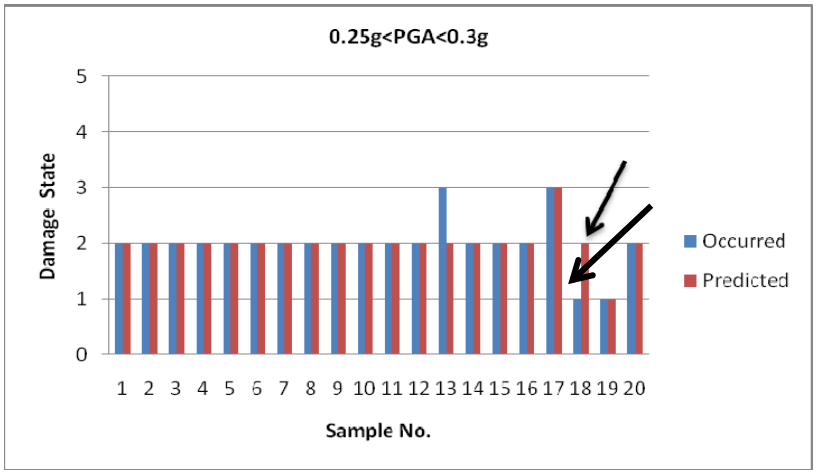


Figure 6
Comparison of ANN prediction to actual damage state for 0.25 g <PGA<0.3 g

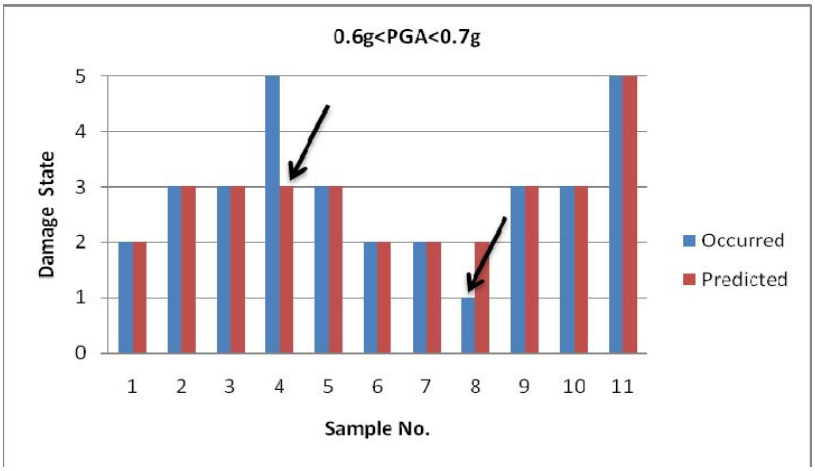


Figure 7
Comparison of ANN prediction to actual damage state for 0.6 g <PGA<0.7 g

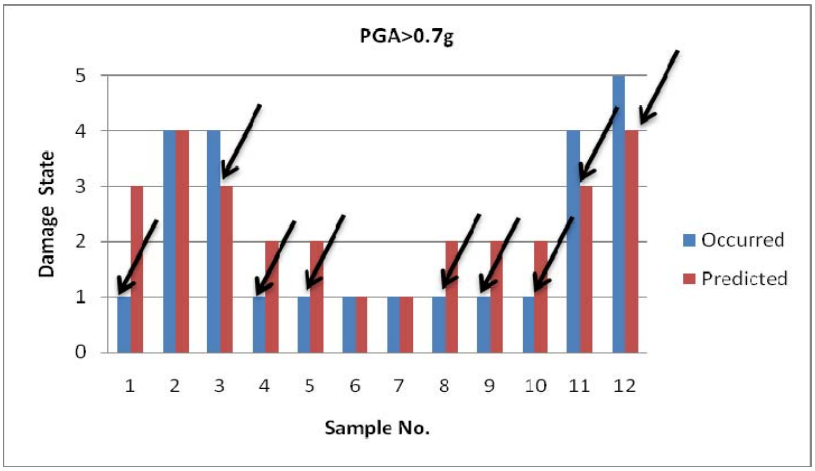


Figure 8
Comparison of ANN prediction to actual damage state for PGA>0.7 g

Conclusions

This study was aimed at investigating the possibilities of adopting artificial neural networks to predict the seismic performance of liquid storage tanks. To this end a data bank of 240 earthquake-affected tanks was selected. Five back propagation ANN of different architectures were designed and trained with 190 data. The main findings of the study are outlined below:

1) The results of this study showed that artificial neural network has acceptable potential for predicting the seismic performance of liquid storage tanks.

2) Based on the results given in table 4 for evaluation of ANN architecture, the best correlation factor (R^2 =0.7793) was obtained from the [6 6 6 1] model with the lowest Mean Square Error (0.101).

3) For PGAs less than 0.3 g, the ANN model accurately predicted the damage state and for 0.6 g <PGA<0.7 g, the ANN predictions are also acceptable, but for PGA>0.7 g, because of the lack of the data in this range for training the model, the predictions in the most cases are higher than the actual damage state, so they were not accurately predicted. It is worth mentioning that seismic events with PGAs higher than 0.7 g are very strong earthquakes and usually have long return periods. In other words these extraordinary earthquakes are rare. Hence the neural network can accurately predict the seismic performance of cylindrical steel tanks for a wide range of PGAs.

The results of this study reveal that an artificial neural network can be used for the development of seismic performance relations (such as fragility curves). In other words, the proposed methodology is a useful tool in seismic risk analysis of tank farms with potential PGAs less than 0.7 g.

Acknowledgements

The authors would like to acknowledge Professor Michael O'Rourke for his support. This study was supported by CCRC of QIAU. The authors are grateful for this support.

References

[1] Brassai, S., Bakó, L., Hardware Implementation of CMAC Type Neural Network on FPGA for Command Surface Approximation, Acta Polytechnica Hungarica, Vol. 4, No. 3, 2007

[2] Bekir, I., Topc, U., Saridemir M., Prediction of Compressive Strength of Concrete Containing Fly Ash Using Artificial Neural Networks and Fuzzy Logic. J Computational Materials Science, 2008; 41:305-311

[3] Raghu, B., Prasad, Eskandari, H., Venkatarama Reddy BV. Prediction of Compressive Strength of SCC and HPC with High Volume Fly Ash Using ANN. J Construction and Building Materials, 2008; 23(1): 117-128

[4] Oravec, M., Petráš, M., Pilka, F., Video Traffic Prediction Using Neural Networks, Acta Polytechnica Hungarica, 2008, Vol. 5, No. 4

[5] You-Po, S., Qing-Jie, Z., Application of ANN to Prediction of Earthquake Influence, Second International Conference on Information and Computing Science, 2009, Vol. 2, pp. 234-237

[6] Chakraverty, S., Marwala, T., Gupta, P., Response Prediction of Structural System Subject to Earthquake Motions Using Artificial Neural Network, Asian Journal of Civil Engineering (building and housing), 2006, Vol. 7, No. 3

[7] Perlovsky, L. I., Neural Networks and Intellect: Using Model-based Concepts. Oxford: Oxford University Press; 2000

[8] Jacobsen, L. S., Impulsive Hydrodynamics of Fluid Inside a Cylindrical Tank and of Fluid Surrounding a Cylindrical Pier. Bull. Seismological Soc. of Am. 1949; 39(3):189-203

[9] ATC. Earthquake damage Evaluation Data for California, ATC13, American Technical Council; 1985

[10] NIBS. Earthquake Loss Methodology. HAZUS 99; 1999

[11] O'Rourke, M., So, P. (2000) Seismic Fragility Curves for On-Grade Steel Tanks. J Earthquake Spectra, 2000; 16(4)

[12] Saridemir, M., Prediction of Compressive Strength of Concretes Containing Metakaolin and Silica Fume by Artificial Neural Networks. J Advances in Engineering Software, 2009; 40: 350-355

[13] Rafig, M. Y., Bugmann, G., Easterbrook, D. J., Neural Network Design for Engineering Applications. J Comp Struct 2001; 79: 1541-52

[14] Lippman, R. P., An Introduction to Computing with Neural Nets. In: Artificial Neural Networks", the computer society theoretical concepts. Washington; 1988

[15] So, P., Seismic Behavior of On-Grade Steel Tanks; Fragility Curves. MSc. Thesis, Rensselaer Polytechnic Institute, Troy, New York; 1999

An Approach for the Empirical Validation of Software Complexity Measures

Sanjay Misra

Department of Computer Engineering Atilim University, Ankara, Turkey
smisra@atilim.edu.tr

Abstract: Software metrics are widely accepted tools to control and assure software quality. A large number of software metrics with a variety of content can be found in the literature; however most of them are not adopted in industry as they are seen as irrelevant to needs, as they are unsupported, and the major reason behind this is due to improper empirical validation. This paper tries to identify possible root causes for the improper empirical validation of the software metrics. A practical model for the empirical validation of software metrics is proposed along with root causes. The model is validated by applying it to recently proposed and well known metrics.

Keywords: Empirical validation; Preliminary empirical validation; advanced empirical validation; software metrics

1 Introduction

The popularity of empirical studies has been rising in the field of software engineering since the 1970s. Its importance for software metrics is pointed out by several researchers [1], [2], [3], [4]. It is one of the major ways through which academicians and scientists can assist industry in selecting new technology. On the other hand, it is a common observation that the standards of empirical software engineering research is not up to a satisfying level [5]. According to surveys on the papers on empirical validation [1], examples of poor experimental design, the inappropriate use of statistical techniques and conclusions that do not follow the reported result were found. Another survey on 600 published papers reported that a considerable amount of research papers lack experimental validation. Further, they use informal (assertion) forms of validation and use case studies approximately 10% of the time; it was also observed that their experimentation terminology was sloppy [4]. In addition, empirical study is often used synonymously with experiments and used in an inconsistent manner [6].

In the case of software complexity measures, the situation is quite similar. There is no match in content between the increased level of metric's activity in academia

and industry [7]. Evidence shows that any successful adoption and implementation of these metrics is limited [8]. We are aware of this fact that there are other factors which impact on the successful metric program in industry, e.g. institutional forces [8]; however, this stage occurs when a metric is applied in the industrial environment. The major problem with the existing metrics available in literature is that they have not been tried to implement in the industrial environment. One can easily find many academic works in the literature of complexity in measures without proper empirical validation. These poor findings are not only due to the lack of clear-cut guidelines and explanations for different types of empirical studies; additionally, the absence of availability of real environments for the implementation of metrics is another issue to consider. In many cases it is a challenging task to identify any direct link between researchers and the software industry, and this makes it difficult to validate the metrics against the projects in software business. As a consequence, researchers try to validate their metrics through other means, such as experiments in laboratories, class rooms or with available data/programs from the Internet. Most of the time, those means can provide only a partial empirical validation. However, for a proper empirical validation, we believe that one must apply the new technology/metric to real data/projects from industry.

Based on the above rationale and motivation from the insufficient discussion on empirical studies in literature, we have developed a model for practical empirical validation. This model is based on the evaluation of the common practices adopted for empirical validation of software metrics. In this model we suggest the application of empirical validation in two parts, namely as preliminary empirical validation, advanced empirical validation and acceptance. The preliminary empirical validation includes the initial validation of the metric by applying it to different test cases and examples. In advanced empirical validation a new metric is tested by using real projects from the industry. Finally, after the replicated experimentations in different environments, the acceptance of the metric(s) by industry is the final step of our model. In fact all the steps given in the model are not new; we have accumulated these different approaches and compiled/presented them in a formal way. We continue this discussion in the next section. The validity of our model is checked by applying it to some well known metrics, e.g. CK metrics suite [9] and entropy metrics [10].

Research Questions and Methodology

In line with what we have presented so far, we have identified that discussions on already proposed metrics may need further exercises for empirical validation. With this motivation, we found this point has received less attention and care in the literature on software metrics; hence it constitutes a potential gap. To address this gap, we have generated the following research questions:

Why have most of the proposed metrics available in literature not received acceptance from the software industry? (1)

Are the researchers following any proper guideline while proposing and validating (empirically) their metrics? (2)

To address these questions, we made an exhaustive literature survey for metrics, as we had already started to build a body of knowledge on measurements in software engineering. With the large number of metrics limiting our survey, we focus on the following two points:

Firstly, we have selected metrics including the CK metric suite and the Entropy metric, whose goals were similar and proposed at the same time. An additional motivation arose, namely that the CK metric suit has considerably higher reputation and adoption in the industry, while the Entropy metric has not achieved such acceptance, although both metrics bear strong background work. This drove us to investigate and evaluate by which practices the CK metric suite has become a benchmark. We are leaving out of the discussion other useful and widely accepted metrics, such as function point analysis, due to the limitations of our research.

Secondly, we have selected metrics which were proposed in last two years (2009 and 2010), which is another limitation within this work.

To answer (1) we have performed the survey on different types of empirical validation techniques which are commonly adopted by the developers of software metrics. We have reviewed all to our knowledge and propose a simple model, which mainly consists of all the important stages required for what we call "proper empirical validation". We validate our model on two metrics.

Again to validate (1) by the model which is developed to overcome the problem, we have analyzed the metrics proposed recently (in last two years).

The paper is organized in the following way. The next section summarizes the different types/ways of the empirical validation which are commonly used for the validation of software metrics. In section three, we introduce the metrics on which we apply our model. We present a model for the proper empirical validation in Section 4. A case study is presented for demonstrating our model in Section 5, which is further validated with newly proposed metrics in Section 6. Some observations and suggestions are given in Section 7. The conclusions drawn from this work are given in Section 8.

Before we go any further, we want to clarify that the terms software metric, software measure and software complexity metric/measure are used synonymously in the literature. Although one can find several definitions [11] for software complexity, we follow the IEEE [12] definition, which defines software complexity as "the degree to which a system or component has a design or implementation that is difficult to understand and verify". Additionally, IEEE [12] also defines metric as "as a quantitative measure of the degree to which a system, component, or process possesses a given attribute".

2 Common Practices Adopted for Empirical Validations of Software Metrics: An Analysis

There are several practices adopted by developers of software metric programs. Common practices adopted for empirical validations of software complexity measures include:

1 Small application programs to demonstrate the metric(s) or measure(s)

2 Formal experiments/examples reported in literature

3 Case studies/surveys available on the web

4 Experiments in laboratories or classrooms

5 Experiments at workplaces in the industry but with users off the subject i.e. who are not potential users.

2.1 Small Application Programs to Demonstrate the Applicability of Metric(S) or Measure(S)

Normally, most of the metrics are demonstrated through implementation on small application program(s). However, if a researcher is applying only this method for their validation, it may not be sufficient even for complete demonstration of the metric(s). For example in [13], the authors claim that their metric provides some indications for the level of coupling; however, from the example program which they provide in the paper, it is not straight-forward to identify how coupling can be estimated. In addition to this issue with this type of practice, if these programs are developed by the developer of the metric(s), these programs cannot be taken even for demonstrating the metric. It is because the developer of the metric(s) knows what he wants to prove with his metric and these programs may be specifically developed for this purpose (i.e. to validate the metric). It is worth mentioning that we do not claim that the researchers are stretching the truth; however, the application of the metric on these programs cannot be justified as they do not provide the reader a transparent implementation. For example, in the validation of improved cognitive information complexity measure [14], the authors themselves developed all the programs for theoretical validation then they claimed that their metric were properly validated; however, there are considerable "dark spots" left to the reader, including how one can prove its practical usefulness without implementing on real projects.

2.2 Formal Experiments/Examples Reported in Literature

The validation of software complexity measures with formal experiments performed on example(s) available in literature provides the only way how to implement the metric, and they stand without practical application. In most of the

cases these inappropriate examples do not satisfy the purpose of real empirical validation of the metrics proposed. Small application programs can only be used to get a preliminary idea about a metric; however, it is hard to accept as a proof of practical applicability as a metric, i.e. this cannot be the way of complete empirical validation. An example of this type of practice can be observed in the proposal of the unified complexity measure [15]. The authors have considered several examples from a book on a programming language.

2.3 Case Studies/Survey Available on the Web

One of the other common practices to validate software metrics/measures is through case studies. These case studies are sometimes small projects reported in literature or on the web [16], [17], or some large programs. Naturally, this way can only demonstrate the implementation of metric on big software products; however, in no case does it represent the practical applicability of the programs. In fact if the proposer of the metric is applying it on some programs/projects available on the web, we can treat it as a survey. One of the examples for this method is the metrics proposed by Aggarwal et al. [18], where the authors have proposed two metrics and validated them with JAVA programs available on the web. Observations show that their process of validation is open to discussion [19]. Further similar to the small application programs, the application of the metrics on such projects available on the web cannot be justified as they do not provide the reader a transparent implementation.

2.4 Experiments in the Laboratory or Classroom

Some authors [20-21] try to validate their metrics with students in classrooms or laboratories. In fact, in software engineering, several researchers suggest performing empirical validation with students in a classroom environment [22, 23]. The examples of experiments in the classroom are: controlled experiments (with graduate students), observational studies (professionals, graduate students) and case studies (projects as part of class work). Although it is arguable that students are the future software developers, experiments with students may reduce the practical value of experiments [1]. A validation process based on such data may be acceptable only for gaining initial knowledge regarding some quality factors, such as understandability.

2.5 Experiments at Workplaces in the Industry but with Users off the Subject, i.e. Who Are Not Potential Users

Sometimes, authors/developers of a new metric program try to validate their work in small and medium scale software industries. This may be a convenient way for academicians whose students are working in those companies. Although there is

no harm in utilizing the available software industries in the vicinity, a worrying issue is that these companies may not really evaluate or validate the metric program. In common practice, it is not common to find such small- and medium-sized software industries who adopt a software metric program in the organisation [24]. Hence, assuming a real validation of a new metric program may be open to discussion.

By adopting all these practices, one can only guess the preliminary idea for the metric. Also it is easier to adopt these practices for a freshman developer of a metric program; however, as a result, they can barely go beyond contributing to a publication. This is the reason we want to point out that, while most of papers in the area of software metrics/measures are gradually increasing, their adoption from industries is limited. In fact this is not only the situation in the case of software metrics, but also a problem in general; i.e. a lack of proper experimentations in software engineering. In an analysis of the eight papers published in IEEE [1] transactions on software engineering, the reviewer found examples of poor experimental design, inappropriate use of statistical techniques and conclusions that did not follow from the reported results. The reviewer further commented that the authors of those papers are well-known for their empirical software engineering work.

3 Chidamber et al.'s Metric Suite and Kim et al.'s Metrics

In this section, we introduce two different metric sets. In one of them, empirical validation is in the core of the development of the metric program, and in the other, the authors have adopted the casual process. We want to show the result of both practices. It is worth mentioning here that we do not discuss those papers in which rigorous empirical validation is the core part of the reported research.

We introduce metrics which were developed by Chidamber and Kemerer [9] and Kim, Shin and Wu [10] for object-oriented (OO) programming. Both groups proposed their metrics approximately at the same time, in 1994 and 1995, respectively. Although both proposals of metrics were introduced for OO systems at the same time and for the same objective (i.e. measuring the complexity of OO systems) Chidamber's and Kemerer's metric suite has gained more popularity in the software industry. On the other hand, the metrics proposed by Kim et al. have not found that much acceptance in the industry and are used only for literature support in research papers. We want to clarify that our intention is not to evaluate or criticize any particular metric(s), but rather to evaluate whether they have success or failure in practice. Furthermore, we will use these metrics as a case study to validate the effectiveness of our model in Section 5.

3.1 The Chidamber and Kemerer (CK) Metrics Suite

Chidamber et al. [9] proposed a suite of metrics which includes five well know metrics: WMC: weighted method per class; RFC: response for a class; NOC: number of children; LCOM: lack of cohesion in methods; and CBO: coupling between objects.

In the WMC, they suggested that one can calculate the weight of the method by using any procedural metric. They used cyclomatic complexity [25] for measuring WMC and assumed the weight of each method to be one.

The second metric, RFC, is defined as the total number of methods that can be executed in response to a message to a class. This count includes all the methods available in the class hierarchy. The depth of inheritance tree (DIT) and the number of children (NOC) are other two important CK measures. The former represents the maximum length from the node to the root of the tree and the latter is the number of immediate subclasses subordinated to a class in class hierarchy.

The LCOM metric is for cohesion and is counted as the number of common attributes used by different methods.

Another metric in their suite is CBO, which measures interactions between objects by counting the number of other classes to which the class is coupled. As stated previously, these are most accepted metrics in the OO domain in the industry; we are not providing the detail of the each metric and refer readers to the original paper [14].

3.2 Complexity Measures for OO Programs Based on Entropy

Entropy [26] is a common concept and applied by several researchers [27-30] to measure the complexity of software. Kim et al. [10] proposed three metrics for OO programs based on the entropy concept. These metrics are: the class complexity, the inter-object complexity, and the total complexity for OO programs. Basically their first metric, class complexity, evaluates the information flows between attributes and functions in a class. Their second metric, inter-object complexity, measures the information flows between objects. The third metric, total complexity, adds the class complexity and the inter-object complexity.

4 A Model for Empirical Validation for Software Complexity Measure

Empirical studies [22] are used to investigate software development and practices for understanding, evaluating, and developing in proper contexts. It allows the analyst to test out the theories with the support of empirical observations. It

includes formal experiments, case studies and surveys observed in industry, the laboratory or classroom [1]. All these different types of empirical validation techniques can also be applied to the validation of software metrics. However, as we have demonstrated in Section 2. All these independent validations are not suitable for the empirical validation of software complexity measures. They provide only a preliminary understanding of the proposed metrics. In other words, all these practices are not without benefit, but they are only suitable for introductory validation. With this point of view, we recommend these practices for preliminary empirical validation.

Ideally, we believe that when a new metric is applied to real projects from industry, its validity should later be evaluated against other similar metrics. However, in many cases, the type of empirical study depends upon situation and circumstances and, in the initial phases of any new proposal, it is not always possible to apply a new metric directly to the real projects in industry. A reason for this might be the following: if the developer of the metrics is an academician and at the particular time does not have access to the proper real (industrial) environment, then he tries to validate his proposal through other means (data and projects on the web).

In considering these practical problems related to empirical validation, the suggested guidelines in the proposed framework are categorized in two major stages as preliminary and advanced empirical validations, which are further classified into different stages. Accordingly, we suggest seven steps in total.

a) **Preliminary Empirical Validation**:

Preliminary empirical validation is divided into four stages. The first phase includes small experiments, case studies, and the comparative study and analysis of work. The second stage includes the application of the metric on real cases from industry.

1 Demonstration of the Metric(s)

This stage is based on short experimentations. The metric(s) should be demonstrated with real example(s). Here the meaning of real example is that the examples must be complete enough to demonstrate each aspect of the metric. In this respect, the example may be developed by the developer of the metric. Further, several programs/examples can be taken, if it is needed for demonstrating the metric. For example, for the demonstration of the cognitive functional size metric [20], the authors have developed three small programs. This is the first stage and most of the developers of the metric complete this step.

2 Case Study

After demonstrating the metric with short example(s), a case study is required. Here the case study refers to a relevant real project to which a new metric program should be applied. In fact, preliminary ideas regarding

the usefulness of a new metric are only validated by applying it to data collected from a real project. One can chose a real project from the web, literature or working projects in the departments, to which he can apply the metric to verify its applicability. In fact this stage is recommended when the data from industry is not readily available.

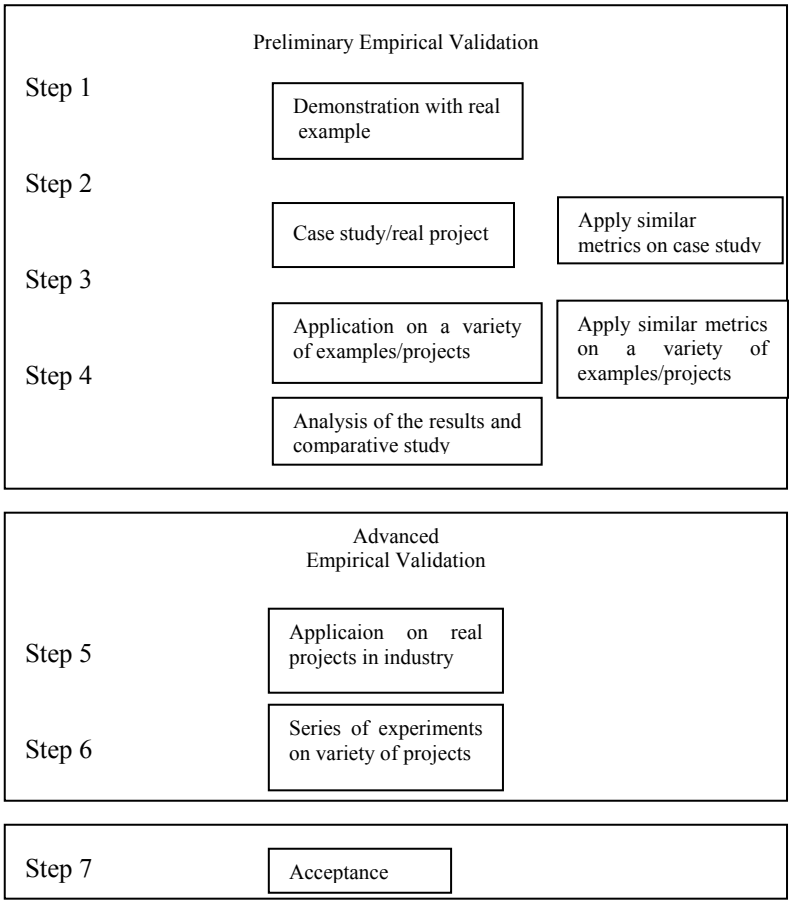


Figure 1
Proposed Model for Empirical Validation of Software

In parallel, similar metrics should also be applied on the case study and the results should be compared with the results of the proposed metrics. If the developer cannot find measureable differences between the proposed metrics and the available metrics, then he can either withdraw his proposal or can improve it.

3 Test Cases

A comparison with similar metrics provides valuable information on the usefulness of the new metric. It will also help in the analysis of the behavior of the metric in a variety of projects. In this stage, the metrics should be applied on a variety of examples/test cases/cases studies, and it should also be applied in different environments, if possible. Here we recommend a variety of examples/test cases/ case studies to check the applicability of the metric(s) on a variety of projects. Further, for comparison, similar metrics must be applied on those projects where the new metric is applied.

4 **Analysis**

The last stage of the preliminary empirical validation is to perform an analysis of the results and the comparative study, which have been done in previous stages. The results of the analysis and comparative study will help the developer to convince industry to apply the new metric program for advanced empirical validation.

b) Advanced Empirical Validation

5 Real Project in Industry

The acceptance of a new software metric is open to discussion if its usefulness is not proved in the software industry. For its acceptance, firstly the proposed metric must be applied by software developers on real working projects. It is worth mentioning here that the program should be applied by the professionals who are working on the projects. This is because they are key informants who can really evaluate the practical applicability of the new metric program.

6 Family of Experiments

The proposed metric must be applied in different projects and in different environments. The reason for this is that, after performing a family of experiments, one can build up the cumulative knowledge to extract useful measurement conclusion to be applied in practice [31].

7 Acceptance

After the series of experiments, the results should be analyzed and compared. If the new metric(s) is proved to be better than the existing metrics/metric programs in an organization it can be considered for acceptance. Otherwise it may require further improvement. After improvement of the metric(s), the same validation process should be revisited from the beginning, i.e. from the first stage of this model.

We have suggested seven steps for the empirical validation of metrics. These steps are not new; however, the proper adoptions by the developers of software metrics are limited. With this point in mind, we have presented them in a formal way in order to provide a straightforward guideline to help developers of software metrics

to gain a clear understanding of the proper empirical validation process. Further, the steps provided in this model are only guidelines and the practical implementation of this model may depend on the actual situation and circumstances. For example, after the demonstration of the metric in the preliminary phase (step 1), if one can find an opportunity to apply the metric(s) program on a project in industry, then the developer does not need to follow the remaining stages of the preliminary validation process. On the other hand, the developer should not forget to apply similar metrics for comparison while applying the proposed metric(s) to an industrial project, since without comparison the usefulness of a new measure cannot be justified.

5 Case study: Applicability for our Model

We have applied our model on the metrics developed by Chidamber and Kemerer [9] and Kim, Shin and Wu [10]. In this section, we want to show that, although both metrics were developed at almost the same time, for evaluating the complexity of OO systems, CK has become a milestone in the field of metrics. Their work is not only used by practitioners and widely adopted by the software industry, but also a simple search by Google can yield 2791 citations. Complexity measures for OO program based on the entropy proposed by Kim et al. is not found to be used by practitioners and has only became a paper which is sometimes used by some new developers of software metrics for citation purposes. We have intentionally used these metrics to evaluate which practices and rules make a metric useful and how our model can be handy while adopting those practices.

We start our evaluation with complexity measures for OO program based on entropy [10]. We have already given introductory information regarding the metrics in Section 3. Here we are evaluating them: how they validated their work? The authors have demonstrated all three metrics: class complexity, inter object complexity and total complexity, with a simple example. This covers the first step of our model.

For the validation of their metrics, they showed that their complexity values fall between $\log_2 n$ and zero [10]. Further, the authors evaluated their metrics theoretically with Weyuker's properties [32]. There is no harm in validating metric(s) theoretically, but theoretical validation only proves that the measure/metric(s) is developed on some sound theoretical base; in no case does it prove any practical applicability in industry. Further, for experimentation, i.e. empirical validation, the authors measured the class complexity for C++ classes using 68 classes that were extracted from classes of user interfaces and data structures [33-36]. We have surveyed these references from books on C++ programming. This part of validation is our third step of the preliminary empirical validation. Furthermore, the authors claim that their experimental results of C++

classes proved the effectiveness of the proposed metrics. They further wrote that in the future, they would gather many OO programs for analysis and for the calculation of the correlation coefficient. This is the summary of their experimentations on their proposed metrics.

As per our empirical validation guidelines, the metrics proposed by Kim, Shin and Wu [10] barely fulfill the preliminary stage of empirical validation. It is because the authors only demonstrated their metric with a small program and then used several classes for showing the implementation of the metrics. Neither did they perform a case study/a real project, nor any comparative study.

Now we evaluate the way the CK metrics suite was introduced. Chidamber et al. [9] firstly provided the fundamental and theoretical background for developing the metrics. All the metrics proposed by Chidamber et al. are straight forward for computing the different features of OO systems, so they are easily countable from any OOS system; e.g. NOC can be easily counted by counting the number of immediate descendants of the class. They compared their metrics with all available OO metrics at that time. If we evaluate their methodology for empirical validation, we find that it closely matches with our advanced empirical validation process. According to Chidamber et al. [9] '*They have applied their metrics through automated tool developed for this research at two different organizations and referred as Site **A** and Site B. Site **A** was a software vendor that uses OOD(object oriented design)..... Metrics data from 634 classes from two C++ class libraries ..... Site B was a semiconductor manufacturer and used the Smalltalk programming language for developing flexible machine control and manufacturing systems. .....Metrics data from 1459 classes from Site B were collected*'.

The quote supports that the metric were developed not only on a sound theoretical background, but also via the adoption of proper validation criteria; i.e. the authors proved the worth of their metrics through the above rigorous empirical validation, which we name as advanced empirical validation in our model. Also, after the development of these metrics, they were adopted by several major organizations, e.g. NASA. Additionally, after gaining popularity, several researchers worked on these metrics and again validated them empirically [37]. This case study proved that the fifth and sixth stage in our model is a necessary requirement for a complete empirical validation process.

6 Validation of the Model with Newly Proposed Metrics

We have applied our model to other recently proposed metrics, reported in 2009 and 2010 for a figure of new proposals. As a start, we divided the metrics into two groups: as the first group, we have only considered those metrics which are

published in highly influential journals, especially in science citation indexed/ expanded Science citation indexed. The second group consists of the papers which were published in conferences and other journals. The metrics, under consideration in the first group are cognitive complexity measure (CCM) [20], [38], (2009,) unified complexity measure (UCM) [15] (2010,) complexity metrics for evaluation of metaprogram complexity [5] (2009), package coupling measurement (PCM) in OO Software [39] (2009), and weighted class complexity [13] (2009). The other groups belong to the metrics: a complexity measure based on requirement engineering (2010) [40], OO cognitive-spatial complexity measures [41] (2009), structured software cognitive complexity measurement, [42] (2009).

Firstly, we evaluate the cognitive complexity metric, which were initially proposed in 2004 as cognitive functional size measure (CFS) [20]; later they were promoted by considering the remaining features (explained in forthcoming lines). For CFS, the authors demonstrated their metrics via a case study (a single program in three different languages) and validated their work by applying CFS on a set of 20 programs which were taken from a book. They have also performed a comparative study with a line of code. Later, they extended their work and proposed cognitive complexity measure [38], which was dependent on architectural and operational complexities. The work was validated via demonstration with a case study, examples and a comparative study. This metric (CFS/CCM) satisfies all the steps of preliminary empirical validation; however, it was not implemented in any industry project. Since this metric satisfies only the first stage of empirical validation, acceptance can only be partial.

The second metric under evaluation against our model is a package coupling measurement in OO software [39]. This metric is a coupling metric and takes into consideration the hierarchical structure of packages and the direction of connections among package elements. The metric was demonstrated with examples, theoretically validated and empirically validated by using 18 packages taken from two open source software systems. The authors claimed that they had found a strong correlation between package coupling and the understandability of the package, and hence the metrics could be used to represent other external software quality factors. As per our model, the metric is well supported (1^{st} stage) with example(s); a case study was performed on a java project (2^{nd} stage); it was applied to 18 packages from an open sources (3^{rd} stage); and it was analyzed properly (4^{th} stage). This metric completed all the steps of preliminary stage of empirical validation, except for a comparative study with similar measures. However this metric is not applied in an industry project.

Another metric under examination is the metrics for the evaluation of metaprogram complexity [5]. The authors proposed five complexity metrics: relative Kolmogorov complexity, metalanguage richness, cyclomatic complexity, Normalized difficulty and cognitive difficulty for measuring complexity of metaprograms at information, metalanguage, graph, algorithm, and cognitive

dimensions. For validation of their metrics, the authors demonstrated their metrics with examples/case studies, then applied their metrics on open PROMOL, a metaprograms created from Altera's library for OrCAD VHDL components library. These metrics also passed all the 4 steps of the preliminary empirical validation; however no evidence is found for the implementation of their metrics in industry.

Table 1
Validation of the proposed model against the newly proposed metrics

	Steps of our model _____ Complexity measure	1	2	3	4	5	6	7(Acceptance)
Complexity measures form SCIE Journal	Cognitive Complexity Metric	Y	Y	Y	Y	N	N	Partially satisfied
	META Program Complexity	Y	Y	Y	Y	N	N	Partially satisfied
	Unified Complexity Metric	Y	Y	Y	Y	N	N	Partially satisfied
	Package coupling Measurement	Y	Y	Y	Y	N	N	Partially satisfied
	Weighted Class complexity	Y	Y	Y	N	N	N	Partially satisfied
Complexity metrics published in conferences and non SCI journals	Complexity metric for req. Engg.	Y	N	Y	Y	N	N	Not Recommended
	Object-Oriented Cognitive-Spatial Complexity Measures	Y	N	N	Y	N	N	Not Recommended
	Towards Structured software Cognitive complexity measurement	Y	N	Y	Y	N	N	Not Recommended

Our next metric under examination is unified complexity measure (UCM) [15]. This metric includes all major factors responsible for the complexity of a program, including cognitive aspects. The authors claimed that the applicability of the measure is evaluated through empirical, theoretical and practical validation processes. The authors performed test cases and a comparative study. According to our model, UCM also satisfies preliminary empirical validation. UCM was demonstrated with an example, validated theoretically and empirically with test cases/ a number of examples (more than 30) and the developers performed a rigorous comparative study with similar measures. However, it was not applied in industry.

The last metric from the first group under examination is the weighted class complexity (WCC) [13]. The metric was proposed to compute the structural and cognitive complexity of class by associating a weight to the class. The authors claimed that the theoretical and practical evaluations based on information theory

had shown that the proposed metric was on ratio scale and satisfied most of the parameters required by the measurement theory. When we evaluated this metric against our framework, we found that WCC is demonstrated with examples (1st stage), case study was performed (2nd stage), it was also applied on a number of classes (3rd stage), and the authors performed a comparative study with CK metrics suite (4th stage). However, they failed to apply it in a real industry environment.

In [31], a complexity measure based on a requirement engineering document, the authors claim that their paper attempts to empirically demonstrate the proposed complexity metric, which is based on IEEE Requirement Engineering Document [43]. However in summary, they demonstrated their metrics with a single program and then applied it on 16 small programs. They also apply the other metrics on these sixteen programs and concluded their results. Furthermore, neither did they apply their metric on a real example/project, nor did they apply it to any industrial project. As result, this metric achieves only three steps, and even could not cover the steps of preliminary empirical validation process. Hence it does not satisfy the preliminary empirical validation.

In [41], OO cognitive-spatial complexity measures, the authors proposed a metric by combining cognitive and spatial aspects of programming. Further, they claimed that the proposed measures were evaluated using standard axiomatic frameworks (which are Weyuker's properties [32] and Briand's framework [44]), and that they were compared with the corresponding existing cognitive complexity measures as well as the spatial complexity measures for OO software; hence their proposed measures were better indicators of the cognitive effort required for software comprehension than the other existing complexity measures for OO software. Using our model: on the validation part, the authors demonstrated their metric by two programs of 21 and 45 lines, and applied similar metrics on those two programs to perform a comparison. They did not perform any of the following: a case study, test cases, and industrial applications. As result, this metric covers only two steps of preliminary empirical validation; hence it does not pass all steps of our preliminary empirical validation.

In [42], towards structured software cognitive complexity measurement with granular computing strategies, the authors integrated the concept of granular computing and cognitive complexity. They claim that they performed the empirical studies, which were conducted to evaluate the virtue of their metric and also the universal applicability of granular computing concepts. In fact, the authors demonstrated their measure with a short program and then applied it on 12 small programs. This measure satisfies three steps from our model. No case study and test cases were performed; hence all the steps of preliminary empirical validation of our model were not satisfied.

The examination of eight recently developed metrics against our model provides valuable information regarding the actual scenery and facts of software metrics.

1 All the metrics which are published in the SCI/SCIE indexed journals performed better in the empirical validation process.

2 All the metrics in the first group (published in SCI/SCIE journals) passed all the steps of preliminary empirical validations.

3 Most of the developers of the metrics in the first group claimed that they had completed the validation process, except the developer of UCM and WCC. Although they only completed the first phase of validation, i.e. preliminary, they do not have the intention to do later steps in future either. This is because most of the developers are academicians and they do not realize the real needs of the software industry. They may defend their positions by applying their metrics on open source projects and a complete validation process may be achieved; however, this is not sufficient for the acceptance of the metric from the software industry.

4 The metrics in the second group, B (conferences and non SCIE indexed journals), were weak in empirical validation; they did not even perform simple test cases/case studies available on the web.

5 One can find the same evaluation results if he applies our model to most of the available metrics of group B.

The above evaluation validates each step of our proposed model. Most of the metrics in group A satisfy all the steps of preliminary validation process. On the other hand, the failure to satisfy the steps of advanced empirical validation is due to a lack of proper guidelines for the complete empirical validation process. Our model is an attempt to fill this gap by mentioning the need of advanced empirical validation in real industrial environments. In fact, the steps in the model are not new; we have formally integrated what we have surveyed, which can act as a guide to a developer of a new metric. With this model, a new developer can understand all the required steps of complete empirical validation, which can help to gain acceptance of his metric or formula within the industry.

It may be too early to assess the future of all the above metrics, because all of them have only recently been proposed. It is possible that some of them may be evaluated by the industry. On the other hand, in the present scenario, by using our our model, it is proved that none of them satisfies the steps of advanced empirical validation; i.e. they have not been evaluated in the industry. In this respect, there is less chance of the adoption of these metrics in the industry, though they do contribute to the community in the form of research papers. These observations proved our statement that most of proposed metrics are inherently irrelevant to industrial needs, which is because of improper empirical validation.

6 Discussion and Recommendation

We want to point out that there is a practical difficulty in advanced empirical process because most of the software industries are likely to be unwilling to apply a new technology/metric since it is difficult to convince them that the metric is more beneficial in comparison to the existing ones. This is one of the reasons why most of the new metrics are not properly empirically validated. Nevertheless, advanced empirical validation is a considerable requirement in validating a new metric and hence we propose the developer of a new metric should follow the following steps:

1 He should prepare a software tool for measuring the complexity value. This tool can be used for the advanced empirical validation as well as for preliminary one. A simulator can also be developed at this stage to help evaluate the new measure before applying it to real industrial data.

2 Based on preliminary observations/result of simulations, a group of practitioners should be appointed for real observations from past sample projects/sub-projects in the industry.

3 This group must apply the proposed technique as well as existing similar metrics on the sample industrial data. It is worth mentioning that once this job is done by practitioners, it will solve the problem of searching for current data from the industry.

4 Further, this group must analyze the results by comparing them with similar metrics. This activity will lead to the evaluation of the proposed metric.

5 At this stage, it is observed that if the developer of the new metric/measure belongs to industry, it is relatively easier for him to follow these steps. On the other hand, if the developer belongs to academia, then again he faces the same practical problem, if the funding is not available. This is actually a common problem, especially in developing countries, and when parental organizations that are not willing to provide any funds for this purpose. Bearing this in mind, we suggest including one practitioner in the starting phase of the proposed metric. His contribution may be in the last phase, which is the most important task for the proof and value of the new metric/measure. It is also worth mentioning that advanced empirical validation may take a few days, weeks or months, depending on the situation (e.g. availability of project) and complexity (number and size) of the proposal.

Conclusion and Future Work

It is generally observed that for most of the new metrics/measures, the developer tries to prove his metric to be the most suitable measure for any particular attributes e.g. [20]. The academicians/developers of the new metrics try to prove their claim by evaluating their proposal through different means. These different

evaluating standards can be theoretical validation (for example evaluation through measurement theory), experimentation in the classroom, case studies, different examples from the web, etc. However, they can be essential but not complete. It is proved that none of the newly proposed metrics are validated against an industry project, and hence the chances for the success of these measures are not promising. As a result, without proper preliminary and advanced empirical validation according to our model, any other criteria for the metric validation cannot be effective.

In general, the empirical methods suggest proposing a model, developing statistical / qualitative methods, applying to case studies, measure and analyzing, validating the model and repeating the procedure [45]. All these forms of empirical validation are recommended for any empirical study in software engineering. Our analysis has suggested that all these steps are required for the proper empirical validation of software metrics. Accordingly, we have accumulated them in our model in order to validate software metrics empirically.

Acknowledgment
The author is thankful to the Editor and reviewers for their valuable comments. I am also thankful to Dr. Tolga Pusatli for improving English of the paper and several rounds of discussions for finalization of the paper.

References

[1] Kitchenham, B. A., Pfleeger, S. L., Pickard, L. M., Jones, P. W., Hoaglin, D. C., El-Emam, K., Rosenberg, J. (2002) Preliminary Guidelines For Empirical Research In Software Engineering, IEEE Transaction on Software Engineering, 28(8), pp. 721-734

[2] Singer J., Vinson N. G.(2002) Ethical Issue In Empirical Studies of Software Engineering, IEEE Transaction on Software Engineering, 28(12), pp. 1171-1180

[3] Brilliant, S. S., Kinght, J .C. (1999) Empirical Research in Software Engineering, ACM Sigsoft, 24(3), pp. 45-52

[4] Zelkowitz, M. V., Wallace, D. R. (1998) Experimental Models For Validating Technology, IEEE Computer, May issue, pp. 23-40

[5] Damaševičius, R., Štuikys, V. (2009) Metrics for Evaluation of Metaprogram Complexity, Journal of Computer and Information science, 16(3), pp. 1-20

[6] Hanny, J. E., Sjoberg D. I. K., Dyba T. (2007) A Systematic review of theory use in software engineering experiments, IEEE Transaction on Software Engineering, 33(2), pp. 87-107

[7] Fenton N. E. (1999) Software Metrics: Success, Failure and New Directions, J.of System and Software, 47(2-3), pp. 149-157

[8] Gopal, A., Mukhopadhyay, T., Krishnan M. S. (2005) The Impact Of Institutional Forces of Software Metrics Programs, IEEE Transaction on Software Engineering, 31(8) pp. 679-694

[9] Chidamber, S. R., Kemerer, C. F. (1994) A Metric Suite for Object Oriented Design, IEEE Transactions on Software Engineering, 20(6), pp. 476-493

[10] Kapsu, K., Shin,Y., Chisu W. (1995) Complexity Measures For Object-Oriented Program Based On The Entropy, In Proceedings of Asia Pacific Software Engineering Conference, 6-9 Dec., pp. 127-136

[11] Zuse, H. (1991) Software Complexity Measures and Methods, de Gruyter Publisher

[12] IEEE Computer Society (1990) IEEE Standard Glossary of Software Engineering Terminology, IEEE Std. 610.12 – 1990

[13] Misra, S., Akman, I. (2008) Weighted Class Complexity: A Measure of Complexity for Object Oriented Systems, Journal of Information Science and Engineering, 24, pp.1689-1708

[14] Kushwaha, D. S., Misra, A. K. (2006) Improved Cognitive Information Complexity Measure: A Metric that Establishes Program Comprehension Effort, ACM SIGSOFT SEN, 31(5), pp. 1-7

[15] Misra, S., Akman, I. (2010) Unified Complexity Measure: a Measure of Complexity' The Proc. Nat. Acad. Sci. India, (Sect. A), 80(2), pp. 167-176

[16] Basci, D., Misra, S. (2009) Data Complexity Metrics for Web-Services, Advances in Electrical and Computer Engineering, 9(2), pp. 9-15

[17] Basci, D., Misra, S. (2011) A Metric Suite for Maintainability of XML Web-Services' IET Software, In press

[18] Aggrwal, K. K., Singh Y., Kaur, A., Melhotra, R.(2006) Software Design Metrices for object orinted Software Journal of Object Technology, 6(1), pp. 121-138

[19] Misra, S., Akman, I. (2008) Applicability of Weyuker's Properties on OO Metrics: Some Misunderstandings", Journal of Computer and Information Sciences, 15(1), pp. 17-24

[20] Wang. Y., Shao J. (2003) A New Measure Of Software Complexity Based On Cognitive Weights, Canadian Journal of Electrical and Computer Eng., 28(2), pp. 69-74

[21] Wang. Y., Shao J.(2006) Psychological Experiments on the Cognitive Complexities of Fundamental Control Structures of Software Systems, In Proc. IEEE ICCI 2006, pp. 1-2

[22] Basili, V. (2007) The Role of Controlled Experiments In Software Engineering Research, Empirical Software Engineering Issues, LNCS, 4336, 2007, pp. 33-37

[23] Zazworka, N., Basili, V., Zelkowitz, M. V.(2008), An Environment for Conducting Families of Software Engineering Experiments, Advances in Computers,74, pp. 175-200

[24] Pusatli, T., O., Misra, S. (2011) Software Measurement Activities in Small and Medium Enterprises: An Empirical Assessment', Accepted for publication, Acta Poletechnica Hungarica, 4, In Press

[25] McCabe, T. J. (1976) A Complexity Measure. IEEE Transactions Software Engineering, 2(6), pp. 308-320

[26] Shannon, C. E., Weaver, W. (1949) The Mathematical Theory of Communication, Urbana, IL: University of Illinois Press, USA

[27] Davis, J., LeBlanc, R. (1988) A Study of the Applicability of Complexity Measures," IEEE Transactions on Software Engineering, 14, pp. 366-372

[28] Etzkorn, L., Gholston, S., Hughes, W. E. Jr. (2002) A Semantic Entropy Metric, Journal of Software Maintenance And Evolution, 14(4), pp. 293-310

[29] Gaffney, J. (1984) Instruction Entropy, a Possible Measure of Program/Architecture Compatibility, ACM SIGMETRICS Performance Evaluation Review, 12(4), pp. 13-18

[30] Basci, D., Misra, S. (2011) Entropy as a Measure of Complexity of XML Schema Documents' Int. A. Journal Of Information Technology, 8(1), 16-25

[31] Serrano, M., Trujillo, J., Calero, C., Piattini, M. (2007) Metrics for Data Warehouse Conceptual Models Understandability. Information and Software Technology, 49(8), pp. 851-870

[32] Weyuker, E. J. (1988) Evaluating Software Complexity Measure. IEEE Transaction on Software Engineering, 14(9) 1357-1365

[33] Pold, I. (1989) C++ for C Programmers, pp. 156-157, The Benjamin Cummings Publishing Company, Inc.

[34] Douglas A. Y., (1992) Object-Oriented Programming with C++ and OSFMotif, Prentice Hall

[35] Robert L. S. (1992), C++ Component and Algorithm", MNT Publishing Inc.,

[36] Stevens, A. I. (1992) C++ Database Development, MIS Press

[37] Subramanyam R., Krishnan M. S. (2003), Empirical Analysis of CK metrics for Object-Oriented Design Complexity: Implications for Software Defects," IEEE Trans. on Software Engineering, 29(4),297-310

[38] Wang Y., Shao, J. (2009) On the Cognitive Complexity of Software and its Quantification and formal methods. Int. Jour. Of Software Science and Computer Intelligence, 1(2), pp. 31-53

[39] Gupta V, Chhabra J. K. (2009) Package Coupling Measurement In Object-Oriented Software. Journal of Computer Science and Technology 24(2), pp. 1-12

[40] Sharma, A., Kushwaha, D. S. (2010) A Complexity Measure based on Requirement Engineering Document, Journal of Computer Science Engineering, 1(1), pp. 112-117

[41] Gupta, V., Chhabra, J. K. (2009) Object-oriented Cognitive-Spatial Complexity Measures, International Journal of Computer Science and Engineering, 3(2), pp. 122-129

[42] Benjapol, A., Limpiyakorn, Y. (2009) Towards Structured Software Cognitive Complexity Measurement with Granular Computing Strategies, In, Proc. 8th IEEE International Conference on Cognitive Informatics, pp. 365-370

[43] IEEE Computer Society (1994) IEEE Recommended Practice for Software Requirement Specifications, New York

[44] Briand, L. C., Morasca, S., Basili, V. R. (1996) Property Based Software Engineering Measurement, IEEE Transactions on Software Engineering, 22(1), pp. 68-86

[45] Basili V. (1993) The Experimental Paradigm in Software Engineering, Lecture Notes in Compputer Science, 706, pp. 1-7

Historical Origin of the Fine Structure Constant

Part II
Subtilis Structurae Constans Inversae Arboris Dei

Péter Várlaki[1, 3], **Imre J. Rudas**[2], **László T. Kóczy**[1, 3]

[1] Széchenyi István University
Egyetem tér 1, H-9026 Győr, Hungary, varlaki@sze.hu, koczy@sze.hu

[3] Budapest University of Technology and Economics
Műegyetem rakpart 1-3, H-1111 Budapest, Hungary

[2] Óbuda University
Bécsi út 96/B, H-1034 Budapest, Hungary, rudas@uni-obuda.hu

Abstract: In this paper we intend to show in some great medieval works that are indeed or very likely linked to St Stephen's court the central role of the number-archetype 137 organizing "fine structures", together with quaternary and denary proto-Kabbalistic "systems", as a possible primordial image and "model" of the quantum-physical fine structure. This is associated with the four quantum-numbers and the fine structure constant (FSC) concept, in the sense that Jung and Pauli discussed similar problems upon the scientific and spiritual history of the Western Thought.

Keywords: fine structure constant; number archetype 137; background processes

1 Introduction

In the introduction of the first part of our article we presented the modern creational and incarnational allegories in connection with fine structure constant (FSC) through mainly Pauli and Mac Gregor's background-physical comments [13, 28]. During the analysis of the collaboration of Jung and Pauli we also showed that both great minds in their own worldview take the religious thesis of spiritual/godly incarnation as an idea of the second creation, in accordance with the Christian traditions (see *Creatio continua* and *Incarnatio continua*). Thus we can talk about a dualistic idea of creation in connection with the interpretation of FSC. We showed the fulfilment of the idea of the incarnation of Christ connected to the number 137 on a "real fine structure", such as the picture of the Emperor Constantine on the Holy Crown (in the role of King Solomon with the face of an

archangel, and he also appears as angel Jophiel too - **יופיאל =137**) that we can, thus, hypostatically take as a forerunner of the above-mentioned ideas of background physics with close connection of the system of meanings of the number 137 on the Holy Crown of Saint Stephen that is the constant primordial image of the fine structure. We put the "fine structure" of enamel pictures and the picture of Constantine (which contains the word *constant*) in the focal point of the allegorical interpretation of the number 137 structure. We exhibited the close connection of the **111+26=137** number composition with the interpretation of Jophiel, the crown angel and heavenly priest, who was regarded as the angel prince of the Torah, the interpreter of 70(72) languages in antiquity; and in the quoted excerpt from medieval times it is the representative of the number 137 and the bearer of the dual feminine-masculine crown, along with personifying the crown consisting 42 letters, that shows the incarnation's basic number [28].

As could be seen, we started using Mac Gregor's double interpretation of the fine structure constant. The first bestows FSC with a creative and incarnate ability. The other interpretation presumes a wider, unknown ability belonging to the constant and the number 137, since he regards it as a governing principle behind the micro-phenomena of the material world [13]. In the first part of our article, we showed the incarnational idea of the number 137 in the assumed primordial model. Now we are going to use other primordial models that come from a similar source to show creation myths, and interpret and illustrate the managing and ordering abilities beside the incarnation abilities. The basis of our analysis is going to be a 137 structured picture of Christ's incarnation tree from *Hortus deliciarum* [7], which can be closely connected to the Holy Crown's enamel picture, which has been already analysed in detail (and which illustrates synchronicity through the "fine structure" of Constantine).

We would like to present the discussion of the Creation together with the idea of the incarnation in the form of depictions in the structure of 137, presupposing the same authorial circle of those "pictures" found in and on the *Book of Bahir*, *Hortus deliciarum*, *the Pala d'Oro in Venice, the Coronation robe (Casula) of St. Stephen and the Holy Crown of Hungary*, all of which have similar meanings and are isomorphic with each other. In the above-mentioned pictures we can talk about 137-structure compositions, and the inverse world tree conception that is related to the creation myth; or rather, the creative and governing primordial models that in a depiction of isomorphic structures appear in almost identical meaning patterns [17]. Next to the 137 structure of the inverted tree, the mythologemes of the creation, the representations of the incarnations' other partially apparent and partially hidden trait is the central role of the angel Jophiel, whose name carries the number 137 [3, 4]. Although he is the angel prince of the Torah in the Judaist tradition [6, 9], he is also the personification of the Sephirothic (inverse) tree, the divine coronation and the crowning of the King Messiah. Thus he meets the requirements of those Christian traditions in which Jophiel is an archangel, the cherub who escorts out Adam and Eve as God has commanded and he who guards

the Tree of Life with his fiery sword [12]. Firstly, while the Tree of Life could become the symbol of the inverted Sephirothic tree, it's guardian, Jophiel, could become the embodiment and personification of the compositions of 137. Secondly, as the personifier of the crown consisting of 42 letters or names, he is the managing (regulating) archangel of the divine incarnation. Lastly, he is the prince of interpretation in the widest sense [6], the hermeneutic explanation and of course the Kabbalah. Thus, who has him living and working inside that is the hermeneutist on the uppermost level or using Pauli's expression, the honoured and chosen "un-detached observer"[1].

In *The Book of Bahir* the Tree of Life and cherub (archangel) coupling appears as Tamar, the date palm tree and probably, in a hidden manner, Jophiel the angel. And does it so in a way that the Tree of Tamar impersonates the decimal Sephirotic tree of the just or true men, while from the twin motif of Tamar's frond, the Lulav, as we could see leads us to the number 137. In the case of Jophiel, this appears inversely, since he represents the number 137 directly, from the "meaning" of his name, which in Bahir and other previous works was stated as 111+26, or more specifically the decomposition of *ALF+IVI* (**אלף יוי**) or *ALF+IHVH,* which is actually the deconstruction letter by letter of the name Jophiel, again leading us to the decimal Sephirotic system or tree. Because, it is Aleph, the first Sephirah, the origin of everything, which is the same as *IVI*, the

[1] Exactly in this hidden and indirect way it is represented in *Bahir*. We find in section 98 the only tree, which encapsulates the holy forms, the date palm tree or Tamar; a closed frond of it is Lulav or the 36 people on the 32 ways of Wisdom. This leads to the number 137 in the dichotomy of Tamar. In the previous, 95th section, it is God's, this only tree, which was planted by God as an inverse Sephirotic tree according to the 22nd section; the author describes it as a 2x36 and 2x32, in other words along with the unity basically a 137 system, a derivation of the Lord's 12 branched Zodiac world tree. (Kaplan here refers to the 12 edges of a cube, thus the number of the geometrical places is 26 and the unit of the cube is 1 [2]). This is why one can construe it with the formula of *IHVH* and *A(LF)* or 111+26). Coming back to the resulting Tree of Tamar in section 98, the author equates it with the Tree of Life through the numbers 32 and 36, next to which a cherub is placed by God as a guardian. The other appearance of the number 137 in *Bahir* (in section 70) is the identification of the name of God and the Alef (Aleph). Numerically, this means the conjunction of the numbers 111 and 26, which equal 137 as a crown on the head of the kings who follow God, according to the verse of Micah 2,13. Thus in section 70, as we showed in part one, the primordial formula of *ALF IVI* creates an anagram of the series of the letters of 137 and gives us the name Jophiel. Since the Tree of Life, as a Sephirotic tree, is characterised by the number 137, it can be presupposed that in *Bahir* a conscious editing took place, (keeping the Christian traditions in mind) in order to identify the angel Jophiel, the personifier of the cherub guarding the tree and the number 137 - in the form of 111+26. (Taking into mind the context of Bahir's complex imagery and meaning system, we support the hypothesis of Neumark as opposed to Schoelem's, since Neumark already presumed the *Bahir* as a very thoughtfully, carefully edited work with a hidden meaning system, without the collective and individual prejudices before the start of academic discussions [17, 28].)

name of God, the "modified" Tetragram, which is already and naturally the representation of the 10 Sephiroth in *Bahir* [2, 14]. It is in a way that the first letter is the letter Yod, the second is a Sephirah, the first He letter is the third Sephirah, the upper Shekhinah, while the second He letter is the 10th Sephirah and the lower Shekhinah. (Heavenly and Earthly Queens). The letter Vav, with a numerical value of six, represents the 6 Sephiroth between the upper and the lower Shekhinahs. Thus comes from the name of Jophiel not only the number 137 but the already structured, well-known denary Sephirotic system as well, so in this interpretation he himself is the prince of the Kabbalah (137!). In addition to this, in the dynamic course of emanation in the first phase of the Creation, the lower 9 Sephiroth unfolds from the Aleph of the uppermost crown in accordance with the system given by the name of the angel. Therefore, the binary and unitary personification of the number 137 resulting from the decimal Sephirotic tree of the Creation and the unfolding of the 10 Sephiroth is the Tree of Life of Tamar and Jophiel, the cherub-archangel. The other decompositional version of Jophiel's name, i.e. *FLA IVI* (**פלא יוי**), means the "Hidden secret of the Lord" which may explain its usual presence in the background (see *ALF* and *FLA* in [2]).

We are going to show these pictures while analysing the creation of the inverse tree in the pictures of the previously listed great works. As we will see, in both cases the dynamic flow and the structure of the creation and the incarnation appear in a uniform representation before us. In the case of the process of the second creation and incarnation, as we observed in the first part, it connects to the dual interpretation of Christ's maternal and paternal origination of the number 137. The paternal origin (the Gospel of Matthew) represents the spiritual and legal incarnation of the Holy Spirit; meanwhile, the maternal origin represents the bodily aspects of the incarnation (The Gospel of Luke). In the pictures of these works, the idea of the inverse cosmic tree with the '137' composition (which includes the 10 Sephiroth), as the primordial image of the fine structure in quantum theory, symbolically carries the arrangement of the spectral lines (in a tree structure-type way) through the "inverse number" of 137, i.e. through the primordial concept of the fine structure constant[2].

[2] In the *Hortus* this archetypal image is completed with the so called embedded "four worlds" of the Kabbalah, where each of them contains the denary ("1-3-7") Sephiroth system. Here, this, together with the 137 structure of the "image", can be interpreted as the primordial model of fine structure related to the "world-building" four quantum numbers.

2 The Theory of the Divine Incarnation in the Portrayal of the Tree of Life in Hortus Deliciarum

2.1 General Abstract

We can see the colourful, beautiful and amazingly interesting illustration of the incarnation of Jesus Christ in *Hortus deliciarum* (Fig. 1) in the picture marked Fol.80v *(Bastard Facs. 8(16))* [7]. If we perceive the picture in the customary way, i.e. that God plants the tree of incarnation, then, of course we talk about an inverse tree reaching down from heaven, from the world of angels and stars, which absolutely fits the inverse tree described in the 21 and 22 sections of *Bahir* [2, 14]. Here, in *Bahir*, this inverse tree in an ambivalent way, even taking the incarnation into account (see §.22 and §.191), and especially matches the Tree of Abraham, which he planted in Beersheba, which was named as Tamarisk (Tamarix) tree (translated from the Hebrew original) at the end of the Classical Antiquity and in the Middle Ages. Here, in the *Hortus* picture we can count the Divine Tree's outer, paired 3-3 branches, which are in accordance with the Sephirotic system, and the central trunk consists of 4 people too, which means the 4 central Sephiroth. Thus the Divine inverse tree that reaches from the stars and the archangels to the Earth is the Living Church, and next to the 10 Sephiroth it shows exactly 137 "persons" (entities). So the image of the Creation, coinciding with the picture of the incarnation of Christ, represents both the 10 Sephiroth and the "137 fine structure" coming from them (as we will see later) in perfect concordance with the Holy Crown's image system symbolising the incarnation and the creation.[3] It is axiomatic that the angel who points at the stars in the

[3] In the first part of our article, our starting point was that by taking the paternal lineage of Luke from Adam to Abraham we get a **61**(62)+**76**=**137** "system" in the Vulgate. We need to assume the otherwise stated order of 3x14 in connection with the number 42 that prevails if and only if, according to one branch of the tradition, we take into account the name of Mary in the number 42, in the 3x14 name structure from Abraham to Jesus, which includes the name of only 41 patriarchs. Structuring of the text from David to Christ and from Abraham and David into the system of 14 notoriously leads to ambivalent interpretations. The most simple case for interpreting the number 137 composition according to the Vulgate in the case of 62+76 would obviously be the giving of the 137 composition by adding the interpretation of the unity of the two Christs. Concentrating on the name of 41 (+20) from the Gospel of Matthew of course the 61+76 dual incarnational structure would, rather simply, lead to the number 137. On the Holy Crown, both can be observed in the numeral systems of the pictures of the two archangels and the Emperor Constantine with Solomon. This is how, for example, we can count 61 white pearls at the edge of the ephod of the two archangels, and these refer directly to Christ's 61 intellectual and legal patriarchs. These archangels, as spiritual-heavenly creatures connected to Mary (one announces while the other protects her) can be viewed as the perfect personifications of the paternal incarnation relayed by the Holy Spirit. The 76 white pearls on the Emperor's robe would of course represent directly the 76 patriarchs of the bodily incarnation. We

picture is no other than the archangel Jophiel. Since, as we saw in the first part, in the already mentioned Kabbalistic traditions he is the "ophen" (**אופן**), the 137 and the Atara (Stephanos!) or the Prince of the Crown. From the 137 people, exactly 26 have a crown. Thus the 111+26, or the decomposition of *ALF IVI* self-evidently bears a crown in the godly name of *IHVH or IVI* in a letter and number correspondence. This is how the 111+26 or the decomposition of *ALF IVI* leads us naturally to the name of Jophiel, identifying thus the archangel Jophiel who explains, interprets and presents the whole picture.

On the picture we can see the angel with the 16 stars and above them we can identify 15 persons as a consistent pictorial structure. The angel is pointing up to the stars while the other accent is on his mouth, because he "says" *"Look now toward the heavens and count the stars, if you are able to count them. And He said to him, So shall your seed be"* (Gen 15,5).

drew your attention to another fine structure, in which on the edge of the ephod of the aforementioned persons, the number of the identical white pearls was 111 while the number of the white pearls outside the robe of the emperor was 26. The former, as we saw, is the aleph, the other is the number of the name of God, which in the form of IVI with the letters of ALF, leads us to the name Jophiel, IVFIAL. In this way, here in the Holy Crown's picture of Constantine, the constant of the fine structure, or its Constantine, is the 137 or the 1/137 that is personified by Jophiel. Thus Emperor Constantine in the role of King Solomon with a face of an archangel appears as Jophiel too, who impersonates creation along with incarnation in addition to Archangels Michael and Gabriel. We could observe the same in the almost identical picture of the Pala d'Oro's Prophet Solomon, where among the 12 prophets, only Solomon's arch's and halo's pattern is the same as the pattern of arches and halos of the 12 archangels in the uppermost line. This significant breaking down of the symmetry (the arches and halos of all the prophets showing a different type but congruent with each other) draws attention to the fact that, taking into account the central letter of 137 as archangel Solomon and the different ornamental patterns (that we will elaborate on later), he appears as archangel Jophiel before us, again. In a bit more complicated way (but these interpretations are possible in that case as well) if we take into account that in the Greek version of the list of names in the genealogy of Luke from Abraham to David, as opposed to the Vulgate and the 14-part original Hebrew list it contains 15 names, because between Esrom and Aram another name is registered. In this case we face of course several variants in the case of the interpretation of the number 137, since the translator could use the Greek list of names, i.e. the predecessor of the Latin version and seen as the original, to complete the list of names given by Matthew, to secure the order of the 3x14. (In this case the second list of names of 14 begins with David so the third list of 14 would come out too.). Without this, however, the 61 Matthew-type and the now 77 Luke-type Greek names can only be interpreted as a 137 structure if we take the Unity of Christ into account. If the combination of the 62+77, in other words considering the name additionally given by Luke in both Greek lists, then the 137 composition can be interpreted either without taking Christ into account or with the uniting of Adam and Christ (taking both as one unit).

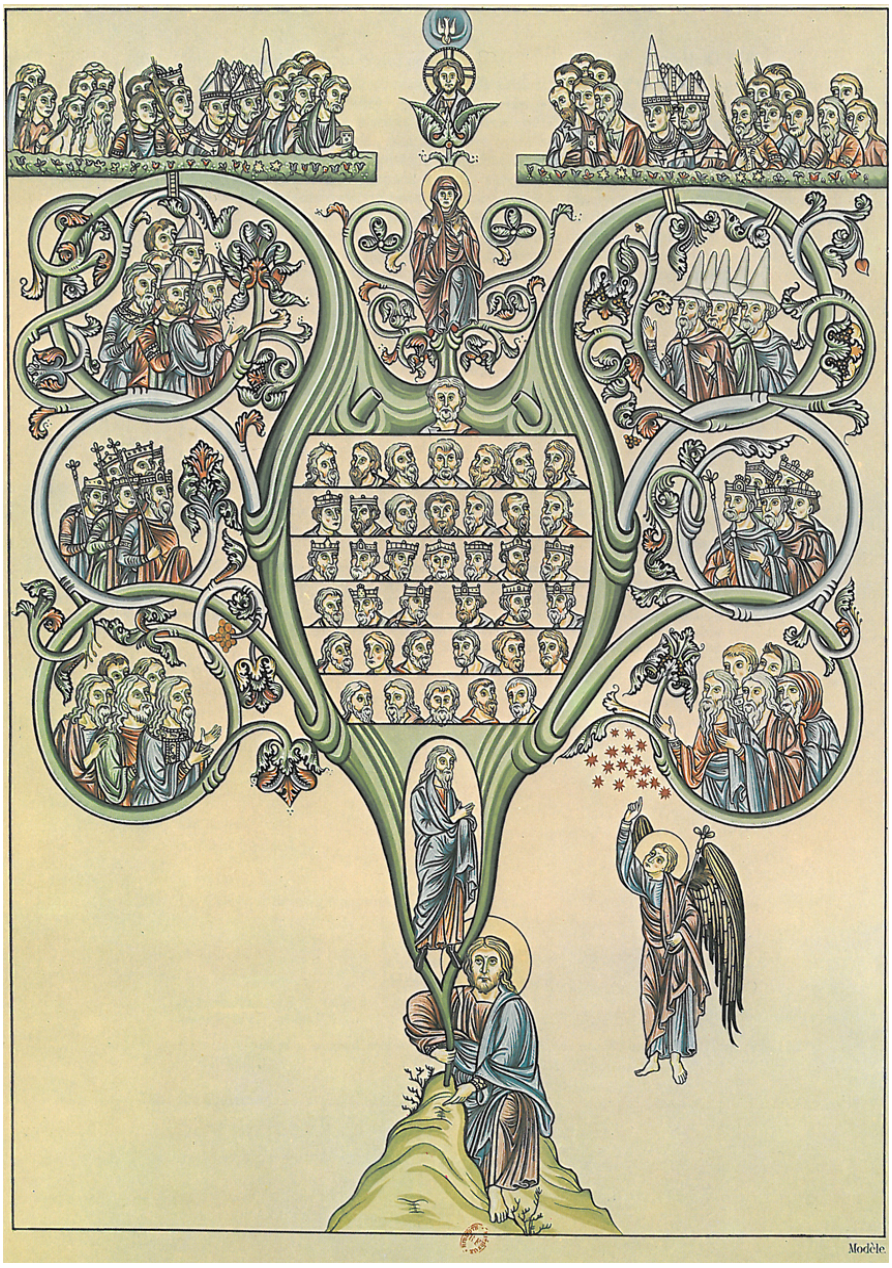


Figure 1
The image of the incarnation of God with the inverse tree of incarnation (Hortus deliciarum, Fol. 80v, Bastard Facs.)

The quotation in this context can be interpreted as an allusion to the significance of counting. The original Hebrew text can be interpreted in such a way that it relates to the number of the seed of the Messiah of Abraham, i.e. if you can count it, i.e. the number 137 on the basis of the picture. With this reference to the number and counting twice in this single sentence can mean to the mystical interpreter that the angel is the prince of numbers, counting and grammar as well. It can be confirmed in the context of *Bahir (section 124, 125)* by the Hebrew anagram of the 10 Sephiroth of God: **סתר יופיאל = י ספירות אל.** The meaning of the above expression is that the God's 10 Sephiroth hide Jophiel, or Jophiel represents the 10 Sephiroth of God, which is in full agreement with the interpretation of Jophiel's Hebrew name. Thus, in the Hebrew transliteration it is the number 16, which is **IV(יו).** This, taken together with the related 15 persons, is 31, which is **AL** (**אל**). If the mouth symbolically represents the angel, whose Hebrew word is **FI (פי),** then we may obtain the name of angel as ***IVFIAL*** (**יופיאל**), i.e. Jophiel (thus its mouth represent the two names of God). Probably it is a result of conscious planning. The other possibility of quotation Gen. 15,5 or 26,4 deals with the multiplication of the stars. It brings up the allusion to the sum 1+…+16=136 which, together with the angel, gives the number 137. The 16 stars, which contain 8 rays, lead to the number 128. It is completed by the 9 decorative elements of the angel's ripidion (The ripidion's Cherub or Seraph has 6 wings and 3 other entities, e.g. hair, head and neck). It would also mean number 137.

As we shall later discuss in detail, the idea of the Hortus incarnation picture is based on the inverse tree of creation from the Slavonic Book of Enoch [17]. The "creational meaning" of the picture is based upon the identification of the four embedded "(1)3+7" Sephiroth structures (where the number 1 is the unity of the Trinity - see e.g. Bahir's §.139 and 140.) in the Incarnation Tree illustrated by Fig. 2. It is a clear isomorphic representation of four embedded worlds of the later Kabbalah (inside of Adam Kadmon), where each world contains its "(1)3+7" own Sephiroth tree.

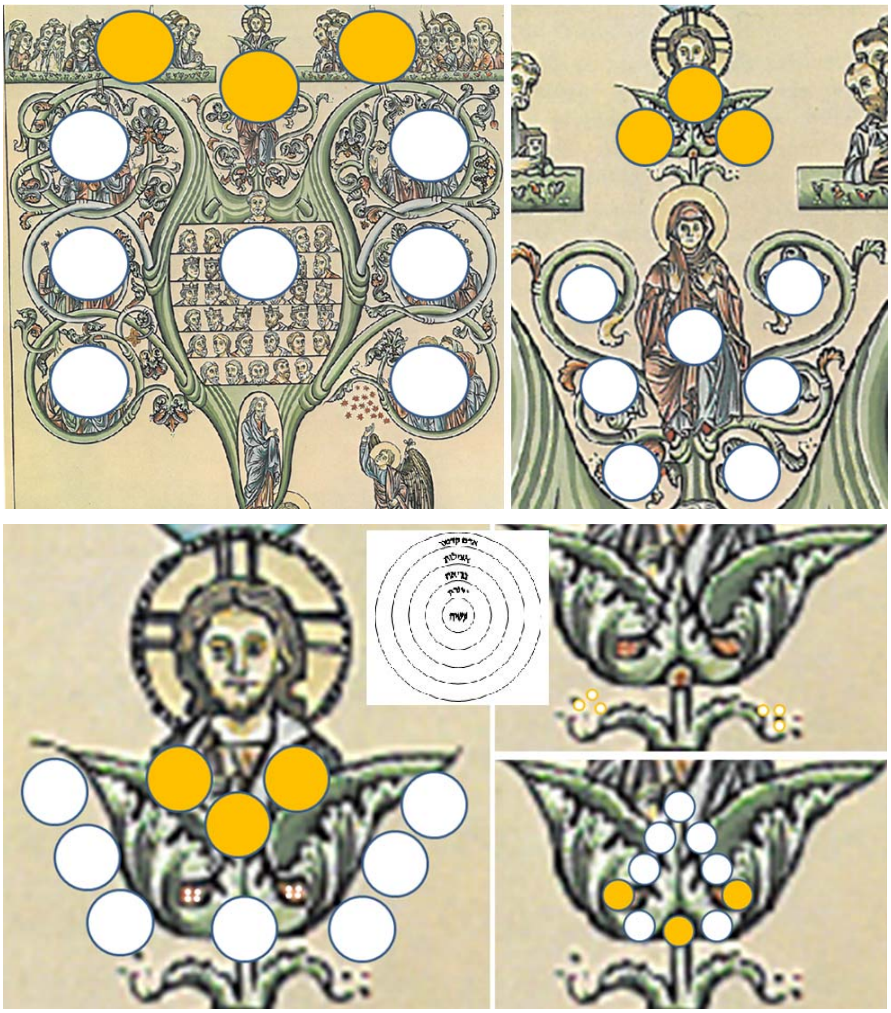


Figure 2

The illustration of the four embedded worlds (with the 10=3+7 Sephiroth systems) on the Hortus Incarnation picture. In the middle of the picture the four embedded worlds of the Kabbala can be seen inside Adam Kadmon (Emanative, Creative, Formative, Active entities respectively – in Hebrew).

Here Adam Kadmon's all-embracing role corresponds perfectly to the unity of the Holy Trinity. Thus, this representation of the pictorial system (in the form of a flower and "plant-symbolic") beside the 137 composition, for the four worlds (emanative, creative, formative, active), contains its "signum sacrum" which is the "(1)3+7" structure hinting at the number 137 as well. In the plant symbolism of the four worlds the first two are mainly tree representations while the other two are represented by the clover archetype with seven parted leaves showing the form of **A(**lpha) and **ω(**mega). The first and last Sephira of the third world is **A** and **ω**,

respectively, while the fourth vertically reversed (i.e. "Active") one, with the "3+7" leaf-motives, constructs the letter-pattern of **A** and **ω.** (Fig. 2) The interesting novelty of this primordial model is that, beside the creation-incarnation motive, it contains (also in plant-symbolism) the "(1)3+7" "Sephirotic" structure of the permanent generation of the Church by Christ using the obviously seven sacraments for the participation in the unifying mystery of the Holy Trinity [7]: "*Ihesus Christus flos florum gignit ecclesiam per babtismum et ceterorum sacramentorum cultum"* (see the same motives in Pala d'Oro on Figs. of Solomon and the "Last Supper" in [28]).[4]

In the picture Abraham is also looking at those 4x4=16 stars that are pointed at by the angel who relays the words of God; what is more, according to the Biblical placement he represents God directly. In this case, however, with the 38 patriarchs in the trunk, and Joseph in the middle below Mary, we can count only 40 forefathers instead of 41. In this case of course, just as the interpreters before did, God or rather Christ or Adam (since God in the Byzantine theology cannot be portrayed) is the planter of the tree. In accordance with this theory, Schmidt (see in [7]) gives a list of names assigned to the heads, the 14 crowned heads kings of Israel from David to Josias. Thus, the lineage from Adam (from the Lord) is symbolically given, and with the exception of specifically one ancestor (who can be replaced by Mary) the 42 forefathers according to Matthew, too. In this particular case, however, the potential of the creative deformation can still be in effect, since we do not get the lineage of ancestry in the form of the 3x14 ancestors specifically stated by Matthew. On the other hand, if we imagine Abraham as the sitting and planting father symbolically representing the Eternal Father and Christ too, then the figure watching the stars must be Isaac, and then in the middle comes Jacob. (It is also "well established" hermeneutical interpretation according to Gen. 26, 4, where Isaac is also watching the numbers of the stars). In this very case Matthew's whole order of 42 is valid, and in the structure of 3x14. Indeed, since from Abraham to David (see Fig. 1), we can count 14 patriarchs without a crown. In this case however, unlike in the previous case, the crowned person is not David but Solomon, whose crown alone bears the sign of the name of God, the 5 dots that is the symbol of the Hebrew letter He (HaShem), since in the Talmud and in the *Book of Bahir* God gave him his own name. This confirms that the first crowned person can really be David or Solomon and the symmetry is

[4] It is known that the origin of the four worlds of the later Cabbala (based on the verse of Isaiah 43,7.) can be found in the Bahir's §, 77, 78. Here, this idea is also related to the Messiah's genealogy concerned from Abraham, where the God's manifestation is between the two living creatures (cherubs) based on the LXX version of Hab 3,2. In this case Abraham is blessed with "All" (in Hebrew Bakol) where the "All" is partly the Sephiroth tree from §.22, partly the "Daughter or Mother" (!) of Abraham. The "All" is related to the God's Glory (emanation) and to the creation, formation and making (action), i.e. the four worlds of the God with the denary Sephiroth trees [2]. Repeatedly, we may conclude that the authors of the Bahir and the original "Incarnation picture" we should search in the same spiritual and intellectual circle.

full with the 14 crowned figures who are followed by another 14 patriarch ending with Christ. Symmetry can be viewed as "complete" in this interpretation too. Here the upper figure interpreted as Joseph can be linked to Mary. The controversy comes from the fact that Isaac is watching the stars as opposed to Abraham, and the row of the kings does not start with David. With the patriarchal equation of Adam-Abraham we can interpret the secretly-present 20 patriarchs from Adam to Abraham (according to Luke).

Naturally, the Biblical basis for the flower and plant allegory of the "Incarnation Tree" (and the sevenfold form of the Holy Spirit), can be found in the verses of Isaiah 11,1-3 *("et eggredietur virga de radice Iesse et flos de radice eius ascendet et requiescet super eum Spiritus Domini*"). Precisely this inverse Sephiroth tree in the reversed form of 7+3 is introduced in the Bahir in §.186, according to Isaiah 11,1-3. This tree of Jesse is the Incarnation Tree's written explanation in the Hortus as well: "*Radicem namque de qua arbour genealogie hujus processit, Dominus manu sua plantavit in terra sua et usque ad perfectionem arbores custodivit, in cujus summitate virga erupit, cujus flori septiformis spiritus insedit. Et quam pulcher ordo! In manu Domini (1) terra (2), in terra radix (3), in radix arbor (4), in arbore virga (5), in virga flos (6), in flore Spiritus sanctus (7)"*(**301** in [7]). Here, we may see the embedded 7 or "3+7" structures, openly because of the "septiformis" of the seventh, i.e. the Holy Ghost. We may find an isomorphic seven inverse tree's layers on the Royal Robe as (1) God's hands, (2) cherubs-seraphs, (3) angels, (4) prophets, (5) apostles, (6) just men, (7) saints. (see Fig.7 and 8.) Furthermore, on the Hortus picture, God with his left hand points the root of the incarnational tree, "quoting" *"ego Iesus misi angelum meum testificari vobis haec in ecclesiis ego sum* ***radix et genus David*** *…". (Rev 22,16)*

On the other hand, as other authors have noted, the folding of the tree and the fact that God or the lowest patriarch is pointing at the root with his left hand represents mostly from the Bible (Isaiah 11.1) the root and the folding of Jesse. Jesse is "*pater filius et mater"* according to the Medieval depicting methods (Fig. 232 in [10]) so the lower figure, in the unity of the Father and the Son, is Jesse and David. The Figure standing on two folds, previously interpreted as Abraham or Isaac, would be in this case David's son, Nathan. With him till Christ, there are 42 persons in the genealogy of Luke, including Mary, which is exactly the same number, given by – reflecting old traditions – the number of patriarchs from Nathan to Christ. Then we can interpret in the lowest central figure the 34 forefathers according to the gospel of Luke from Adam to David through Jesse, who represents God as a Father. In this case we take into account the tradition that the list of names from Luke shows the lineage of Mary (who is substituted by her husband, Joseph). Then the head under the central figure of Mary shows not Joseph but her father, Heli, according to this tradition. Here the productive controversy comes from the fact that we cannot interpret personally the 38 patriarchs in the middle according to the characteristics of their representation (e.g. their crowns).

The lineage system is enforced by the fact that on Christ's halo, which is inseparably linked with the Holy Spirit appearing in the form of a dove, we can see 42 ornamental entities (Bastard Facs.) 20 on the right hand side and 20 on the left and two at the top (Fig. 1). The hand of Christ, the *gloriole* and the Holy Spirit creates a trinity. The base is Christ's right hand that, as if at the same time holding the *gloriole*, is pointing at the dove. The Hebrew word for hand or palm is at the same time the name of the letter *Kaf* (כף = כ) which, through the letter-number, also means the number 20. Thus, the probable allusion of 20+42=62 with which the Holy Spirit crowns Christ, strengthens the hypothesis of the 137 structure of the incarnation. Through this interpretation Mary and the bole underneath her represents the maternal lineage from Nathan to Mary with 41 names while below patriarch Jesse *pater-filius* represents in one person the shared system of lineage from Adam to David. The role of the archangel Jophiel is self-evident, since he ensures and arranges that the Jophiel crown (*offen*, *atara*), which consists of 42 letters (names), rises onto the Creator's (Christ's) head.

Note: If we want to interpret the 42 Greek names in the incarnational tree of the *Hortus* without counting Mary in the 3x14=42 structure, then obviously because of the ambivalent supposition God, as the identity of the Father, can be personified only by Abraham, according to the well-known tradition of the restriction of the depiction of God too. In this case there is the 42 paternal forbearers from Abraham to Christ and thus the interpretation of the 137 incarnational structures can be given by using the above-mentioned version. It is also obvious that this case is not the base situation or the primal occurrence since in the trunk the 14 regal fathers from the gospel of Matthew create a unity with the 14 crowns, in which the first crowned person is David and the last one is the king is Josias. Here the conscious edition is so strong that the act of giving the 7, 5, and 2 crowned personages line by line refers to the 10 Sephiroth, the golden crown and the Hebrew word of Gold, *ZaHaB (*זהב*)* that represents the regal bride. This plays an important role on the Holy Crown as well as on other pictures of the *Hortus* too, in accordance with the conception written in the 52-56 sections of the *Book of Bahir*. In not the main interpretation, based on the correction of Greek names by Luke, King David doesn't belong to the crowned heads; he in a way precedes them. The number 14 not only the word *Zahab* but as is widely known they are the numbers of the name of David, so King David can be interpreted as part of the 14 than the first item of it but he can also precedes the 14. In this very case the following 14 kings generates the name of David.

The planting of the inverse tree in the *Hortus* picture can be interpreted, based on the particular depiction of God's (or the patriarch Abraham's) leg and mountain ridge, that the inverse tree of creation and incarnation sprouts from the navel of God or the loins of Abraham. The inverse tree as the inverted cross tree (Tree of Life) can be viewed in a very similar depiction of the Holy Crown's picture called the Pantocrator, where the cross tree coming from the navel of God [5] is pointing at our world in an inverted fashion. It is self-evident that, as we are going to

present in details in the followings parts, both depictions can be traced back to the picture of creation of the book of Slavonic Enoch where the ancient Aion sprouts from the stomach (navel) of the God named Adoil (Adoel). Scholem writes about it*: "This aeon bears the inexplicable name Adoil; (**עדואל**=111=**אלף**=Aleph!) the proposed etymology 'aeon of God' would, in any case, be very poor Hebrew." However this* great Aeon or the inverse tree growing out of the abdomen or the groin of God was precisely match-able with name ***Adoil*** in the given hermeneutical circle, completely independently from its former real Hebrew form, which is unknown even nowadays. We are going to analyse this important idea in detail in the following sections, and in part three of the paper as well.

2.2 Analytical Description with Notes

The two leaders of the Church (St Peter and St Paul) situated on either side, in the middle and in an uppermost position in the *Hortus* picture represent the dual structural order in effect in the church in a way that their attention is supervising the other structural order taking effect. Similarly to the Royal Robe's and Pala d'Oro's pictorial systems it contains also symbolically the *two cherub systems*. Here, they are realized by two books showing the symbols of the four Evangelists (four cherubs) in the forms of the quincunx. They are held by the two Church Princes, Peter and Paul, mirror symmetrically in the right and left side of the picture, respectively. In this part of the picture we can see that Christ probably is born between two leaves of the cabbage. It is an unpaired pictorial-linguistic play because the Hebrew name of the cabbage is the same as the word of "cherub" (**כרוב**). Thus, this allegory means, according to the Judeo-Christian tradition, that God's presence will appear between the two cherubs (see e.g. its illustration in the *Hortus*) [9]. Furthermore, the two leaves of the cabbage (see on Fig. 3) is closely related to the two books with the symbols of the two cherub-systems (and we may still mention that the cabbage is the plant of the health and life in the Talmud). The 2x4 white points in the red circles of the two cabbage ("cherub" in Hebrew) leaves, on the right and left side of "Alpha and Omega", strongly confirms the symbolic presence of the two (four-faced) cherubs. (Fig. 3)

The trifolium with centre of "**A**,**ω**" together with the six branches of the Holy Virgin's tree, as a denary system, is an isomorphic pictorial map of the Greek crown, while the blessing Christ with the distinguished four-four apostles on his right and left side (according to the whole view and the "direct" position to Christ) can be accepted as an isomorphic pictorial map of the Latin crown of the Holy Crown of Hungary, where the cross is symbolized by the dove-image of the Holy Ghost. Here the sun is signified by Christ's golden halo and the moon by the "circle" of the Holy Ghost.[5]

[5] There is another decisive and entirely unique representational isomorphy between the Hortus incarnation picture and the Holy Crown of Hungary. Namely, on the Holy Crown's Pantocrator picture's quadratic frame we may find 12 white pearls and 12

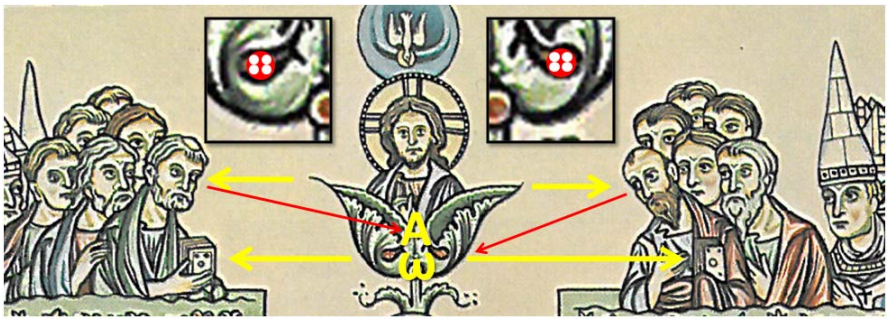


Figure 3

The illustration of the two cherub (twice four evangelists) systems on the Hortus Incarnation picture

This corresponds exactly to the textual description which ensues that the way the Lord procreated his church ("*Jhesus Christus flos florum gignit ecclesiam*") is based on the very same incarnational order in which the Lord is personified in Christ, which is being represented and supervised by the two heads of church together in a twin-like complementing unity. This order of the 137 structure can be counted right there on both sides of the tree by listing the seven persons on the

red gems around Christ's monogram X. On the four hoops of the crown we can identify a 5+8+10+5 "pearl-gem structure" (15 white pearls, 5 horizontal and 8 vertical posed red gems) for each of them, which could be a map of the Hebrew word of the living creature (5+8+15=**החיה**) or cherub. It is a unique (pictorial-linguistic) abstract representational form of the Lord's throne picture, in the style of Ezekiel vision, from the Apocalypse (4,4-6 and 5,6-8.): "*Around the throne are ... twenty-four elders, dressed in white robes, with golden crowns on their heads... and on each side of the throne, are four living creatures."* In this 112+24+1= 137 composition, where 137 is the linguistic-numeral symbol of the "throne-wheel" from the Ezekiel vision, the 112 is the number of the "Lord (is) God" in Hebrew (112=**יהוה אלהים).** The pair-less symbolical representation of the "picture" is that the 12 white pearls is the symbol of the 12 apostles and the "Mercy" and the 12 red gems is the symbol of the 12 patriarchs and the "Law" (Judgment), both 12 "elders" from the Apocalypse (e.g. 7,4 and 21,12). Namely, the full number of white pearls is 72 (the number – and colour – of "Mercy" - *Chesed*) and full number of the red gems is 64 (the number – and colour – of the "Judgment" - Din). In the Hortus incarnation picture no considering the hidden ("singular") young king among the martyrs we may count beside king David (as the symbolic representation of the Lord) 24 kings with crowns, 12 on the right 12 on the left side of the picture. Here, the king Joatham ("God's perfectness" in Hebrew!) in the proper centre of the picture, we count on the left side of the picture which in this case, paradoxically (see Bahir §.52 where the God's right hand is the left one), contains (with him) 72 entities. On the right side (counted the other "central" forefathers to the left side) then, we may find 64 entities. In this case Christ is the 137th one. This 112+24+1=137 composition with the above details, both in structure and meaning, is equivalent with the discussed representational system of the Holy Crown. Concerning in the central symmetric position 12 apostles and 12 prophets on the Royal Robe (Casula) with the Christ on the throne we may obtain another 112+24+1=137 composition (counting two cherubs).

central axis in the larger group of the incarnational order by Luke. Then by adding the 4x4=16 stars on the left hand side of the tree structure, meaning the descendants of Abraham and Christ, together we can count 75+1=76 people in the form of 34+41 (19+15=34 and 16+17+8=41). On the right hand side we find 61 people, again the structural form of 34 (17+17) + 27. Thus, minus the only Christ, we get the many-times-mentioned and analysed structure of 61+1+75=137. Within this we differentiate between the 34+28 and the 34+42 pieced and branching two incarnational orders too, which is also the order of the Church procreated by Christ. The twin-shaped depiction as a flower on the incarnation of Tamar or the Tree of Life requires us to consider the twin-state together with unity. In this latter case we get the structure of 137 as the meaning of unity. The consistency is almost perfect and complete in relation to the pictorial and meaning structure of the Holy Crown. With the 7 figures on the right hand side on the central axis without Christ we can validate the 68+1+68 (already analysed) composition of Tamar. Counting the four-four figures on the inside and the outside of the axis on the right and the left hand side, 64 +1+72 or 65+72 =137 respectively, we can interpret compositions of vital importance. From the richness of structural solutions we are going to highlight two, related structures. One of them considers the number of hands (palms). The number of these, including the angels', is 32. On the uppermost level together with Jesus Christ's visible right hand we can count 12, on the branch on the right hand side 7, on the left 6 and below, in the case of the three figures, 5 hands. Adding Mary's two palms to 7, we get the Hebrew form of the name Tobias [27]. On one hand he is the one who caught the great fish, on the other hand he is the maker of the dual pontifical and regal crown. In the *Admonitions (as in § 134. of Bahir),* the structure of 32 given in Hebrew letters is the *atarah*, connected to the word crown [27]. Similar 32 structures can be identified at several other places, like on the Pala d'Oro as well. Its interpretation is "the good of the Lord" or the good of the Lord's decimal order [27]. The 10 of course stands for the crown compiled for the 10 Sephiroth.

The other structure comes from the number of the crowns. In the trunk we can count 14 crowned figures among the kings of Israel. From David through Solomon and King Ezechias to Josias all can be identified based on the second, incarnational component of 14 from the gospel of Matthew. On the right hand side we find 6 crowned figures while on the left there are 5. From these on the left the first is probably Constantine the Great, based on the depiction of his crown as the *Basileus (Βασιλευσ*) of Solomon or David, the creator of peace, builder of the church of Sophia, just like Solomon. As we have already analysed, his crown bears the sign of God just as Solomon or David did, which is his privilege only.

This sign of God can be found in the books held by St. Peter and St. Paul as the symbol of Christ, the four evangelists or the four cherubs. On the other side the leading crowned figure with the marked regal sign could perhaps symbolise the empire, personified most probably by Charlemagne.

Figure 4
The hypothesis: St Emerich among the virgin martyrs holding the palm branch of victory

As in the uppermost high priest branch, the Latin and Greek priests greet each other on either sides and the ruler of the two sides in the middle do the same. Thus we can identify 16 crowned people/kings, but we cannot identify 9, while we can see the 26th (or in reverse order the first) more or less hidden amongst the young virgin martyrs holding the palm branch of victory in the uppermost level of the right hand side. He too, is holding a palm branch, as the symbol of victory and perhaps martyrdom, while paying attention to the Holy Spirit descending onto the head of Lord in the form of a dove. (Fig. 4) His crown is as simple as possible; its shape is the same as the crown of one of the kings of Israel. There are no more crowns like this in this picture. (The difference is that on the crown from the Old Testament the ornamental system we can observe is 3+7 while on his it is 3+6. This could mean the non-fulfilment, and our hypothesis is that this regal figure is St Emerich (Emericus, Imre), son of St. Stephen, as the contributor to the creation of the name of God, maybe strengthening his position in martyrdom of life. The 26 crowns as a result of this and the 26 crowned figures refer to the 111+26 structure we analysed in our previous article, where the number of the name of God (YHVH) while the 111 is the number of the Hebrew word *Aleph* (ALF), the natural meaning of which is 1000; but it could also mean the letters of *Aleph*, so thus the number 1, and furthermore, it could mean Keter's uppermost crown in the Kabbalah [2]. In this way, the name of God forms a crown in multiple ways, since 26 is the name of God and according to Hagigah 13b makes a crown. This is supported by the interpretation of the quote of Micah in the 70th section of *Bahir* that the name of God is on the heads, or in the Aleph which, amongst the many possible interpretations, could be taken that true kings (or rather in general the true) wear God's name as a crown on their heads. The number 137, as we have mentioned several times, means God's crown consisting of the 42, feminine and masculine or the Atarah and Keter names (letters) together through the number 137 of Jophiel and Ofen, the wheels of the throne chariot in the Eleazar fragment [3, 4]. Since Christ's *gloriole* or crown, the simplest symbol of the incarnation, consists of 42 names, the symbolism of the 137-crown with its 42 letters or names in the Eleazar fragment stands visibly very close to the symbolism of the *Hortus* and the Holy Crown of Hungary in their meaning systems (together with the 42 words of the Admonition's "*Capitulatio*" [27]).

In the Christian interpretation these symbolic contents together mean the resurrection of the Lord after a thousand year (*resurrectio prima*), re-incarnation (spiritual incarnation) on the occasion of the 137 number-archetypical, Sephirothic or teurgic, dominical and incarnational crowning of the new, temporal king. Here, in the worldview of the "author" in the "background" such a – methodically pioneering, mirroring the modern probability theory – idea that is composed by Jung 900 years later in his famous *"Answer to Job"*, with full identity of the medieval coronation idea of King:

"Although the birth of Christ is an event that occurred but once in history, it has always existed in eternity. For the layman in these matters, the identity of a nontemporal, eternal event with a unique historical occurrence is something that is extremely difficult to conceive. He must, however, accustom himself to the idea that "time" is a relative concept and needs to be complemented by that of the "simultaneous" existence, in the Bardo or pleroma, of all historical processes. What exists in the pleroma as an eternal process appears in time as an aperiodic sequence, that is to say, it repeated many times in an irregular pattern."(See e.g. the time and frequency domain representations of stochastic processes.)

Back to the hypothesis of St. Emerich (*Emericus*), we feel that the truth of this strengthens the fact that the Holy Crown was a Roman (Latin and Greek) regal crown made for St. Emerich which we can refer to with the names father and son together, rightly so. This is made probable by the highlighted appearance of St. Stephen as a Proto-martyr with the face of an angel and with an apex crown on his head. Because he indeed comes right after to the apostles on the Greek side, with 8 ornaments resting on 6 utterly special and unique basic decorations on his head. These together show probably the verse 6, 8 in the Acts of the Apostles, in which it is stated how great and wonderful deeds St. Stephen did. This crown of a main deacon as a headdress of a hidden figure appears on the other side in mirror-symmetry, most probably denoting St. Lawrence, as Rome's other major main deacon. The number of the days between their feast days (the 10th of August and the 26th of December) is 137!

The symbolism of the name of God and the crown makes understandable the inseparable unity of Solomon and Constantine "Basileus" (through the representation of the crown angel-priest Jophiel), similarly to the inseparability of the Old and the New Testaments on the community of the number 137. They both wear God's name; Solomon, according to the Bahir (65. §), is the king who peace is attached to and in his name the word 'peace' is connected to the letter *He*, meaning the abbreviation of the name of God, with the numeral value of 5. The Constant, the attribute of permanent-ness, unchanging-ness; the *UNI T* makes understandable the permanence and invariant state of the only sign of redemption in the name of Constantine.

As we have repeatedly mentioned, peace belongs to both of them, Israel's peace of the Messiah and the Pax Romana of Christ; and St. Sophia personifies both their

churches and temples.[6] So the significant number 5 on both their crowns refers to the name of God, *H=HSM=YHVH*, while the crown itself, as we saw, is ALF or the Aleph; thus the two together *ALF YHVH* shows us the number 137 again, as the numeral archetype of the divine universal crown. (We should not forget that the name of Solomon in Hebrew also means robe). Another crucial element carrying importance is that his mother, in a transferred meaning by Sophia, crowned King Solomon with the crown of atarah. The 3,11 verse of the Song of Songs is treated by Great St Gregory as the incarnation or the symbol of the divine personification in which the Virgin Mary is crowning the Lord with her body (womb). Thus, in the Christian interpretation, the triple (!) crown of Solomon is the symbol of the incarnation as well (**726** in [7]). Thus the crown of Constantine –identified with him – is the crown of Mother Sophia too, the symbol of the incarnation of God in the crowned ruler. While Solomon and the hypothetical Constantine look in the same direction with the same crown, the face of the hypothetical Charlemagne, which resembles strongly "David's face" on the Hortus picture, looks in the other direction, as a mirror image of the other ruling pair. This is most natural, since while Constantine the Great identifies himself with Solomon in his antitype, Charlemagne does this with David on purpose, as he is, according to Melchizedek, the high priest king who also identifies himself as the king of peace based on his name (as *rex Salem*). As is commonly known, Alcuin and his academic circle called Charlemagne David.

3 The Cosmic Birth of 137 in the Book of Bahir and in the"Creation Image" of the Holy Crown

In the Pantocrator picture of the Holy Crown of Hungary as well as in the *Hortus* the incarnational drawing above and in the textual part of the *Bahir* (sections 21 to

[6] The conscious intention of the identification of Solomon and Constantine can be observed in the shared meaning systems of the Pala d'Oro and the Holy Crown. In the Pala d'Oro, Solomon is standing in a chief position, citing the verse 9,1 of the book of Proverbs in which Sophia is building a temple with seven pillars. She gives wine and bread to those who approach her. In the picture of the Last Supper, the table shows the schematic picture of the temple of Sophia with seven pillars founded by Constantine. The picture is natural because the Last Supper symbolises the eternal service of wine and bread in the church. Thus, bearing in mind the identical apse decoration of Solomon and the archangels, Solomon, Michael and Gabriel archangels on the Pala d'Oro and Emperor Constantine with the (Jophiel's) angel-faced King Solomon with the two archangels on the Holy Crown carry the same meaning system. In the Pala d'Oro all of this is intensified by placing peace and a princess called Irene (i.e. Peace in Greek - Ιρηνη) in the same line with Solomon, while in both, the coronation robe and the Pala d'Oro, the unique system consisting of two cherubs can be observed - which of course is tied to the Ark of the Covenant and the Temple of Solomon [9, 23].

23) in an almost identical depiction, the image of the Creation appears, based on the Bible but interpreted mythologically.

This is most obvious in the script of *Bahir* where the Lord himself creates the world from the primordial chaos without archangels Michael and Gabriel and it is said that no one was with him in this primordial creational phase. According to the text he was the one who planted the inverted tree that is the core of everything and which he names wholeness in Hebrew; this is the ancient Aeon that is the origin of everything. The actual text is the following:

"*It is thus written Isaiah 44:24), "I am God, I make all, I stretch out the heavens alone, the earth is spread out before Me. [Even though we read the verse "from Me" (May-iti), it can also be read] Mi iti –"Who was with Me? I am the One who planted this tree in order that all the world should delight in it. And in it, I spread All. I called it All because all depend on it, all emanate from it, and all need it. To it they look, for it they wait, and from it, souls fly in joy. Alone was I when I made it. Let no angel rise above it and say, "I was before you." I was also alone when I spread out My earth, in which I planted and rooted this tree. I made them rejoice together, and I rejoiced in them."*[2]

This is then in the *Bahir* the ancient Aion or the creation image of the emanation of the 10 Sephiroth. Since no one or nothing was with Him, according to the given allegorical image or primordial model, obviously the tree had to originate from Him. This type of creation image of the *Bahir* can be read in the next section. Here a king would like to plant a tree, to which he is looking for a source of water. We find exactly the same image in the 191st section of the *Bahir*, where substituting God and from his order, Abraham is the creator in the world, personifying mercy. "*All this Abraham did as it is written (Gen. 21,33), 'and he planted a Tamarisk in Beersheba, and he called them in the name of the Lord, God of the world'.(**בשם יהוה אלהים עולם**) He would share his bread and water with all the people in the world*."[2]

So in this image Abraham is named as the substitute of God, the personification of the divine Mercy, who symbolically plants the inverse world tree by the fountain of "7" (**שבע**), whose name in Latin (Tamarix) in the Medieval era might become associated with the name Tamar and the numbers 10 and 1 (**I,X**). He names the tree as the God of the World and, with this, proves that the tree in question is the tree of the 10 Sephiroth, which was made tangible by the Lord's Tetragram. Of the latter we know that in the context of the *Bahir* it can be seen as the symbol of the 10 Sephiroth, as we have already discussed.

In sections 117 and 119 the description of the (now) Tamar tree is continued (bearing in mind the already-examined allegorical pictures of sections 95 and 98). Here too, the inverse tree personifying the bottom seven Sephiroth originates from God, who has become a human, and from whom the inverse tree is growing! Since section 117 is based on verse 15,3 of Exodus, where God is the man of war. (**יהוה איש מלחמה** or using the gematria **יהוה דלת** [27]) This, in any case, is the symbol of

the God turning into a human. The Hebrew word AIS also means human as well as man. In the mystic interpretation the *Aleph* is the uppermost crown (or rock), the letter *Yod* refers to the source of the rock, while the letter *Shin* means the root of the inverse tree in the soil of God. This is where the inverse tree originates from, which symbolises the seven lower Sephiroth. Since the meaning of AIS is human, the human turned into God, in the anthropomorphic interpretation, the inverse tree grows from Him. (In another interpretation of an important *Bahir's fragment* [17] this means the three crowns united into one in an isomorphic way with the idea of the triple crowning of Solomon in the *Hortus* (**725** in [7]*)*. Here the crown means the man and the completeness of the world altogether too [17]. We learn in the 119th section that just and true people's spirits would stick to this tree, those who come through the tree's water source and stick to its trunk. In the space-time world the just and the good people mean the inverse tree's living flowers in the tenth Sephirah (or the Church). If there are some, the Shekhinah descends into the world (or unites with the people) while the just, based on their deeds, can be found in God's lap (bosom). In this exceptional allegorical image the good and just, or true people, do not lie in Abraham's lap but in God's[7]. Scholem, very rightly so, identifies this allegorical image with the inverse tree model and links it to the "picture" from Slavonic Enoch, in which the inverse World Tree grows out of God's body [17]. We can see the strong ties of this image with the Biblical picture shown in the 191st section and with the interpretation of the image when Abraham plants the tree (the inverse World Tree) by the fountain of "7" (**שבע**), the tree of the world's God, which is identified with the name of God. (A similar allegory can be found in "St Paul's joint olive tree" of the Church and Israel).

[7] Scholem writes about this part of the *Bahir*: *"The totality of the (10) powers of God thus constitutes a cosmic tree that is not only the tree of souls from which the souls of the righteous fly out (flourish!**פורח**) and to which, apparently, they return, but a tree that also depends upon the deeds of Israel." "then the root (of the tree) is the third sefirah, the "mother" in the language of the Bahir."* [17] (Compare this with the inscription of Incarnation picture of *Hortus*: "*Ihesus Christus flos florum gignit Ecclesiam*"). This part of the *Bahir* and the mentioned picture of the *Hortus* is deeply related to the Psalm 91, 13-14, where *"The righteous flourish like the (date) palm tree...they are planted in the house of the Lord"* and *"Iustus sicut palma (Thamar i.e. date palm) florebit (**צדיק כתמר יפרח**) ...plantati in domo Domini... nostri florebunt."* Furthermore, *"The trunk of the tree, which in section 85 grows out of the root, corresponds to the image of the spinal column in man, above all in sections 67 and 104. If Israel is good, God brings new souls out of the place of the seed, which corresponds to the great channel (of the tree) of section 85. The manner in which the myth of the tree is varied here (as well as in sections 104 and 121) corresponds to the interpretation given by section 15 to its oldest form, as we encounter it in section 14." (Scholem*[17]) This conclusion of Scholem corresponds to the main conception of the structural and meaning system of the Royal Casula as well (see later).

In essence this is what we can see in the picture (Figure 6.) of the upper Pantocrator of the Holy Crown, where the slanted cross[8] in a syncretistic way could be the symbol of *Aleph* and the number 10 (10 Sephiroth) too. Here grows from the stomach of God, who is pictured with a double beard, the primordial Aion, the slanted cross (or in the equivalent form of the letter *Taf (Tav)* in the Old Hebrew, which is the sign of the salvation – Ez. 9,4), which already has been identified with the letter Aleph, which as we saw contains the name of God. Then obviously it is also the cross of Christ that, according to the unbroken Christian tradition, can be perceived as the Tree of Life and the tree of the cosmic world too. It may be sufficient here just to refer to the interpretation of the quote from the *Hortus* in footnote number 9.

The cross, as the tree of Life[9], and the symbol of salvation of course refers to the initials of Christ, and so to Christ himself as well, and as a Latin numeral, the 10 Sephiroth in the form of the *Aleph*. Thus the *Aleph*, and the name of God it contains, are a strong allusion to the number 137. The X, i.e. the Latin ten, as a slanted cross may signify Hungary too, as the kingdom of Ten Tribes (Hungary = Latin Ungaria = Turk Onnogur = Ten tribes, which is a generally accepted scientific opinion for the etymology of the word Onnogur, as 3 Kabars and 7 Magyars tribes.). This primordial creation appears in the incarnation picture of the *Hortus*. If we carefully observe the figure on the pictorial throne of the mountain peak (who can be interpreted as God, Abraham or, in other interpretations, the forefather Jesse) it seems as if the shape of the seat would show that the tree, the flower of which being the forefathers and the flower of flowers, would come from the lap of Christ himself, the divine figure as the number 137th.[10] Thus it meets

---

8 On the basis of Ferencz's careful and entirely convincing proof of the originally homogeneous and uniform planning of the "triple" Holy Crown (Greek crown-1, Latin crown-2 and the Pantocrator picture with the "slanted cross"-3) originating probably in the court of St Stephen (according to his opinion as well) and consciously formed into a pair less asymmetric structure to ensure the specific shape of the slanted cross [5], we may recognize an "incarnation" of a grandiose "hermeneutical creative deformation" with its rich interpretational potentiality. (Still see Solomonic triple crown in Hortus later).

9 We can read about the cross defying death (the holy cross drawing out the Leviathan) as the tree of life in Hortus: "*Postquam primus parens per lignum in pelagus hujus saeculi quasi in verticem naufragus corruit, atque avidus Leviathan seva morte totum genus humanum absorbuit, placuit redemptori nostro vexillum sancte crucis erigere, et hamo carnis sue squamea hostis guttura constringere, ut cuspide* ***vitalis ligni*** *perfossus evomeret quos per vetitum lignum improbus predo devorasset. Hec* ***sancta crux*** *est nobis lampas lucis eterne in hujus vite caligine, que suos sequaces ducit ad celestia, suis amatoribus confert gaudia angelica."* [7]

10 In *Hortus'* Incarnation picture the embracing arms of God is remarkably specific. Most probably it symbolises the promise of his own Incarnation (the conception by the Holy Spirit) according to traditions in connection with Habakkuk. So bearing in mind the image of the tree embraced by God, the promise of the "embrace" also means that this tree originates from God himself. Thus Habakkuk and the inverse tree

with the depiction of the cross tree in the Holy Crown's Pantocrator picture, as well as with the image of coming of the lap of Abraham or Jesse's roots (seeds) (see the picture with Abraham's lap in Fig. 9). Thus these two types of allegorical systems of depiction or primordial models can be taken as isomorphic with each other. We can amplify this with the identification of the whole ornamental system of the picture Pantocrator. According to this, we can separate 32 entities with the Moon, the Sun and the red "sparkles" (scintilla) in which inside two circles (not counting the inner circle of the Sun), based on the numbers (see Fig. 5), there lies the name of God, the IHVH.

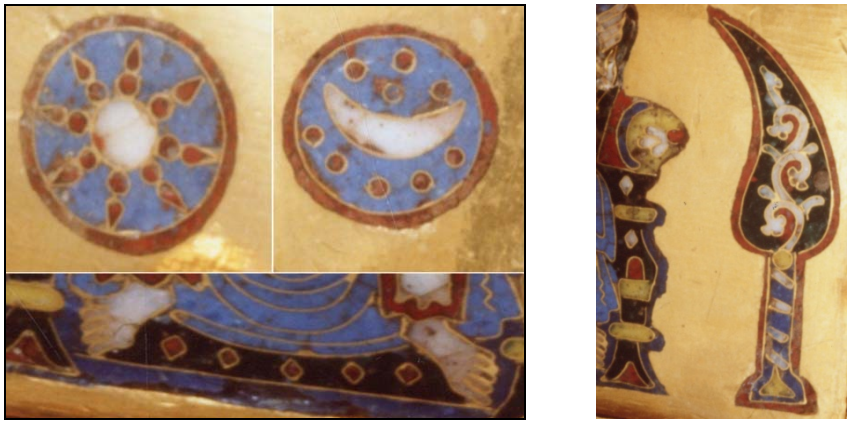


Figure 5

(a) The 32=30+2 system of sparks (scintillae) with the symbolic name of the Lord (10+ 6, 5+5 = יהוה) in the circles of the sun and moon, on the Pantocrator enamel picture of the Holy Crown. (b) The illustration of God's tree with the detail of the God's throne.

in this hermeneutic circle can be seen as belonging together. This meaning identification is further strengthened by the fact that in the picture of crucifixion, in addition to the two quotes from Habakkuk (3,4 and 3,8) there is the expression "*revertere Sunamitis*" [7]. This seems to be an obvious reference to the fact that the interpretation of the image is done by Habakkuk (according to the well-known tradition Habakkuk is the son of Sunamitis). Here even the "embrace" has semiotic meaning, the name of which in Hebrew also חבק ("Habakk"). In *Hortus*, in the picture of the prophet Habakkuk next to the famous verse (2,3) there is verse 3,3, according to which the Saint (Christ) "arrives" or "grows" from the mountain of Pharan. (In the *Hortus* the mountain of Pharan plays an important role in other places too related to the Ark of Covenant i.e. the two cherubs). Verse 3,8 combines the vision of the divine throne chariot (with a horse, see exactly the same motif in the picture of "Crucifixion" with the personified Church) with Salvation, the word of which in Hebrew can be translated as Jesus (See also in verse 3,13 the line „in salutem cum Christo tuo" and in 3,18 "exultabo in Deo Iesu meo"). Thus probably the Biblical basis of the picture is Habakkuk's 3th verse.

This system of 30 without the Sun and the Moon, according to the layout of the figure, gives the name of Judah. The slanted cross, as a Greek and Latin number together could mean the number 610, which with the numerical value of 30 of the name Judah together shows the number 640, which is the numerical value of the Hebrew name of Tamar (and the word for Sun, **שמש**). Here, based on the 197th section of the *Bahir* similarly to **306** section of the *Hortus* [7], Zara corresponds to the sun, and Phares to the Moon and as the twins of Judah and Tamar, who appear as Christ's primogenitor and progenitress.

This all is a natural and powerful (crowning) symbol of the marriage of Christ with his bride the Church because in the well known medieval allegory, Judah symbolizes God (Christ) and Tamar is the personification of the Church (**563** in [7]). It is interesting that, according to one of the traditions of the Talmud, Tamar originates from Melchizedek, thus seeming to represent alone the humanity not coming from Noah. This tradition, by the way, points to her special angelic and messianic origin, so according to Melchizedek it bonds David and Christ's high priesthoods together strongly with her own progenitrix personality, or at least to the thorough, mystic observer experienced in traditions.

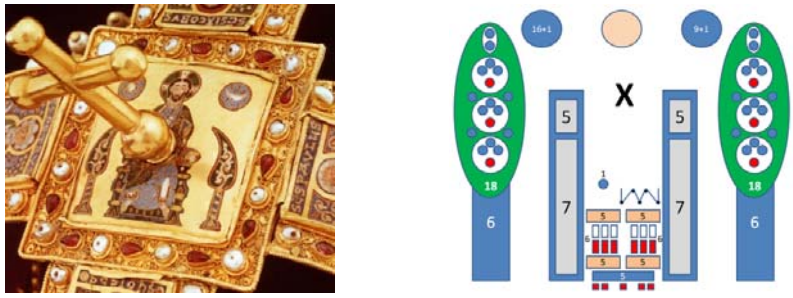


Figure 6
The interpretation of the slanted cross [5], as the inverse cosmic tree (tree of Life), on the Holy Crown with the possible 137 compositions

The resulting number 32 can be further expanded with the two "ornamental systems for the tree", the systems of 2x24 or 12+36. These together as the number 48 could point to the Hebrew word for star. On the right and the left hand side the 2x24 can be extended with the 24 items of the decoration of the throne (Fig. 6). This order of 3x24 corresponds to the primordial image of the archangelic order described in the 108th section of the *Bahir*, in close connection with section 22. (According to the *Bahir* the "holy forms" existed before the cherubs who were to personify them - 99.§.) The resulting 72+32=104 items could refer to the Hebrew word *kokavon* (**כוכבון**), which corresponds to the Greek *asteriskos,* which could be not only the sign of Christ but also the image of the ancient Aeon. Even the form of the Holy Crown is an asterisk. The third ornamental group of 32 above the 104 is made of the partly "hidden" representation of the Lord's tassel (tzitzit). Before the shin, the 2x3+2x3 equaling 12 ornamental entities indicate still 4x5 geometrical points as well on God's praying robe (they are meaningfully

isomorphic with the 10-10 fingers and toes of God). Thus, this 32 composition refers to a consciously edited 20+12=32 structure (Fig. 6).[11]

This latter, (as 10+22=32) in a well-known way, Sefer Yetzirah (The Book of Creation) means the 32 ways of Wisdom, where the 22 means the Hebrew alphabet's 22 letters which build the world, while the number 10, the Ten Sephiroth (the 10 fingers of the Hands of God) again means a world-building ten numbers (decimal system) and the 10 Sephiroths (as "living numerical beings"), of the pleromatic world [17, 19]. (We can observe this on the coronation robe by the hands of God on the left and right side; one of them has the Sun, which most probably refers to Zara, while the gap or opening by the other hand refers to Phares). In the *Book of Bahir* the ten fingers of the hands of God at the same time means the Ten Commandments as well, as it is the primordial image of the 10 Sephiroth of the Pleroma (See section 124). The resulting number of 136 is made up to be 137 by a probably ornamental item of the eye of a "fish head" on the mantel of the Lord. This is, by the way, in the Judaist tradition, the eye of the great fish of Jonah, or the Leviathan, through which the secrets of the abyss of the sea or the unconscious are revealed. The eye of the fish can be seen in a similar interpretation in the St John picture of the Holy Crown. The 137 composition as a result, of course by taking into account the plenty of other 137 interpretations as well, strengthens the interpretation of the slanted cross as the *Aleph* along with as the interpretation of the name of God with the number 137. The horizontal structure with the composition of 72+1+64=137 definitely proves the conscious planning. The Lord's left hand itself forms the Holy Scripture; more precisely his five fingers are the five books of Moses, the five ray of lights according to verse 3,4 of Habakkuk in the context of *Bahir*: "*There are five rays. This are the five fingers on man's right hand"* (section 188). In *Bahir* clearly, the Habakkuk-ian rays of light emerge from the fingers of God and they mean the Torah in an archetypal interpretation. The very same quote from Habakkuk can be read in the picture of the Crucifixion in the *Hortus* above the right hand of Christ. *Note*: The

[11] Consequently, the 32 structure of the Lord's tassel (tzitzit) can be identified by the 4x3=12 (!) triangles where each of these four "triangle-groups" indicates five geometrical points. This can be considered as the symbolic representation of the 4x8 (8=3+5) = 32 threads of the tzitzit. (Fig. 6) This "symbolic model" of the 32 structure for the Lord's tzitzit entirely corresponds to the *Bahir's* interpretation of the 32 threads of the tzitzit, where the 32 threads correspond to the 32 paths of the Lord's garden, where the cherubs are watching the Tree of Life (see §.92 and 98). Consequently, above and under of the throne, the 32 "star-entities" (together with the sun and moon) may represent another 32 "composition". Similarly, the two trees of the garden contain 24-24 ornamental entities together with the 12-12 ornamental elements of the throne on the right and left side, respectively. This ornamental composition can be considered as the 36-36 structure. Thus, it is equivalent to the 2x36+2x32 composition of the frequently discussed white pearls and red gems structure. The other 72+1+64 composition of the white pearls and red gems on the Latin crown's hoops, corresponds to the left- and right-hand ornamental 72+1+64 compositions of the throne image. (See in detail in part three.)

validity of the interpretations of the 24-type is confirmed by the frame's 12+12=24 pearl and gemstone structure too. The whole system mirrors the depictions of the creation-image of the Book of of Slavonic Enoch and of Sefer Yetzirah, as they can be "seen" through the lenses of *Bahir*. As we can observe in many symbolic depictions, because of the prohibition of the pictorial depiction, Abraham too, in addition to Christ, can symbolically personify God here. Thus, in a pictorial depiction, the unity of the white and the black beard can symbolise well the Father and the Son.

According to this, the representation can refer to God, Adam, Abraham, Isaac, Judah, Perez (Phares) and of course to Jesse and David too. In the Christian representative tradition we can see especially the duo of Abraham and Isaac, or the Jesse *Pater et filius*, Father-Son connection as the allegorical representation together of the divine "Father and Son". Even St Paul in his letters to the Galatians takes Abraham as the personifier of the Father; here the symbolic figure of Christ in the allegory is Isaac, while Sarah embodies the Libera, the new Heavenly Jerusalem, the Church of Christ and thus the bride of Christ too. We saw in the incarnation picture of the *Hortus* that the Divine father can be (besides the obvious Christ depiction) substituted by Abraham and Jesse too, since both of them can be perceived as the planters of the Divine tree. In another allegory of the *Hortus,* Judah secures the personification of the God as the father, while at this point the equivalent of the Church, the bride of Christ is Tamar (see **563** in [7]). In the Pantocrator picture, the significantly pointy triangular beard as the symbol of the son can be referring to Perez (Phares), who divides (in a sense 'breaks through') the white beard of the father (*ambivalently represented as beard or neck*), and whose name in Latin is *Divisio* [7], which can usually be translated with the words erupt or break through. The already discussed numeral systems of the light items also spell out the Hebrew word *Yehuda,* and with the interpretation of the *Bahir* and the *Hortus* of the Sun and the Moon, as the symbols of Perez and Zerach. The Moon as the symbol of the Church is the exact counterpart of the "divisio" with its black and white divided circular parts, just as the *Hortus* too discusses, where the white part could mean the Church converted to Christ while the black could be the symbol of the mass of the non-converted people (see **306** in [7]) The Sun is of course in the *Hortus* too the Zerach (Zara), just as in the *Bahir*. Naturally, in the former case it is as the symbol of Christ. It seems as if the author of the *Bahir* (§. 197) had known this 9th c. Latin interpretation of Perez (Phares), Zerach (Zara) and of course Tamar. Here the metaphor of Strabo is based on verse 4,1 of Malachias; Tamar impersonates Iudicium and Iustitia as the antitype of Mary. On the basis of the texts cited in the *Hortus deliciarum*, Thamar is a perfect Old Testament pre-figuration of the Sun woman. We may confirm it by the above-mentioned text of Strabo or the *Hortus* as well as by §.197 of the *Book of Bahir*. In both cases the Messiah twins, Phares (Perez) and Zara (Zerach), are symbolized by the Sun and Moon. The Sun woman is threatened by the red serpent while, according to a possible interpretation of the Hebrew text of the Bible, Thamar is threatened by a fiery serpent **(שרף,תשרפ)** or the fiery dead.

In addition to the Abraham-type allegory, the Pantocrator picture refers to another important allegorical picture of the Sefer Yetzirah. According to this, God makes an alliance with Abraham between the ten fingers of God's hand and the ten toes of his feet (see §.58 in the Bahir [2]). In this case, the symbol of the alliance is an "omphalic" cross that points to both the Old and the New Testament. Underneath there is a shape that can be supposed as a prayer tassel or *tzitzit*, its 6+6 tassels being the symbol of the 12 tribes and 12 Apostles of Israel. The *tzitzit* is a Hebrew word, the numerical value of which is 600, and its Greek letter is **X**. The number of the threads of the God's tassel (*tzitzit)* is 32 (in the Bahir), of which, as we mentioned above, we can see 12 here, containing 4x5 "geometrical points". This emphasises in the given allegorical system is the recognition of the 2x10 digits together in the 32-type. The Hebrew word of "*keret*" (**כרת**), as in alliance, can be rewritten as *Keter* (**כתר**), which is the uppermost crown in the *Bahir*, plus the other name and symbol of the *Aleph* and which corresponds to, in the meaning context of the given hermeneutic circle, the slanted cross.

4 The Creation-Incarnation Symbolism with the 137 Compositions Found on the Coronation Robe

4.1 About the Casula Representational System

It seems that the most important features of the theoretical and representational system of the original Coronation Robe (Casula) are the particular theological concepts of androgyny and twin-ness and special pictorial realization of these. The fact itself that the coronation mantle can be divided into two parts, as left and right side, carries rather particular and mystic theosophical notions, and this representational solution along with the Pala d'Oro in Venice is unmatched in the history of culture [29, 30]. The same applies to the division of the Robe from a front view and a back view. As a result we get a quaternal (mirror-symmetric) system in which the idea of androgyny dominates. On the coronation robe the first sign of this is the depiction of God's right and left hands. By the neck part under the collar, which was subsequently (probably when it was cut apart) placed onto the Robe, the depiction of the right hand is completed with a round sun disk and the depiction of the left hand (which is specifically limned as a left hand) with a diamond shape. The diamond shape clearly refers to the female principium as opposed to the circle's masculine symbolism. The wide usage of the Du-Parcufim *(***דופרצופים***)* principle (as appearing in the mystical theosophy and being described with the notion of androgyny) in documents written in Hebrew [9] regards the sun and the moon as the masculine and the feminine principium in a way that it is represented in the picture of God's left and right hand side on top of the crown. Other such depictions of God's hand referring to the feminine (diamond shape)

and masculine (sun) on the left and right side at the same time are not known in the history of the canonical Christian fine arts. It is almost certain that there was not and could not be an example for this in the first half of the Middle Ages. (We do not know about such an analogy in the early Gnostic depictions of fine art either, though there the anthropomorphic theosophical theory seems to be natural.) The two hands can be in close connection with the prophet Habakkuk and his book, which play a major role in the Robe's spiritual and hermeneutical system; in this, God's hand emerges from the abyss (Hab 3.10, 11) together with the rising sun and the "remaining" moon, and from his fingers light with great shining breaks through for the world (Hab 3, 4). Based on the picture this is self-evident, since the hands point clearly to the mandorlas of the Heavenly King on the right shoulder and the Heavenly Queen's on the left shoulder. In this way the rays of the left hand are mirrored on the *imago* of the Heavenly Queen, who – according to the description- shines in the sky (*emicat in celo*). So the original light from the primordial light casts its rays upon the Heavenly Queen via the God-hands; it is not She who shines but her notion, her picture, her *imago* based on a particular, neoplatonic and mystic viewpoint (*Sanctae Genitricis Imago* [22, 30]). The solution reminds us most of the 147th section of the *Book of Bahir* with Habakkuk in the main role. (Margalioth [14]). The light symbolism thus is intertwined with the depiction idea of the Du-Parcufim principle (the two-facedness idea) since on the left shoulder of the robe of the high priest king, who represents cosmic anthropos, the picture of the Heavenly Queen follows the cherubs' and seraphs' quaternary system's Du-Parcufim principle. It is worth noting that the cherubs and the seraphs appear in a dynamic aspect; we can see them holding up the mandorlas of the Heavenly King and the Heavenly Queen, praying next to them or kneeling by them. The resulting image of the 4 cherubs and 4 seraphs on the two sides cannot be compared to the usual representation of the four evangelists, but rather it reminds us of the doubled depiction of the original Ezekiel version's four living creatures, the cherubim. We note again that it is an unparalleled mode of representation in the Christian intellectual and cultural history, since the Heavenly Queen bears the attribute of divine dignity with the 4 cherubs (4 evangelists) independently from the Heavenly Father. So this is connected with God's masculine right and feminine left hands, according to the geometrical ordering too. The cherubs and the seraphs such androgynous, or Du-Parcufim, way of representation, apart from the Pala d'Oro of Venice and the *Hortus*, again is an unmatched example in the history of Christian culture. The androgyny of the cherubs, their masculine and feminine being appears early in the Talmud referring to those cherubs which can be found on either sides of the Arch of the Covenant and the Temple of Solomon. The Hebrew tradition really talks about masculine and feminine cherubs as personifications of the two names of God, i.e. **יהוה אלהים** (Idel, 1988 [9]). The royal maker when creating the robe could have learned about this idea from Origen or Philo as well, who wrote more about the cherubs as masculine and feminine creatures and the symbolic idea of their *hieros gamos* (hierogamy) as comes in effect in Judaism [9].

4.2 137 Composition on the Coronation Robe

In the representational and hermeneutic system of the Robe, similarly to the Holy Crown (and Saint Stephen's other known and supposed creations) the numeric (number) archetypes play an important basic role, particularly the different compositions of the numerical structures of 137. The Casula's (the mantle's) essential structural and hermeneutical rendering feature is shown by the two decisively important rubric systems in the 137-structure.

The rubrics of the coronation robe consisting of 137 entities are summarized below together with the rubric of the four mandorlas forming the letter Tau in the 65+72=137 composition in footnote 12.

Figure 7

The 36 deans or just men (including separately the 4 arch-deans) on the Casula

The 72 characters of the right side of the caption going round with the 64 characters on the left hand side together with the shared ("doubled") letter R creates again a 137 composition. (see in Appendix 2 of the first part of this paper[12])

The third basic 137-structure on the robe is shown by the 136 (68+68) figures found on the conical shaped robe in the "du-parcufim" symmetrical system with the interpretation of the centre as a unity. Here we can supplement the closely analyzed 72 (36+36) system above with the 24 holy, pious and large sized birds found in between the 12 saints.

[12] EMICAT IN CELO SANCTAE GENITRICIS IMAGO / DAT SUMMO REGI FAMVLATVM CONCIO CELI (65) HOSTIBVS EN XPISTVS PROSTRATIS EMICAT ALT X / SESSIO REGNANTEM NOTAS ET XPM DOMINANTEM (72). (R)EGINA HEC CASVLA OPERATA ~ ET DATA ECCLESIAE SANCTA€ MARIAE SITAE IN CIVITATE ALBA : · (72 letters and characters) ANNO INCARNACIONIS XPI : M : XXXI : INDICCIONE XIIII : A STEPHANO REGE ET GISLA R (65 letters and characters)

Figure 8
The reconstructed Casula (Royal Robe) of St Stephen of Hungary (1031)

If we identify them with swans it could be an allusion to the Dioscuri through Leda in complete concurrence with the twins of Thamar. In the case of the interpretation as storks for the birds similarly to the Latin name (in their Hebrew name they are too hassids, or in the Latin translation in the Vulgate: saints), just like the other 12 crowned saints with whom naturally they create a 36, "chassidic" order according to the Judaic interpretation, which here coincides with the stork's image as a bird of the ancient Romans. Of course with the identification of the holy birds with different types of birds can be part of the complex representational and hermeneutic system. The additional supplementation of the 72 basic elements can be meant by the counting of the 8 (4+4) representations of the cherubs (evangelist symbols) as separate entities. Thus, taking into account the theriomorphic symbolism, we can count another 32 (24+8) entities in addition to the 72. The resulting 104 figures are completed to a total of 136 by the 32 "just (true) men" standing above the apostles on the bastions of the heavenly Jerusalem. Regarding as a unity the four central mandorlas of the salvation history of Christ we obviously get to the number 137. The 32 just (true) men as the 3 faces of the 12 apostles' zodiac order together with the four highlighted prophets in a circular (the 12 or zodiac pattern not being part of the division (as seen in Fig.7), as 4 main deans or main just men, give altogether 36 deans, thus making up the system of 36 righteous (true men) that governs and upholds the world (Ameisenowa 1948 [1], Bahir [2]). Saint Emerich (Imre) in this structure is crowned by "137" and thus he takes part in the "reconcilement" of the number 72, the attribute of mercy, and the number 64, the attribute of judgment. The resulting 2x32=64 system on the Robe with the theriomorphic symbolism can be clearly interpreted as the symbol of the Divine throne chariot represented in a numeric (number) archetype. Thus to a very great extent, in its structure too it is identical to the structure connected with the 72 (36+36) true men and the entity of the Divine throne chariot as 64 (32+32), from the 95th paragraph of Bahir, which are obviously connected by the identity of the Messiah. The identification of the 64 with the Holy Virgin, the "Judgment"

(*Iudicium*) that personifies the Church, can be found in an identical form in *Hortus deliciarium* [7]. Here the "house" of the Church (and the Holy Virgin), in one of the most important pictures, appears as a throne chariot with the cherubs, in which we can count exactly 64 people (Fol. 225v). Just like in the picture of the crucifixion, the Church, personified by the crowned Holy Virgin, according to the 3.8 verse of Habakkuk, is sitting on a horse that is depicted with 4 cherubs' heads and cherubs' legs. (Fol.150r). Since 72 is the number of Mercy and God's name, 72+64 means the nuptials of Christ and the Church, while an additional entity as the 137th is the natural symbol of the reception or re-incarnation of God in the given hermeneutic circle.

4.3 Further 137-Type Interpretations

The number of the main figures or persons found on the Robe can be interpreted, in addition to the above-mentioned identification, in another two ways. In the first case the holy, pious birds supplementing the 12 saints to a 36 group will be still counted in but the so far separately counted 4 cherubs each that is similar to the structural order of the Pala d'Oro, we take as two united cherubs (a cherub system). In this case on the right and the left hand side the 65-65 figure depiction can be supplemented with the already closely analysed 6 persons from the centre's 4 mandorlas' (the four main figures and the two archangels [29, 30]). Here the seventh can mean Saint Emerich (Imre) himself as number 137th. In the previous case when we treated as a unit of the four mandorlas depicting the salvation history of Christ when adding it to the 68 figures each, the resulting number, the 137, was the number archetype of the crown of *Atara* of Solomon ***(עטרה)***. Since according to the Eleazar fragment, the wheel of Ezekiel's chariot or in Hebrew the *Ophen* ***(אופן),*** whose value is 137, the archetype of the crown itself or the numeric archetype of it (Dan 1968, 1982 [3, 4]). So the number 137 in the interpretation of the given representational system seems to be crowning, probably by Jophiel, Saint Emerich himself, who, in this type of interpretation, means the only "singular" entity besides the number 137.

If we take a look at the joint system of "personalities" of the Holy Crown and the Robe (the Casula) in way that we do not take into account the 24 saintly, pious birds, only the human figures, then we get to the 137+1 system as well. Since the 112 (136-24) people remaining in this case (when the 4+4=8 cherubs are taken into account as evangelists) are supplemented with the "persons" seen on the 19 pictures of the Holy Crown and the repeatedly mentioned 6 persons of the 4 central mandorlas'. Thus we get 112+19+6=137 number which again as a unity crowns Saint Emerich as seen on the Robe. Here another important possible interpretation emerges. Since the Crown's pictures and the figures of the 4 central mandorlas that are equivalent make up a separate 26-order with the prince. Apart from this there are 112 people in a 56+56-structured division on the robes left and right sides. But we have already seen the number 137's sum of 111+26

composition, which the letter *Aleph* that means the Hebrew number 1000 and the number of God's name as well (Bahir §.70). This can be interpreted here too, if we count the prophet Habakkuk, given as the only "angelic prophet" in a special position, probably in the role of Jophiel (see *Bahir* §'s 68-70), as a narrator, an interpreter, and we get the 111+26=137 structure. (see the picture below Fig. 9). We can see exactly this solution in the inverse Divine tree that describes the incarnation in *Hortus deliciarium* as can be seen in Fig. 1 (Fol.80v in [7]). Here the 16 descendants of Abraham, or saints, shown as stars by an angel next to Christ we can count 121 persons altogether (since the descendants of Abraham would be as many as the stars, so the stars signify people too) we can interpret 137 entities or people in addition to the angel. What is surprising is that this too is in the 111+26 composition, since exactly 26 crowned figure, or 26 kings, can be seen in the picture. The sum here, as we mentioned above, also shows the 25+1 structure, since the only beardless prince takes his place among the young, beardless, virgin martyrs holding a palm tree branch within the depiction of the saints of the church. So he can be identified precisely as Saint Emerich, just as he is a singular entity on the Robe, while the 19 pictures of the Crown with the 6 persons attached to it are a matched to the gathering of the crowned personages in the *Hortus* picture. So in the *Hortus* as well as on the Robe in this case an angelic figure points to the 137 incarnation and (following Habakkuk) the spacious-temporal personification of the angel Jophiel (**יופיאל**) the heavenly priest, where the **111+26** structure of the number 137 is found in the especially important **אלף יוי** or **אלף יהוה= 137** composition.

It is worth mentioning that on the uniformed "Cloak-picture" of the Holy Crown's two archangels and emperor, the white pearls form exactly the same 111+26 composition, in the way that there are 111 white pearls on the periphery, while 26 white pearls are in the centre. The right side/left side divisions of the 26 white pearls are exactly equal to the 14+6+5+1 division system of the crowned heads of Hortus deliciarum. The division of the other 14 ornamental items found here represent with their number the Hebrew name of David, and with their shape his star (his shield); and so symbolically it shows the Messiah-king successor (see Fig. 7 in the part one) the same in the picture of "On the lap of Abraham" in Hortus deliciarum (Fol.263v, see Fig. 9). All the above strongly establish through the central archetype role of the number 137 that the three grand works of art originate from the same royal court and workshop. The specific numbering order confirms the reconstruction of the vision of "Woman dressed in the Sun" (with the two archangels) on the central, most important mandorla up in the front. Jophiel, as we have already discussed, is the prince or the archangel of the Torah, and the interpretation of the Torah (in expanded meaning, the God's Book), even according to ancient traditions. The 111+26 composition together with the anagram–model (the temura) of the Hebrew-letter interpretation (**אלף יוי**) indicates the Angel Jophiel (**יופיאל אלף יוי**). Thus, next to the incarnation inverse tree of *Hortus deliciarum*, the angel from the composition of 26 number of the crowns could be identified as Archangel Jophiel.

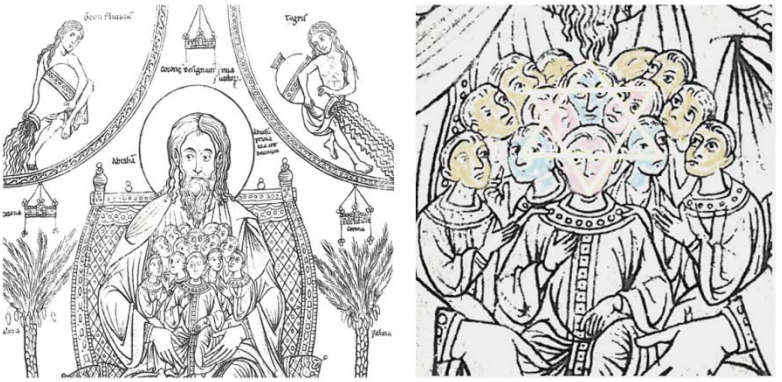


Figure 9

On the bossom of Abraham with the name (4-6-4=**דוד**) and star (6) of David (Fol. 263v)

This 111+26 composition is explained by Prophet Habakkuk in the 70th paragraph of Bahir's Book. The above is completed later, in paragraph 95 of *Bahir* with the inverse tree introduced in paragraph 21-23, and to which he assigned the number 137.[13] Because the name of the prophet Habakkuk, just like the name of the angel Jophiel, can be connected to the crown of *Atara* through the number of their names, thus ***Habakkuk*** in the angel-like depiction on Robe can be identified as the angel Jophiel, based on the number-composition of the *Bahir* and on the 137 composition of the "Casula". Since both of their names (**חבקוק הנביא** = **עטפה** = 284) refer to the word "Atara" crown – the Greek translation of which is "Stephanos" – so they are the symbolic prototypes and the

[13] It may be worth mentioning Jung's opinion of this question: "*I'm rather certain that the sefiroth(ic) tree contains the whole symbolism of Jewish development parallel to the Christian idea (concerning the incarnation of God). The characteristic difference is that God's incarnation is understood to be a historical fact in the Christian belief, while in the Jewish Gnosis it is an entirely pleromatic process symbolized by the concentration of the Supreme triad of Kether, Hokhmah and Binah in the Figure of Tifereth. Being the equivalent of the son and the Holy Ghost, he is the Sponsus bringing about the great solution through his union with the Malkuth (Atarah). This union is equivalent to the Assumptio Beatae Virginis, but definitely more comprehensive then the letter as it seems to include even the extraneous world of the Kelipoth. X (probably Scholem) is certainly all wet when he thinks that the Jewish Gnosis contains nothing of the Christian Mystery. It contains practically the whole of it but in its unrevealed pleromatic state." (Jung Letters, Vol. II. Letter to E. Neumann)*

helpers in the current, time- and age-related interpretation of St. Stephen, the linguist and hermeneut [27]. (The angel shape and the angel-face likeness is typical not only of Habakkuk and Jophiel, but obviously, through the angel-faced St. Stephen proto-martyr, it could symbolically be considered as the king).

In verse 3.1-3 of Isaiah the expression "Sar Homasim" (**שר חומשים**) appears, which is the prince or the archangel of the Five or the Fifty in the interpretation of Talmud. First, the Talmud interprets the Five as the Torah (as the Five Books of Moses); and thus it is obviously about the archangel or the prince of the Torah, who is archangel Jophiel in the mystic traditions [6, 9]. In the following Talmudic interpretation, the Fifty (50) instead of the Five (5) means the monarch or archangel interpreter. This part of the Talmud allows us to regard archangel Jophiel not just as the prince of the Torah, but as the archangel of the interpretation, in the most common meaning. Because the Holy King in the 8th Caput of the Admonitions defines himself as the prince (monarch) of translation and interpretation [27], the joint and (identifiable as each other) symbolic perception as the prince or the angel of interpretation in the broadest meaning of Jophiel archangel, Prophet Habakkuk, and his own name is understandable.

All of the above are confirmed by our hypothesis (see [25, 27]) that we put in writing years before the above thoughts arose. According to that hypothesis, the Old Testament bases of the Admonitions are the verses 3.1-3.3 of Isaiah (and its interpretation by Talmud) with the 18 threats addressed to the child kings. In the Admonitions the 8th Caput refers to the prince of the Five and the Fifty, which in this context means the interpretative prince of God's Book (the Bible), or – in the broadest meaning (but primarily between the Latin and Greek languages and traditions) – the ruler of interpretation. These are the traditions of the great king, without which his son would stay a child king and would not be able to reign successfully in his kingdom. The Hebrew word for obedience (**משמע**) means both understanding and interpretation, so according to the king, the disobedient boy disperses the flowers of the crown. *("Spiritus inobediantiae dispergit flores coronae").* It is also worth mentioning that in after the Talmudic interpretation of the admonitions addressed to the child kings by Isaiah in the Hagigah 12.a, we read about the angel who is weaving – as if from flowers - the crown of the Creator from the prayers of good (obedient) Israel, and that that crown rises on the head of the Creator by saying God's secret name. It seems that the 8th chapter of the Admonitions was edited by St. Stephen with an eye to the two concurrent parts of the Talmud, which is normal after all, since the great king called the whole of the Admonitions also as the royal crown ("*superius libata regalem componunt coronam").* So the authentic interpreter king can symbolically be identified with the archangel Jophiel.[14]

[14] In this article [25,27] at the centre of the hermeneutic circle, we place as an "interpreter" the scientific enthusiast of language (The legend of Hartwick, 6th caput of the Admonitions), Saint Stephen, who sees himself as such in the 8th caput. (See Gy. Kapitánffy: Hungaro-byzantina, Typotex, Budapest, 2002). At some time in the

Conclusions

In this part of our paper we discussed the primordial creation images together with the idea of the incarnation in the form of depictions in the structure of 137, presupposing the same authorial circle of those pictures and images found in the *Book of Bahir*, some important pictures of *Hortus deliciarium*, the *Pala d'Oro in Venice*, the *Coronation mantle (Casula)* and the *Holy Crown of St. Stephen*, all of which have similar meanings and are isomorphic with each other. In the above-mentioned pictures we can talk about 137-structure compositions, and the inverse world tree conception related to the creation myth; or rather, the creative and governing primordial models that in a depiction of isomorphic structures appear in very similar meaning patterns. In the pictures and the texts of these works, the idea of the inverse cosmic tree with the '137' composition with the 10 Sephiroth, as the primordial image of the fine structure in quantum theory, symbolically carries the arrangement of the spectral lines, in a tree-structure-type way, through the "inverse number" of 137, i.e. through the primordial concept of the fine structure constant. Finally, following Jung, we have intended to show in the discussed (proto-Kabbalistic) "primordial models" that the Sephiroth tree contains the entire symbolism of the Christian idea of the incarnation of God. In the space-temporal process, God's incarnation is understood to be a historical fact, while in the pleromatic process it is symbolized by the concentration of the Supreme triad of Kether (the supreme masculin Crown), Hokhmah (Father) and Binah (Mother) in the central Figure of Tifereth. Being the equivalent of the Son and the Holy Ghost, he is the Sponsus bringing about the great solution through his union with the sponsa who is Malkuth (Atarah, the feminin Crown) the last (10^{th}) Sephira. According to Jung, this union is equivalent to the *"Assumptio Beatae Virginis*". The "archetypal (eternal) approach" contains practically the whole of the structure and meaning of the Christian Mystery, but in its unrevealed pleromatic state (see [14] in the Part three of the paper). As we have tried to prove, this double (temporal and archetypal) incarnation and creation "process" is carried out by the ordering principle and structure of the number-archetype 137 which is here the *sine qua non* of the realisation of "*creatio et incarnatio continua*" similarly to its role in the modern physics.

References

[1] Ameisenowa, Z.: Animal-headed Gods, Evangelists, Saints and Righteous Men. Warburg Institute, London, 1948, pp. 21-44

past it became evident that the task of the hermeneutics was to address the particular conditions under which a text was interpreted. The original model for this was the interpreter of God's will, who was able to interpret the language of the oracle. But to this day, not a single of the interpreter's tasks has been achieved by merely giving back that which the speaker – whom the interpreter translates– in reality said, but rather, the interpreter must alone validate the speaker's thoughts according to the real situation of the conversation, because only the interpreter knows *both languages* of the discussion. (Gadamer, H. G.: The truth and the method, 1982)

[2] The Book Bahir, (ed. A. Kaplan) Samuel Weiser, INC., York Beach, Maine, 1989

[3] Dan, J.: 'The Emergence of Mystical Prayer', Studies in Jewish Mysticism, Proceedings of Regional Conferences held at the University of California, Los Angeles, and McGill University, eds. Joseph Dan and Frank Talmage, Cambridge Mass: Association for Jewish Studies 1982, pp. 85-120

[4] Dan, J.: **תורת הסוד של חסידי אשכנז ירושלים**, pp. 119-122, 1968

[5] Ferencz, Cs.: The crown of St Stephen of Hungary, Budapest, Heraldika, 2002 (In Hungarian)

[6] Gaster, M.: „Hebrew Visions of Hell and Paradise," in the Journal of The Royal Asiatic Society, London, 1893

[7] Green, R. (ed.): Herrad of Hohenbourg: Hortus Deliciarum, (Commentary and Reconstruction) I-II. London, 1979

[8] Hahnloser, R - Polacco, R.: Una nuova lettura della Pala d'oro.Canal & Stamperia. Venezia, 1994

[9] Idel, M.: Kabbalah New Perspectives, Yale Univ. Press, 1988

[10] Jung, C. G.: Psychologie und Alchemy, Walter Verlag, Olten (1972)

[11] Jung, C. G: Mysterium Conjunctionis, Princeton Univ. Press., 1977

[12] Lawrence, R. M. The Magic of the Horse-Shoe, With Other Folk-Lore Notes (1898)

[13] Mac Gregor, Malcolm H., The Power of Alpha, World Scientific, Singapore, 2007

[14] Margalioth.: (The Book Bahir), **קוק הרב מוסד מרגליות הבהיר ספר ,ירושלים**

[15] Pauli, W. (eds. Enz, C., Meyenn, K.V.): Writings on Physics and Philosophy, Springer (1994)

[16] Ricoeur, P.: Structure et herméneutique, In: Le Conflit des Interprétations, pp. 31-63, Seuil (1969)

[17] Scholem, G.: Origins of the Kabbalah, Princeton Univ. Press, 1990

[18] Scholem, G.: On Mystical Shape of Godhead, Schocken, New York, 1996

[19] Scholem, G.: Major Trends in Jewish Mysticism, Schoken, New York, 1983

[20] Straub, A., Keller, G.: Hortus Deliciarum, Strassbourg, 1879-1900

[21] Thorndike, L.: A History of Magic and Experimental Sciences, Columbia Univ. Press, 1952

[22] Tóth E., Szelényi K.: The Holy Crown of Hungary, Budapest, 2002

[23] Várlaki, P., Kóczy, L.: A Comparative Study of Pictures from Pala d'oro in St. Mark Catherdral of Venice and from the Holy Crown of Hungary, In: Proc. of Int. Conference on Genealogy and Heraldry, pp. 131-171, 2006 (in Hungarian)

[24] Várlaki, P., Nádai, L., Bokor, J.: Number Archetypes and "Background" Control Theory Concerning the Fine Structure Constant. Acta Polytechnica 5 pp. 71-104 (2008)

[25] Várlaki P., Kóczy L. T. Genealogical Myth and Allegory in the Pala d'Oro of Venice and Saint Stephen's Royal Mirror, Int. conf. of the Hungarian Heraldic and Genealogical Society, pp. 89-124 (in Hungarian)

[26] Varlaki, P., Kóczy, L., Kiss G.: Analysis of the Enamel Pictures on the Holy Crown of Hungary, Int. Conf. of the Hungarian Heraldic and Genealogical Society, pp. 101-130 (in Hungarian)

[27] Várlaki P, Bokor J Number Archetypes, Symbolic Coding Letters and "Background Communication Theory" in Saint Stephen's Royal Mirror.In: IEEE 7th International Conference on Computational Cybernetics, 2009, pp. 201-211

[28] Várlaki, P., Rudas, I., Kóczy, L.: Historical Origin of the Fine Structure Constant. Part I. St Stephen's Crowning Achievement, Acta Politechnica Hungarica, Vol. 7. No. 1, pp. 119-157

[29] Várlaki, P., Rudas, I.: Comparative Structural and Hermeneutical Analysis of the Hungarian Coronation Robe (Casula), Symp. on the1010 anniversary of St Stephen's Coronation and 600 anniversary of the Foundation of the Óbuda University", p. 45 (in Hungarian)

[30] Várlaki, P., Kéri, Á.: A Reconstruction and Hermeneutical Representation of Coronation Robe of St Stephen of Hungary, "Országépítő", 2005, No. 1, pp. 13-18 (in Hungarian)