

# Contents

<i>WE ARE TWO YEARS OLD...</i>	1
<b>Zoltán Csernátony</b> On the efficiency of multi-domain routing	2
<b>Albert Mráz, Tamás Zámbó, Sándor Imre</b> Efficient dynamic resource management for OFDMA-MIMO wireless transmission	8
<b>Gergely Öllös, Rolland Vida</b> Adaptive sleep scheduling protocol in wireless sensor networks	20
<b>Miklós Kuczmann</b> Modeling feeds of antennas by Finite Element Method	31
<b>Tibor Takács, László Vajta</b> Novel image similarity measurement in Automated Optical Inspection	36
<b>Pál Varga, László Gulyás</b> Traffic analysis methods to support decisions at the knowledge plane	50

---

## Editorial Board

### Editor-in-Chief:

CSABA A. SZABÓ,  
Dept. Telecomm., Budapest Univ. Technology and Economics (BME)

### Chair of the Editorial Board:

LÁSZLÓ ZOMBORY,  
Dept. Broadband Communications and Electromagnetic Theory, BME

ISTVÁN BARTOLITS,  
National Communications Authority

ISTVÁN BÁRSONY,  
Institute of Technical Physics and Material Science,  
Hungarian Academy of Sciences (MTA)

LEVENTE BUTTYÁN,  
Dept. Telecommunications, BME

ERZSÉBET GYŐRI,  
Dept. Telecommunications and Media Informatics, BME

SÁNDOR IMRE,  
Dept. Telecommunications, BME

CSABA KÁNTOR,  
Scientific Association for Infocommunications

LÁSZLÓ LOIS,  
Dept. Telecommunications, BME

GÉZA NÉMETH,  
Dept. Telecommunications and Media Informatics, BME

GÉZA PAKSY,  
Dept. Telecommunications and Media Informatics, BME

GERGŐ PRAZSÁK,  
National Council for Communications and Information Technology

ISTVÁN TÉTÉNYI,  
Computer and Automation Research Institute, MTA

GYULA VESZELY,  
Dept. Broadband Communications and Electromagnetic Theory, BME

LAJOS VONDERVISZT,  
National Communications Authority

## International Advisory Committee

VOLKMAR BRÜCKNER,  
Hochschule für Telekommunikation Leipzig, Germany

MILAN DADO,  
University of Zilina, Slovakia

VIRGIL DOBROTA,  
Technical University Cluj, Romania

AURA GANZ,  
University Massachusetts at Amherst, USA

EROL GELENBE,  
Imperial College, London, UK

BEZALEL GAVISH,  
Southern Methodist University, Dallas, USA

ENRICO GREGORI,  
CNR IIT Pisa, Italy

ASHWIN GUMASTE,  
IIT Bombay, India

LAJOS HANZO,  
University of Southampton, UK

ANDRZEJ JAJSZCZYK,  
AGH University of Science and Technology, Krakow, Poland

MAJA MATIJASEVIC,  
University of Zagreb, Croatia

VACLAV MATYAS,  
Masaryk University, Brno, Czech Republic

OSCAR MAYORA,  
CREATE-NET, Italy

YORAM OFEK,  
University of Trento, Italy

ALGIRDAS PAKSTAS,  
London Metropolitan University, UK

JAN TURAN,  
Technical University Kosice, Slovakia

GERGELY ZARUBA,  
University of Texas at Arlington, USA

HONGGANG ZHANG,  
Zhejiang University, Hangzhou, China

---

## Protectors

GYULA SALLAI – president, Scientific Association for Infocommunications

ÁKOS DETREKÓI – president, National Council of Hungary for Information and Communications Technology

---

## We are two years old...

*szabo@hit.bme.hu*

**2010** has been our second year since the launch of Infocommunications Journal, an international companion publication of a 65 years old Hungarian language journal “Communications Technology”, published by the Hungarian Scientific Association for Infocommunications, a Sister Society of IEEE. I would like to express my sincere gratitude to our Hungarian Editorial Board and in particular to the International Advisory Committee for their continuing support throughout these two years. May I also commemorate one of the Committee members, Professor Yoram Ofek, a recognized expert in broadband communications and networking and a dear colleague, who passed away in 2010.

In 2010 we published three outstanding special issues and it is my pleasure to thank the Guest Editors for their dedicated work. In January, we had a Special Issue compiled from selected and peer-reviewed papers of the “First Hungarian–Japanese Joint Conference on Future Information and Communication Technologies”, our Guest Editors were Werner Klaus (NICT, Japan), Gábor Magyar and Robert Szabó (Budapest Univ. of Tech. Econ., Hungary). In April a “Special Issue on Telecommunications Network Strategy and Planning” was put together by Tibor Cinkler (Budapest Univ. of Tech. Econ., Hungary), Oscar Gonzalez-Soto, (ITU, Spain), Gyula Sallai (Budapest Univ. of Tech. Econ., Hungary), Rati C. Thanawala (Alcatel-Lucent, USA) and Andy Valdar (University College London, UK). Our Special Issue in July was dedicated to “Novel Solutions for the Next Generation Services” edited by Maja Matijasevic (Univ. of Zagreb, Croatia) and Sandor Imre (Budapest Univ. of Tech. Econ., Hungary)

The quality of our journal is determined to a great extent by the diligent and rigorous work of the reviewers, let me thank them too by including the list of our reviewers in 2010 on the inner back cover.

It is not easy to get a new journal accepted and recognized by the international research community. One of the almost necessary conditions to achieve that is to get listed and indexed by one of the international indexing services. We submitted applications to Scopus, EI/Compendex and Inspec which are currently under review, let me hope that I can report on a positive outcome soon.

So much about the past, and after two really interesting years we will be facing new challenges in 2011. We had very useful discussions with IEEE COMSOC leaders earlier this year and obtained their advice on how to further improve the quality of the journal, one of their Sister Society publications, which is also linked from COMSOC portal. We are starting to move along a roadmap that would eventually lead to a technical sponsorship by IEEE and inclusion of our content in XPloré. As a first step of this roadmap, we are going to have a new international editorial board from 2011. We will also publish the papers in IEEE format, so I kindly ask our prospective authors to submit their manuscripts and final papers according to the formatting instructions available on [http://www.ieee.org/publications\\_standards/publications/authors/authors\\_journals.html#sect2](http://www.ieee.org/publications_standards/publications/authors/authors_journals.html#sect2), where you can find a document “Template and Instructions on How to Create Your Paper”.

The technical quality of our journal is primarily depends on the quality submissions. I would like to invite researchers from our Central and Eastern European region, but of course also from all over the world to submit their results to our journals. Topics of interest include:

- Data and network security
- Digital broadcasting
- Infocommunication services
- Internet technologies and applications
- Media informatics
- Multimedia systems
- Optical communications
- Society-related issues
- Space communications
- Telecommunication software
- Telecom. economy and regulation
- Testbeds and research infrastructures
- Wireless and mobile communications

Theoretical and experimentation research results achieved within the framework of European ICT projects are particularly welcome. From time to time we publish special issues and feature topics so please follow the announcements. Proposals for new special issues and feature topics are also welcome.

Finally, let me wish to our readers, authors, reviewers and committee members a happy and prosperous New Year!

**Csaba A. Szabó, Editor-in-Chief**

*Budapest University of Technology and Economics, Department of Telecommunications,  
Hungary*

# On the efficiency of multi-domain routing

ZOLTÁN CSERNÁTONY

*Budapest University of Technology and Economics,  
Department of Telecommunications and Media Informatics  
csernatony@tmit.bme.hu*

*Keywords: multi-domain, aggregation, routing, scalability, efficiency*

**In order to cope with scalability issues, large networks are often divided into several domains where routing is performed only on the basis of their aggregating structures. Although scalability can be assured this way, further inaccuracies are introduced in the routing information affecting eventually the efficiency of routing. In our paper, a novel statistical threshold-based aggregation model is suggested with which we compare existing routing methods on the basis of the efficiency they provide. The method itself is a statistical extension of the standard poly-line segment method where re-aggregation is triggered on the basis of measurement data. According to the simulation results, our statistical method outperforms existing methods.**

## 1. Introduction

In large communication networks, scalability issues occur as (1) traditional routing algorithms perform inefficiently in the presence of large numbers of nodes, and (2) distributing a large amount of routing advertisement might be infeasible due to the concerning instabilities [1] and convergence criteria of certain routing protocols [2].

In order to assure a scalable networking operation, nodes are grouped into smaller *domains* (sub-networks), and the topologies of these domains are *aggregated* to simpler structures with domains only distributing (routing) advertisements concerning only their aggregating representation. By applying such a technique, a scalable *multi-domain* network can be obtained which is actually a hierarchical structure consisting of a lower layer meant by the original network with an upper layer of the aggregating topologies<sup>1</sup>.

Although topology aggregation assures scalability, it also has an unpleasant side-effect as it introduces certain inaccuracies in the routing advertisements<sup>2</sup>. Eventually, these inaccuracies might result in inefficient path selections due to the loss of information which the routing itself is executed on. As the premises suggest, a general property of applying topology aggregation is that the amount of advertisement information gets traded with *routing efficiency*.

Although there have been important researches done on topology aggregation and its effects on multi-domain routing, the direct relationship of scalability and routing efficiency is yet to be examined. As routing efficiency is related to the accuracy of aggregating information, contributions relevant to our concentration have been already made. An earlier study shows that, given the inaccuracy of aggregating representations, finding paths

that are likely to satisfy requirements for parameters of additive metrics is NP-hard [3]. For a directed graph, [4] presents a compact aggregating representation that bounds the aggregation-induced inaccuracies by a worst-case distortion factor in case of a single additive parameter. Either for a single additive or for a restrictive parameter, [12] presents a distortion-free representation applying full mesh and spanning tree, respectively.

Although [5] demonstrates that even though topology aggregation reduces the routing information to a large extent, it does not necessarily diminish routing efficiency as considerably. In [6] and [15], a line-segment approach is introduced for topology representation by which the possible bandwidth and delay values are represented by a single linear approximation. Further improvements to this method apply estimations by poly-lines and (higher-degree) curves instead of single lines [7,8], among which poly-lines proved to be the most efficient. Another approach of aggregation captures the statistical properties of the original network by histograms with probabilities in order to handle uncertain parameters [9]. Study [10] employs a spanning tree method by which the topology is aggregated incrementally according to the histogram consisting of measurement data, while [11] applies partial link-based advertisement which is controlled by monitoring only the portion of the physical with the largest contribution to link-state variation.

In this paper, we suggest a novel threshold-based statistical aggregation model which is actually a measurement-driven extension of the poly-line segment method which has proven to be a highly efficient method [8]. For the purpose of evaluation, scalability and routing efficiency are evaluated in the case of four different aggregation methods. On the basis of the efficiencies achievable by them, we make comparisons by means of simulation.

<sup>1</sup> Due to the simpler topology (i.e. sharing less sensitive information), an increased level of confidentiality is an additional benefit of aggregation.  
<sup>2</sup> Although the stochastic nature of traffic introduces further inaccuracies, that particular aspect is out of our study's scope.

The rest of the paper is organized as follows. In Section 2, we present the multi-domain network model, the aggregation method and the routing model, while Section 3 introduces functions for evaluating the scalability and routing efficiency. In Section 4, we evaluate these efficiency quantities by means of simulation, and finally, Section 5 summarizes the paper.

## 2. Topology aggregation and routing model

Aggregation models provide more accurate aggregating representation for a larger amount of information. Consequently, scalability is expected to be traded with routing efficiency. In this section, we give a formal foundation of the applied aggregation and routing models.

### 2.1 Modeling multi-domain networks

In this paper, we use similar basic assumptions in the network modeling as there have been in [6]. A multi-domain network is regarded as several interconnected domains where the terminating nodes of an interconnecting edge are called *border nodes*. We represent a domain's graph by a tuple  $(V, B, E)$  where  $V$  denotes the set of nodes,  $B$  the set of border nodes ( $B \subseteq V$ ) while  $E$  denotes the set of edges.

Two traffic attributes are considered in this paper: *bandwidth* and *delay*, which are of *restrictive* and *additive* metrics, respectively. The operation by which the resultant value of a path  $p$  is to be computed determines the metrics of attributes, which are in our case

$$b[p] = \min_{e \in p} b[e]$$

$$d[p] = \sum_{e \in p} d[e]$$

where  $b[e]$  and  $b[p]$  denote the bandwidth of edge  $e$  and path  $p$ , respectively, and we use  $d$  for denoting delay. Fig. 1 illustrates the operations according to *restrictive* and *additive* metrics.

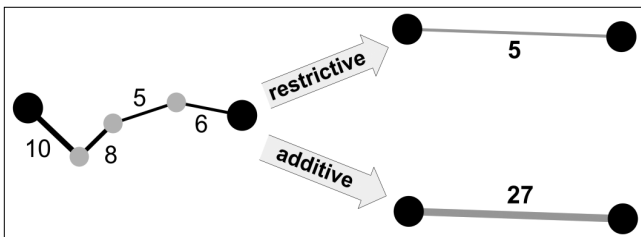


Figure 1. Resultant path values for restrictive (bandwidth) and additive (delay) attributes

### 2.2 Aggregation models

In this section, on the basis of their resulted efficiencies, we compare our *Statistical Threshold-based Model* (STM) with the classical *Poly-line Segment Method* (PLSM) for topology aggregation and the rather extreme *Best point* (BP) and *Worst point* (WP) approaches. These models provide aggregating topologies which are simplified abstractions of the physical topologies as they contain only the border nodes with aggregating edges (“lo-

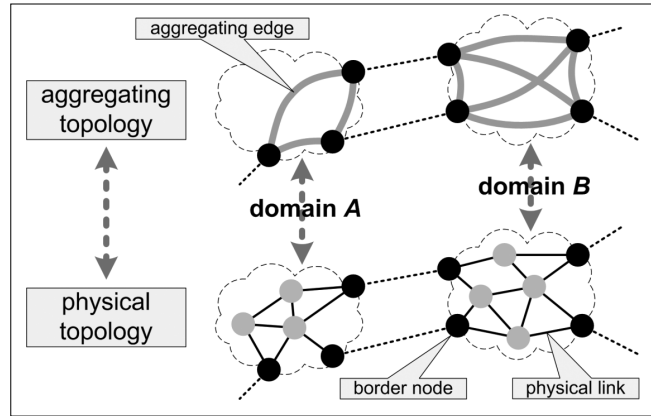


Fig. 2. Topology aggregation of domains: an illustration (two-layer model)

gical links”) running between them  $(V, B, E) \Rightarrow (B, E_{agg})$ . This way, a two-layer network representation can be obtained as being illustrated in Fig. 2.

**Poly-Line Segment Method (PLSM):** This method [7,8] estimates the delay-bandwidth function by fitting (by least squares method) line segments onto such points on the bandwidth-delay plane that consist of the minimum delays for all possible bandwidth values. These points are called *representatives* and they form a staircase on the bandwidth-delay plane [8]. The bandwidth axis is then divided into  $L$  disjoint intervals and a single line is fit to the representatives within each of these intervals.

#### Best Point (BP) and Worst Point (WP) approaches:

These are used for benchmarking purposes as they mean the two extremities of aggregation policies: BP uses the maximum bandwidth and the minimum delay values, while WP uses the maximum bandwidth and the maximum delay of all representatives [6,8].

**Statistical Threshold-based Model (STM):** Contrarily to PLSM, it follows a different concept as – instead of the representatives – the line segments are fit onto the *average* (measured or estimated) *delay values*, triggered by predefined (*aggregation*) *thresholds*. The model addresses bandwidth intervals

$$[0, l_L], [l_1, l_L], [l_2, l_L], \dots, [l_{L-1}, l_L]$$

by (*bandwidth*) *levels*

$$0 < l_1 < l_2 < \dots < l_L$$

The resulting aggregating topologies are denoted by

$$(B, E_1), (B, E_2), \dots, (B, E_L)$$

where in order to support connection requests with the maximum possible bandwidth requirements (i.e. the maximum bandwidth link capacity),  $l_L = \max\{b(e) : \forall e \in E\}$  should hold.

More detailed (thus more accurate) representation can be made by involving more levels in the aggregation. In case of a certain pair of border nodes, for interval  $[l_j, l_{j+1}]$ , the aggregation method consists of two steps:

#### 1) Determining topology $(B, E_j)$ :

All the edges with insufficient bandwidth values get filtered out according to criterion  $b(e) < l_j (\forall e \in E)$ . There is an aggregating edge between two border nodes only if there exists a path between them after the edge-filtering.

2) *Determining the linear delay-bandwidth function:*

A linear function is fit (by least squares method) on the actual delay values of all the connection requests with bandwidth demand  $l_j \leq b \leq l_{j+1}$  that have traversed them<sup>3</sup>. Each domain is assumed to store the actual (outcome) bandwidth request-delay data points of previous connection requests for each pair of border nodes in a simple First-In First-Out (FIFO) storage of length  $K_{FIFO}$ . Aggregation (line fitting) is performed once the number of successful connection requests exceeds predefined *threshold*  $K_{aggr}$  so that  $k \bmod K_{aggr} = 0$

is satisfied. The initial data points are determined by PLSM (i.e. based on static values).

In Fig. 3, STM is illustrated. Note that bandwidth is represented through the intervals, implicitly.

2.3 Routing model

Bandwidth and delay are taken into consideration separately by the following steps:

- 1) *Composing the aggregating view of the network:* Selecting the aggregating topology for each domain that corresponds to the minimum bandwidth level in accordance with the current bandwidth request:  $\min\{j: l_j \geq r, j=1, \dots, L\} \Rightarrow (B, E_j)$ . In case of BP and WP,  $L=1$  holds formally.
- 2) *Logical routing:* Routing on the basis of the network consisting of the selected aggregating topologies. Dijkstra's algorithm is performed on the graph according to the selected level with its corresponding aggregating (advertised) edge delay values.
- 3) *Actual (physical) routing:* Routing by Dijkstra's algorithm by each selected domain on the basis of their physical topologies and link delay values between its selected pair of border nodes.

As a feedback, each domain is assumed to get notified of the outcome values of all the transfers that have been initiated by it.

3. The efficiency of multi-domain routing

In order to compare different aggregation schemes, there should be functions defined for "measuring" the *efficiency* they provide. This section applies two concepts for efficiency evaluation: the first one aims to measure the information saving due to aggregation, while the second one is to express the optimality of path selections. Section 3.1 and 3.2 discuss the former and latter ones, respectively.

3.1 Aggregation efficiency

The amount of information to be advertised is closely related to scalability. Upon this fact, it is important to represent the (relative) amount of advertisement information that can be saved by applying topology aggregation. We use the following definition for expressing *aggregation efficiency*:

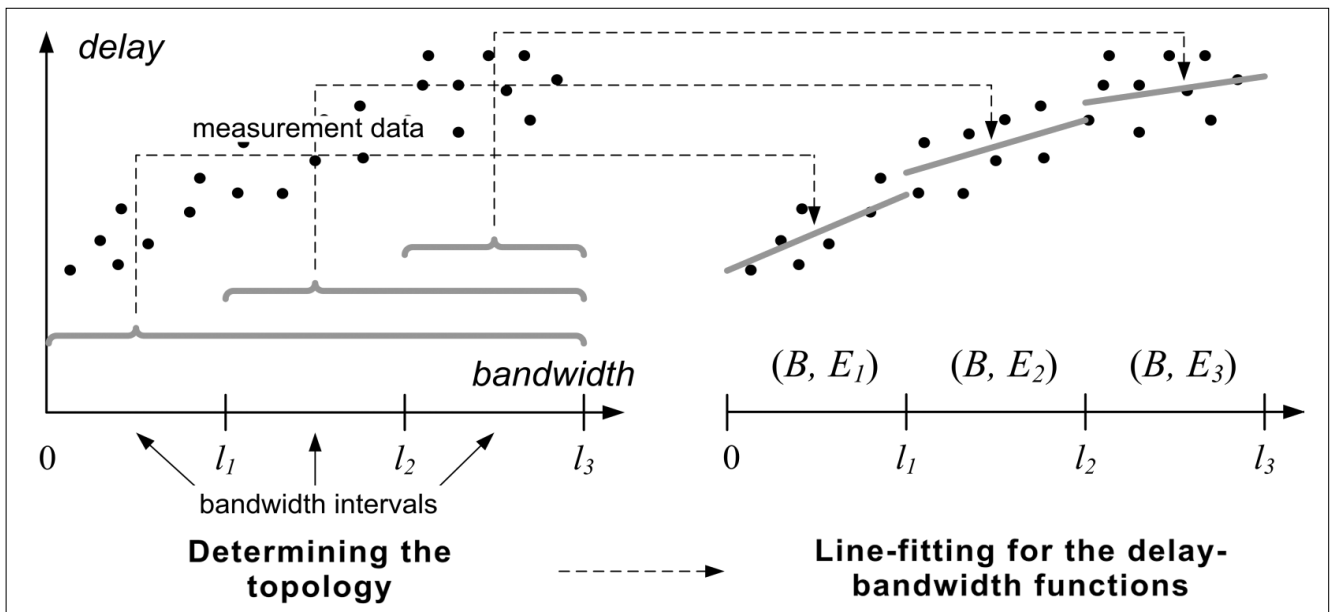
$$\eta_{AGGR} = \frac{I_{orig} - I_{aggr}}{I_{orig}} \tag{1}$$

where  $I_{orig}$  and  $I_{aggr}$  denote the amount of information meant by the original and aggregating topologies, respectively. Since a graph can be described by its edge set, assuming that each of the two attributes means a unit of information per edge, the following expressions hold:

$I_{orig} = 2 \cdot |E|$  and  $I_{aggr} = \sum_{j=1}^L |E_j|$ . By evaluating them into (1), we obtain

$$\eta_{AGGR} = 1 - \frac{I_{aggr}}{I_{orig}} = 1 - \frac{1}{2} \frac{\sum_{j=1}^L |E_j|}{|E|} \tag{2}$$

Fig. 3. Statistical approximation via the threshold-based model by fitting line segments on measurement data for different bandwidth intervals (the case of three bandwidth intervals)



<sup>3</sup> Consequently, this model requires the deployment of such a monitoring method along with a database which keeps track of the observed traffic data. In this paper, however, we do not consider the realization of such requirements.

Note that, on the one hand, due to inequality  $l_{aggr} \geq 0$ , upper bound  $\eta_{AGGR} \leq 1$  holds. Also note that by applying refined aggregation models (e.g. star topology instead of meshes), the aggregation gain can be further increased, however this issue is out of our study's scope. Note that this result applies to both PLSM and STM since they only differ in the input data the aggregations are executed on.

### 3.2 Routing efficiency

Besides scalability, another important aspect is to evaluate the efficiency of routing. For that purpose, we introduce a *routing efficiency* function for both bandwidth and delay. We define these functions to be of normalized values for the range  $[0,1]$  and to be measurable by a domain on the basis of advertised and outcome values.

In the case of bandwidth (i.e. bandwidth sufficiency), we define *success ratio* as:

$$\eta_{succ} = \frac{\text{total number of accepted requests}}{\text{total number of requests}} \quad (3)$$

In the case of delay – as an inverse quantity of delay deviation [9] – we consider *delay efficiency* as the distance of the outcome (i.e. actual) values from the advertised values for the selected paths:

$$\eta_d = 1 - \frac{1}{N} \sum_{n=1}^N |a_{out}(n) - a_{adv}(n)| \quad (4)$$

where  $a_{adv}(n)$  and  $a_{out}(n)$  denote the advertised and outcome summarized values along the path selected at the  $n$ -th routing occasion, respectively. Since both  $a_{adv}(n)$  and  $a_{out}(n)$  are normalized for interval  $[0,1]$ , bounds  $0 \leq \eta_d \leq 1$  hold as well.

## 4. Simulation studies: Evaluating aggregation and routing efficiencies

In this section, we examine scalability and routing efficiency in case of PLSM and STM by evaluating the efficiency functions presented in Section 3.1 and 3.2. For the simulation testbed, we apply a combination of real-life approximating generated networks.

### 4.1 Simulation testbed

The simulation *testbed* is a two-layer multi-domain network whose inter-domain topology is generated by the *power-law model* [13] and the intra-domain topologies are generated based on the *Waxman model* [14]. The power model determines the number of degrees of a domain (i.e. the number of connected inter-domain links) according to formula

$$P(k) = c \cdot k^{-\gamma} \quad (5)$$

where  $P(k)$  denotes the probability of a node having  $k$  degrees and  $c$  is a normalizing constant that ensures  $\sum_{k=1}^{D-1} P(k) = 1$ , where  $D$  is the number of domains (there is no domains with zero degree (i.e. isolated domains)). On

the other hand, Waxman model determines the probability that nodes  $u$  and  $v$  are connected by a physical link:

$$P(u, v) = b \cdot e^{-\frac{d(u, v)}{\beta L}} \quad (6)$$

$$L = \min_{\forall u, v, u \neq v} d(u, v)$$

where the (physical) distance of nodes  $u$  and  $v$  is denoted by  $d(u, v)$ , while  $L$  denotes the minimum distance of any two nodes and  $b$  is a constant for normalization. The network consists of 20 domains with 30 nodes per domain. The parameters are set as been in [6], where parameters  $\beta = 0.6$  and  $\gamma = 2.2$  are applied. Values from the normalized interval  $[0,1]$  are randomly assigned to each edge as bandwidth capacity (i.e. max. bandwidth) and delay:  $b(e), d(e) \in [0,1], \forall e \in E$ .

In this paper, we apply a uniform distribution for the bandwidth demands of connection requests on the interval  $[0,1]$ . A dynamic traffic shape model is employed in which arriving times and holding times are distributed

Fig. 4. Bandwidth intervals vs. success ratio and aggregation efficiency

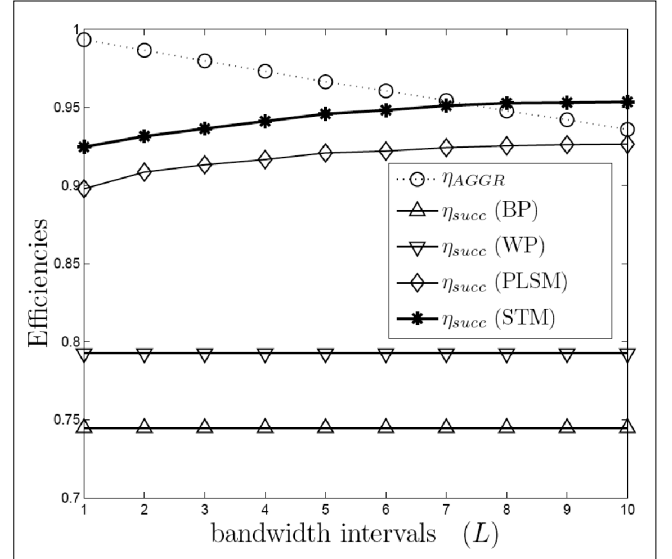
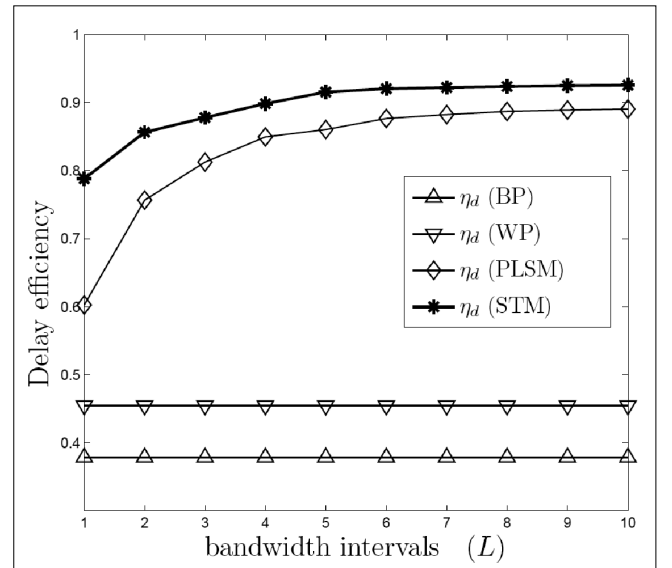


Fig. 5. Bandwidth intervals vs. delay efficiency



uniformly. For each number of bandwidth intervals and aggregation schemes, 1000 connection requests have been generated and averaged. In case of the two aggregation models PLSM and STM, for the sake of simplicity, the bandwidth levels (thus the intervals themselves) are distributed uniformly on the bandwidth axis. We use one up to ten different bandwidth levels  $L = 0, 1, \dots, 10$ . For STM, the value of the aggregation threshold is set to  $L \cdot 20$  and FIFO storage sizes (see Section 2.2) are set to  $L \cdot 10$ .

#### 4.2 Simulation results

Our simulation results for BP, WP, PLSM and STM are presented in Fig. 4 and 5. Fig. 4 shows the evaluation of aggregation efficiency and success ratio for the four different aggregation schemes, while Fig. 5 shows the corresponding delay efficiencies depending on the number of bandwidth intervals. All efficiency values are computed as the average of efficiencies achieved by all domains. Note that in case of BP and WP, all these efficiencies are independent from its value (see Section 2.2) as their concerning aggregation models do not consider multiple bandwidth intervals.

Our results reveal the tradeoff that is between the efficiencies of aggregation and routing either when PLSM or STM is applied. Given the uniform distribution of edge and bandwidth level values, the aggregation efficiency shows a linear decrease in all cases. On the other hand, success ratio and delay efficiency show a considerable improvement as the number of bandwidth intervals increases especially for lower numbers, as the length of the bandwidth intervals decreases more significantly at smaller numbers of bandwidth levels.

Our results also show that STM outperforms the results that can be achievable by BP, WP and PLSM. It is also important to recognize that STM has a load balancing feature which helps keeping the blocking ratio low (thus the success ratio high). The reason for this property is the fact that the more heavily an aggregating edge is utilized, the higher average delay values its corresponding actual (physical) paths have. STM deals with this problem as it re-aggregates paths on the basis of measurement data, once the given aggregation threshold is exceeded. The extensive study of this property is a matter of further investigations. As a consequence, between scalability and routing efficiency, better compromises can be made by applying STM comparing to the cases when other methods are applied.

### 5. Conclusions

In our study, multi-domain routing has been discussed with focusing on the effect that topology aggregation has on different efficiency-related quantities. The qualities of scalability and routing efficiency have been considered through the evaluation of functions characterizing them. A Statistical Threshold-based Model (STM) has been introduced for topology aggregation which has out-

performed the – otherwise highly efficient – Poly-line Segment aggregation Method (PLSM). Our simulation results show that there is indeed a tradeoff between scalability and routing efficiency, which can be improved by applying our aggregation model.

There is a number of important advantages of STM:

- more accurate representation thus higher routing efficiency can be achieved,
- the computational time of the re-aggregation can be reduced significantly as the representatives do not need to be computed, and
- the method is flexibly tunable by setting the values of threshold parameters and the length of the FIFOs.

However, it also has a few drawbacks:

- additional storage requirements (FIFOs), and
- novel management tasks are necessary to implement.

Future directions of research should consider the optimal threshold values and size of data storage for such measurement-based statistical aggregation models. Examining the possible adaptation of this model within the framework of real networks and protocols is also of fundamental interest.

#### Author



**ZOLTÁN CSERNÁTÓNY** received his MSc degree in Electrical Engineering (with a final grade “excellent”) in 2010 from the Budapest University of Technology and Economics (BME). His research interests include the scalability, efficiency and economical aspects of routing in multi-domain networks. Since his graduation, he has been admitted for PhD studies at the same institute to start working on his degree from September 2010 onwards.

#### References

- [1] Mishra, “Scalability in communication networks”, *IEEE Network*, 16(4), August 2002.
- [2] T. Korkmaz and M. Krunz, “Source-oriented topology aggregation with multiple QoS parameters in hierarchical networks,” *ACM Trans. Modeling Comp. Simulation*, Vol. 10, No. 4, pp.295–325, October 2000.
- [3] R. Guérin and A. Orda, “QoS-based routing in networks with inaccurate information: theory and algorithms” *IEEE/ACM Trans. on Networking*, Vol. 7, pp.350–364, June 1999.
- [4] B. Awerbuch and Y. Shavitt, “Topology aggregation for directed graphs,” *IEEE/ACM Trans. on Networking*, Vol. 9, pp.82–90, February 2001.

- [5] F. Hao and E. W. Zegura,  
“On scalable QoS routing: performance evaluation of topology aggregation,”  
In Proc. IEEE INFOCOM 2000, pp.147–156.
- [6] Lui, K., Nahrstedt, K. and Chen, S.,  
“Routing with topology aggregation in delay-bandwidth sensitive networks”,  
IEEE/ACM Trans. on Networking,  
Vol. 12, Issue 1, pp.17–29.  
February 2004.
- [7] R. Hou et al.,  
“Aggregation-Based QoS Routing in the Internet,”  
Journal of Communications,  
Vol. 5, No. 3, pp.239–246,  
March 2010.
- [8] Tang Y., Chen S., Ling Y.,  
“State aggregation of large network domains”,  
Computer Communications,  
Vol. 30, Issue 4, pp.873–885,  
February 2007.
- [9] L. Xiao, J. Wang, K.-S. Lui, K. Nahrstedt,  
“Advertising Inter-domain QoS Routing Information,”  
IEEE Journal on Selected Areas in Comm.,  
22(10): 1949–1964,  
December 2004.
- [10] D. Jurca and R. Stadler,  
“H-GAP: Estimating Histograms of Local Variables with Accuracy Objectives for Distributed Real-Time Monitoring”,  
IEEE Trans. on Network and Service Management,  
Vol. 7, Issue 2, pp.83–95,  
June 2010.
- [11] Y. Yu, et al.,  
“On the efficiency of inter-domain state advertising in multi-domain networks”,  
In Proc. of the 28th IEEE Conf. on Global Telecomm.,  
2009.
- [12] W. Lee,  
“Topology aggregation for hierarchical routing in ATM networks,”  
In ACM SIGCOMM Comp. Comm. Rev.,  
Vol. 25, pp.82–92,  
April 1995.
- [13] M. Faloutsos, P. Faloutsos and C. Faloutsos,  
“On power-law relationships of the internet topology,”  
In Proc. ACM SIGCOMM 1999, pp.251–262.
- [14] B. M. Waxman:  
“Routing of multipoint connections,”  
IEEE Journal on Selected Areas in Comm.,  
Vol. 6, pp.1617–1622,  
December 1988.
- [15] J. Zhang, Y. Han, L. Wang,  
“A New Topology Aggregation Algorithm in Hierarchical Networks”,  
ISECS Intl. Coll. on Comp., Comm., Ctrl. and Mgmt.,  
2008, pp.179–183.



# Efficient dynamic resource management for OFDMA-MIMO wireless transmission

ALBERT MRÁZ, TAMÁS ZÁMBÓ, SÁNDOR IMRE

*Budapest University of Technology and Economics, Department of Telecommunications  
{mraz, imre}@hit.bme.hu*

*Keywords: OFDMA, MIMO, carrier allocation, scheduling, antenna selection*

**Efficient radio resource allocation algorithm is proposed for multiuser MIMO-OFDMA environment providing proportional fairness. Familiar MIMO radio channel model is extended for OFDMA transmission, to exploit multiuser diversity. Antenna selection is realized in order to maximize the overall data rate based on MIMO channel estimation. Adaptive M-QAM modulation is performed on selected antennas on each subcarrier, taking into account singular value decomposition based channel gain values arisen from the analytical calculation of MIMO channel capacity. Transmit power control is included over subcarriers and transmit antennas.**

## 1. Introduction

The 4G mobile telecommunications systems are inspired to treat raising user demands in terms of data rate in 3G standards. The key advantages of 4G systems lie on the background of the physical layer access; OFDM (Orthogonal Frequency Division Multiplexing) modulation means an efficient and robust technique which is capable to combat fast variations and the frequency selectivity of the radio channel. OFDMA (Orthogonal Frequency Division Multiple Access) provides further advance in terms of spectral efficiency through location dependent multiuser-diversity [1], and represents the most efficient multiple access technique based on OFDM modulation [2].

Applying MIMO (Multiple-Input and Multiple-Output) antenna solution within a radio communication link is – in principle by the fruition of certain channel properties – able to multiple the attainable data rate by a factor which is determined by the number of transmit and receiver antennas. Hence, the combination of OFDMA and the MIMO antenna solution could step forward to be the key element of 4G radio telecommunication systems physical access, as it is mentioned in WiMAX [3] and LTE [4].

In the current work we are dealing with adaptive downlink radio resource allocation in MIMO-OFDMA systems. The subcarriers defined by the OFDMA scheme are shared among multiple users within the system, aiming maximal system channel capacity along such contradicting requirements like user fairness and maximal available transmit power, while the potential of multiple antennas was also taken into account. The optimization possibilities for a typical MIMO-OFDMA scheduling algorithm are the subcarrier allocation, the adaptive transmit power control over subcarriers, antenna selection at the transmitter side, transmit power control for antennas on selected subcarriers [5,6]. The optimization task is similar to OFDMA physical resource management, except the appearance of additional dimensions in the optimization task, generated by the transmitter and re-

ceiver antennas. The multiple antenna transmission enhances the efficiency of the communication, and provides further exploitable optimization gain in the overall transmission rate with the possibility of antenna selection. However, the appearance of a new dimension, the MIMO-OFDMA optimization task becomes to a more difficult problem than ‘simple’ OFDMA optimization, which is an *NP-hard* computational problem in itself [7].

A suboptimal transmit power control (scheduler) algorithm addressing the described problem, published in [6] was taken as a basis. As the new technical content of this paper, we came up with improvements in terms of transmit power control for previously selected active transmitter antennas on selected subcarriers. The scheduler algorithm outputs transmit power values for each subcarriers and active antennas. Simulation results show that in high SNR (Signal-to-Noise Ratio) range about 10% gain can be achieved in channel capacity by extending the power control to antennas. Furthermore adaptive *M*-QAM modulation per subcarrier for Rayleigh-fading channel model was introduced allowing better prediction of realistic channel capacity.

Along this paper a brief summary of OFDM and OFDMA downlink transmission is presented, a draft of the so called water-filling (or water-pouring) transmit power allocation algorithm and its complexity is shown. The MIMO radio access technique is examined regarding the overall system capacity and its optimization tasks in Section 3. The MIMO-OFDMA system model is proposed for locating partial tasks in transmission architecture. A reference MIMO-OFDMA RRM (Radio Resource Management) solution is introduced in Section 4, which allocates radio resources in case of MIMO-OFDMA transmission, providing proportional fairness among users. Practical improvements on the reference algorithm described above are presented in Section 5, where adaptive power control for antennas is discussed in details in Section 6. Section 7 deals with simulation results followed by the conclusion in Section 8.

## 2. Multiple antenna transmission

The multi-antenna transmission- and reception schemes have the name MIMO in the related literature. MIMO schemes can be characterized into three main groups: beam forming solutions, frequency- and space-diversity schemes and *spatial multiplexing* (SM). Spatial multiplexing is often referred as MIMO in itself. In this paper we will be consider spatial multiplexing exclusively.

### 2.1 Spatial multiplexing

During the spatial multiplexing mode, different data symbols are transmitted on the radio link by different antennas on the *same frequency and time interval*. Multipath propagation is assumed for the efficient operation of spatial multiplexing, i.e. rich multipath scattering increases the capacity of the MIMO channel, and LOS (line of sight) propagation degrades strongly the performance of MIMO [8]. The MIMO transmission can be characterized with a time-variant channel matrix according to (2.1)

$$\mathbf{H}(\tau, t) = \begin{pmatrix} h_{1,1}(\tau, t) & h_{1,2}(\tau, t) & \cdots & h_{1,N_R}(\tau, t) \\ h_{2,1}(\tau, t) & & \ddots & \\ \vdots & & & h_{n_T, n_R}(\tau, t) \\ h_{N_T, 1}(\tau, t) & \cdots & & h_{N_T, N_R}(\tau, t) \end{pmatrix}$$

where  $h_{n_T, n_R}(\tau, t)$  represents the complex time-variant channel transfer function at the path between the  $n_T$ -th transmitter antenna and the receiver antenna  $n_R$ .  $N_T$  and  $N_R$  represent the number of transmitter and receiver antennas respectively.

#### 2.1.1 Capacity of MIMO channel

Derived from Shannon's law, the following expression was proven in [8] and [9] for the capacity of a MIMO channel

$$C = \max_{\text{tr}(\mathbf{R}_{ss}) \leq P} \log_2(\det(\mathbf{I} + \mathbf{H}\mathbf{R}_{ss}\mathbf{H}^H)), \quad (2.2)$$

where  $\mathbf{H}$  denotes the channel matrix defined in (2.1),  $\mathbf{H}^H$  denotes the transpose conjugate of  $\mathbf{H}$ ,  $\mathbf{I}$  represents an identity matrix with a flexible size, and let denote  $\mathbf{R}_{ss}$  the covariance matrix of the transmitted signal vector  $\mathbf{s}$ . However, (2.2) has a considerably uncomfortable form for practical applications. Hereinafter, we will investigate the transformation of the MIMO channel into parallel SISO (Single Input, Single Output) channels in order to apply a similar expression to calculate the equivalent capacity like Shannon's formula.

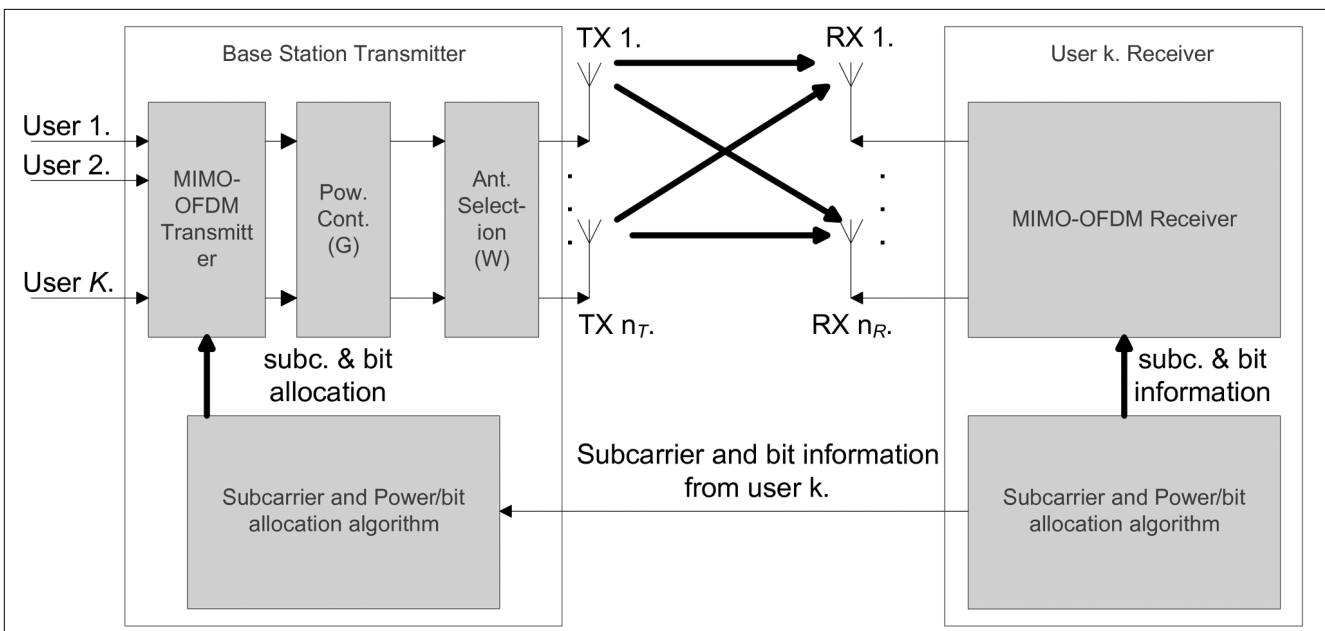
#### 2.1.2 Capacity of parallel equivalent SISO channels

As it is shown in [10], if singular value decomposition (SVD) [11] is applied to MIMO channel matrix, expression (2.2) can be transformed into a more convenient form, which is equivalent to calculate the aggregate Shannon-like transmission rate (in bit/s) of multiple independent SISO channels.

$$C = B \cdot \sum_{i=1}^M \log_2 \left( 1 + \frac{\varepsilon_i^2 \cdot p_i}{N_0} \right). \quad (2.3)$$

In the expression above  $B$  and  $N_0$  represent the channel bandwidth and the spectral power density of the additive white Gaussian noise (AWGN) respectively.  $\varepsilon_i$  denotes the  $i$ -th singular value of the  $\mathbf{H}$  channel matrix, where  $i \in \{1, 2, \dots, M\}$ . The  $\varepsilon_i^2$  values are equivalent to the eigenvalues of the matrix  $\mathbf{H} \cdot \mathbf{H}^H$ , often referred in literature as  $\lambda_i (= \varepsilon_i^2)$ .  $M = \min(N_T, N_R)$  denotes the number of independent SISO channels and  $p_i$  represents the transmit power allocated on the  $i$ -th equivalent SISO channel. The (real) values of  $\varepsilon_i^2$  contain both the effects of the average channel gain, and the spatial correlation of the channel paths within the MIMO constellation. Note, if the MIMO channel paths are independent, and the capacity will be maximal, and  $\varepsilon_i^2 = \varepsilon_j^2$ , for  $\forall i \neq j$ .

Figure 1. MIMO-OFDMA System Model



### 3. MIMO-OFDMA system model

In this section, a model will be provided for MIMO-OFDM radio transmission, a MIMO-OFDM block scheme and the optimization potentials will be summarized.

#### 3.1 MIMO-OFDMA

A system model of downlink MIMO transmission combined with OFDMA multiple access is illustrated at Fig. 1. A subcarrier- and power allocation algorithm is performed at the transmitter side (assuming the availability of CSI information). Each of the transmitter and receiver antennas are assigned to *individual* OFDM transmitters and receivers.

##### 3.1.1 MIMO-OFDMA channel model

For the characterization of a multiuser MIMO radio channel, some transformations should be performed on a channel model given in section 2.1. Since the radio channel has a frequency selective feature along the subcarriers, and different users have channel paths with different conditions toward the base station, the MIMO channel – defined in (2.1) – needs to be *distinguished* along each  $N$  carriers and  $K$  users. Completed with the assumption, that the radio channel will be considered constant in time and frequency within the scheduling time intervals, and along the bandwidth of the different subcarriers.

Expression (3.1) describes the the channel matrix – according to (2.1) – for subcarrier  $n$ , and user  $k$ .

$$\mathbf{H}_{k,n} = \begin{pmatrix} H_{1,1}^{(k,n)} & H_{1,2}^{(k,n)} & \dots & H_{1,N_R}^{(k,n)} \\ H_{2,1}^{(k,n)} & \dots & \dots & \dots \\ \vdots & \dots & H_{n_T,n_R}^{(k,n)} & \vdots \\ H_{N_T,1}^{(k,n)} & \dots & \dots & H_{N_T,N_R}^{(k,n)} \end{pmatrix}, \quad (3.1)$$

where  $n_T \in \{1, 2, \dots, N_T\}$  and  $n_R \in \{1, 2, \dots, N_R\}$  are the indexes of transmitter and receiver antennas, and  $H_{n_T,n_R}^{(k,n)}$  denotes the complex channel samples according to the *scheduling period* for the different antenna-paths.

During a MIMO-OFDMA transmission it can be easily shown, that for characterization of the *overall* radio channel, the introduction of a 4-dimensional  $\mathbf{H}_{\text{tot}}$  hypermatrix is needed with the size of  $K \times N \times N_T \times N_R$ , according to

$$\mathbf{H}_{\text{tot}} = \begin{pmatrix} \mathbf{H}_{1,1} & \mathbf{H}_{1,2} & \dots & \mathbf{H}_{1,N} \\ \mathbf{H}_{2,1} & \dots & \dots & \mathbf{H}_{2,N} \\ \vdots & \dots & \mathbf{H}_{k,n} & \vdots \\ \mathbf{H}_{K,1} & \dots & \dots & \mathbf{H}_{K,N} \end{pmatrix}, \quad (3.2)$$

where  $\mathbf{H}_{k,n}$  elements are arisen from (3.1).

#### 3.2 MIMO-OFDMA system capacity optimization

The capacity optimization task of a MIMO-OFDMA transmission can be generalized as follows (3.3)

$$C = \arg \max_{C,p} B \cdot \sum_{k=1}^K \sum_{n=1}^N \rho_{k,n} \left[ \sum_{i=1}^{M_{k,n}} \log_2 \left( 1 + \frac{\lambda_{k,n}^{(i)} p_n}{N_0} \right) \right],$$

where  $\rho_{k,n}$  represents the elements of the  $K \times N$  *subcarrier assignment* matrix, where  $\rho_{k,n} = 1$ , if subcarrier  $n$  is assigned to user  $k$ , and  $\rho_{k,n} = 0$  otherwise. A requirement has to be set for the transmit power allocation:

$\sum_{n=1}^N p_n \leq P_{\max}$ ,  $p_n \geq 0$  for all  $k$  and  $n$ , where  $P_{\max}$  denotes the maximal transmit power of the MIMO-OFDMA transmitter. The value of  $\lambda_{k,n}^{(i)}$  represent the eigenvalue  $i$  of  $\mathbf{H}_{k,n} \mathbf{H}_{k,n}^H$  matrix, where  $i = 1, \dots, \text{rank}(\mathbf{H}_{k,n})$ .

##### 3.2.1 Subcarrier allocation

In case of MIMO-OFDMA access, in terms of maximization of the overall system capacity, user  $k$  is assigned to subcarrier  $n$ , for which the following expression will be maximal

$$k_n = \arg \max_k \prod_{i=1}^{M_{k,n}} \left( 1 + \frac{\lambda_{k,n}^{(i)} p_n}{N_0} \right). \quad (3.4)$$

The method above follows the simple OFDMA subcarrier allocation, since subcarrier to be allocated are characterized with a scalar value, however  $k_n$  values contain  $\lambda_{k,n}^{(i)}$  eigenvalues, which are determined both by the gains of the MIMO path of the MIMO channel indexed by  $k$  and  $n$ , and the degree of linear independence of MIMO paths on the same  $k$  and  $n$  indexes.

The capacity of OFDMA access can be – in principle – multiplied with a factor, which can be expressed as

$$M_{k,n} = \text{rank}(\mathbf{H}_{k,n}) \equiv \min(N_T, N_R). \quad (3.5)$$

Note that, in case of MIMO-OFDMA there exist  $K \times N$  pieces of channel matrices, and the singular value decomposition and transformation into parallel SISO channel (as seen in 2.1.2) are executed  $K \times N$  times.

The practical realization of subcarrier allocation above follows to a computationally complex task. The following two criterions are suggested in [12] to simplify subcarrier allocation, which lead to suboptimal resource allocation

$$k_n^P := \arg \max_k \prod_{i=1}^{M_{k,n}} \lambda_{k,n}^{(i)} \quad (3.6)$$

$$k_n^S := \arg \max_k \sum_{i=1}^{M_{k,n}} \lambda_{k,n}^{(i)} \quad (3.7)$$

called *product*- and *sum* criterions. Product criterion is more efficient in high SNR domains than sum criterion, and vice versa. However, application of MIMO technique is usually relevant by high SNR values, hence product criterion has greater importance in MIMO resource management.

##### 3.2.2 Transmit power control

A further requirement to maximize the overall system capacity is the allocation of  $p_n$  transmit power values for subcarriers, which are the roots of the following equation system

$$\sum_{i=1}^{M_{k,n}} \frac{\lambda_{k,n}^{(i)}}{\lambda_{k,n}^{(i)} p_n + N_0} + \alpha = 0, n \in \{1, 2, \dots, N\}, \quad (3.8)$$

where  $k_n$  is the index of user, which owns the carrier  $n$ ,  $\alpha$  satisfies the requirement  $\sum_{n=1}^N p_n = P_{max}$ . The method above is referred as the *multi-dimensional water-filling* algorithm, namely for  $M_{k,n}=1$ . we get the form of 'common' water-filling solution [13].

### 3.2.3 Antenna selection

In MIMO systems the antenna RF (radio frequency) circuits can also be considered as a resource factor during the capacity optimization. The market rates of RF chains do not follow the general rules in computer studies (Moore's law). Thus, by the manufacturing of MIMO radio devices is a main aspect to reduce the number of necessary RF circuits. This consideration gives the reason of existence of so called *antenna selection* algorithms if a MIMO device has multiple transmitter antennas, but there are only  $L \leq N_T$  available RF chains. The task of a transmitter side antenna selection algorithm is to select  $L$  pieces of antennas, on which the most efficient radio transmission can be realized, based on channel properties on selected chain paths [14].

In optimal case, the expression of the channel capacity is needed to be calculated for  $\binom{N}{L}$  cases before each symbol transmission however, this optimal technique can lead to an unacceptable computational complexity likewise. A common algorithm for antenna selection produces near optimal results for transmit side antenna selection with linear operations called Gorokhov's method [15].

## 4. MIMO-OFDMA RRM solution

A reference method [6] will be proposed in this section, which provides a suboptimal solution for the maximization of MIMO-OFDMA capacity expression (3.3) by providing proportional fairness. Subcarrier allocation and power control is realized for selected subcarriers. Listed partial tasks are handled separately to give a treatable computational complexity and thus, suboptimal results. The performance (spectral efficiency) of the algorithm is evaluated by Shannon's formula, giving a loose upper bound for system capacity, which is not feasible for analyzing the performance of the algorithm in a standardized telecommunication network.

### 4.1 Objectives of the reference algorithm

Product criterion – defined in (3.6) – is used for subcarrier allocation purposes. Proportional fairness is ensured among users, which means, that a defined quotient  $\{\gamma_k\}_{k=1}^K / R_k$  should be equal for each user, where  $\{\gamma_k\}_{k=1}^K$  represents a vector, containing the transmission rate relations between different users, and  $R_k$  denotes the actual transmission rate of user  $k$ , which can be calculated according to (3.3) with the Shannon's formula.

$$R_k = \sum_{n=1}^N \rho_{k,n} \left[ \sum_{i=1}^{M_{k,n}} \log_2 \left( 1 + \frac{\lambda_{k,n}^{(i)} \cdot p_n}{N_0} \right) \right] \quad (4.1)$$

The partial tasks of the proposed algorithm will be detailed in the following subsections.

### 4.2 Subcarrier pre-allocation

During this phase, the  $N_k$  number of allocated subcarriers will be calculated for each user based on  $\{\gamma_k\}_{k=1}^K$ . This pre-allocation step will be executed according to a calculated  $\bar{H}_k$  average channel gain values for each user according to

$$\bar{H}_k = \frac{1}{N_T \cdot N_R} \frac{1}{N} \sum_{n=1}^N \sum_{n_T=1}^{N_T} \sum_{n_R=1}^{N_R} |H_{k,n}|, \forall k \quad (4.2)$$

where  $H_{k,n}$  is already defined in (3.1). Equal power distribution is assumed on subcarriers in order to reduce complexity, and an approximated average transmission rate can be calculated for each user

$$\bar{R}_k = N_k \log_2 \left( 1 + \frac{|\bar{H}|^2 \cdot \bar{P}_k}{N_0} \right) \quad (4.3)$$

An iterative  $N_k$  calculation will be executed. In each step  $\bar{R}_k$  is updated, and the user with the lowest  $\{\gamma_k\}_{k=1}^K / R_k$  quotient will obtain a new subcarrier.

### 4.3 Subcarrier assignment

After determining the number of subcarriers for each user, the effective subcarrier assignment will be executed. Users are divided into two groups according to their  $\bar{H}_k$  average channel gain values.

$$\bar{H}_1 \leq \bar{H}_2 \leq \dots \leq \bar{H}_k \leq \dots \leq \bar{H}_K \quad (4.4)$$

At first, users with low average channel gain have the possibility to choose the 'best' subcarriers characterized with the  $k_n$  lambda product values (3.6). This step allows the compensation of low channel gains for a subcarrier which has high lambda product value, since eigenvalues of MIMO paths are carrying the degree of linear dependency of antenna paths as well. Therefore a reliable transmission can be kept up by a relative low channel gain, the degree of the linear dependency among MIMO paths is relatively low. The subcarrier allocation happens as described in section 4.2, refreshing  $R_k$  according to expression (4.1), till each user obtains the  $N_k$  number of subcarriers calculated beforehand. After allocation of subcarriers, the transmit power allocation will be performed on selected subcarriers based on (3.8).

## 5. Improvements on MIMO-OFDM RRM

In this section, extensions will be proposed to enhance the efficiency of the reference algorithm, in terms of systems' transmission rate and to make the results to be interpretable for a practical telecommunication system (LTE, WiMAX, etc.). Adaptive  $M$ -QAM modulation is realized on the selected subcarriers, based on the instantaneous channel state information. Antenna selection and power control for selected antennas will be performed over the subcarriers.

**5.1 Adaptive M-QAM modulation in Rayleigh-fading channel**

For analyzing the operation of a MIMO-OFDMA resource management algorithm in a real communication system, a capacity calculation of a specific modulation method is needed. In 3GPP LTE and WiMAX M-QAM adaptive modulation is performed on groups of neighboring subcarriers. However – in principle –, there is a possibility to perform adaptive modulation for each subcarrier.

The task of adaptive modulation can be summarized as the definition of signal-to-noise-ratio (SNR) domains, in which the performance of the selected modulation method (or level) will be maximal in terms of spectral efficiency. For solving the problem above, we calculate the channel capacity for different M levels of QAM modulation based on analytical bit error results on Rayleigh-fading channel.

*5.1.1 Error rates of M-QAM in Rayleigh-fading channel*

Analytical bit error probability expression for M-QAM modulation in Rayleigh-fading channel is available according to [16] and [17]:

$$P_b = \frac{2}{\pi\sqrt{M}\log_2\sqrt{M}} \sum_{k=1}^{\log_2\sqrt{M}} \sum_{i=0}^{(1-2^{-k})\sqrt{M}-1} \left\{ (-1)^{\lfloor \frac{i2^{k-1}}{\sqrt{M}} \rfloor} \times \left( 2^{k-1} - \left\lfloor \frac{i2^{k-1}}{\sqrt{M}} + \frac{1}{2} \right\rfloor \right) \right\} \quad (5.1)$$

$$\times \left( \prod_{l=1}^{\frac{\pi}{2}} M_{\gamma_l} \left( -\frac{(2i+1)^2 3 / (2(M-1))}{\sin^2 \theta} \right) \right) \quad (5.2)$$

where  $M_{\gamma_l} = \frac{1}{1-s\gamma_l}$  (5.2)

represents the moment generating function (MGF) of l-th branch of applied diversity method in case of Rayleigh-fading channel, and

$$M_{\gamma_l} = \frac{1+K}{1+K-s\gamma_l} e^{\left[ \frac{Ks\gamma_l}{(1+K)-s\gamma_l} \right]} \quad (5.3)$$

denotes the MGF for Rician channel.

*5.1.2 Channel capacity*

For a given bit error probability, the capacity of the radio channel for a given SNR can be calculated in closed form pursuant to the following steps. We invoke the expression of the conditional entropy according to [18] and [19]. The spectral efficiency of a binary channel (in bit/s/Hz) expressed with the conditional entropy function, can be calculated as

$$C_{M-QAM} = \log_2 M \left( 1 - H(X|Y)_{M-QAM} \right), \quad (5.4)$$

where

$$H(X|Y)_{M-QAM} = -[(1-P_b)\log_2(1-P_b) + P_b \cdot \log_2(P_b)], \quad (5.5)$$

denotes the conditional entropy function.

*5.1.3 Adaptive modulation on subcarriers*

The task of adaptive modulation is the calculation of SNR levels for switching between modulation levels. The essential parameter for bit error probability calculation is the SNR.

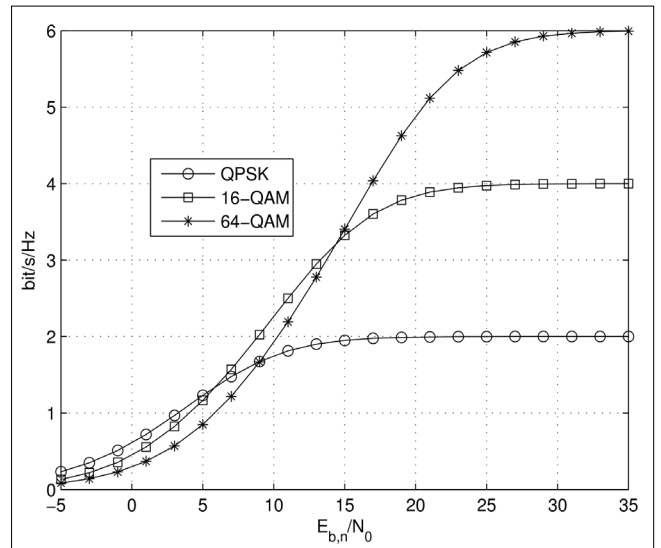
Let us denote it with  $\gamma_n$ , and expressed for subcarrier n as

$$\gamma_n = \frac{|H_n|^2 P_n}{N_{0,n} \cdot \Delta f_c} = \frac{E_{s_n}}{N_{0,n}} = \frac{E_{b_n} \log_2(M_n)}{N_{0,n}} \quad (5.6)$$

where  $P_n$  denotes the allocated transmit power (in W),  $E_{s_n}$  and  $E_{b_n}$  represent the received symbol- and bit energy for subcarrier n respectively,  $N_{0,n}$ ,  $\Delta f_c$  and  $M_n$  nominates the spectral power density of the AWGN, the subcarrier spacing (in Hz) and the modulation level on subcarrier n.

After that, the modulation levels can be determined obviously for each SNR (or transmit power) domain. At Fig. 2, the capacity curves of QPSK (4-QAM), 16-QAM and 64-QAM are plotted for different  $E_{b_n}/N_0$  levels, and the intersections of the curves determine the modulation switching points, to provide the maximal spectral efficiency.

Figure 2. M-QAM levels for adaptive modulation by different  $E_{b_n}/N_0$  levels



**5.2 Antenna selection at the transmitter side**

As it was exposed in section 3.2.3,  $L = \min(N_T, N_R)$  number of transmitter antennas should be assigned to available transmitter RF devices transmission for maximizing capacity, if  $N_T \geq N_R$ . For realizing computationally efficient antenna selection method, antenna specific channel state information is needed as the input of the algorithm. Similarly to expression (4.2), the average channel gain values can be derived from  $\mathbf{H}_{k,n}$  matrices given in (3.1).

At the first step, let define an  $N_T \times N_R$  average  $\mathbf{H}_{aver}$  channel gain matrix according to

$$\mathbf{H}_{aver} = \frac{1}{K} \frac{1}{N} \sum_{k=1}^K \sum_{n=1}^N |\mathbf{H}_{k,n}|. \quad (5.7)$$

A further transformation will be accomplished on  $\mathbf{H}_{aver}$  to obtain  $N_T$  pieces of scalar values, characterizing transmitter antennas.  $\mathbf{H}_{aver}$  will be transformed into a  $\mathbf{h}_{n_T}$  vector with

$$\mathbf{h}_{n_T} = \frac{1}{N_R} \sum_{n_R=1}^{N_R} \mathbf{H}_{aver}, \quad (5.8)$$

with that, the antenna selection algorithm selects the first  $L$  largest elements from  $\mathbf{h}_{n_T}$  in each scheduling interval

$$\arg \max_{n_T} \frac{1}{N_R} \sum_{n_R=1}^{N_R} h_{n_T, n_R}, \quad (5.9)$$

realizing statistical antenna processing gain in terms of spectral efficiency compared to the random selection of transmitter antennas. (See simulation results in Section 7.)

## 6. Adaptive power control for antennas

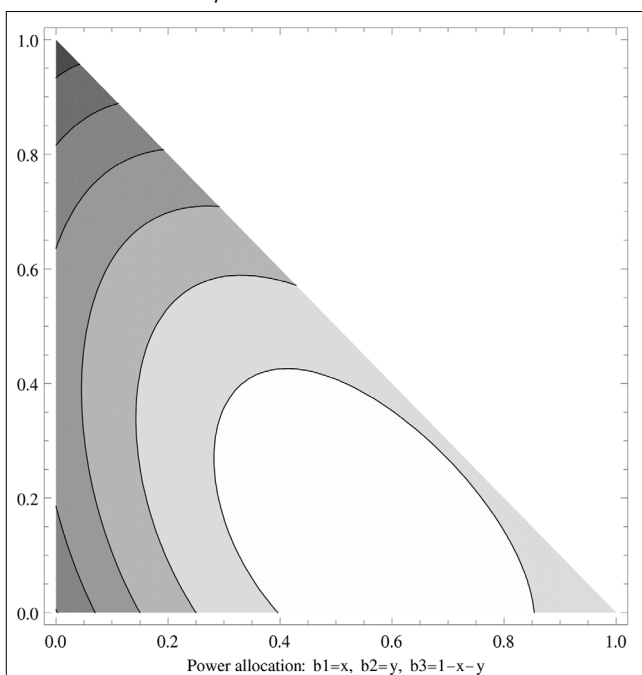
The reference method – presented in Section 4 – stops at the power control for subcarriers, and assumes equal power distribution along the transmitter antennas. However, the allocated power values for subcarriers can be divided further, based on the eigenvalues of  $\mathbf{H}_{k,n}$  matrices. Since the rank of  $\mathbf{H}_{k,n}$  equals with the number of selected  $N_T$  antennas, eigenvalues can be ordered to the corresponding transmit antennas. The power control task could be naturally contracted, however in order to hold the optimization task in computational complexity, it is practical to separate the two power control tasks after all. In this case, the elements of a  $\beta_n^{(i)}$  vector shall be calculated during the following formalized task according to Shannon's formula (6.1)

$$\arg \max_{\beta \in \mathbb{R}^{N \times M_{k,n}}} \sum_{k=1}^K \sum_{n=1}^N \rho_{k,n} \left[ \sum_{i=1}^{M_{k,n}} \log_2 \left( 1 + \frac{\lambda_{k,n}^{(i)} \beta_n^{(i)} p_n}{N_0} \right) \right],$$

$$\text{so that } \sum_{i=1}^{M_{k,n}} \beta_n^{(i)} = 1, \forall n \quad \leq \beta_n^{(i)} \leq 1/\text{hile } \forall i, n.$$

values are independent over the carriers, the optimization problem can be separated into  $N$  pieces of maximization tasks. It is noticeable, that the optimization

Figure 3. Capacity function of 3-antenna system by different power allocation



task defined in expression (6.1) is very similar to the general water-filling problem with the modification, that subcarriers' channel gain vector is transformed into SISO  $\lambda_{k,n}^{(i)}$  eigenvalues, and the number of subcarriers will be represented by  $M_{k,n}$ . At Fig. 3, the capacity of a 3 antenna scenario is illustrated along different  $\beta_n^{(i)}$  power distributions. The reason of existence of antenna power allocation can be well observed, because the maximum of capacity is not located at  $\beta_n^{(1)} = \beta_n^{(2)} = \beta_n^{(3)} = 1/3$ .

The water-filling solution is available, and is formalized in Equation (3.8). After the appropriate transformations mentioned above, the antenna power allocation algorithm can be formalized as follows

$$\beta_n^{(i)} + \frac{N_0}{\lambda_{k,n}^{(i)} \cdot p_n} - \alpha = 0, i = 1, 2, \dots, M_{k,n}. \quad (6.2)$$

The defined method has the advantage, that the computation can be parallelized, resulting significant acceleration in calculation. Let note, that maximizing Shannon's capacity does not provide the optimal solution in a communication system, which operates with real  $M$ -QAM modulation. Realization of so-called bit loading water-filling [13] exceeds the confines of this paper.

Parameter	Notation	Value
Subcarrier spacing	$\Delta f_c$	15 kHz
System bandwidth	$W_{\text{OFDM}}$	5 MHz
Number of subcarriers	$N = \lfloor W_{\text{OFDM}} / \Delta f_c \rfloor$	333
FFT size	$N_{\text{FFT}}$	512
Sampling frequency	$f_s = N_{\text{FFT}} \cdot \Delta f_c$	7.86 MHz
Symbol period	$T_s = 1 / \Delta f_c$	66.7 $\mu\text{s}$
Carrier frequency	$f_c$	2.5 GHz
AWGN spectral density	$N_0$	-174 dBm/Hz
Cell radius		500 m
Base station height		30 m
Mobile station height		1.5 m
Pathloss model		COST-HATA-231
LTE signaling overhead		27.29 %

Table 1. Simulation parameters

## 7. Simulation results

The performance results, and plots are mainly based on analytical calculations (Shannon's capacity and  $M$ -QAM capacity), however some of the system parameters – like user's position within cells, complex Rayleigh-fading channel coefficients, and users' fairness vector – were randomly generated.

### 7.1 System Parameters

For analyzing performance of scheduler algorithm LTE radio access, parameters were considered in Table 1.

### 7.2 Illustration of operation

Fig. 4 illustrates a scheduling cycle in case of  $K=6$  users. Channel gain parameters are randomly generated (first row). For the sake of simple illustration, fairness parameters are set to  $\gamma=(1,1,1,1,1,5)$ . The allocated carriers to users can be observed at the third row, to provide

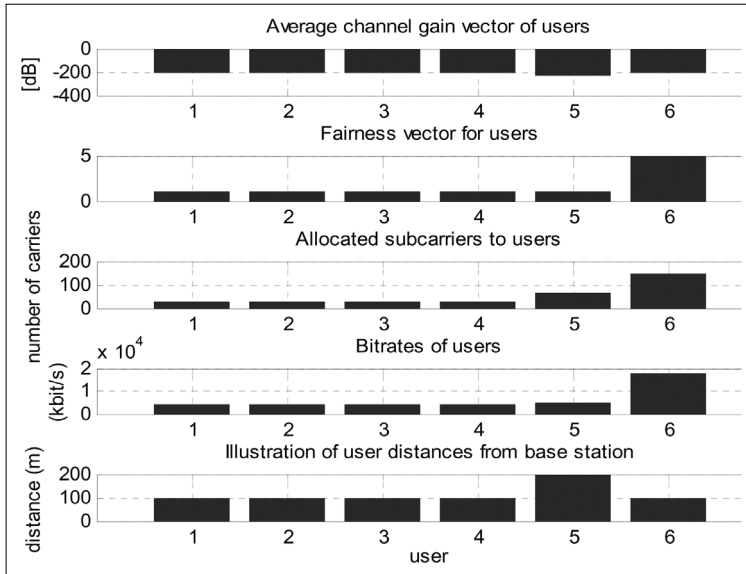


Figure 4. Illustration of a scheduling cycle

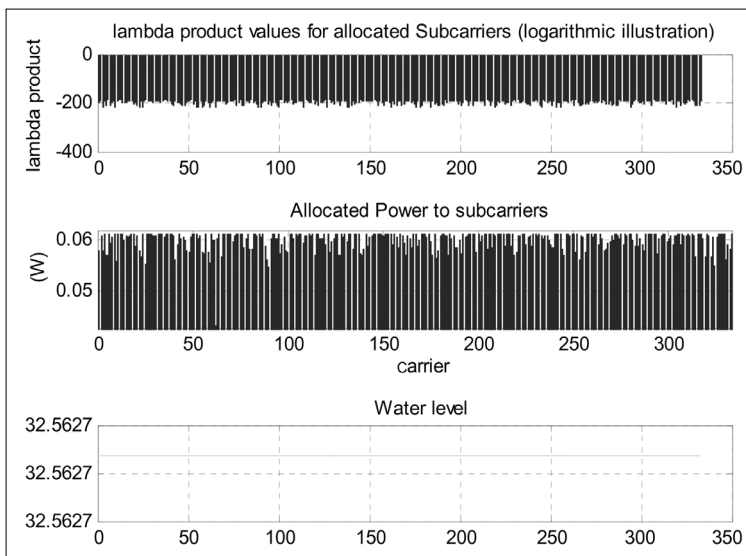


Figure 5. Adaptive Transmit Power Control over subcarriers

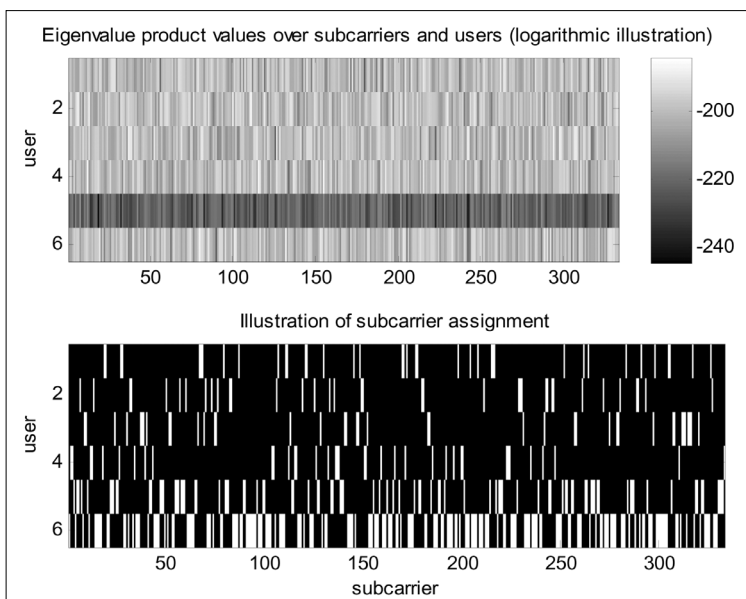


Figure 6. Illustration of MIMO-OFDMA singular value products, and subcarrier assignment

achievable transmission for users. The fulfillment of fairness requirements is visible, even if channel gain distribution has a large deviation for certain subcarriers.

The results of the multi-dimensional water-filling (defined in section 3.2.2) are illustrated on Fig. 5. The proper operation of the algorithm can be checked with equal  $\alpha$  ‘water level’ for each subcarrier.

The results of the subcarrier allocation – based on lambda product values – are depicted at Fig. 6. At the upper diagram, the lambda product values are shown for each subcarrier and user (in logarithmic scale). User 5 (which has the largest distance from base station) has the minimal average  $\prod_{i=1}^{M_{k,n}} \lambda_{k,n}^{(i)}$  value (row 5).

Lower part of Fig. 6 contains the illustration of the subcarrier allotment. It seems that user 6 (which has the largest proportional fairness parameter) has obtained the main part of subcarriers. In addition, user 5 (which has equal fairness parameter with first 4 users) have received more carriers than first 4 users to balance the penalty resulting from its lower average channel gain values.

### 7.3 Performance analysis

In the following steps, the performance analysis will be summarized of the reference method completed with new elements, discussed in Section 5. We have illustrated the spectral efficiency as a function of the  $E_s/N_0$  quotient at the transmitter-side. The set-up and the *randomly generated* channel coefficients implied a path attenuation of approximately 100 dB, which the transmitted signal was exposed to in our simulations.

#### 7.3.1 User distances and System capacity

In Section 4.1, the overall system capacity and fairness were presented as contradictory aspects. At Fig. 7, at first equal (100 m) user distance were set as an input of the scheduling, resulting 81.82 Mbit/s of overall system transmission rate. However, this situation can be considered fairly infrequent in real conditions. On the right side of the figure, the first user’s location were set to 400 m from the base station.

In this case (maintaining fairness) the scheduler has allocated almost all of the subcarriers for that user, resulting significant degradation in overall (76.61 Mbit/s) transmission rate.

#### 7.3.2 Fairness analysis

The widely used Jain’s index [20] means a feasible method for analyzing the degree of fairness. Jain’s index provides a value between 0 and 1.

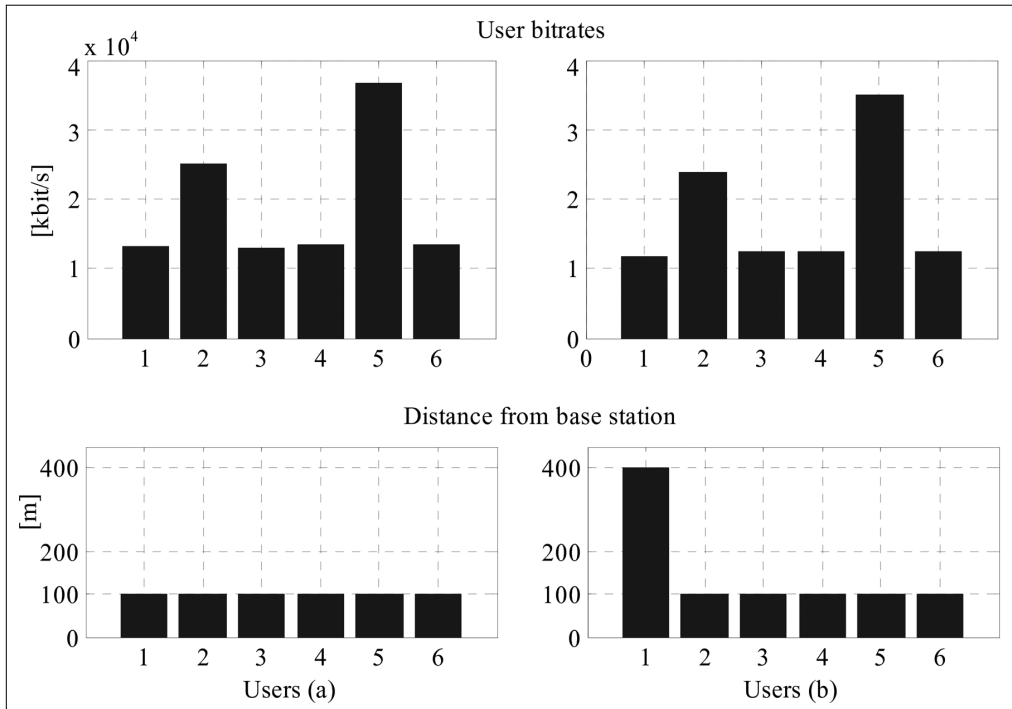


Figure 7. Effect of user distances on system capacity

$$f(\bar{r}) = \frac{\left(\sum_{k=1}^K \frac{R_k}{\phi_k}\right)^2}{K \sum_{k=1}^K \left(\frac{R_k}{\phi_k}\right)^2}, \quad (7.1)$$

where  $\phi_k$  could be an amount of money, which user  $k$  pays for the service, proportional to its desired transmission rate. At Fig. 8, the Jain's fairness index is illustrated by growing number of users.

Curves were calculated by fixed distances, but different channel coefficients. It can be observed, that growing antenna number produces decreased fairness, even by extended system bandwidth (20 MHz) and transmit power (20 W). A reason for lower fairness values by growing antenna numbers can be that the effect of the distances has a greater influence on lambda products.

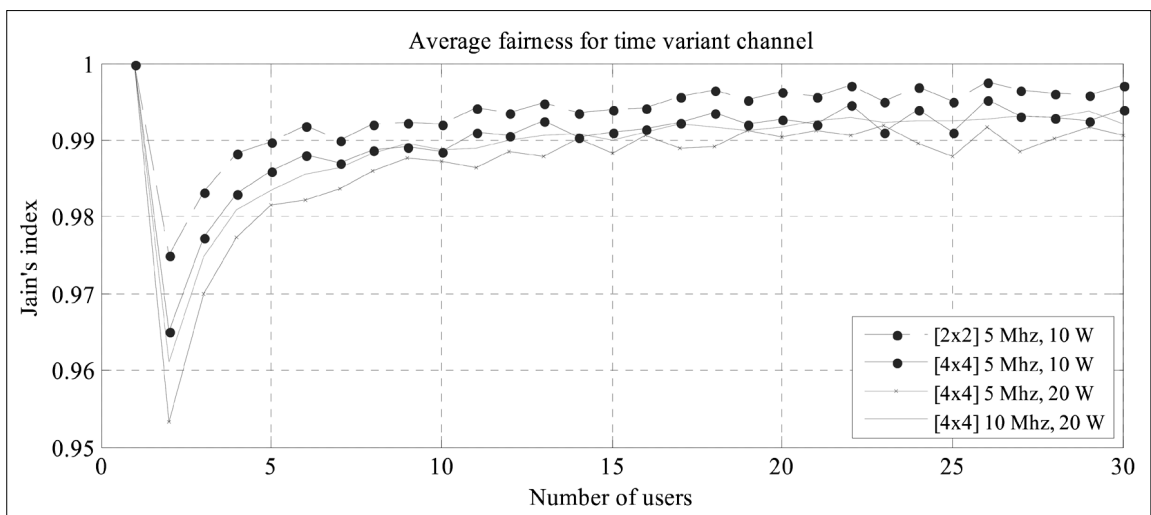
### 7.3.3 MIMO performance

Fig. 9 illustrates the spectral efficiency of MIMO by different transmit power levels and antenna arrangements.

In literature of MIMO technology a common statement holds, that performance of MIMO transmission grows with SNR, and application of MIMO is worthy by high SNR values. Two kind of propagation environments (rural and urban) are simulated in terms of propagation loss (differences between two environments in terms of complicated scattering environment is not modeled). In terms of propagation loss, rural (dotted lines) environment provides better SNR results. The lower part of Fig. 9 illustrates the fraction of 4x4 MIMO and SISO channel for the rural case. The curve of the rural environment is located on the top for all transmit power values. This can be explained only with lower propagation loss (higher SNR) in rural environment.

In a real scattering environment this effect should be turned over, because the degree of scattering significantly affects the capacity of a MIMO channel. However, exact modeling of different MIMO environments could be the subject of a distinct work.

Figure 8. Illustration of fairness by different MIMO scenarios





7.3.4 Adaptive modulation

Fig. 10 shows the spectral efficiency of different MIMO antenna scenarios calculated from Shannon’s capacity, and adaptive *M*-QAM capacities. The application of *M*-QAM modulation does not provide any spectral efficiency gain of course, however it can illustrate the performance of the MIMO transmission (and scheduling algorithm) in a realistic transmission system with adaptive *M*-QAM modulation. During the simulation, 50 cycles were averaged for alternating MIMO channel coefficients. Transmit power was divided equally among subcarriers and antennas.

7.3.5 Power Control

Fig. 11 contains the results of spectral efficiency of power control, calculated by Shannon’s formula. Three cases have been distinguished for each antenna configurations: equal power distribution for subcarriers and antennas (continual and dashed lines with no marking), adaptive power control only on subcarriers (cont. and dashed lines marked with ‘x’), and adaptive power control for subcarriers and antennas alike (cont. and dashed lines marked with diamonds). On the left side, two MIMO configurations are illustrated in lower transmit power domain, at the right side both configurations seem in higher domains. In higher SNR domains full (both over carriers and antennas) power allocation curve accommodates to the curve of the equal power distribution, and it’s visible that power control provides a 10-12% gain at lower SNR values.

At Fig. 12, the same simulation is illustrated with the difference, that spectral efficiency has been calculated

according to the capacity expression of *M*-QAM given in Section 5.1. The ‘full’ power allocation method provides spectral efficiency gain at low SNR values likewise, while in high SNR domains the performance of power control will turn over. In a practical realization adaptive power control should be turned out in high SNR domains.

8. Conclusion

As an improvement of the referenced MIMO-OFDMA RRM solution defined in Section 4, a new downlink resource allocation algorithm was developed, which maximizes the overall spectral efficiency of the system with certain constraints.

The constraints were selected with respect to practical applications: proportional fairness is a suitable representation of QoS the mobile stations can be provided with. The proposed algorithm results in power allotment over subcarriers which allows the base station to meet the requirements, while the overall system capacity is kept closer to the optimum, than it was in the proposed reference work. Planning phases were interpreted, which separate the complex (NP-hard) optimization problem into distinct – independently executable – sub-tasks (i.e subcarrier allocation, adaptive modulation, antenna selection, transmit power control over subcarriers and antennas). Interpreted separation results the manageability of the proposed optimization problem. Nevertheless, the execution of separated optimization tasks is only able to result suboptimal radio resource management. The current work is not supposed to quantitati-

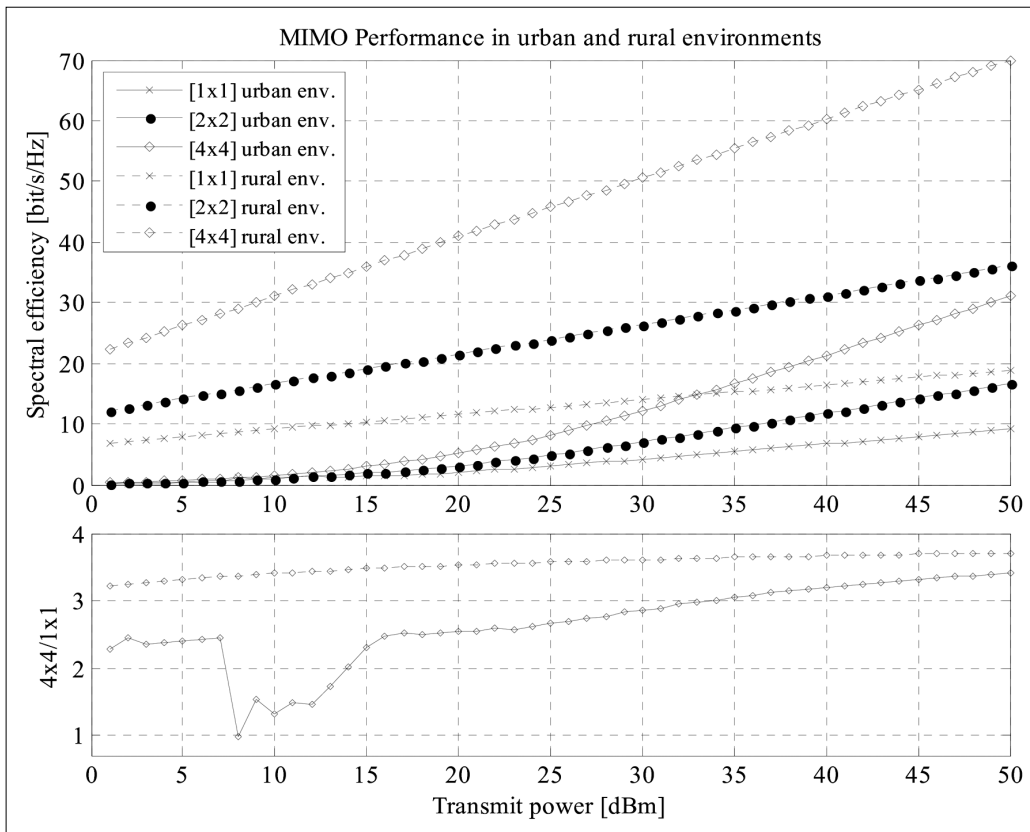
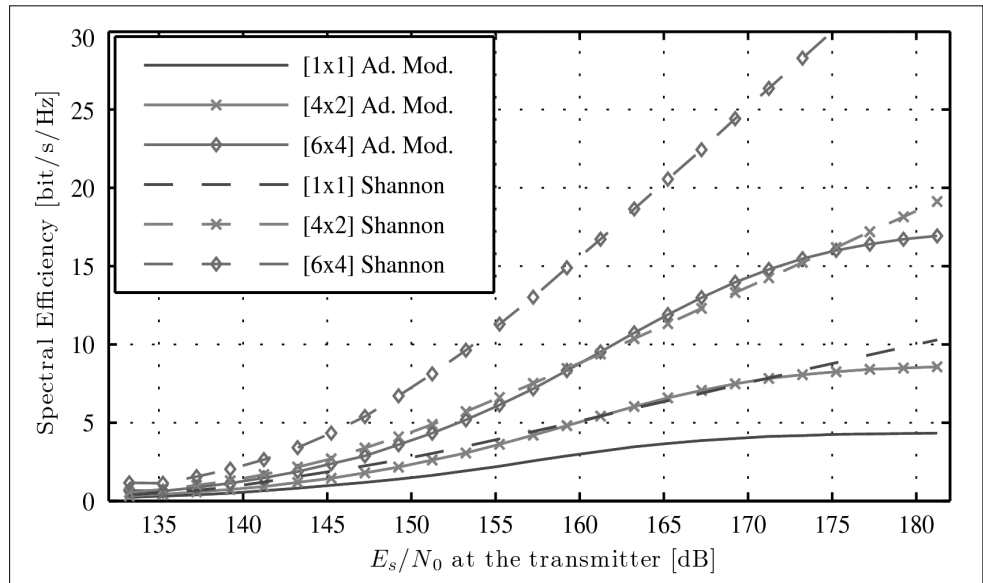


Figure 9. MIMO performance by different propagations environment

Figure 10. M-QAM adaptive modulation's capacity compared to theoretical capacity



vely analyze the measure of how far the resulted power distribution is from the optimum solution. However we can declare a couple of things qualitatively.

Splitting the set of users up into two groups along the users' measured channel parameters is a fairly rough approach, but this step significantly cuts down on complexity. A refined division of still acceptable complexity

could take closer to the optimum. Our proposed amendment however did not lower the complexity either, in comparison with the reference method the additional power allocation step requires more computational capacity at each base station.

Introducing antenna selection causes further divergence from optimum, however efficient selection of an-

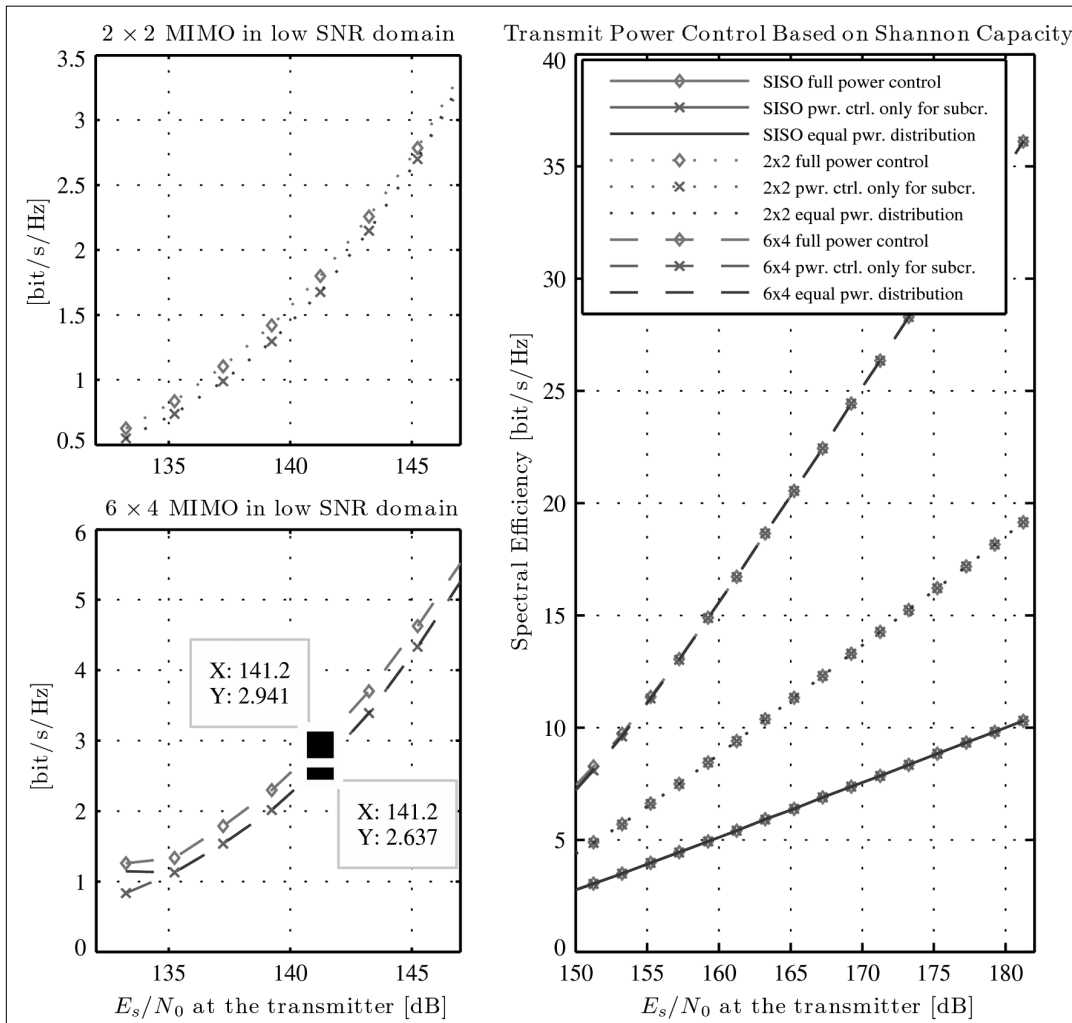
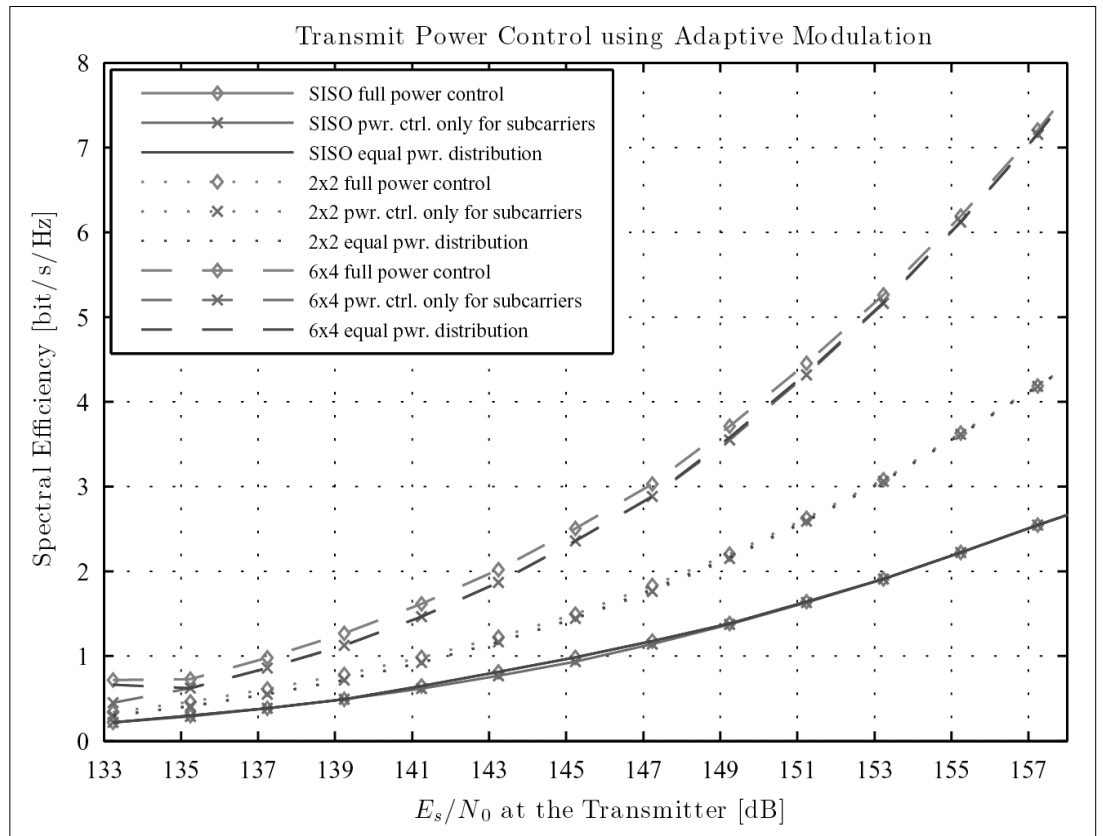


Figure 11. Shannon capacity of MIMO power control

Figure 12. *M*-QAM capacity of MIMO power control



tennas to be cut off could enable significant savings on RF circuits and power demand. A field of improvement of the current work is to evaluate algorithms to be used for antenna selection.

In return for more computational cost with the proposed power control scheme over transmitter antennas the per user transmit rate can be increased by about 10%. Since considerable benefits come forward in the lower SNR region, end user satisfaction can be levered in such situations where the channel properties are less favorable e.g. far away from the base station or in rarely covered area.

The antenna selection and antenna power control mean additive components to the reference RRM algorithm. In addition, the performance evaluation of the proposed algorithm happens assuming adaptive *M*-QAM modulation over the subcarriers, which enables better convergence to Shannon-capacity in multiple SNR domains, than in a case of applying single *M* level of QAM modulation, as well as adaptive modulation allows a more realistic performance evaluation in a standardized communications system. Those simulation results showed which capacity level can be most likely achieved in practical applications.

Of course the performance and thus the actual channel capacity is hardly dependent on the accuracy of Channel State Information fed back to the transmitter. We did not consider channel estimation and equalization methods – inevitable in practical applications – in order to get more realistic results, which is a possible direction to head for.

Authors



**ALBERT MRÁZ** was born in 1979 in Budapest, Hungary. He received the MSc degree in electrical engineering from the Budapest University of Technology and Economics (BUTE) in 2005. He is currently working toward the PhD. degree at BUTE. His research interests include the modeling and resource management of broadband wireless networks (3G and 4G).



**TAMÁS ZÁBÓ** received the M.Sc. degree in Computer Science from Budapest University of Technology and Economics (BME) in 2009. He has been taking part in researches at BME Department of Telecommunications. His main field of research interest is 4G mobile telecommunications. At the moment he is employed at Ericsson Hungary as Tool Expert.



**SÁNDOR IMRE** was born in Budapest in 1969. He received the M.Sc. degree in Electrical Engineering from the Budapest University of Technology (BUTE) in 1993. Next he started his Ph.D. studies at BUTE and obtained Dr.Univ. degree in 1996, Ph.D. degree in 1999 and DSc degree in 2007. Currently he is carrying his activities as a Professor and a Head of Department of Telecommunications at BUTE. He is a member of Telecommunication Systems Committee of the Hungarian Academy of Sciences. He participates in the Editorial Board of two journals: Info-communications Journal and Hungarian Telecommunications. He was invited to join the Mobile Innovation Centre as R&D director in 2005. His research interests includes mobile and wireless systems. His main research interests and contributions are in the areas of various wireless access technologies, mobility protocols and reconfigurable systems.

## References

- [1] D. Tse,  
"Optimal power allocation over parallel gaussian broadcast channels," Proc. of International Symposium on Information, p.27, 1997.
- [2] H. L. Guoqing Li,  
"On the optimality of the OFDMA network," IEEE Communications Letters, No. ISSN 1089-7798., pp.438–440, 2005.
- [3] I.M.T. IEEE Computer Society and T. Society,  
"IEEE standard for local and metropolitan area networks," IEEE Std 802.16, Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems, February 2005.
- [4] J. S. Erik Dahlman, Stefan Parkvall and P. Beming,  
3G Evolution, HSPA and LTE for Mobile Broadband. Academic Press, 2007.
- [5] I. S. Maung Sann Maw,  
"Resource allocation scheme in MIMO-OFDMA system for user's different data throughput requirements," IEICE Transactions on Communications, Vol. E91-B, No. 2, 2008.
- [6] W. P. Jian Xu, Jongkyung Kim and J.-S. Seo,  
"Adaptive resource allocation algorithm with fairness for MIMO-OFDMA system," In VTC Spring, No. 63, pp.1585–1589, 2006.
- [7] H. Yin and H. Liu,  
"An efficient multiuser loading algorithm for OFDM-based broadband wireless systems," IEEE Globecom, Vol. 1, pp.103–107, August 2000.
- [8] A. van Zelst,  
"MIMO-OFDM for wireless LAN," Ph.D. dissertation, Eindhoven University of Technology, April 2004.
- [9] J. G. D.S. Shiu, J. Foschini and J. Kahn,  
"Fading correlation and its effect on the capacity of multielement antenna system," IEEE Transactions on Communications, Vol. 48, p.502, 2000.
- [10] G. Tsoulos,  
MIMO System Technology and Wireless Communications. CRC Press, ch. Theory and Practice of MIMO Wireless Communication Systems, pp.32–34, 2006.
- [11] E. Telatar,  
"Capacity of multi-antenna gaussian channels," European Transactions on Telecommunications, Vol. 10, No. 6, pp.585–596, 1999.
- [12] G. L. H. Liu,  
"On the optimality of OFDM in multiuser multicarrier MIMO systems," In Proc. VTC-2005-Fall Vehicular Technology Conf., No. ISSN 1090-3038, pp.2107–2111, 2005.
- [13] J. Bingham,  
"Multicarrier modulation for data transmission: An idea whose time has come," IEEE Communications Magazine, pp.5–14, May 1990.
- [14] M. Z. W. A. F. Molisch,  
"Mimo systems with antenna selection," IEEE Microwave Magazine, Vol. 5, No. 1, pp.46–56, March 2004.
- [15] A. Gorokhov,  
"Antenna selection algorithms for MEA transmission systems," In IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP, Ed., Vol. 3. IEEE, pp.2857–2860, 2002.
- [16] Y. D. Cho K.,  
"On the general ber expression of one- and two-dimensional amplitude modulations," IEEE Transactions on Communications, Vol. 50, No. 7, pp.1074–1080, July 2002.
- [17] M. Simon and M. Alouini,  
Digital communication over fading channels: a unified approach to performance analysis. John Wiley and Sons, 2002.
- [18] T. M. K. Granino A. Korn,  
Mathematical Handbook for Scientists and Engineers: Definitions, Theorems, and Formulas for Reference and Review. Dover Publications Inc., March 2003.
- [19] C. Arndt,  
Information Measures: Information and its Description in Science and Engineering. No. ISBN 3-540-41633-1, ch. Channel Information, pp.370–373, Berlin, Springer, 2001,
- [20] K. Fall,  
"EECS instructional and electronics support," Online lecture.

# Adaptive sleep scheduling protocol in wireless sensor networks

GERGELY ÖLLÖS, ROLLAND VIDA

Budapest University of Technology and Economics,  
Department of Telecommunications and Media Informatics  
{ollos, vida}@tmit.bme.hu

Keywords: wireless sensor networks, distributed sampling, adaptive regression

**Energy efficiency in wireless sensor networks is a major issue, since the sensors usually have limited and irreplaceable power sources. Sleep scheduling solutions proved to be exceptionally effective strategies to achieve this goal. Numerous such algorithms have been proposed and examined, but virtually without any considerable support for dynamic systems. In this paper we propose and analyze an adaptive, soft-state, fully distributed and robust sleep scheduling method that can easily cope with frequent node failures. The proposed scheme can dynamically eliminate the redundancy and estimate the deficient data based on learned relations in a way to ensure low and balanced energy consumption. This is done without the need for offline pre-computations, dedicated phases, time synchronization, localization, or base station assistance. We compare our technique with deterministic clustering methods, provide parameter sensitivity analysis and discuss the simulation results.**

## 1. Introduction

Distributed sensor networks were in the focus of researchers since the early 1990s. There was a trend to move from centralized, highly reliable, powerful but expensive platforms to a large number of cheap, decentralized and potentially unreliable components that as a group are capable of far more complex tasks than any individual super-node. Wireless sensor networks (WSNs) are formed by one or more base stations (sinks), where the collected data is sent, and a large number of sensors distributed over the monitored area and connected through radio links. Sensors are low-cost and low-power tiny nodes equipped with limited sensing, computing, and radio communication capabilities. They typically have irreplaceable power sources, designed for single usage, and are deployed in an unplanned manner.

There is an essential difference in our terminology, as compared to the usual one, related to cluster definition. By cluster we indicate a subset of entities that could be *potentially* monitored (e.g., a set of coordinates where sensors could be placed), and not a subset of sensor nodes; thus, in our terms, cluster formation mainly depends on the environment and the physical phenomena in which we would like to find the redundancy. The nodes can move over those clusters, which are slowly changing in time. In order to better understand our model, we introduce some basic definitions. Let  $F$  be a set of entities that could be potentially monitored. Then,  $f_i \in F$  is the  $i$ -th cluster, i.e., a subset of  $F$  in which each of the entities can be mutually described based on another arbitrary entity in the same set, within a user specified error bound. Thus, we need to sample only one of the entities in the cluster, and then can estimate any other entity in the same set. When cluster  $f_i$  is monitored using  $k$  nodes, we call it  $k$ -coverage, where the redundancy

is  $1:k$ ; thus,  $k-1$  nodes can be sent to sleep mode. The number and the topology of the clusters  $f$  depend on several factors such as the monitored physical phenomena, the environment, or the error bound. The clusters might also dynamically change in time. If we have two clusters  $f_i$  and  $f_j$ , and we manage in a way to estimate any of the entities in  $f_i$  based on the readings of any entity in  $f_j$ , the two abstract clusters will merge.

When  $\forall f_i \in F$  is monitored by one and only one node  $n_i$ , we call the coverage perfect. This can be achieved only if  $N \geq |f|$  where  $|f|$  is the number of clusters and  $N$  is the number of nodes. Since the structure of the clusters is unknown, we overdeploy the field  $F$  in order to increase the probability  $P(\forall j \exists i: n_i \rightarrow f_j)$ , where  $n_i \rightarrow f_j$  means that node  $n_i$  measures one of the entities in cluster  $f_j$ . Therefore, the global lifetime of the network  $GL$  (i.e., the time until  $\forall j \exists i: n_i \rightarrow f_j$  holds) can be easily computed:

$$GL = \min_{\forall j} \left\{ \sum_{i=1}^{|M_j|} \frac{B_{M_j^i}}{C} \right\} \quad (1.1)$$

where  $M_j$  is a set of nodes that measure entities in cluster  $f_j$ ,  $M_j^i$  is the  $i$ -th node's index in this cluster,  $B_i$  is the battery capacity of node  $n_i$  (in [mA/h]), and  $C$  is the power consumption of a particular node in [mA] (i.e., we suppose that all of the nodes are similar and have the same power requirements).

In this paper we propose thus a dynamic sleep scheduling protocol that aims at maximizing global network lifetime while ensuring that all clusters are monitored by at least one awake node all the time.

This paper is organized as follows. Section 2 discusses the related work, while in Section 3 we show a short real-world case study to emphasize the linear association between temperature measurements; based on the described properties, we propose a simple model for

the analysis that is done in the next sections. In Section 4 we describe and analyze the adaptive regression method, which is the main component of the Adaptive Sampling Protocol (ASP) we propose. Then, in Section 5 we present the complete, fully distributed network level solution. In Section 6 we show the simulation results, define the deterministic clustering to which we compare our solution, and investigate the power balancing capabilities of the protocol. Finally, in Section 7 we conclude the paper.

## 2. Related work

The early papers on energy efficiency were discussing fault tolerance [2] or energy-efficient routing [3], but sleep scheduling, i.e., sparing the energy of the network by placing a subset of nodes into sleeping mode, is a relatively new approach [4]. It is true that sleep based protocols are common in wireless networks generally, as battery energy can be significantly preserved if the mobile device is in sleep mode. However, sleep scheduling in wireless sensor networks is a much more sophisticated problem. Sensors are not standalone devices, they are responsible together for the monitoring task. Therefore, a sleep scheduling protocol should enable sensors to take turns in sleeping and preserve their energy while ensuring however, that the monitoring quality is not affected. In the last few years, many papers discussed a wide range of sleep-scheduling solutions. For instance [5] discussed localized sleeping algorithms based on distributed detection for differential surveillance, [4] discussed system issues and focused on prototyping, while [6] focused on the detection of rare events. However, all of these solutions are based on static, and not adaptive methods. In this paper we discuss an application layer approach, as opposed to many other methods and protocols that achieve higher energy efficiency operating on lower layers, like the MAC – medium access control layer [7-9]. Similar ideas to our method are also explored in [10].

In [11] authors proposed a similar coverage-preserving node-scheduling scheme which can reduce energy consumption and therefore increase system lifetime by turning off some redundant nodes. They presented a basic model for coverage-based off-duty eligibility rule and then extended it to several different scenarios. Each node in the network autonomously and periodically makes decisions on whether to turn itself on or off, using only local neighbor information. To preserve sensing coverage, a node decides to turn itself off when it discovers that its neighbors (sponsors) can help in monitoring its whole working area. To avoid blind points, which may appear when two neighboring nodes expect each other's sponsoring, a backoff-based scheme is introduced to let each node delay its decision with a random period of time. This method assumes however that nodes know their position and sensing range, which in addition is circular and has the same radius for all nodes. Further, this

method can not fully exploit the linear correlations between the measurements and it is not able either to balance the available power levels in the network.

In [12] authors proposed a scheme in which the lifetime of a sensor node is divided into epochs. For each epoch, the base station computes a minimum set of active nodes, based on the current level of coverage requirement, i.e., each sensor samples the field only if it is chosen by the base station to do so. In [13] the authors' approach has two phases. The first one is the development of models (offline) for predicting the measurements of one sensor using data from other sensors. The second is the creation of the maximal number of subgroups of disjoint nodes so that for each such subgroup the measured data is sufficient to recover the measurements of the entire sensor network. For prediction of the sensor measurements, the authors introduced a new optimal non-parametric polynomial time isotonic regression. To capture the evolving dynamics of the instrumented environment, they monitor the prediction errors occasionally to trigger adaptation of the models.

These schemes usually assume a static level of coverage, but even if adaptivity is ensured, either the parameters of the adaptive model are computed offline, or the adaptive algorithm is controlled by a central base station, or the used model is too restricted. There are three main disadvantages to an adaptive, but centralized approach. First, there is a significant communication overhead. Second, the response time to dynamic events might be unacceptably high. Third, if the base station is temporarily unavailable, the sleep scheduling on the whole network is disrupted; as a result, the sensor network cannot continue to function efficiently.

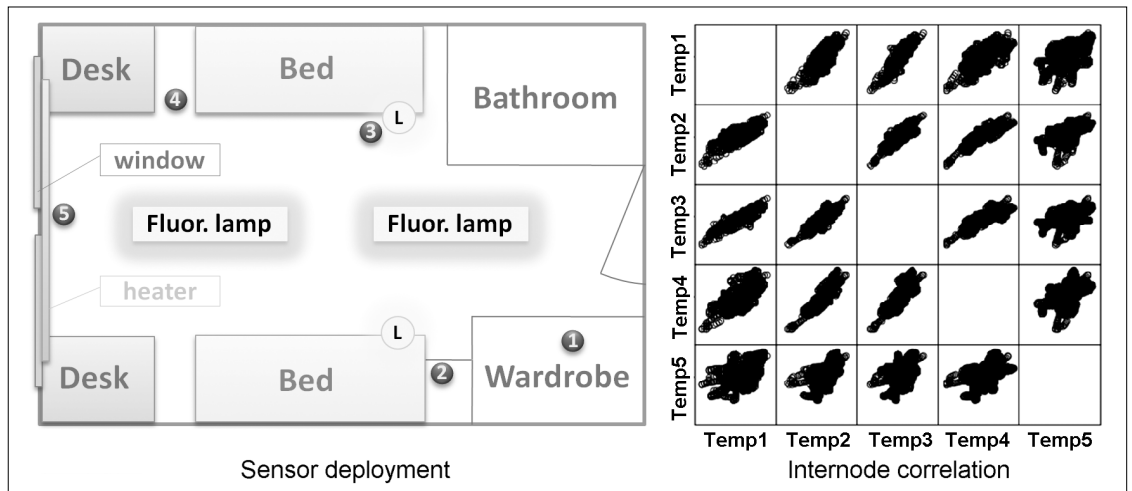
Our proposed method differs from existing works, since the adaptation to dynamic events is online and continuous; there is no need for dedicated phases, nor for base station assistance, since the measurements of sleeping nodes are approximated locally. Also, our proposed ASP protocol can support mobility, it is fully distributed, can be gradually enabled on the network, does not need position information, the model has no restriction on sensing range and finally it is a robust solution in terms of node failures.

## 3. Short case study

In this section we will shortly describe the correlation and statistical properties of temperature and luminosity samples, in order to support our model.

In a dormitory room we placed five sensors that measured temperature and luminosity for three days. The room residents were living their normal daily life without any interruption or alteration. We used identical sensors, deployed as shown by the numbered points in *Fig. 1 (left)*. We can see that node  $n_5$  was close to the heater and to the windows, it had therefore the biggest temperature interferences. On the same figure we can see the ceiling floor lamps and the reading lamps (marked with  $L$ )

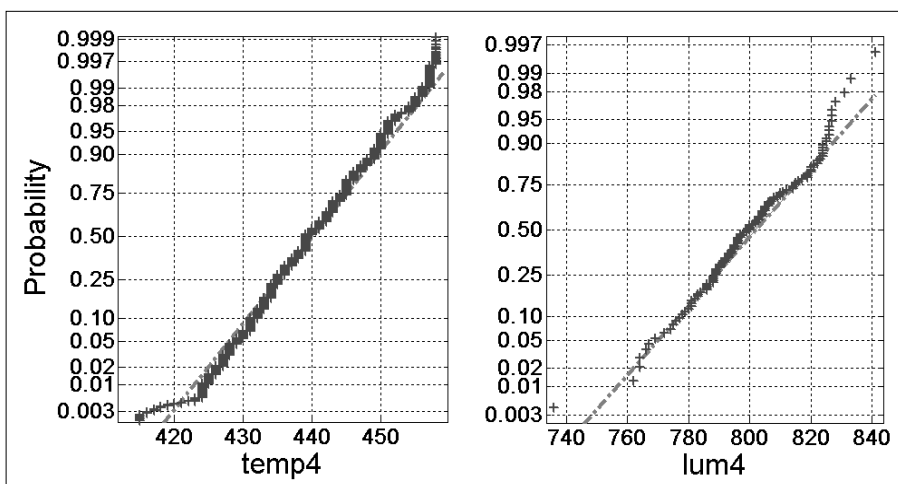
Figure 1. Deployment (left) and internode correlation (right)



as well. We can thus see that node  $n_3$  had the largest light interference caused by a reading lamp. Naturally, the room residents caused additional interferences. This arrangement ensured real-world measurements. The used nodes were Crossbow MICA-Z motes [14], running the Tiny OS [15] operating system with Zigbee stack. Motes were equipped with an ISM radio transceiver (2.4 GHz IEEE 802.15.4) with a maximum data rate of 250 kbps, and had 4 Kbytes of internal memory. The sensor and data acquisition card plugged into the processor radio board collected light and temperature measurements with a one minute sampling rate. In short, we collected temperature and light samples from five different sensors over three days, sampled every minute without interruption.

In Fig. 1 (right) we see the correlations between measurements, represented on specific graphs for each pair of nodes. For instance, let's take the graph that describes the relations between temperature readings of nodes  $n_3$  (Temp3) and  $n_2$  (Temp2). For any given time  $t$  there is a corresponding point on the graph, with the readings on node  $n_3$  represented on the x axis, and the readings of node  $n_2$  on the y axis. As it can be seen, every graph has its inverse, which does not hold extra information, because the sub-graphs are symmetric, i.e. [Temp3, Temp2]

Figure 2. Two typical normal probability plots of temperature and luminosity samples



is the exact inverse of [Temp2, Temp3]. We can see that node  $n_2$  is well correlated with the measurements of node  $n_3$ , because they are close to each other and are far from the windows. By contrast  $n_3$  and  $n_4$  are closer to each other, but  $n_4$  is close to the window as well, and it is thus exposed to significant temperature disturbances (the room residents often ventilated the room). In our measurements a significant spatial correlation can be observed, which we exploit (among others) in the Adaptive Regression procedure our protocol builds on (details will be given later).

Fig. 2 presents the sampled temperature (the Mote's ADC output can be converted to degrees using the Steinhart-Hart equation) and luminosity data; superimposed on the plot is a line joining the first and third quartiles of the samples. This line is extrapolated out to the ends of the sample to help evaluate the linearity of the data. The purpose of a normal probability plot is to graphically assess whether the data could come from a normal distribution or not. If the data are normal, the plot will be linear, while other distribution types will introduce curvatures in the plot. We can see that our samples approximately follow a normal distribution; thus, we assume normality of samples originating from short sampling intervals (window size of several minutes).

Since the strong linear correlation in an over-deployed sensor network is typical, especially when monitoring temperature, humidity, light, etc., we have chosen a simple linear regression model. Another reason is the well known fact that if the samples are normally distributed then the relation between the measurements could only be linear.

We will generate a number of artificial samples for two nodes of different (linear) correlation structure, in order to observe and compare the adaptation response and behavior. The first node's ( $n_i$ ) measurements  $X \in N(\mu_2, \sigma_2)$  are modeled by normal distribution. The sec-

ond node's ( $n_j$ ) measurements  $Y$  are modeled through a linear relation, as follows. Let  $Z \in N(\mu_1, \sigma_1)$  then:

$$Y = (1 - k)(aX + b) + kZ \quad (3.1)$$

where  $a, b$  are the parameters that affect the linear relation, while  $k$  affects the strength of the relation between  $X$  and  $Y$ . If (3.1) is true then:

$$Y \in N \left( a(1 - k)\mu_2 + k\mu_1 + b(1 - k); \sqrt{a^2(1 - k)^2\sigma_2^2 + k^2\sigma_1^2} \right) \quad (3.2)$$

The  $\alpha$  and  $\beta$  parameters of the linear regression ( $Y \approx aX + \beta$ ) are as follows:

$$\begin{aligned} \alpha &= \frac{\text{cov}(X, (1 - k)(aX + b) + kZ)}{\sigma^2(X)} \\ &= \frac{\text{cov}(X, (1 - k)(aX + b)) + \text{cov}(X, kZ)}{\sigma^2(X)} \\ &= \frac{\text{cov}(X, aX(1 - k)) + \text{cov}(X, kZ)}{\sigma^2(X)} \\ &= \frac{a(1 - k)\text{cov}(X, X) + k\text{cov}(X, Z)}{\sigma^2(X)} \\ &= \frac{a(1 - k)\sigma^2(X)}{\sigma^2(X)} = a(1 - k) \end{aligned} \quad (3.3)$$

$$\begin{aligned} \beta &= E(Y) - \alpha E(X) \\ &= E((1 - k)(aX + b) + kZ) - a(1 - k)E(X) \\ &= E(X(a - Ka) + b - kb) - a(1 - k)E(X) \\ &= a(1 - k)E(X) + b(1 - k) - a(1 - k)E(X) \\ &= b(1 - k) \end{aligned} \quad (3.4)$$

Further the Pearson product-moment correlation coefficient for this model is:

$$\begin{aligned} r &= \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)} = \frac{\text{cov}(X, (1 - k)(aX + b) + kZ)}{\sigma(X)\sigma(Y)} \\ &= \frac{a(1 - k)\sigma^2(X)}{\sigma(X)\sigma(Y)} = a(1 - k) \frac{\sigma(X)}{\sigma(Y)} \\ &= \frac{a(1 - k)\sigma_2}{\sqrt{a^2(1 - k)^2\sigma_2^2 + k^2\sigma_1^2}} \end{aligned} \quad (3.5)$$

This coefficient is zero (no relationship) if  $k=1$  and there is an exact linear relationship ( $Y=aX+b$ ) if  $k=0$ :

$$r_{k=1} = 0, \quad \sigma_1, \sigma_2 \neq 0 \quad (3.6)$$

$$r_{k=0} = \begin{cases} 1, & a > 0 \\ 0, & a = 0 \\ -1, & a < 0 \end{cases} \quad \sigma_1, \sigma_2 \neq 0 \quad (3.7)$$

This model will be assumed throughout the analytical evaluations of the proposed protocol (or for sample generation during simulations if not stated differently).

#### 4. The adaptive regression method

In this section we describe the main component of our Adaptive Sampling Protocol (ASP), the adaptive regression (AR) method. This method is used to track other

nodes in the network and interpolate the measurements of those nodes if needed. Thus, the AR method deals only with two nodes: the local node that executes the AR method and a distant node being tracked by the local node.

The adaptation, i.e., the adaptive regression method is simple. The main idea is as follows. In each iteration (when a sample arrives from the distant node) the local node samples the environment as well and pushes the sample pair (local and distant measurement) into a sample FIFO buffer. We use this buffer for estimating the linear regression parameters and the expected error. For each monitored neighboring node, the local node has separate FIFO buffers, and each newly received distant sample is pushed into the proper buffer, along with the latest local measurement. Then, the node recalculates the parameters of the linear regression when needed.

The length of the FIFO in an ideal situation is 2 since this is enough to determine the linear relation. However, in real world measurements there is a significant noise; therefore, we need to have more than two samples (typically 10-30). The optimal number depends on the amount of noise present (lower bound) and on how fast the correlation structure changes (upper bound). Basically, the length of the FIFO should be determined empirically; however, a 20 unit long FIFO is a good trade-off between correlation detection time and noise immunity in most cases regarding temperature or humidity measurements.

The method can dynamically determine if two different clusters  $f_i$  and  $f_j$  can be merged together (if there is a strong linear relation), switch off the redundant node and therefore prolong the global lifetime  $GL$  of the network. In the beginning, we assume that  $\forall j \exists ! i: n_i \rightarrow f_j$ ; the coverage is perfect,  $N \geq |f|$  is satisfied. Let  $x[t]$  be a sample from one of the entities in cluster  $f_i$  (a realization of  $X$ ), sampled by node  $n_i$  at moment  $t$ . Similarly, let  $y[t]$  be a sample from one of the entities in cluster  $f_j$  (a realization of  $Y$ ), sampled by node  $n_j$  at the same moment. Two clusters  $f_i$  and  $f_j$  can be merged at moment  $t_k$ , for a period  $t_p$ , if  $\exists a[t_k], b[t_k]$  so that:

$$\frac{1}{t_p} \sum_{t=t_k}^{t_k+t_p} \{y[t] - a[t_k]x[t] - b[t_k]\}^2 \leq U_{err} \quad (4.1)$$

where  $U_{err}$  is the user specified mean square error (MSE). Naturally, we have to know the MSE of our model before we send nodes  $n_i$  or  $n_j$  to sleep mode for a time interval  $t_p$ . We continuously estimate the mean square error of our model, and if (4.1) is satisfied, we presume that the process is stationary for another time interval  $t_p$ . Then, we send one of the nodes to sleep mode for  $t_p$ , while the awakened node will regress the sleeping node's measurements (based on the estimated regression line) and send them to the sink, on behalf of the sleeping node. The parameter estimation in case of linear regression [16, 17] is well known, so we only summarize the equations.

Let  $\{x[t_k], y[t_k], \{x[t_k+1], y[t_k+1]\}, \dots, \{x[t_k+t_p], y[t_k+t_p]\}$  be the discrete samples from clusters  $f_i$  and  $f_j$ , sampled by nodes  $n_i$  and  $n_j$ .



$$\text{If } a[t_k] := \frac{S_{xy}}{S_x^2} \approx a = \frac{\text{cov}(X, Y)}{\sigma^2(X)} \quad (4.2)$$

$$\text{and } b[t_k] := M(\underline{y}) - a[t_k]M(\underline{x}) \approx b \\ = E(Y) - \frac{\text{cov}(X, Y)}{\sigma^2(X)} E(X) \quad (4.3)$$

the sum specified in (4.1) will be minimal; thus, the linear model is optimally set. In our algorithm we are continuously pushing the  $[x, y]$  pairs to a FIFO queue, and with each new learning pair we update the latest  $a, b, a_{inv}, b_{inv}$  parameters. For all of the monitored clusters  $f$ , node  $n_i$  can have separate and independent FIFO queues (as in the ASP protocol described later).

The sum specified in (4.1) is the estimation of the expected error. For simplicity let  $\{x[t_k], y[t_k], x[t_{k+1}], y[t_{k+1}], \dots, x[t_k+t_p], y[t_k+t_p]\}$  be the discrete samples from clusters  $f_i$  and  $f_j$ , sampled by nodes  $n_i$  and  $n_j$ .

$$\frac{1}{t_p} \sum_{t=t_k}^{t_k+t_p} \{y[t] - a[t_k]x[t] - b[t_k]\}^2 \\ \approx E[(Y - aX - b)^2] \quad (4.4)$$

**Lemma 1:** If (4.2) and (4.3) is true then

$$E[Y - aX - b] = 0 \quad (4.5)$$

**Proof:**

$$E[Y - aX - b] = \\ = E\left[Y - \frac{\text{cov}(X, Y)}{\sigma^2(X)} X - E[Y] + \frac{\text{cov}(X, Y)}{\sigma^2(X)} E[X]\right] \\ = E[Y] - \frac{\text{cov}(X, Y)}{\sigma^2(X)} E[X] - E[Y] + \frac{\text{cov}(X, Y)}{\sigma^2(X)} E[X] = 0 \quad (4.6)$$

The expected error calculated by the adaptive regression method is:

$$E[(Y - aX - b)^2] \\ = E^2[Y - aX - b] + \sigma^2(Y - aX - b) \\ = \sigma^2(Y - aX - b) = \sigma^2(Y - aX) \\ = \sigma^2(Y) + a^2\sigma^2(X) - 2acov(X, Y) \\ = \sigma^2(Y) + \frac{\text{cov}^2(X, Y)}{\sigma^2(X)} - 2\frac{\text{cov}^2(X, Y)}{\sigma^2(X)} \\ = \sigma^2(Y) - \frac{\text{cov}^2(X, Y)}{\sigma^2(X)} \quad (4.7)$$

As we pointed out earlier, if the expected error is lower than the error specified by the user, one of the nodes goes to sleep mode, depending on which node has less energy remaining; this will ensure proper power balancing, which extends the GL (1.1) global lifetime of the network.

## 5. The Adaptive Sampling Protocol

Each sensor node that executes the Adaptive Sampling Protocol (ASP) operates in three phases: adaptation, bargaining, and interpolation. Moreover, the bargaining phase has three steps: interpolation request, interpolation response, and election. The protocol is composed of (and can be well described by) three distinct extended finite state machine (EFSM) models. However, because of space limitations, we focus only on the overall behavior of the ASP protocol. Our solution exploits a potential that we pointed out in Section 3, namely that the nodes close to each other are usually well correlated. We also assume that these close by nodes can hear each other as well.

In Fig. 3.1 we can see the adaptation phase. We will focus on node  $X$  and its neighboring nodes  $Y_n$ . Each node in the adaptation phase grabs packets from its surroundings. In this example nodes  $Y_1, Y_2, Y_3$  can receive the packets transmitted by node  $X$  and execute a copy of the adaptive regression method for node  $X$ . On the other hand, let's say that node  $Z$  receives the packets as well, but it does not monitor node  $X$ , since it does not have enough resources (the maximum number of monitored nodes has been already reached). The unlabeled nodes can't receive the packets of node  $X$ , since they are too far away. With each new sample that  $X$  sends to the base station, the  $Y$  nodes adapt their model as described in the previous sections.

In Fig. 3.2 we can see the interpolation request step of the bargaining phase. In order to decrease the communication overhead, the interpolation request is implicit; it is thus embedded into the packet which carries the latest measurement. The nodes are sending interpolation requests by uniform distribution. The probability of sending the request could be increased based on the available power levels, gradually enabling there-

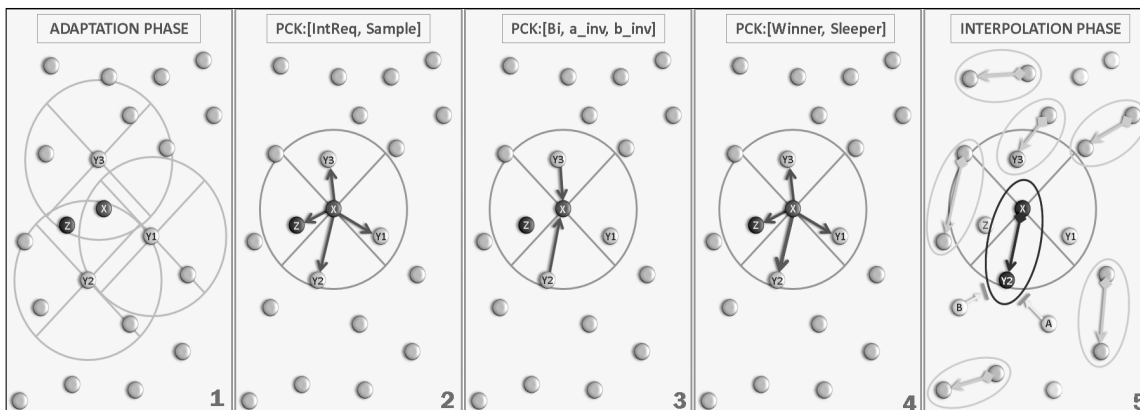


Figure 3. Adaptation (1), bargaining – interpolation request (2), interpolation response (3), election (4) – interpolation (5) phases

fore the sleep scheduling in the network. Since the interpolation request is included in a packet that carries the samples, only those nodes can hear the request which have heard the samples and potentially adapted their model to node  $X$ . An interpolation request does not generate any unnecessary overhead (since it is embedded), because if some node replies to it, an interpolation phase will certainly follow (i.e., one of the nodes will certainly sleep).

Since the energy spared during sleeping is much more than the energy spent for replying the request (maximum each neighbor can send one single reply, where the number of neighbors is typically below five to ten), increasing the probability to send a request results in more energy savings. When this probability is 1 (constantly sending the request), then the ASP protocol spares as much energy as the user specified quality threshold and the correlation structure of the measurements permits.

In Fig. 3.3 we can see the interpolation response step of the bargaining phase. Only those nodes (from the vicinity of node  $X$ ) answer the interpolation request whose model's expected quality of regression is below the user specified threshold of expected quality  $T_{eq}$ . Here we extended the adaptive regression method with a quality definition. The expected quality is a weighted sum of the expected error (as described in Section 4) and the average age of the sample pairs in the sample FIFO queue, based on which we calculate the parameters of the linear regression. In distributed environments this extension is necessary since we have more than two nodes in the network and they can constantly move as well. It can easily happen that the monitored node  $X$  moves out of the receiving range of nodes  $Y_n$ , and then it moves back after a while; this would cause the samples to be fragmented in time in the sample buffer of nodes  $Y_n$ . In other words, the samples in the FIFO queue will either be too old or current ones, but nothing between them; the samples will not be uniformly distributed, which distorts the regression. We can detect and control the effects of this distortion since the average age of the samples is increasing with the dispersion of the age of sample pairs.

In this example, let say that the dispersion of the age of sample pairs in the buffer of node  $Y_1$  is too high; thus, the expected quality of regression is high as well, which results in node  $Y_1$  not answering the request of node  $X$ . Please note that the expected quality metric is inverted (the lower the expected quality, the better the extrapolation). Since node  $Z$  did not monitor the measurements of node  $X$ , it does not answer the request either. Nodes  $Y_2$  and  $Y_3$  answer the implicit interpolation request of node  $X$  with an interpolation response that carries the particular node's actual power level and its inverse regression parameters.

In Fig. 3.4 we can see the election step of the bargaining phase. After node  $X$  received the actual power level and inverse regression parameters, of the candidate nodes ( $Y$ ), it selects the node with the minimal po-

wer level (in our example  $Y_2$ ). At this point, we determined which pair to involve in the regression phase. However, node  $X$  has also to decide which of the two nodes ( $X$  or  $Y_2$ ) will go to sleep. The decision is simple: the node with less energy remaining. After the decisions, node  $X$  informs the winner candidate, and the selected node goes to sleep mode for a predetermined time interval  $T_p$ .

In Fig. 3.5 we can see the interpolation phase. In our example node  $Y_2$  goes to sleep for a predetermined interval  $T_p$ , and during that period node  $X$  interpolates its samples and sends them to the base station on behalf of node  $Y_2$ . In order to interpolate the measurements of node  $Y_2$ , node  $X$  uses the inverse regression parameters that  $Y_2$  has sent in the interpolation response step of the bargaining phase. Nodes  $X$  and  $Y_2$  are in interpolation and sleeping mode, respectively. During this state, sleep request from other nodes (in this example nodes  $A$  and  $B$ ) are ignored.

Since the protocol is soft state and fully distributed, it can handle frequent node failures as well. If a node fails during the adaptation or interpolation phase, this is equivalent with the situation when the failed node doesn't have a good model for interpolating the requesting node's measurements (i.e., it does not answer the request in any way). If a node fails in any other phase, the worst case scenario (when the node doing the interpolation fails) is that for a single  $T_p$  time interval (which is measured in seconds) we lose the measurements of that cluster (two nodes). After that, the protocol naturally recovers from the failure through the next interpolation request.

The ASP protocol takes into account node power levels, in order to ensure proper power balancing, as well as the expected quality of the interpolation. Since the expected quality, as a statistical measure, is much less reliable, and the power balancing is an important task (to extend the global lifetime of the network), the protocol ensures the node that goes to sleep is always the one which has the less energy remaining, in the vicinity of node  $X$  (including node  $X$  itself). In the same time, only those nodes will answer the interpolation request of node  $X$  for which the expected quality (composed of expected error, as described earlier, and the average age of the samples in the buffer) is less than a user specified expected quality bound. This allows us to influence the interpolation error. Furthermore, the sleep scheduling protocol can be gradually enabled on the network, as described earlier.

In the ASP the expected quality of the regression is calculated as a weighted sum of the expected error and the average age of the samples. Since the expected error is not symmetric, we have to calculate the expected quality on both sides of the regression. The average age of samples is symmetric, since we are calculating the average age of learning points from the same buffer, for each side. The symmetry of the average age of samples is straightforward; however, the asymmetry of the expected error needs a minor explanation.

**Lemma 2:** If the standard deviations of the measurements of two nodes  $X, Y$  are finite, nonzero, different, and

the Pearson product-moment correlation coefficient is not  $\pm 1$ , then the expected error (as we defined it in Section 3) is not symmetric (which is usually the case):

$$E[(Y - aX - b)^2] \neq E[(X - mY - n)^2] \quad (5.1)$$

**Proof:**

$$E[(Y - aX - b)^2] = E[(X - mY - n)^2] \quad (5.2)$$

$$\sigma^2(Y) - \frac{cov^2(X, Y)}{\sigma^2(X)} = \sigma^2(X) - \frac{cov^2(X, Y)}{\sigma^2(Y)} \quad (5.3)$$

$$cov^2(X, Y) \left( \frac{1}{\sigma^2(Y)} - \frac{1}{\sigma^2(X)} \right) = \sigma^2(X) - \sigma^2(Y) \quad (5.4)$$

$$\frac{cov^2(X, Y)}{\sigma^2(X)\sigma^2(Y)} = 1 \quad (5.5)$$

$$R^2(X, Y) = 1 \quad (5.6)$$

$$R(X, Y) = \pm 1 \quad (5.7)$$

This is the reason why we need to send the  $a_{inv}$  and  $b_{inv}$  parameters in the interpolation response phase along with the battery status.

## 6. Simulation results

In this section we analyze and compare the Adaptive Sampling Protocol with the deterministic clustering approach. First we discuss the sample generation process and its statistical properties. Then, we introduce the reference model to which we compare the proposed method throughout this paper. We provide then performance and overhead analysis, and finally we discuss the power balancing property of our protocol, which significantly extends network lifetime.

### A) Sample generation and analysis

We measured the properties of the samples in the following manner. First we generated 25x10000 samples for 25 nodes (for each node 10000 samples). The sampling frequency of the nodes was 1Hz and they were able to execute 100 logical operations in 1 second (100 state transitions per second in the EFSMs). Thus, for 100 seconds of simulation time, we needed 100x100=10000 samples, where each sample represents 10 ms holdup in time. The typical sample's buffer length was 20, there was thus a 20-entries long FIFO in which we shifted the samples (pushed one to the top, and discarded one from the bottom of the FIFO, in each second), while we continuously computed the statistical properties of the samples in the FIFO. We have randomly chosen two nodes, and plotted the result in Fig. 4 as illustration.

As we can see, the covariance, the dispersion and the expected error on both nodes are continuously changing (as we discussed in Section 3). The range and profile of the curves are similar between each pair of measurements, but the maximum and the inflection points are differently situated. What is typical for each pair is the decreasing determination index (or coefficient). The reason why we generated such samples needs a short explanation. The variation of the samples is made up of

two parts: the part that can be explained by the regression equation (this is the determination index) and the part that can't be explained by the regression. The determination index can have many different definitions, depending on the class of problem. In our case (linear regression) the determination index (or coefficient) is exactly the square of the Pearson product-moment correlation coefficient.

According to **Lemma 1** and the definition of the expected error (in Section 4):

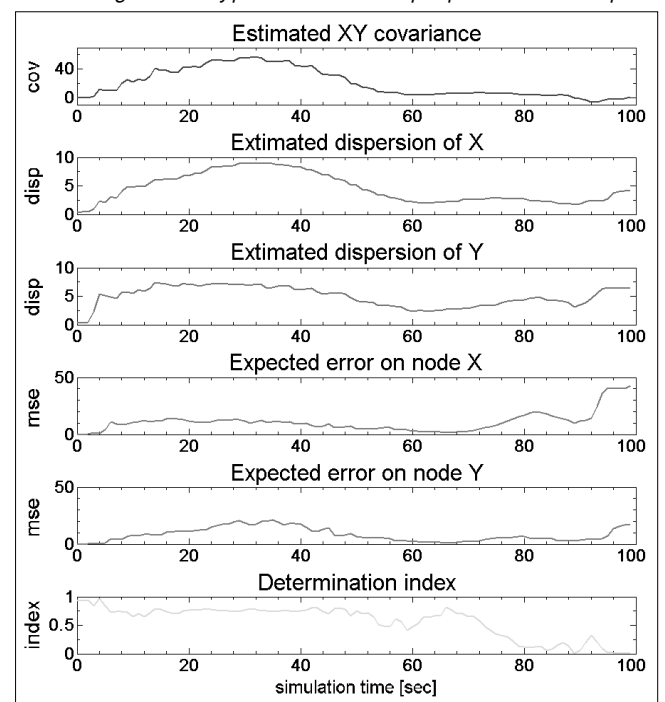
$$E[(Y - aX - b)^2] = \sigma^2(Y) - \frac{cov^2(X, Y)}{\sigma^2(X)} \quad (6.1)$$

Since the regression function  $f(X)$  in our case is linear  $f(X) = aX + b$ :

$$\begin{aligned} R_f^2 &= 1 - \frac{E[(Y - f(X))^2]}{E[(Y - E(Y))^2]} = 1 - \frac{E[(Y - aX - b)^2]}{\sigma^2(Y)} \\ &= 1 - \frac{\sigma^2(Y) - \frac{cov^2(X, Y)}{\sigma^2(X)}}{\sigma^2(Y)} \\ &= \frac{\sigma^2(Y) - \sigma^2(Y) + \frac{cov^2(X, Y)}{\sigma^2(X)}}{\sigma^2(Y)} \\ &= \frac{cov^2(X, Y)}{\sigma^2(X)\sigma^2(Y)} = R^2(X, Y) \end{aligned} \quad (6.2)$$

Therefore, if the determination index is 1, there is no introduced error by the adaptive regression (there is an exact linear relation between the measurements of the local and the distant node). As the determination index is decreasing the performance of the adaptive regression (AR) is declining as well. If the determination index is zero then the linear regression can't explain the variance of the extrapolated node by definition. If the samples are

Figure 4. Typical statistical properties of samples



normally distributed (as we presumed and demonstrated in our scenario) they are independent as well.

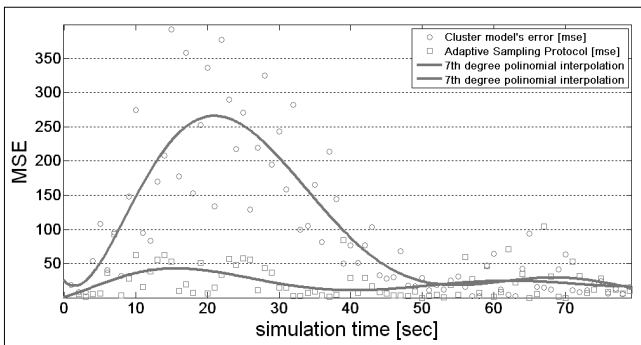
**B) Deterministic clustering**

The deterministic clustering approach is (implicitly) widely used in existing sleep scheduling protocols. The main idea is that if we form the clusters in an over-deployed network based on topological distances (the nodes that are close to each other form a cluster), then with the measurements of one node we can approximate the other nodes in the same cluster (they are measuring approximately the same values, as discussed in Section 3). We select a cluster head which measures the environment, and the rest of the nodes are going to sleep mode until the cycle ends. In each cycle a new node (from the same cluster) assumes the role of the cluster head, this node being chosen in a deterministic or probabilistic manner. These protocols are however unaware of the current measurements, and the clustering is static; they do not support dynamic environments, where the correlations between nodes are changing.

Our reference model (referred as deterministic clustering) is as follows. We divide the network into two-node clusters (if there is an odd number of nodes, then the last cluster consists of three nodes) in order to be comparable with the ASP protocol which dynamically creates two-node clusters as well. In each cluster one of the nodes is always sleeping, while the awoken node samples the environment and sends the measurement to the sink on behalf of both nodes. In each cycle (sampling period) the nodes assume reverse roles, in order to sustain the network's power dispersion. The error that this model makes is the squared measurement difference between the measurements of the awake and the sleeping node. If we compute the average error in the network for each cluster, we get the mean squared error (MSE) of the deterministic clustering in a given cycle.

In Fig. 5 we can see the comparison between the MSE obtained for the deterministic clustering protocol (not aware of measurements) and for the Adaptive Sampling Protocol, which is aware of the measurements and the correlation structure. The reason for the high MSE values between seconds 10 and 40 is that the dispersion of the measurements between the samples is higher (due for example to node or event mobility). After 60 seconds of simulation, the dispersion is significantly smaller,

Figure 5. Deterministic clustering versus ASP



which means that the nodes are measuring similar values (that's why the blind deterministic clustering performs so well). As you can see, the ASP protocol can adapt to dynamic environments. In the rest of the analysis we will usually compute the average MSE of the protocol per simulation run. In this case, for 80 seconds of simulation (one run) the average error made by ASP is approximately 27, while the average error made by the deterministic clustering scheme is approximately 103.

**C) Performance of the ASP protocol and comparisons**

In this section we compare and analyze the Adaptive Sampling Protocol. First we provide a parameter sensitivity analysis for the length of the sample buffer and the user specified threshold of expected quality; then, we also discuss the protocol's overhead.

In Fig. 6 we can see the protocol's behavior if we change the length of the sample FIFO (queue) buffer, as well as its effect on the protocol overhead. As it can be expected, if we increase the size of the sample buffer, the response time of the ASP protocol (for correlation changes) increases. In other words, if the correlation structure between two monitored nodes changes (relatively) quickly, the expected quality will not decrease below the user specified threshold (TEq) so rapidly; therefore, the interpolation request will be rejected and there will be thus less sleep cycles (in general), as the figure shows. The protocol's overhead is measured in the number of extra packets sent, which is strongly correlated with the number of sleep cycles. In order to avoid redundancy, we will discuss this issue in detail later.

If the number of sleep cycles is decreasing then naturally the power consumption is increasing. Fig. 7 shows

Figure 6. Sample buffer size sensitivity (sleep cycles)

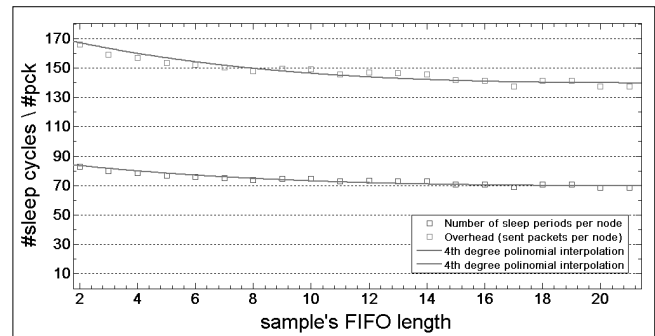
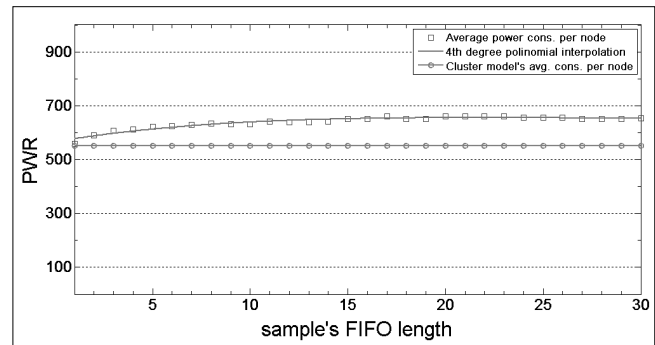


Figure 7. Sample buffer size sensitivity (power consumption)



the average power consumption per node, for both the ASP protocol and the deterministic clustering. We simulated both protocols with the following parameters: there were 25 nodes in the network (5x5 grid), the simulation time was 1000 seconds, and the speed of the EFMSs was 100 ticks/sec. For a sleeping node we chose the power consumption to be 0.001 units/tick and for the awake node 0.01 units/tick. Given the parameters, it is easy to compute the average power consumption of a node for the

Figure 8. *TEq* sensitivity analysis (estimation error)

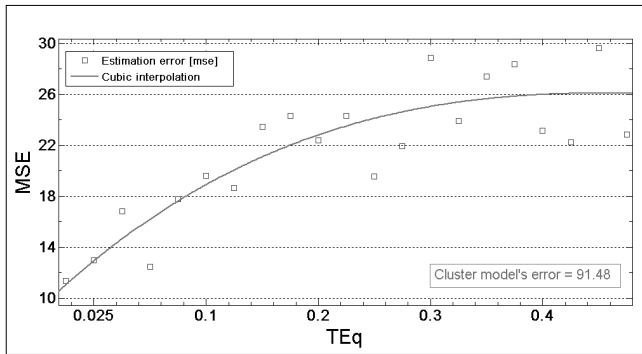


Figure 9. *TEq* sensitivity analysis (power consumption)

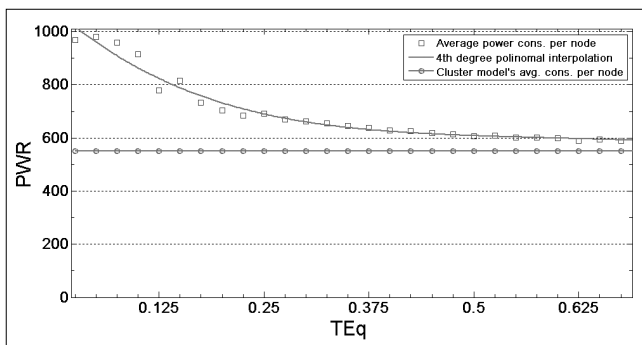


Figure 10. *TEq* sensitivity analysis (sleep cycles)

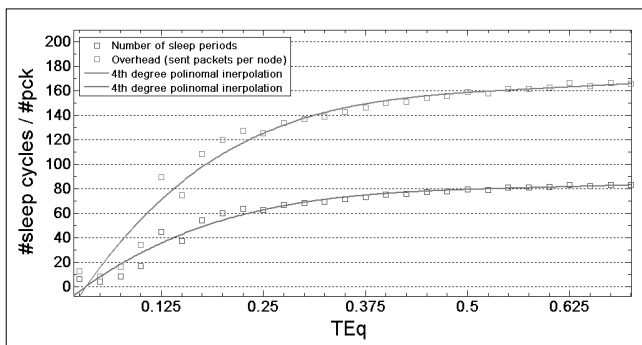
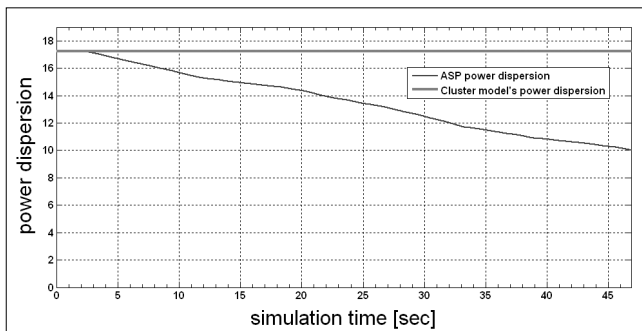


Figure 11. Power dispersion comparison



deterministic clustering: it is 550 units in this scenario ( $550 = 500 \cdot 100 \cdot 0.01 + 500 \cdot 100 \cdot 0.001$ ) since each node is sleeping in half of the time. Given that the deterministic clustering has no sample FIFO, this consumption is independent of the FIFO length; that's why in the figure it appears as a straight line. If we increase the length of the samples FIFO, the adaptation to correlation changes is slowing down; many of the old samples are still in the FIFO, they overweight the new samples, and thus the protocol can't exploit short term correlations. This means that the number of sleep cycles is decreasing (as we can see in Fig. 7) and, therefore the power consumption is slightly increasing.

In Fig. 8 we can see that if we increment the user specified threshold of expected quality, the extrapolation error is increasing as well. The simulation configuration is as we described earlier. There are 25 nodes (in a 5x5 grid arrangement) in the network, and they are not moving. The samples that are fed to the network are as we described in Section 6/A. Each node can track 8 nodes and in the 5x5 grid each node has maximum 8 neighbors. The simulation time is 1000 seconds and the sample buffer length is 15. Like in the previous comparison, in this static environment the ASP protocol outperforms the deterministic clustering roughly 3-4 times regarding the estimation error. Please note that the samples between nodes are virtually not correlated in 30% of the simulation time, given that the typical determination index is decreasing (as we discussed in Section 6/A.).

As we mentioned it earlier, the blind deterministic clustering scheme results in the theoretically minimal energy consumption. Fig. 9 indicates how much does the power consumption of the ASP converge to this minimum. However, as the power consumption is decreasing, the extrapolation error is increasing (Fig. 8).

In Fig. 10 we can see the average number of sleep cycles per node and the average number of sent messages per node, as we change the *TEq* (threshold of expected quality) parameter. As we pointed it out earlier, if the *TEq* parameter increases, the number of sleep cycles increases as well, and thus the power consumption decreases. Before each sleep cycle, there is a three step negotiation, with the first step (interpolation request) being implicit (carried in the packet along with the sample). The remaining two steps result in overhead packets, the average overhead per node (in sent packets) is therefore strongly correlated with the number of sleep cycles, and is increasing as the *TEq* parameter is increasing. The number of overhead packets is approximately equal to the number of sleep cycles times two; however, this relation is strongly varying from node to node, although in average (per node) this is a close estimation, as shown in Fig. 10.

This overhead could slightly increase if the neighborhood of nodes is dense, since in this case more nodes can apply for the competition (send an interpolation response). The increase in node density means that a node will probably have more neighbors. Since the number of nodes that a particular node can track has a fixed upper bound, the number of answers to interpolation requests

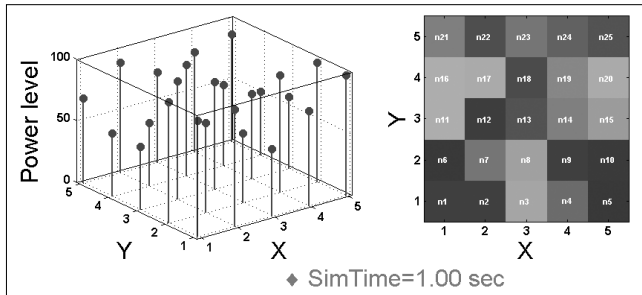


Figure 12. Network power level status in the first second

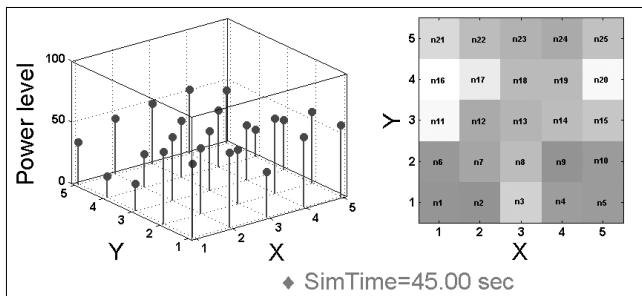


Figure 13. Network power level status after 45 seconds

in the network has an upper bound as well. When we simulated the network's behavior, we have set this parameter to be equal to the number of neighbors, maximizing thus the overhead in the described scenario. This means that if the density of nodes increases with the number of interpolation requests remaining constant, the overhead will not be significantly higher. Of course, this could increase the radio interference on the MAC layer, but because of the distributed nature of the protocol this would not affect significantly the overall behavior of the ASP protocol.

#### D) Power balancing

In this section we discuss the power balancing capabilities of the Adaptive Sampling Protocol.

In Fig. 11 we can see the comparison of power balancing capabilities of the ASP and the deterministic clustering. The deterministic functioning of the cluster based approaches assures the detection time of events. This capability usually infers the constant power dispersion in the network through time, which could significantly decrease the global lifetime of the network. As we can see in Fig. 11 the ASP balances the energy reserves of the network, and increases thus the global lifetime of the network (1.1).

Fig. 12 is a snapshot of energy reserves in the network in the first second of the simulation time. As we mentioned it earlier, the nodes were arranged into a 5x5 grid. During this simulation, each node could monitor maximum 4 nodes and each node had maximum 4 neighbors. The user specified threshold of expected quality was 0.15, the samples buffer size was 20, and the initial power dispersion was random (uniform distribution). As we compare the snapshot from the first second with the energy snapshot after 45 seconds (Fig. 13) of simulation time, we can see that the nodes with higher energy reserves consume more energy than the others.

In other words, the ASP has the ability to logically transfer energy reserves between nodes so as to extend the global lifetime GL (1.1) of the network.

## 7. Conclusion

This paper proposes the Adaptive Sampling Protocol, a fully distributed WSN protocol. Some of the applications of the proposed method are target tracking, environmental monitoring, surveillance, early warning systems, etc.

In ASP the nodes in the network are monitoring each other's measurements, dynamically learn the linear relations among them (if any), eliminate (send to sleep) the redundant nodes, and estimate the deficient data without the need for offline pre-computations, dedicated phases, or base station assistance. There is no need for time synchronization or localization. The algorithm is based on continuous correlation monitoring and estimation, where the extrapolation error can be influenced by a user specified threshold of expected quality. The ASP protocol can be gradually enabled on the network, i.e., from a deterministic functioning, when the detection time is guaranteed, to the fully adaptive mode, when ASP spares as much energy as the correlation patterns and the user specified threshold permit. Another advantage of the ASP protocol is the strong energy balancing capability which could significantly extend the lifetime of the network.

The ASP protocol is designed to support adaptive environments and as the survey [1] indicates, it's a first of its kind. ASP is a robust protocol and can function even if the network has broken up to isolated segments; it can easily cope with frequent node failures as well. Further, the protocol overhead is well correlated with the number of sleep cycles, which can be influenced by the  $TEq$  parameter (as simulations showed). Since the communication is local, the power requirements for the overhead frames are minimal. Also, there are no network level interferences introduced, as opposed to the base station centralized approaches. If the measurements are not correlated, then the ASP protocol switches back to deterministic mode, but only on that part of the network where the linear association is under the threshold. The disadvantage of our protocol is that the power consumption only converges to the theoretical minimum, but never reaches it. Further, one node can monitor multiple distant nodes but can interpolate only one at a time. This means that the protocol can dynamically create only two sized clusters and thus the energy spared is limited to 50%.

The main component of the ASP protocol is the adaptive regression core, which we first discussed separately. We supported our assumptions with a short case study, discussed the properties of the measurements and based on this knowledge, we generated the samples for simulations. The results show that the ASP protocol generally outperforms the clustering approaches and it converges to the theoretically minimal energy consumption. We showed that the power balancing capabilities of the ASP protocol are strong. Furthermore, we showed that

with the user specified threshold of expected quality, the real estimation error can be well influenced.

In the future we will work on a distributed model which can predict various occurrences of discrete events in dynamic environments, based on a fully distributed (neuro) Fuzzy architecture.

**Acknowledgement**

The work described in this paper has been (partially) supported by EC FP7 EFIPSANS project (INFISO-ICT-215549).

**Authors**



**GERGELY ÖLLÖS** is a Ph.D. student at the Budapest University of Technology and Economics. He obtained his MSc degree in Technical Informatics from the same university in 2008. He was project leader at WetCom s.r.o., Bratislava, Slovakia, in 2008 where he designed and built a laser stabilization controller for accurate positioning along with its communication firmware. He worked at Nokia, Komárom, Hungary in 2008 as High Frequency Radio Engineer, P2P and C lecturer. In the last few years he worked on a neural network based navigation system and built a robot for demonstration for which he received diplomas in Prague and Budapest in 2008 and 2007, respectively. Now his interests lie predominately in the area of low-power wireless networking and machine learning..



**ROLLAND VIDA** is Associate Professor at the Budapest University of Technology and Economics. He obtained his BSc and MSc degrees in Computer Science from the Babes Bolyai University, Cluj-Napoca, Romania, in 1996 and 1997 respectively, and his PhD degree from the Université Pierre et Marie Curie, Paris, in 2002. Between 2003 and 2005 he obtained the György Békésy Postdoctoral Fellowship, and in 2007 the János Bolyai Research Fellowship. In the last five years Dr. Vida has acted as organizer, TPC member or reviewer for more than 30 international conferences, participated in several national and European research project, and taught different networking courses at universities in Romania, Slovakia and Hungary. In 2008 he was elected as Chair of International Affairs of the Scientific Association for Infocommunications, Hungary.

**References**

[1] L. Wang and Y. Xiao, "A survey of energy-efficient scheduling mechanisms in sensor networks" In *Mobile Networks and Applications*, 2006.

[2] F. Ye, G. Zhong, S.L.J. Cheng, and L. Zhang, "Peas: A robust energy conserving protocol for long-lived sensor networks," In *Proc. of ICDCS 2003, 23rd IEEE International Conf. on Distributed Computing Systems*, Providence, USA, May 2003, pp.28–37.

[3] D. Ganesan, R. Govindan, S. Shenker, and D. Estrin, "Highly resilient energy-efficient multipath routing in wireless sensor networks," *ACM SIGMOBILE Mobile Computing and Com. Review*, Vol. 5, April 2001, pp.11–25.

[4] C. Hsin and M. Liu, "Network coverage using low duty-cycled sensors: random and coordinated sleep algorithms,"

In *Proc. of IPSN 2004, The 3rd International Symp. on Information Processing in Sensor Networks*, Berkeley, USA, April 2004, pp.433–442.

[5] T. Yan, T. He, and J. Stankovic, "Differential surveillance for sensor networks," In *Proc. of ACM Sensys 2003*, Los Angeles, USA, November 2003, pp.51–62.

[6] Q. Cao, T. Abdelzaher, T. He, and J. Stankovic, "Towards optimal sleep scheduling in sensor networks for rare event detection," In *Proc. of IPSN 2005, The 4th International Symp. on Information Processing in Sensor Networks*, Los Angeles, USA, August 2005, pp.20–27.

[7] W. Ye, J. Heidemann, and D. Estrin, "An energy-efficient mac protocol for wireless sensor networks," In *Proc. of the IEEE INFOCOM*, June 2002.

[8] S. Du, A. K. Saha, and D. B. Johnson, "Rmac: A routing-enhanced duty-cycle mac protocol for wireless sensor networks," In *Proc. of the IEEE INFOCOM*, May 2007.

[9] T. van Dam and K. Langendoen, "Adaptive energy-efficient mac protocol for wireless sensor networks," In *Proc. of ACM SenSys*, Los Angeles, USA, November 2003.

[10] F. Ye, G. Zhong, S. Lu, and L. Zhang, "Conserving protocol for longlived sensor networks," In *Proc. of the 23th IEEE International Conf. on Distributed Computing Systems (ICDCS)*, Providence, Rhode Island, USA, May 2003.

[11] D. Tian and N. D. Georganas, "A coverage-preserving node scheduling scheme for large wireless sensor networks," In *Proc. of the 1st ACM International Workshop on Wireless Sensor Networks and Applications (WSNA)*, New York, USA, 2002.

[12] H. Ling and T. Znati, "Energy efficient adaptive sensing for dynamic coverage in wireless sensor networks," In *Proc. of WCNC 2009, Budapest*, April 2009.

[13] F. Koushanfar, N. Taft, and M. Potkonjak, "Sleeping coordination for comprehensive sensing using isotonic regression and domatic partitions," In *Proc. of IEEE Infocom 2006*, Barcelona, Spain, April 2006.

[14] Crossbow technology, micaz 2.4ghz mote (Online). Available: <http://www.xbow.com/Products/productdetails.aspx?sid=164> [retr. Okt. 2009]

[15] Open technology alliance, tinyos (Online). Available: <http://www.tinyos.net> [retr. Okt. 2009]

[16] J. Cohen, P. Cohen, S. West, and L. Aiken, *Applied multiple regression and correlation analysis for the behavioral sciences*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 2003.

[17] N. Draper and H. Smith, *Applied Regression Analysis*. Wiley Series in Probability and Statistics, 1998.

# Modeling feeds of antennas by Finite Element Method

MIKLÓS KUCZMANN

Széchenyi István University, Department of Telecommunications, Laboratory of Electromagnetic Field  
kuczmann@sze.hu

Keywords: antenna feed models, Finite Element Method, monopole antenna, antenna measurements

The simulation of the input impedance and the radiation pattern of an antenna is very important in the design stage of antennas when engineers use computer aided design software. Other parameters can be calculated from these data.

The simulated input impedance depends on the applied feed model that is the reason why it is so important to know the behaviors of the different feeding models. The most frequently used models are prescribed in the paper in the frame of the Finite Element Method. The models can be used accurately in other numerical techniques as well.

## 1. Introduction

The most important measured parameters of an antenna are the input impedance and the radiation pattern. Other parameters, for example the reflection coefficient or the voltage standing-wave ratio can be calculated from the input impedance, the directivity as well as the gain can be obtained from the radiation pattern. The simulated input impedance depends on the applied feed model that is the reason why it is so important to know the advantages and the disadvantages of the feeding models. The most frequently used models are the current probe model, the voltage gap generator, the magnetic frill generator and the waveguide port.

This paper presents the above mentioned approaches through a monopole antenna situated above a ground plane. The Finite Element Method (FEM) has been used in the numerical field analysis, which is a widely used technique to solve partial differential equations obtained from Maxwell's equations. Here, the Helmholtz-equation for the magnetic field intensity is studied in two dimensions supposing axial symmetry. First, the problem and the corresponding equations are shown, and then the four feeding models are described. After the presentation of numerical results, a short discussion and summary close the paper.

## 2. Finite Element Method (FEM) in antenna simulation

The FEM is a widely used numerical technique in computer aided design of electrical engineering problems. Only a brief introduction can be provided here, a detailed description can be found in [1-3].

The basis of the technique is the discretization of the problem region by simple elements. These finite elements are the triangle and the quadrangle in two dimensional problems, or the tetrahedral, hexahedral and prism

elements in three dimensional problems. The system of equations to be solved for the potentials or for the field quantities can be assembled after obtaining the weak formulation of the partial differential equations and the boundary conditions of the problem. The latter equations can be derived from Maxwell's equations [1-4].

The problem to be presented here is a monopole antenna situated above a ground plane [1]. The body of the antenna is the inner wire of a coaxial transmission line as it can be seen in Fig. 1. The following Maxwell's equations must be solved in the domain  $\Omega$  [4]:

$$\nabla \times \mathbf{H} = j\omega\epsilon\mathbf{E}, \quad (1)$$

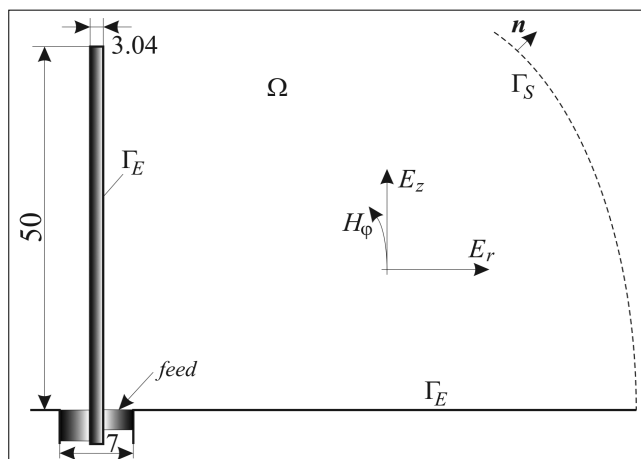
$$\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H}, \quad (2)$$

$$\nabla \cdot \mathbf{H} = 0, \quad (3)$$

$$\nabla \cdot \mathbf{E} = 0, \quad (4)$$

where  $\mathbf{H}$ ,  $\mathbf{E}$ ,  $\omega$ ,  $\epsilon$ , and  $\mu$  are the magnetic field intensity and the electric field intensity, the angular frequency of excitation, the permittivity and the permeability, respectively. The phasor representation has been used, because of the time-harmonic representation (the excitation is supposed to be sinusoidal), i.e.  $j = \sqrt{-1}$  is the imaginary unit.

Figure 1. The geometry of the monopole antenna





It is well known that the electromagnetic field of the monopole antenna is transverse magnetic (TM) [1,2,4], i.e. the magnetic field has only one component in the  $\varphi$ -direction, and the electric field has two orthogonal components, as it is denoted in Fig. 1.

The electric field intensity must be normal to the surface of the ground plane and the surface of the antenna, i.e. the boundary condition

$$\mathbf{E} \times \mathbf{n} = \mathbf{0} \quad (5)$$

can be supposed on  $\Gamma_E$ . Here  $\mathbf{n}$  denotes the outer normal unit vector.

On  $\Gamma_S$ , absorbing boundary condition must be prescribed to absorb the electromagnetic energy [1,2],

$$\lim_{r \rightarrow \infty} r [\nabla \times \mathbf{H} + jk_0 \mathbf{n} \times \mathbf{H}] = \mathbf{0}, \text{ on } \Gamma_S, \quad (6)$$

which can be approximated by the first order absorbing boundary condition

$$\mathbf{n} \times [\nabla \times \mathbf{H} + jk_0 \mathbf{n} \times \mathbf{H}] = \mathbf{0}, \text{ on } \Gamma_S, \quad (7)$$

where  $k_0^2 = \omega^2 \mu \varepsilon$  is the wave number in free space ( $\mu = \mu_0, \varepsilon = \varepsilon_0$ ). This models the unbounded space. The calculation domain must be truncated somehow, because the discretization cannot be performed at infinity, so the condition (7) on  $\Gamma_S$  is available to decrease the domain volume. The efficiency of absorbing the electromagnetic energy along the boundary  $\Gamma_S$  can be increased by applying a perfectly matched layer (PML) which outer boundary has been assigned as the absorbing boundary [1].

Finally,  $H \times n = 0$  must be satisfied along symmetry planes (along the line  $r=0$  in axial symmetry situations).

It is evident that the application of the magnetic field intensity as the primary variable results in the most economic formulation. The partial differential equation to be solved for the magnetic field intensity here is the following [1,4]:

$$\nabla \times \nabla \times \mathbf{H} - k_0^2 \mathbf{H} = \mathbf{0}, \quad (8)$$

and  $E = (\nabla \times \mathbf{H}) / j\omega \varepsilon$  is the electric field intensity from (1) and (2). After some mathematical manipulations and using (3), the following partial differential equation can be obtained for  $H_\varphi$ :

$$\Delta H_\varphi + k_0^2 H_\varphi = 0, \quad (9)$$

which is a scalar Helmholtz-equation of the magnetic field intensity.

### 3. Feeding Models in FEM

The feeding models of antennas are applied to take the input of the antenna into account. The most widely used feeding models are shown in Fig. 2 [1,5-7].

#### 3.1 The current probe model

The most widely used current probe model is a short current with a delta function, e.g.

$$\mathbf{J}(x, y, z) = \mathbf{e}_z I_0 \delta(x - x_f, y - y_f), \quad 0 \leq z \leq d. \quad (10)$$

It models a wire with zero diameter,  $x_f$  and  $y_f$  are the coordinates of the current  $I_0$  ( $x_f = 0$  and  $y_f = 0$  in Fig. 2/a),

and  $J$  has only one component in the  $z$  direction. This infinitesimal dipole can be generalized in any direction of the space. The electromagnetic field is singular in the vicinity of the probe [1]. This is the reason why it is more convenient to prescribe the magnetic field intensity on the surface of the antenna wire as it is represented in Fig. 2/a. The  $\varphi$ -component of the magnetic field intensity can be calculated by

$$H_\varphi = \frac{I_0}{2a\pi}, \quad 0 \leq z \leq d, \quad (11)$$

and  $a = 1.52$  mm is the radius of the antenna. The length of the probe in the  $z$  direction should be as small as possible, but it can be concluded that  $d \ll \lambda$  must be specified, and  $\lambda$  is the wavelength of the electromagnetic wave in vacuum,  $\lambda = c/f$  ( $c$  is the speed of light and  $f$  is the frequency of excitation).

Once the electric field  $E$  is determined by the applied numerical method, the voltage across the probe can be computed as

$$U = -\int_0^d E_z(r = a) dz, \quad (12)$$

and the input impedance of the antenna is

$$Z = U / I_0. \quad (13)$$

The current distribution along the antenna can be calculated by the following form of Ampere's law:

$$I(z) = 2a\pi H_\varphi(z). \quad (14)$$

#### 3.2 The voltage gap generator

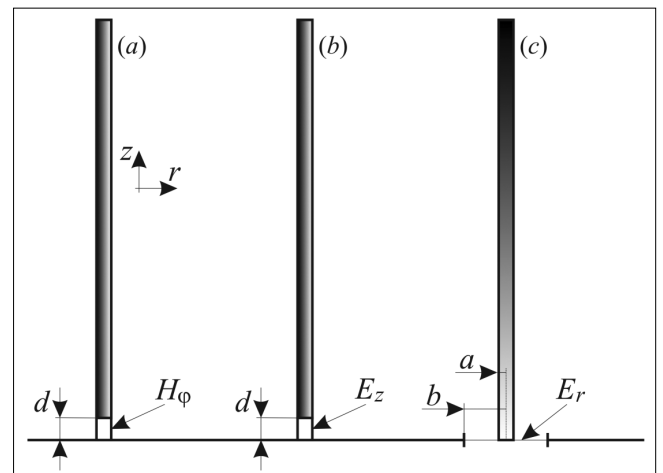
This model is basically used in the Method of Moments (MoM) [5], and it is based on the potential difference  $U_0$  between the antenna and the ground plane as it is presented in Fig. 2/b. The  $z$  component of the electric field can be obtained by

$$E_z = -\frac{U_0}{l} \quad (15)$$

in the gap ( $0 \leq z \leq d$ ), and  $d$  the length of the gap. The electric field intensity is prescribed along the line  $r = a$ , and  $0 \leq z \leq d$  (see in Fig. 2/b).

The current in the feeding point then can be calculated by (14) substituting  $z = 0$ , i.e.  $I = 2a\pi H_\varphi(z = 0)$ , then the input impedance can be obtained by  $Z = U_0 / I$ . The current distribution along the antenna can be simulated by (14).

Figure 2. Feeding models



### 3.3 The magnetic frill generator

The magnetic frill generator is a model of the antenna input fed by a coaxial line [5]. The following electric field intensity can be supposed in the radial direction by assuming purely TEM mode inside the coaxial transmission line (see in Fig. 2/c):

$$E_r(r) = \frac{I[V]}{2r \ln(b/a)}, \quad a \leq r \leq b, \quad (16)$$

where  $a$  and  $b$  are the inner and outer radius of the coaxial line ( $a = 1.52$  mm and  $b = 3.5$  mm). The current distribution along the antenna can also be simulated by (14), and the input impedance can be calculated in the same way as presented in Section 3.2.

### 3.4 The waveguide port

The waveguide port model is more accurate and is a more efficient approach in general case. This is based on the weighted sum of TEM, TE and TM waveguide modes, and the weighting coefficients are collected in tables [1]. This model has been implemented in Comsol Multiphysics [8]. The scattering parameter (reflection coefficient)  $S_{11}$  can be extracted from the simulated electric field, finally, the input impedance can be obtained as [1,8]

$$Z = Z_0 \frac{1 + S_{11}}{1 - S_{11}}, \quad (17)$$

where  $Z_0 = 50 \Omega$  is the characteristic impedance of the waveguide. The reflection coefficient is calculated automatically in Comsol when applying waveguide port.

## 4. Simulation results

The problem has been solved by the functions of the Radio Frequency module of Comsol Multiphysics [8]. This software is a very efficient FEM design environment. The aforementioned feeding models can be implemented and tried out in an easy way. The models can be downloaded from the author's homepage [9].

The  $\varphi$ -component of the magnetic field intensity has been simulated by the TM Electromagnetic Waves application mode, and two dimensional axial symmetry plane has been analyzed for simplicity, because the aim is the study of the different models. Second order Lagrange shape functions have been used to approximate the unknown field quantity.

After some trials, 55296 triangles have been used to mesh the geometry (Fig. 3), and it results in 111329 unknowns. This is a very dense mesh. The convergence of the simulated input impedance can be seen in Fig. 4, where the measured impedance is also shown. Measured data are from the paper [6]. The variation of input impedance is practically the same when applying finer and finer mesh. There is a permanent difference between measured and simulated data.

The geometry of the antenna has been subtracted from the calculation domain, because it is supposed to be made of ideal conducting material, i.e. discretization is not necessary inside the wire. The same mesh has

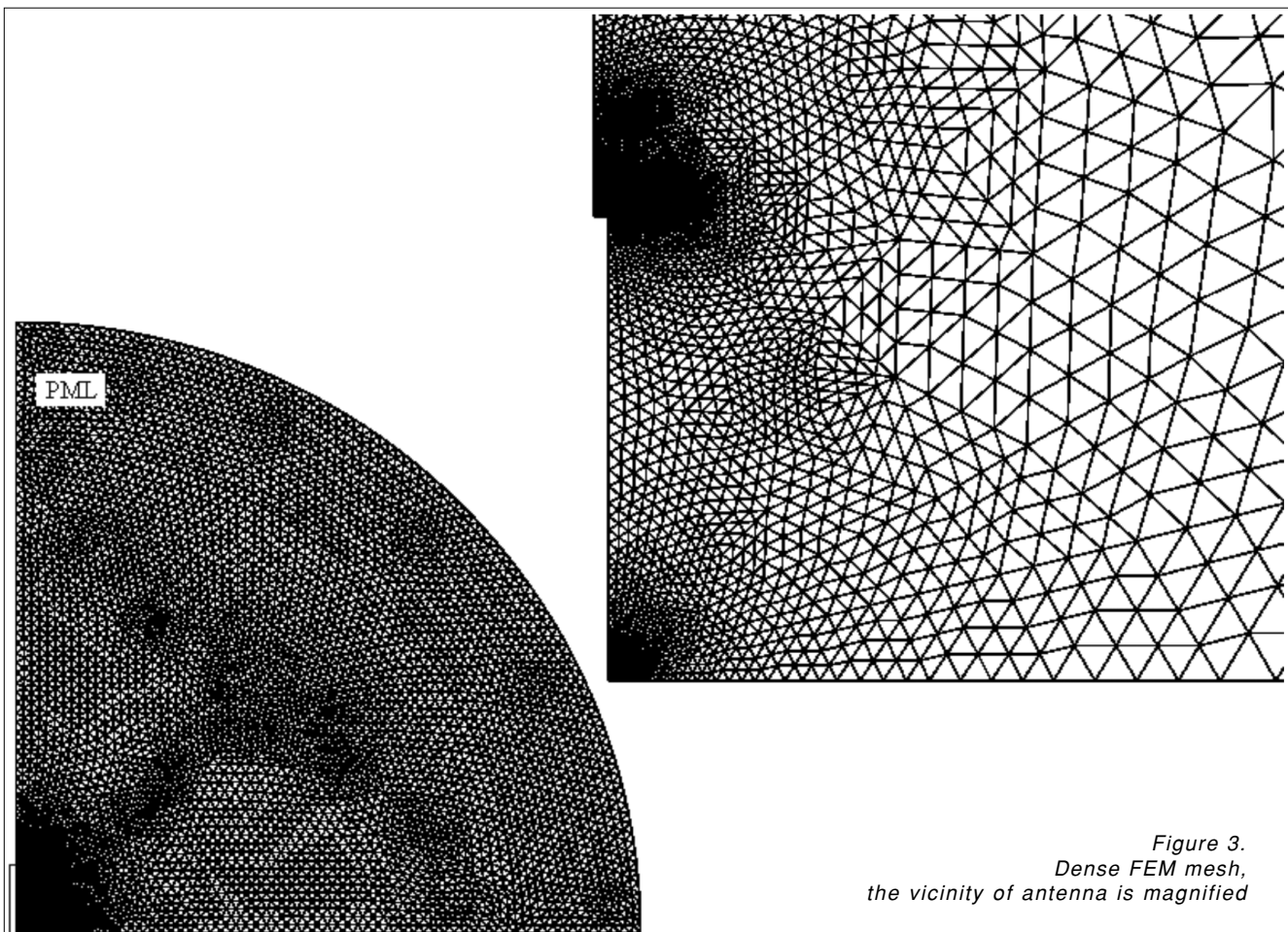


Figure 3.  
Dense FEM mesh,  
the vicinity of antenna is magnified

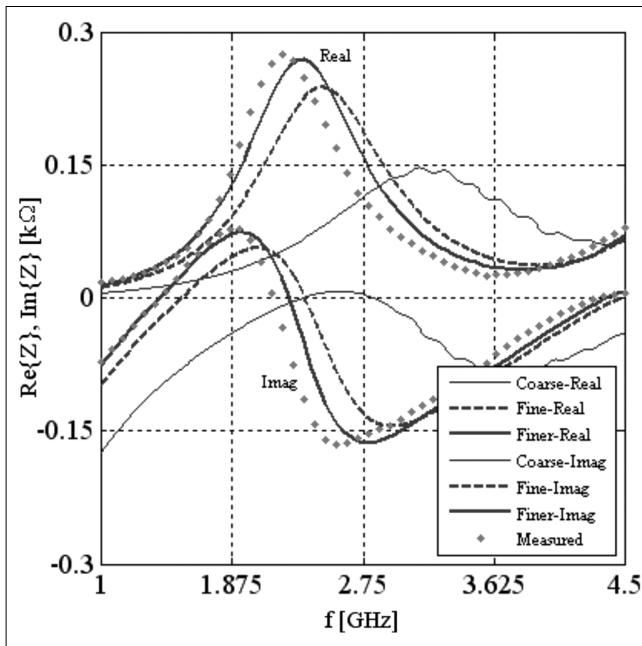


Figure 4. Convergence of the solution vs. number of triangles

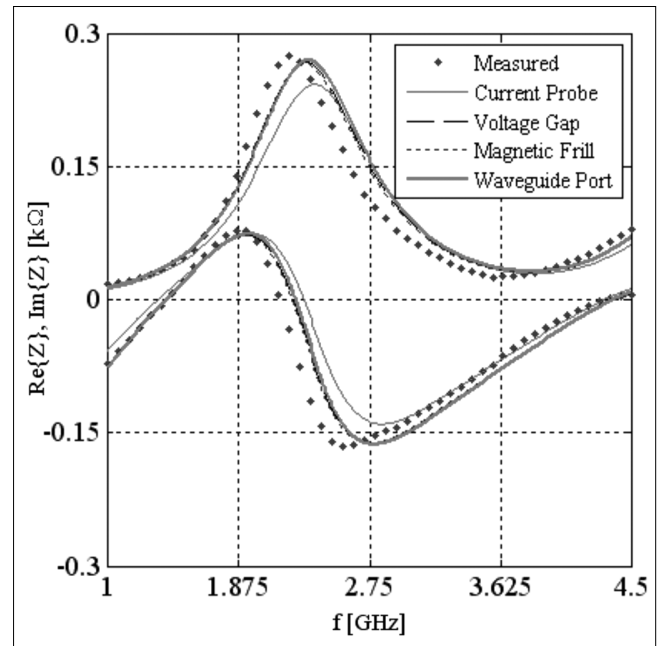


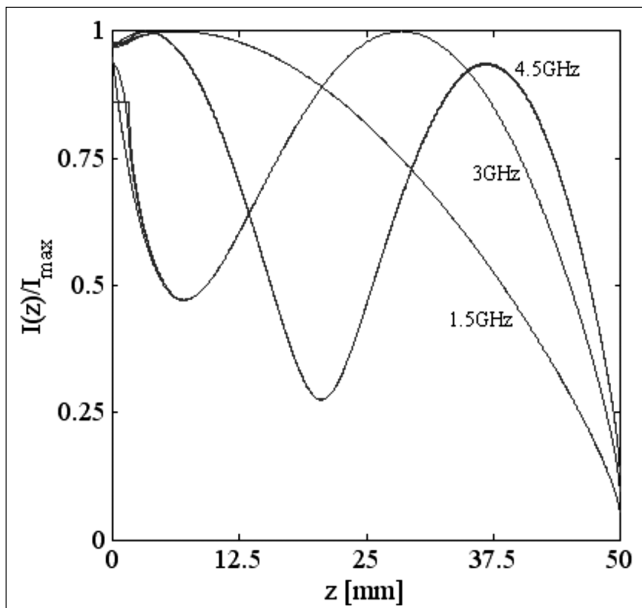
Figure 5. Comparison of the input impedance of the monopole antenna

been used in all the frequency during the frequency sweep in the range of 1 GHz and 4.5 GHz. A PML layer has been inserted to improve the absorption of electromagnetic field, and the radius of the computational domain is 1 m.

Fig. 5 shows a comparison between measured input impedance and simulated ones. The application of current probe model results in the weakest approximation, the approximated value obtained from the other models are practically the same.

The current distribution along the antenna is a very important input data to calculate other important quantities.

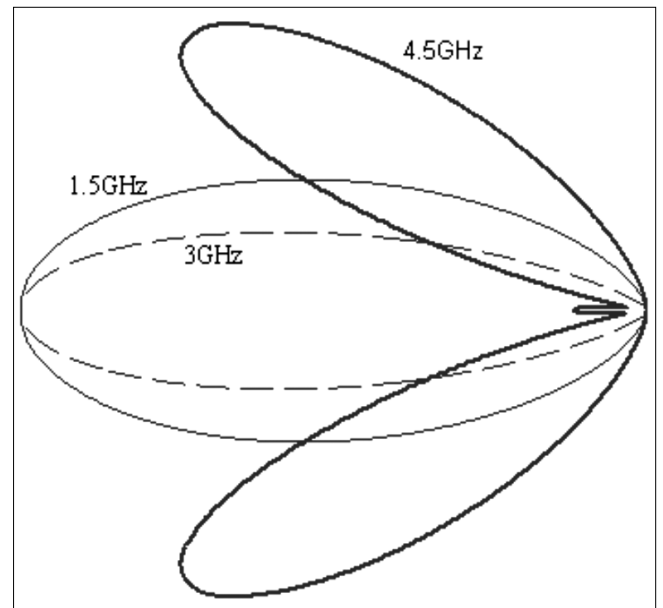
Figure 6. Normalized current distribution along the antenna at three different frequencies



A comparison between the obtained currents simulated by the above mentioned feeding models can be seen in Fig. 6 at the frequencies  $f=1.5$  GHz, 3 GHz and 4.5 GHz. The results are practically the same, but a small difference can be seen in the vicinity of  $z=0$  (the feeding point), and it is the effect of the different feeding models.

Fig. 7 shows the simulated field pattern of the monopole antenna at three different frequencies. The characteristics have been mirrored to the plane. It is noted that the characteristics simulated by the different feeding models are practically the same, i.e. the far field region is not depending on the applied feeding model.

Figure 7. The normalized far field characteristics at three different frequencies



## 5. Summary

Feeding models of antennas have been presented in the frame of FEM. The input impedance, the current distribution and the characteristics of a monopole antenna on a ground plane have been simulated and compared with measured data. The next step of the research work is to apply the feeding models in the case of more complex antennas in 3-dimensional situations, and to compare the results with other numerical techniques, e.g. with MoM.

### Author



**MIKLÓS KUCZMANN** was born in Hungary, 1977. He has become M.Sc. in Electrical Engineering in 2000, and Ph.D. in Electrical Engineering in 2005 at the Budapest University of Technology and Economics, Department of Electromagnetic Theory. He is Associate Professor at the Department of Telecommunications, Széchenyi István University, Győr, Hungary, where he is the Head of Laboratory of Electromagnetic Fields since 2005. Dr. Kuczmann has won the Bolyai János Scholarship from the Hungarian Academy of Sciences in 2006, the "Best PhD Dissertation" award in 2006 from the same Institute, and the Award of Academic Press in 2010.

### References

- [1] J.-M. Jin, D. J. Riley, Finite Element Analysis of Antennas and Arrays, Wiley, IEEE Press, 2009.
- [2] J.-M. Jin, The Finite Element Method in Electromagnetics, Wiley, IEEE Press, 2002.
- [3] M. Kuczmann, A. Iványi, The Finite Element Method in Magnetics, Academic Press, Budapest, 2008.
- [4] K. Simonyi, L. Zombory, Theoretical electromagnetics, Műszaki Könyvkiadó, Budapest, 2000.
- [5] W.C. Gibson, The Method of Moments in Electromagnetics, Chapman & Hall/RCR, 2008.
- [6] T. Hertel, G. Smith, On the convergence of common FDTD feed models for antennas, IEEE Trans. Antennas Propag., Vol. 51, No. 8, pp.1771–1779, 2003.
- [7] Z. Lou, J.-M. Jin, Modeling and Simulation of Broad-Band Antennas Using the Time-Domain Finite Element Method, IEEE Trans. On Antennas Propag., Vol. 53, No. 12, pp.4099–4110, 2005.
- [8] <http://www.comsol.com>
- [9] [http://maxwell.sze.hu/~kuczmann/Korszeru\\_antenna/Antenna\\_lap.htm](http://maxwell.sze.hu/~kuczmann/Korszeru_antenna/Antenna_lap.htm)

# Novel image similarity measurement in Automated Optical Inspection

TIBOR TAKÁCS, LÁSZLÓ VAJTA

Budapest University of Technology and Economics,  
Department of Control Engineering and Information Technology  
{takacs, vajta}@iit.bme.hu

Keywords: image processing, automated optical inspection (AOI), image similarity measurement

**Image similarity measurement is one of the most important topics in industrial image processing systems. In automated optical inspection (AOI) in electronic device manufacturing, the widely used methods are built mainly on separated analysis of gray-scale images, and do not apply the high similarity between captured pictures. This paper presents a new methodology to measure the relative similarity of AOI images. Our method utilizes and satisfies the special conditions and requirements of AOI systems. The need for human intervention (parameter adjustment, calibration) is almost totally eliminated. As our experiments show, our novel techniques classify more than 98% of images in the perfect classes which makes the techniques built on this similarity measurement method entirely useful in industrial applications already at this stage of our research.**

## 1. Introduction

Today's electronic device manufacturing factories are unimaginable without modern automated optical inspection (AOI) systems. Humans are expensive and too slow to keep up with the frenetic beat rates of today's assembly lines. Their eyes are not good enough to match the output of ultra-fine manufacturing processes, and we have difficulty keeping focused on very repetitive tasks [14].

Nowadays, the most widespread optical inspection systems in electronic device manufacturing are based on analysis of high-resolution gray-scale camera pictures created by monocular matrix camera. (In this article, we focus mainly on system built up on this principle, especially on post-soldering inspection systems.)

The images created by AOI devices have a very special feature which is useful in accurate and reliable quality measurement systems. Namely, the pictures about the same component type are either very similar to, or very dissimilar from each other. Generally, the components with good quality have very similar appearances; however the bad images differ in a high degree (Fig. 1).

This article is about the similarity of AOI images. First, we illustrate the possibilities and difficulties of similarity measurement in optical inspection. Next, we explain our novel similarity measurement method based on image processing, statistical, outlier detection and classification techniques. Finally, we show two important examples in AOI systems (image database filtering, decision support system for human operators) where our algorithm provides an excellent solution.

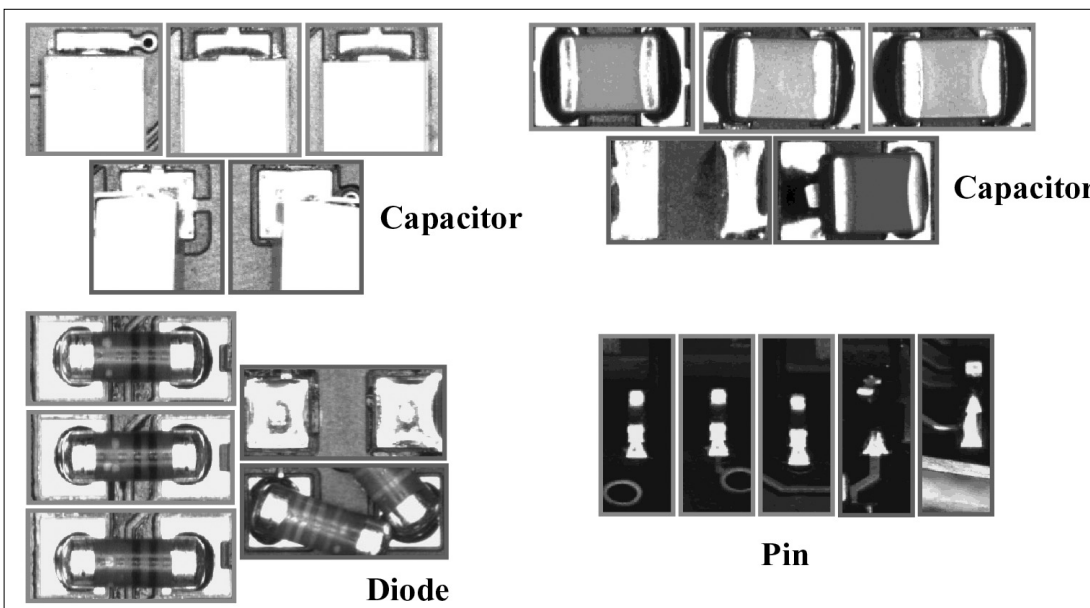


Figure 1.  
Similarity of AOI images

## 2. Image similarity measurement in AOI systems

### 2.1 Process model of AOI system

First of all, we briefly present a generalized ideal model of automated optical inspection systems of post-soldering printed circuit board (PCB) inspection (*Fig. 2*). By means of this simplified model we explain the environment where our techniques and algorithms can be placed. We present also several keywords which constitute the basic terminology of this article (marked as *italic*).

After a manufacturing step, the PCB is put in the AOI device specified to inspect the errors of previous processes (1). The AOI device loads the parameter values according to the type of the actual PCB (2) and executes the optical inspection procedure (*AOI algorithm, AOI macro*). (3): the AOI device creates several images, performs image processing algorithms and classifies the components as “good” or “bad”. The images classified as good are stored in the “good” part of the image database (4). If the PCB contains some errors, it is sent to the repair station with *error information* like error codes, locations and *error images* containing components classified as bad (5). The human operator at the repair station *re-inspects* the received errors. If an error is a “real error” (namely the AOI device is right), the PCB will be removed from the manufacturing line (sometimes the human operator can repair the error) and the image containing the bad component is stored in the “error” part of the image base (6).

If the AOI machine failed (*false alarm* or *false-call*), the image is sent to the “false-call” group in the image database (7). It is also possible that a bad component is not recognized by the AOI devices and it will be identified only at a later manufacturing process or at usage (*“slippage”*). The slipped images are also stored in the “error” part of image base later (8).

If the AOI system does not work with appropriate quality (i.e. too many false calls occur, or a slippage is iden-

tified), an AOI process engineer optimizes the AOI inspection algorithms (9): he or she tunes the macro’s parameters or changes the working process of the algorithm. To verify the optimization’s result, he runs the adjusted macro on the old images stored in the image database. If the new algorithm has better inspection quality than the old one, the old will be replaced with the new.

### 2.2 Background

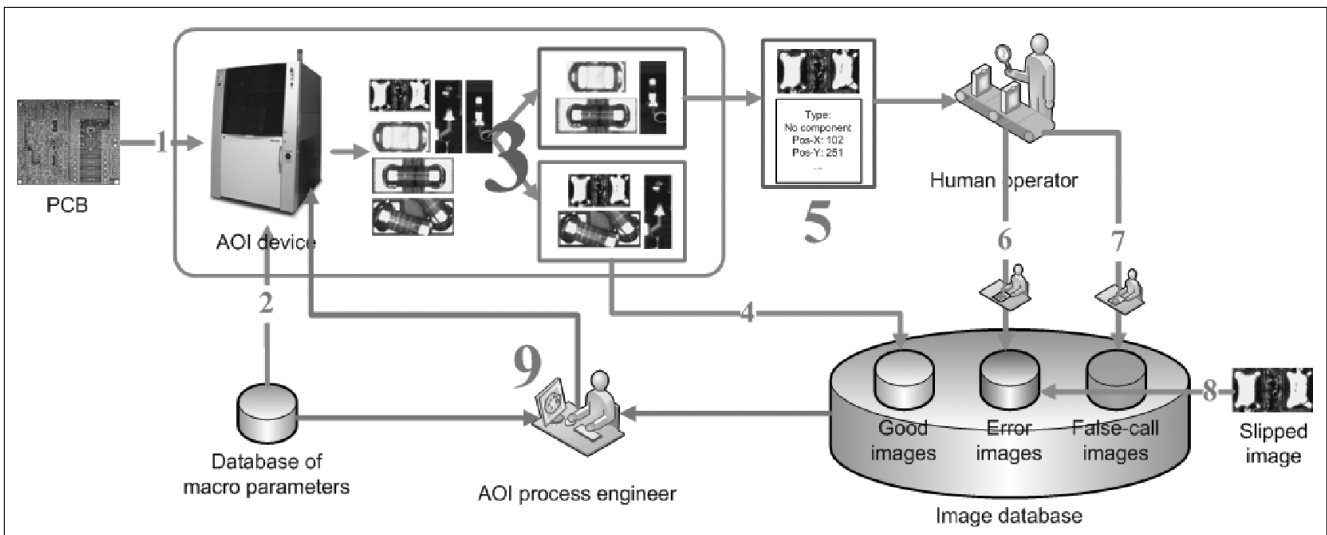
As mentioned in the Introduction, the images containing good components are very similar to each other, but the real error pictures differ from them (and also from each other) to a high degree. This occurrence motivated us to apply image similarity measurement as classification method. Our algorithm compares the image under inspection with a special image database – which contains good and bad reference images –, and the proposed class is calculated by means of this similarity information.

At this point, the following important question arises: when is an image “enough similar” to a set of good images? Similarity is only a qualitative mark, but for classification in AOI system it is necessary to use quantitative metrics. To determine an exact classification threshold, which is valid for all types of components, is a quite impossible task.

The basis of our technique is that we consider the similarity as a **relative**, but quantitative measurement value. The algorithm calculates a *similarity degree* for all images, also for references and also for actual inspection image. The classification decision can be executed depending on the actual situation (actual similarity degrees of reference images). This principle precludes the mentioned definability question of similarity.

In the following, we present first an overview of the related work. Next we introduce the requirements about a similarity measurement system in field of AOI. Finally, our novel method will be explained how can the similarity degrees of images in an image-database calculated.

Figure 2. Generalized ideal model of AOI systems



### 2.3 Related works

In automated optical inspection systems, the image similarity measurement is used to find the images containing faulty components. In the literature and industry, there are two main directions in optical quality measurement.

In the first group, where the most widespread methods are located, the captured images are separately analysed by means of image processing and classification algorithms using several predetermined parameters. They process only the actual images, and they do not take into account information about earlier results.

Wang et al. [22] proposed a new approach using accelerated species based particle swarm optimization for multi-template matching. The method completed the general PSO algorithm to allow the particles to search for multiple optima (either local or global) simultaneously to overcome the deficiency of the original PSO in multimodal optimization problems.

The paper written by Mar et al. [17] presented a computer vision system for automatic detection, localisation and segmentation of solder joints on PCBs under different illumination conditions. After an illumination normalization approach, the PCB image is transformed from an RGB colour space to a YIQ colour space for the effective detection of solder joints from the background. Finally, by thresholding and region filling, the solder joint are segmented and classified.

An interesting solution was proposed by Kong and Wang in 2007 [12]. Their paper deals with the reconstruction of the solder joint's surface in PCB based on shape from shading technology which is an important non-contact measurement method.

The second frequent way in optical inspection of PCBs is the usage of special computational models like artificial neural networks. Acciani et al. in 2006 [1,2] developed a solder joint classification method which uses multiple neural networks. Five different levels of solder quality in respect to the amount of solder paste have been defined. Two feature vectors extracted from the images' region of interest feed the neural network system for the classification.

A complete system to detect mounting defects in the circuits in presented in [5]. The authors processed the AOI images using wavelet transform and neural networks, for low computational cost and acceptable precision.

An artificial neural network (ANN) was used by Ong et al and presented in [18]. They combined orthogonal and oblique gray-level images at pixel level which were then directly input into an ANN for processing, eliminating the need to determine heuristic features. Learning vector quantization architecture was used as the classifier.

These approaches can solve several important problems in field of automated optical inspection, but they have also limitations. Namely, there are several fields of quality inspection in electronic manufacturing where special requirements need to be considered and the current approaches cannot satisfy them.

Our novel similarity measurement technique was developed based on these requirements. In next section, we will present four general conditions which are very important in case of optical inspection algorithm. In Section 3 (3.1, 3.2), we will propose two new approaches built on novel similarity metric where we will show the details of current approaches' limitations and present what the proposed method adds to the state of the art.

### 2.4 Special requirements

Our main goal is to estimate the appropriate class of an inspection image only by means of a reference image database using image similarity measurement. To produce a well usable classification system, our method needs to satisfy the following requirements:

- **Flexibility.** The general AOI inspection algorithms are rigid measurement methods. After adjusting their parameters, they execute the same steps with same conditions, and they do not react to the manufacturing changes directly and immediately. This is one of reasons of the relative high false call rate. Because our method aims at reducing this problem, it needs to adapt itself to the actual conditions and environment continually and automatically.

- **No parameters.** Our method contains image processing, outlier detection, clustering and statistical algorithms which have several parameters. Our purpose is that the values of these parameters would be calculated by the algorithm during runtime and not by human experts during implementation time. This is also a precondition of flexibility described in previous point.

- **Transparency from image type.** An AOI device inspects several different components creating very different images. We aimed at developing an algorithm which can work on several types of images equally well.

- **No a-priori information about AOI inspection algorithm.** An AOI macro inspects specified region of images and executes special transformations. Our decision support algorithm does not consider this a-priori information to avoid the dependency on the settings of the AOI macro.

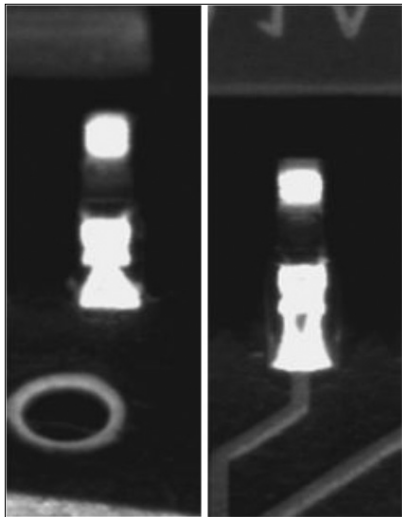
### 2.5 Calculation of similarity degree

The basic idea behind our similarity degree calculation in AOI systems is quite simple. The calculation time of similarity degree is very critical because of high productivity of the modern manufacturing lines. Therefore to compare all error images with all corresponding images is an impossible task in inspection process. This urged us to develop a novel method which can estimate the similarity degree without numerous compare calculations.

To solve this problem, the algorithm creates first a **golden template image** which represents a general case of ideal (good) images. The similarity degree is inherited from the difference between the image and this golden template.

Behind this simple theory, a complicated algorithm lies with the following steps:

Figure 3. Differences in position of images



1. Iterative translation compensation (pre-processing)
2. Calculation of golden template
3. Split into sub-images
4. Calculation of difference profile
5. Estimation of similarity degree

In the following the details of the algorithm will be presented. The results of the processing steps will be illustrated by means of image database about SOIC (small-outline integrated circuit) pins. The dimension of images in image-space is 242x104 pixels, which shows a pin with size 1x0.5 mm.

### 2.5.1 Iterative translation compensation (pre-processing)

The images showing good components contain very similar features, but unfortunately the location of the important objects can be varied to some degree (see Fig. 3). The reduction of position disparity makes the classification methods faster, easier and more accurate, because similarity measurement can suggest that similar image-features have similar positions. Therefore, as the first step, the position disparities of the actual error image will be compensated.

First, the compensation algorithm looks for the *reference points* of the image. This reference point is a special feature point on the image: centre of mass of the whole image; centre of mass of a specified feature etc. [6]. The choice, which feature point is the best to use, depends on the (component) type of the actual image. (A future research could be to determine a general reference point.) Fig. 4 illustrates an example of the calculated reference points. In this case, the centre of mass of the white object in the upper part of the image (in green circle) represents the reference point. Therefore the different appearance of wires does not influence the result.

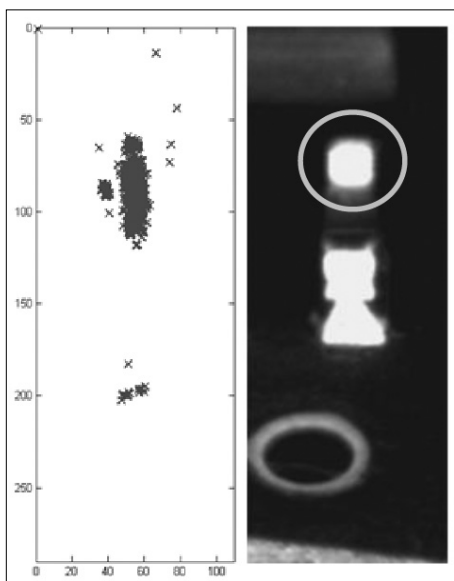


Figure 4. The positions of the reference point in SOIC image base

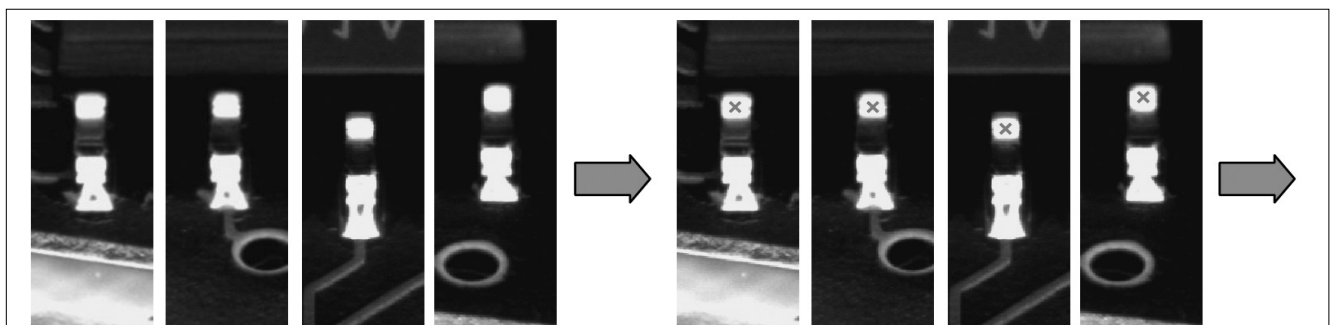
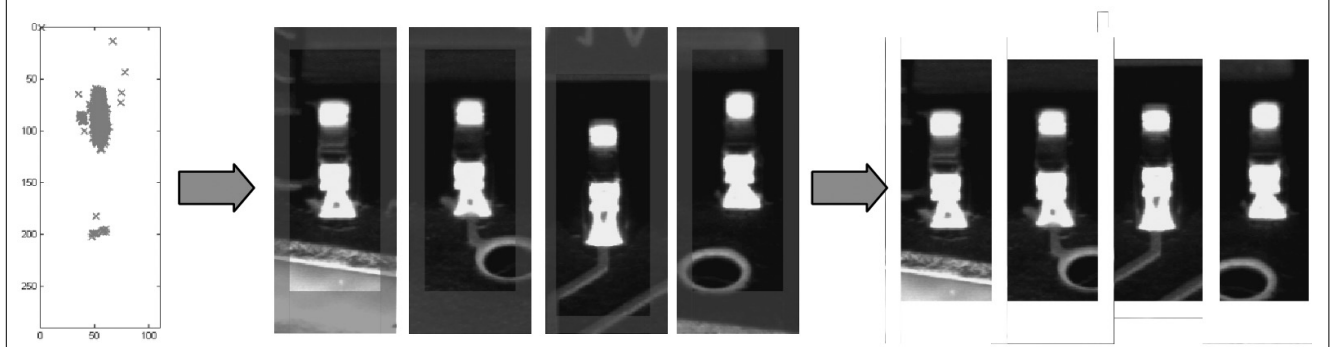


Figure 5. Working process of translation compensation





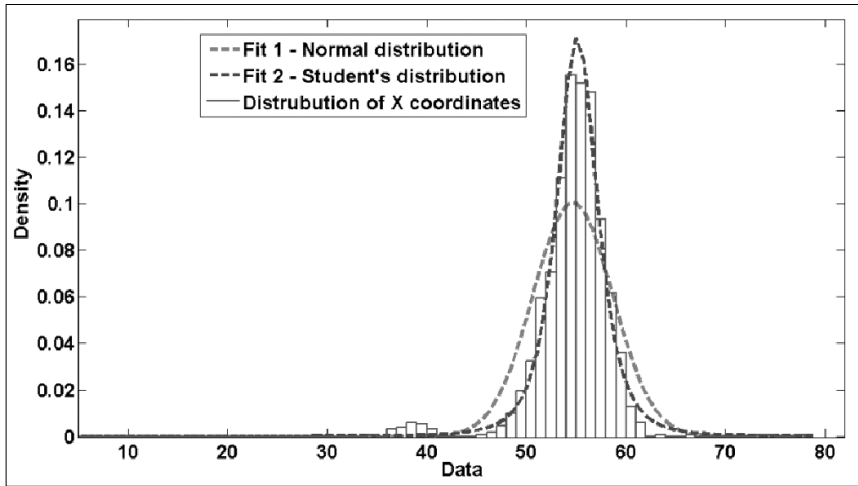


Figure 6. Distribution of X coordinates

In the ideal case, the reference points have the same position on each image, but in a real situation their locations vary to some degree. The goal is to remove these disparities to get the reference points overlapped, resulting that the features on the images will cover each other as well.

Next, the algorithm calculates the expected value of reference points and cuts the sides of each image to move the image's references over the position of the expected value (Fig. 5). This results that the reference points of all images have the same location. The size of cutting depends on the distance between the actual reference and the expected value of all points therefore the bigger this distance the smaller the resulted picture size becomes. Because the images need to have the same size after translation compensation, only one outlier reference point, which is very far from the expected value, produces very small result images losing the most relevant information. Therefore the maximal cutting sizes in both directions were limited to the average distance from the expected value.

The determined bound was a good choice to avoid too small result images while several images' positions were compensated. Fig. 6 visualizes the distribution of reference points' coordinates. It is interesting to note that the coordinates are not normally distributed. Our experiments show that distributions with best fits are the scaled and translated Student's t-distribution [11] or logistic distribution [3].

The previous steps (determination of reference points, calculating the expected value, cutting image sides to reduce the disparity) can be repeated until the disparities of the reference points fall below a specified limit. Our research showed that after 4 iterations the position improvement of position optimization stopped.

The graphs on Fig. 7 illustrate the efficiency of this method. They show the standard deviation of pixel values before and after compensation. It is easy to see that after translation compensation the variance of pixels decreased in high degree (most at edges

of objects) because the images become more similar to each other.

After translation compensation, the differences between features' positions of different images are roughly eliminated. Small differences in size and location are certainly possible, but large variations are only accepted in case of real error images. This characteristic of compensated image base is used to determine the images' similarity degree.

### 2.5.2 Calculation of golden template

Next the golden template is calculated which is an "ideal" good image. By its calculation, only the good (and false-call) images are taken into account from the reference image database, because the similarity degree means how the actual image is similar to the good images.

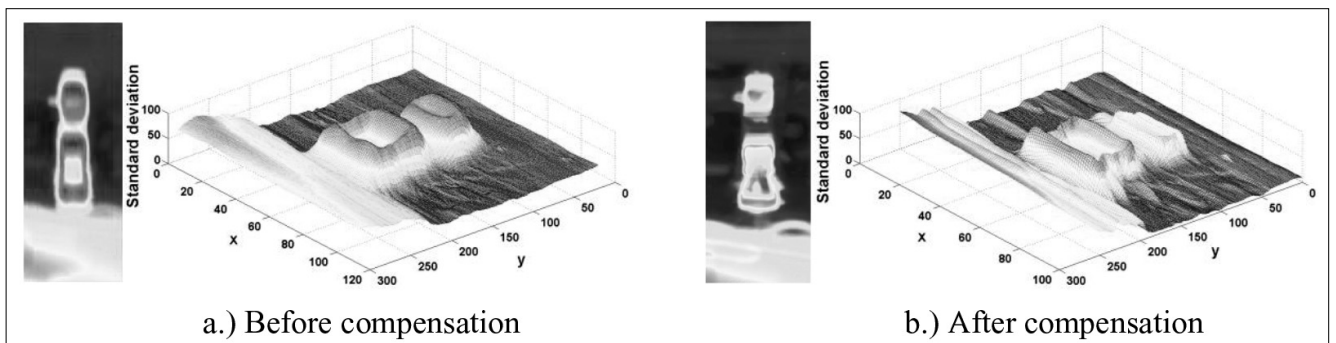
The images can be considered as the samples of a random variable matrix:

$$I = (\xi_{ij}), \forall \xi_{ij}: \Omega \rightarrow \mathbb{R} \quad (1)$$

where  $i$  and  $j$  are pixel indices,  $\xi_{ij}$  is the random variable,  $\Omega$  is the space of possible gray-scale values, and  $\mathbb{R}$  is the real numbers representing the gray-scale values (typically integer numbers between 0 and 255). The distributions of these variables show the probabilities of the pixels' gray scale values.

The comparison algorithm (as we will introduce soon) is executed on pixel level, therefore it is a reasonable choice to use a matrix created from the expected values of the random variable matrix as golden template [7]:

Figure 7. The effect of iterative translation compensation algorithm (the standard deviation of pixels decreased substantially)



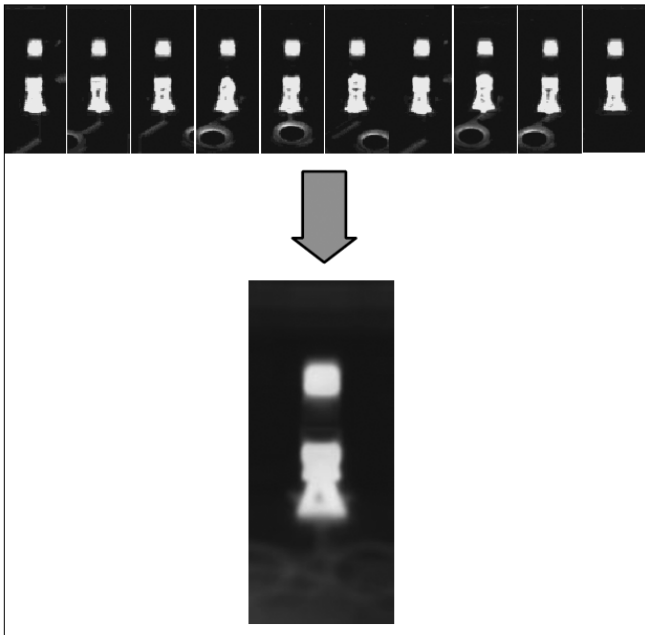


Figure 8. Example for a golden template (SOIC)

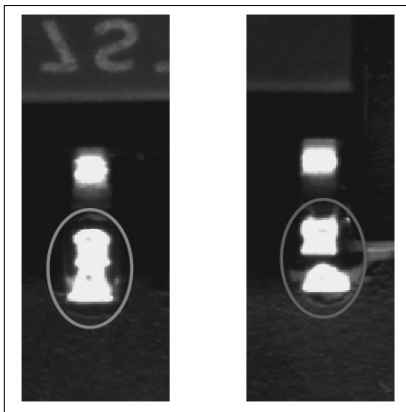


Figure 9. Illustration of small differences between good and bad images

$$gt(x, y) = (E(\xi_{xy})) \quad (2)$$

Certainly we have to estimate the expected values with the sample means therefore the resulted template image is the mean of all good images (Fig. 8).

### 2.5.3 Split into subimages

As a next step, the algorithm determines the similarity between the images and the golden template. The differences between good and bad images can be occasionally very small and hard to detect (see Fig. 9). Using general techniques by comparing whole images, these tiny (with size about 4x4 pixels) but very important dissimilarities get lost, implying that the class of error image cannot be found out.

Our new specialized technique splits the images into several small sub-images and the comparison method is executed on them separately and independently. The sub-images' size is much smaller than the original images' therefore the tiny errors contained only the good images become detectable on sub-images. The images are split with four different grids (Fig. 10) to avoid an important area being always divided between more sub-images. As our experiments show, the ideal size of a sub-image is about 8x8 pixels.

### 2.5.4 Calculation of difference profile

The split of an image results an image matrix which elements are the created sub-images. Next each sub-image will be compared with the corresponding area of the golden template generating the *difference profile* of the actual image.

In the simplest case, all pixels are taken into account with the same weight when computing the differences. But in a real situation, this assumption would distort the calculation. At some pixels of the image plane, the gray scale value of different images varies to a high degree. This means that several images differ very much from the golden template (from the expected value) at these regions. These high differences increase the values of the difference profile without reason, because the classification decision cannot be made by means of pixels where the gray scale value does not depend on the type of the image (namely it is bad or good).

This diversity of the gray scale values can be represented by standard deviation (or variance) of pixels' random variable. Large variance means that there are a lot of images in the actual image base which differ from the golden template at these regions to a high degree. Fig. 11 illustrates two main places where the pixels' variance has huge value: at the sides of objects and at unimportant areas (i.e. not inspected by AOI macros). First instance occurs because the size and position differs to some degree despite earlier translation compensation therefore around the edges the pixels' value varies. At an unimportant part (second case) the AOI mac-

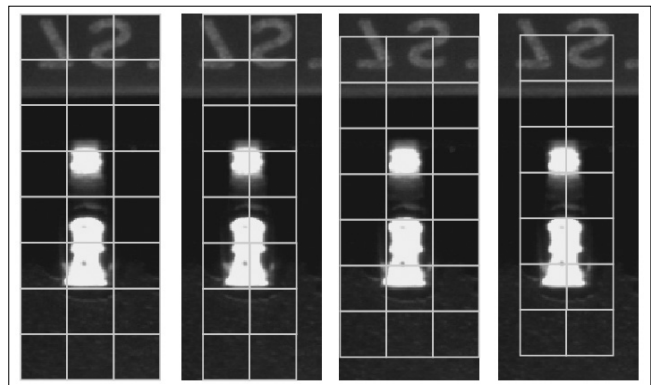
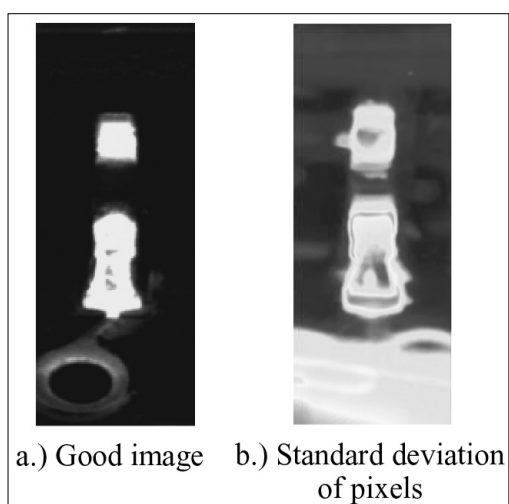


Figure 10. Grids to creates subimages

Figure 11. Illustration of standard deviation's role



ros do not inspect the images therefore at these areas also the good images can contain various pixel values causing high variance.

Our algorithm does not have any a-priori information about the images and about the AOI inspection macros. Hence it is not possible to remove these unnecessary regions before executing the comparing method. Therefore we used weighted distance metric to calculate the difference profile which considers also the variance of pixels.

Our method eliminates this problem by using weighted Euclidean distance where the weights are derived from the standard deviation of the actual sub-image. The greater the standard deviation of a pixel's random variables is, the smaller its significance. Therefore the weight function of  $i$ th and  $j$ th sub-image is the standard deviation matrix multiplied by minus 1 and scaled between 0 and 1:

$$W_{i,j}(x,y) = \frac{-\sigma_{i,j}(x,y) + \max_{0 \leq k < N_{i,j}; 0 \leq l < M_{i,j}} \{\sigma_{i,j}(k,l)\}}{\max_{0 \leq k < N_{i,j}; 0 \leq l < M_{i,j}} \{\sigma_{i,j}(k,l)\} - \min_{0 \leq k < N_{i,j}; 0 \leq l < M_{i,j}} \{\sigma_{i,j}(k,l)\}} \quad (3)$$

where  $\sigma_{i,j}(x,y)$  is the standard deviation function (matrix) of the sub-image,  $N_{i,j}$  and  $M_{i,j}$  are the dimensions of the sub-image. Fig. 12 illustrates the result of this equation.

The scaling depends on the minimal and maximal level of the local (not the global) standard deviation field (namely only the actual sub-region of the golden template is considered). This means that the comparison is executed on each sub-image independently and the scaling takes the relative differences of a sub-region's variances into account.

We have tested several metrics like Euclidean distance, Manhattan distance, correlation etc. Our experiment shows that best choice is to use Euclidean distance.

As a result, the  $i$ th and  $j$ th element of the difference profile, namely the distances between the sub-image and golden template, can be calculated as follows:

$$D(i,j) = \sqrt{\sum_{k=0; l=0}^{N_{i,j}-1; M_{i,j}-1} \left( (gt_{i,j}(k,l) - subim_{i,j}(k,l))^2 \cdot W_{i,j}(k,l) \right)} \quad (4)$$

where  $gt_{i,j}(x,y)$  is the covered region of golden template,  $subim_{i,j}(x,y)$  is the sub-image and  $W_{i,j}(x,y)$  is the weight function determined earlier. Fig. 13 illustrates two examples for the created difference profile.

### 2.5.5 Estimation of similarity degree

Difference profile contains the error values of image's sub-regions. We have deduced the similarity degree from analysing this error field.

We have considered several metrics of difference profiles, like maximal and minimal values, average value, variance (standard deviation) etc. Our experiments show that it is hard to distinguish the sets of good and bad images by means of only one feature.

Fig. 14 illustrates that neither maximal error value, nor the standard deviation of the difference profile, as the two most relevant metrics, can separate bad images from good images by themselves. But the simultaneous usage of both techniques yields a 2-dimensional metric which combines their advantages and creates a simi-

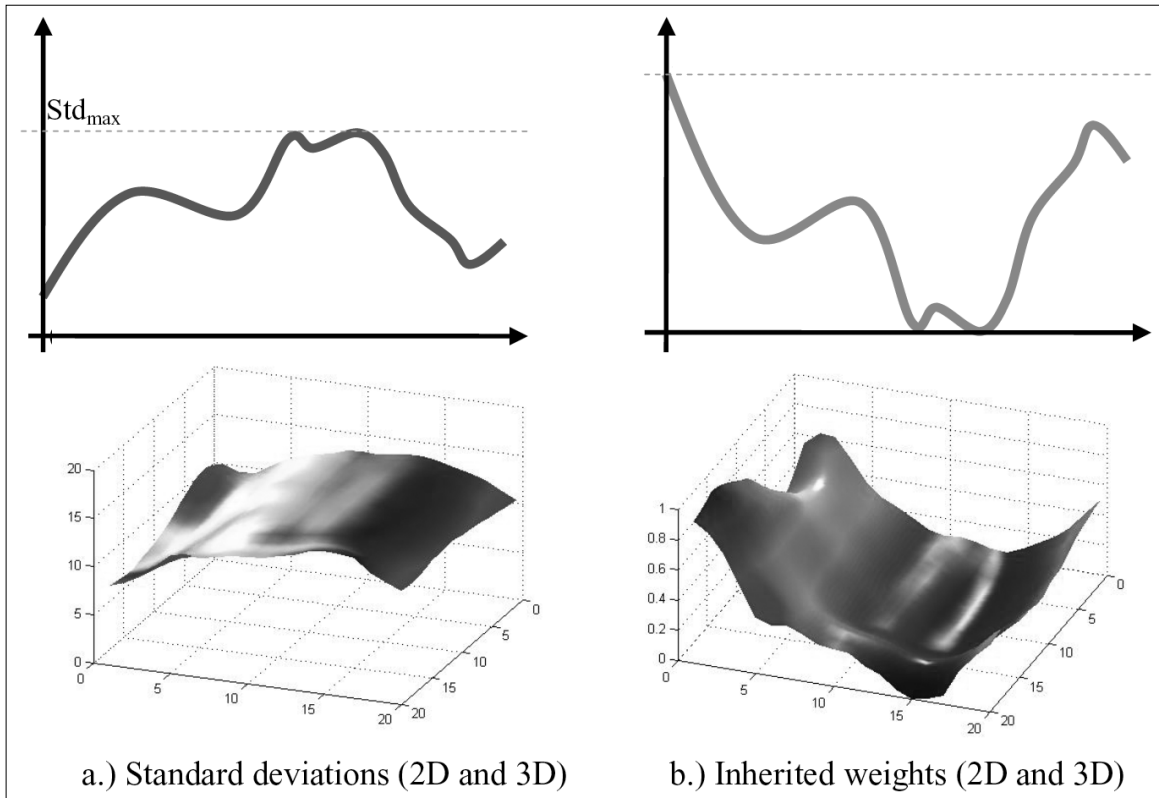


Figure 12. Examples for weighting functions

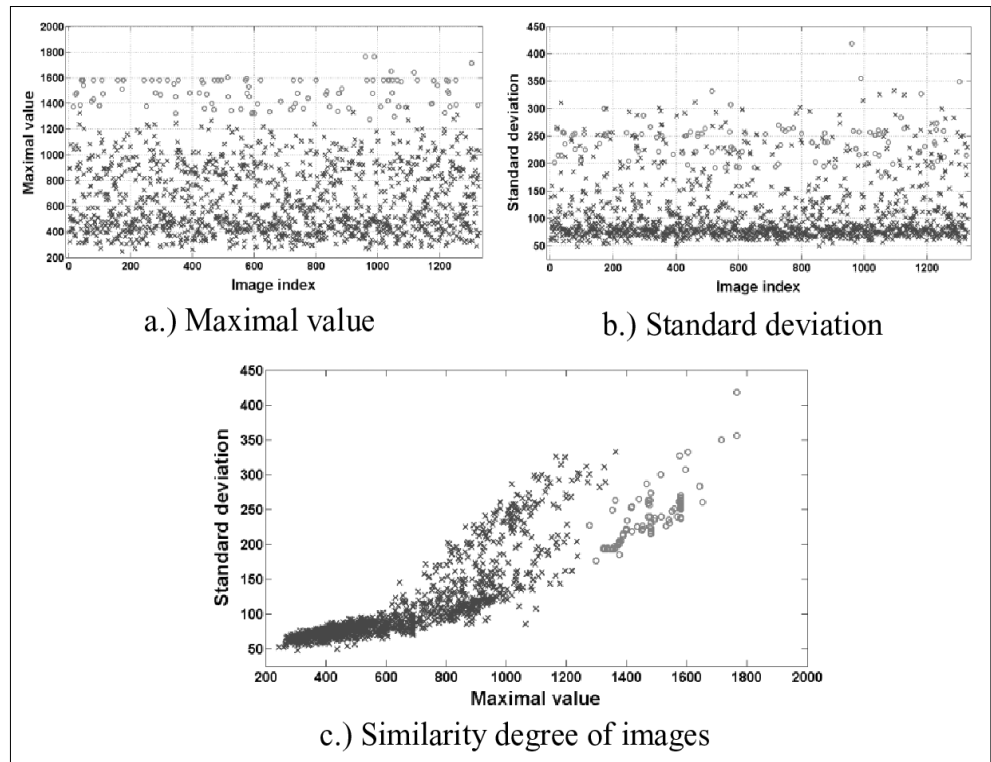
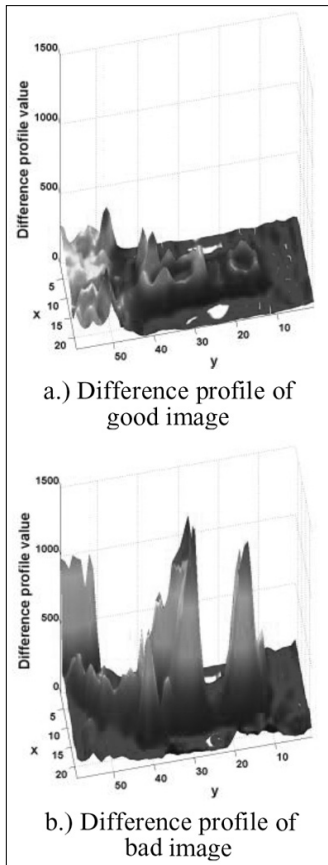


Figure 13. Examples for difference profile (SOIC)

Figure 14. Examples for similarity degrees (SOIC)

similarity degree field where the two image groups can be relatively well separated (Fig. 14). This field gives an excellent basis for the next classification method.

### 2.6 Summary

The flowchart in Fig. 15 summarizes the working process of our similarity degree calculation. First the positions are equalized and the golden template is calculated which is the expected value of good images. Next the images are split into small sub-images, otherwise the small but very important errors on images would get lost by comparing the whole images. At the third step, we calculate the difference profile which is the weighted Euclidean distance between the sub-regions of the images and the golden template. The weights are inherit-

ed from the standard deviation of the local region. The similarity degree of an image is a 2-dimensional vector derived from the maximal difference value and the standard deviation of the difference profile.

### 3. Experiments

In previous section, we explained the calculation process of similarity degrees. After computation, a 2-dimensional similarity degree field is generated where all images is projected in. Analysing this vector field, important AOI problem can be solved without serious additional work. In this paper, we will present two important areas where we used our method with significant success (see Fig. 16).

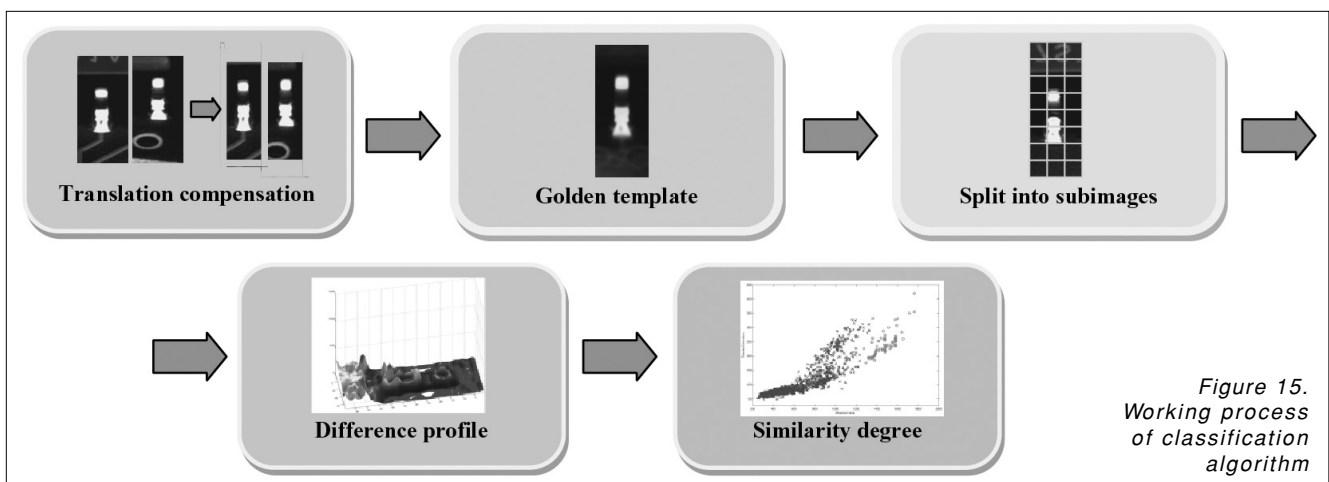


Figure 15. Working process of classification algorithm

**3.1 Decision support in re-inspection phase**

In Section 2.1, we introduced that, after the AOI inspected the PCB, a human operator checks again the determined errors. This is because of small modification in manufacturing process generates false alarms (namely the AOI device classified a good component as faulty) which are revised by human experts. This human-machine cooperation would create a great inspection system because the devices are very fast and accurate, while the humans are not. But unfortunately, the need of human factor decreases the reliability and productivity of the manufacturing and the inspection process. Therefore it is aimed to develop techniques to assist the human operators' work in re-inspection process. Our purpose was to create a re-inspection system which compares the error images (on which a bad component was found by AOI device) with a reference database, and by means of the comparing information takes a proposal for the human operator about the appropriate class of the image (see Fig. 16).

The technique of "learning and recognizing" in AOI systems appeared already in earlier publications. In general, the researchers use neural networks in inspection systems [1,2,5,18]. In spite of this, we would like to avoid the utilization of artificial neural networks in automated optical inspection. Although the neural networks make exact, repeatable calculations and in general do not contain any stochastic steps, they are "black boxes" from the point of view of the engineers who are responsible for quality inspection in the whole factory, because it is hard or impossible to convert the connection forces (weights) between neurons in a trained neural network to exact physical meaning of optical inspection and measurement.

The training of neural networks raises also an interesting question. Our task is to pre-classify the error images in two groups: good images (false-calls) and bad images (real errors). There are several samples in the first group, but much less real images exist. Until the false-

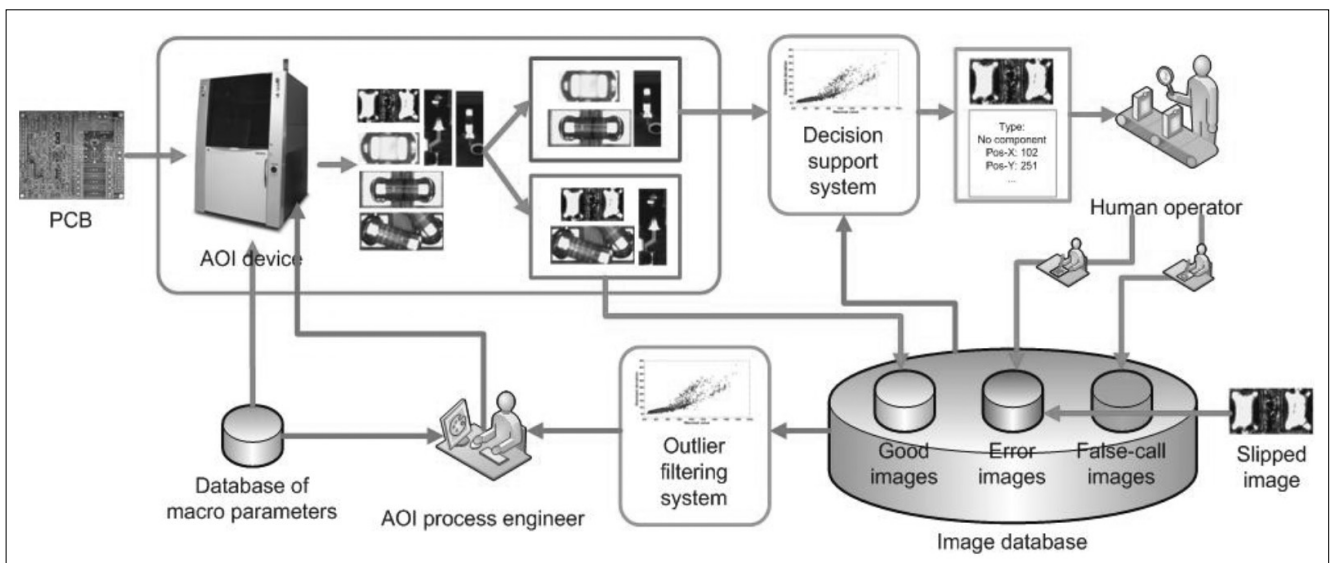
call images are relative similar to each other, the real errors have numerous appearances. Therefore in field of the training process and working quality, more difficulties and challenges occur.

These facts motivated us to advance in another development direction. The similarity degree analysis is a very good basis of a decision support system. The algorithm is based on the fact that false-call images cannot satisfy the serious requirements of AOI macros, but they are very similar to normal (good) and to other false-call images. Therefore if an error image seems like stored good images, it is probably also a good image; if the image is dissimilar, the algorithm suggests that it shows a real error. By means of the similarity measurement field it is possible to decide about a new error image, that it contains a good or a bad component.

At calibration, the golden template and the reference images' similarity degrees are calculated. At re-inspection process, the algorithm determines the similarity degree of the actual error image (by means of earlier calculated golden template) and places the value on the 2-dimensional similarity degree field. The proposed class of error image is determined by three different factors: class of nearest image, class of image-group belong to the smaller average distance, clustering result (details are explained in Section 3.2). The classification method calculates all of these metrics and the voting result determines the final class of error image: if already one metric signs that the error image contains real error, the image is classified as bad.

We will present the most important results in Section 3.2. The decision support system based on our novel similarity measurement method can separate the good and bad pictures with very small error rate, independent from the type of images. The method does not use image base specific settings (except the type of reference point at translation compensation), and it works without any modification on several AOI image bases equally well.

Figure 16. Model of AOI systems completing with our novel methods



### 3.2 Outlier filtering in AOI image databases

The presented cooperation between machines and humans implies special difficulties in the topic of optical inspection. An interesting and important challenge is the faulty content of image databases storing previously created inspection images, namely during image collection – mainly caused by human factor – some falsely classified images can be put in the image database. This means that image bases, which normally include pictures showing only good components, contain also images with bad components. This occurrence makes, for example, the optimization process impossible because it is always assumed that the training image bases are homogenous. It causes less accurate AOI systems, higher optimization cost and prevent the development of fully autonomous AOI devices.

The image database filtering algorithm is typical outlier detection task [7,8,10]. There are several methods published in the topic of image processing outlier detection [13,15,19-21,23]. Most frequently, it is used in medical imaging and analysis of satellite images, but it has high significance in other image processing areas as well. Compared to these methods, the environment of our algorithm has very special requirements. First of all, the usage of training samples is entirely intolerable because creation of training samples and keeping them in good condition are hard and very time-consuming processes in case of manufacturing. Another very important condition is that the whole image – not only some features – need to be considered by identifying the outliers, because the AOI images contain very important information on the full image plane (except at narrow region of images' side).

The last important difference between this algorithm and earlier published methods is that our technique does not know what kind of outliers can occur on falsely classified images. This means, that our algorithm gets an image database as input, without any information about features why an image is an outlier. Our method “finds out” these differences in case of each AOI image set separately, apart from the component what the images contain. This is a novel aspect in the area of outlier detection in image processing systems.

The similarity degree field gives a very essential solution for this application. The similarity degree belong to all the images in the database under filtering will be calculated and by analysing the produced similarity vector field, the outliers can be detected.

It was introduced in Section 2.5.2 that the golden template is really the mean of all good images. Certainly, this creation method is impossible because the usage of training samples is excluded. Therefore by calculating the mean image, all the images are considered as well. The number of outlier images is much less than the number of well-classified images therefore the mean image cannot be damaged much by the outlier images. Fig. 17 shows a mean image created with (a) and without (b) outlier images. It is apparent that the differences between the images are negligible (c).

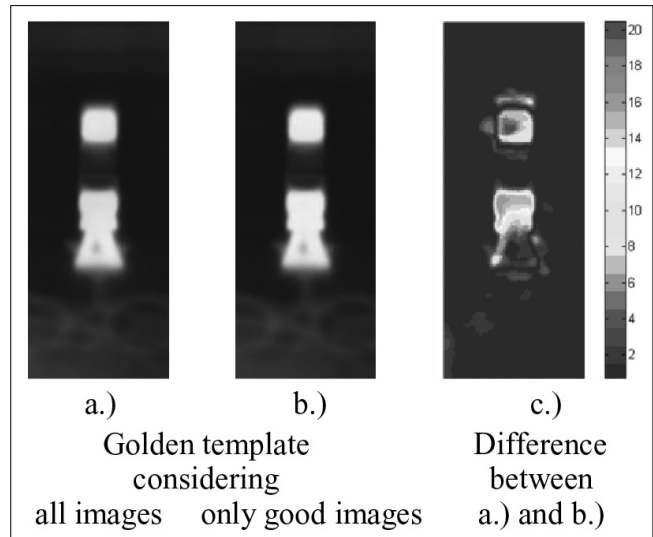


Figure 17. Effect of outlier images in golden template

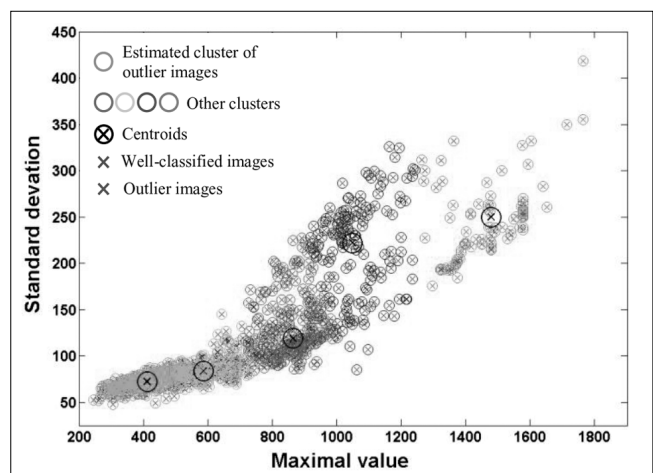
After calculating the similarity degree (by means of the special golden template), it is necessary to separate the whole set into two subsets (well-classified and outlier images). We applied the well-known k-means algorithm [9,16] to cluster the similarity degree field in two groups. In our method, it is created 5 clusters ( $k=5$ ), because of special locations of well-classified and outlier points. Because the clustering depends on the starting parameters of the k-means method, the clustering is executed more times with different starting arguments to find all possible outliers.

In Fig. 18, a clustering result is shown (SOIC). The graph illustrates that all of the outlier image were found, and only 7 well-classified images are classified badly, which is smaller than 0.005% error rate. (The detailed results are explained in next section.)

### 3.3 Results

We tested both algorithms on several image databases which were created at an industrial production line. Each image database contains good (well-classified) and bad (outlier) images. Table 1 presents the image databases used.

Figure 18. Result of clustering algorithm in case of 5 clusters



We used the same image databases in both algorithms. In case of decision support system, half of images were the training samples, the other half of them were the testing elements. The quality factor was that how many testing image are classified well. In outlier filtering method, all the images in a database were created the original (unfiltered) image set and the purpose was to filter only the bad images. It is important to note that on every image base the same algorithms (with the same parameters) were executed. There were no special settings or adjustments for single image sets.

Table 2 shows the result of 7 image databases. First it is important to note that the qualities of two algorithms are very similar: at same databases both methods have quite the same results. This is because both techniques are built on the same basis (on analysis of similarity degree field) therefore this identity is not surprising.

The results in the table show that both methods were able to execute their tasks with very small error rate: the general hit rate is greater than 98%. At decision support system, in five cases all real error images were found, in the remaining two image sets only some special error could not be identified well. In case of outlier filtering method, all outliers were found in five databases, in the other two image bases the classification rates were above 96%.

The best results were noticeable in both cases in image database of C0805, C1210, SMCTAB and SOIC. The algorithms could separate the good and bad images almost perfectly. The result speaks for itself, it is not necessary to analyse this conclusion further.

At the MELF and SMCTAC image bases, “only” 94-97% of real errors or outliers were detected. Fig. 19 shows that very small regions contain the differences, and at these areas, the pixel's standard deviation has a relatively high value. In spite of these difficulties, our method assigns relatively low similarity degree to these pictures, but at those points in similarity degree plane, the density of good images is higher and the classification algorithm classifies these pictures in the false group.

In case of R0805, the number of badly clustered good images is relatively high (~10%). The reason is that in this image base, the good images are very different from each other (see Fig. 20). Therefore it is quite hard to determine the regions where the good and bad images differ which caused the high rate of false positive errors. Here the question about the usability of the algorithm in industrial production can occur, but it is important to note two facts. First, all real error images or outliers were detected, which more important task in case of manufacturing. Second, looking at Fig. 20, we can see the high heterogeneity of the good images which is a very serious challenge for all classification methods (an also for human experts). The 10% false positive rate in this adverse environment is an acceptable result, also in industrial production.

An important feature of this algorithm is that the classification decision is executed without using any threshold parameters, but it uses a standard clustering technique. Moreover, the method does not use image base specific settings (except the type of reference point at translation compensation), but it calculates the inner



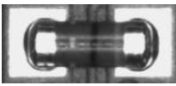




<b>C0805</b>	Capacitor with standard size 0805		Number of images: <b>1342</b> Good: <b>1221</b> (90.98%) Bad: <b>121</b> (9.02%)
<b>C1210</b>	Capacitor with standard size 1210		Number of images: <b>1405</b> Good: <b>1306</b> (92.95%) Bad: <b>99</b> (7.05%)
<b>MELF</b>	Metal electrode leadless face diode		Number of images: <b>1576</b> Good: <b>1451</b> (92.07%) Bad: <b>125</b> (7.93%)
<b>R0805</b>	Resistor with standard size 0805		Number of images: <b>1086</b> Good: <b>1011</b> (93.09%) Bad: <b>75</b> (6.91%)
<b>SMCTAB</b>	Surface mount capacitor with size „B”		Number of images: <b>1174</b> Good: <b>1119</b> (95.32%) Bad: <b>55</b> (4.68%)
<b>SMCTAC</b>	Surface mount capacitor with size „C”		Number of images: <b>1384</b> Good: <b>1274</b> (92.05%) Bad: <b>110</b> (7.95%)
<b>SOIC</b>	Small-outline integrated circuit		Number of images: <b>1329</b> Good: <b>1221</b> (91.87%) Bad: <b>108</b> (8.13%)

Table 1. Image databases for testing

Image type	Decision support system		Outlier filtering method
	Good image	Bad image	
<b>C0805</b>	Well classified: <b>609</b> (99.84%) Badly classified: <b>1</b> (0.16%)	Well classified: <b>60</b> (100%) Badly classified: <b>0</b> (0%)	Found outliers: <b>121</b> (100%) Well-classified images in good cluster: <b>1218</b> (99.75%) Well-classified images in outlier cluster: <b>3</b> (0.25%)
<b>C1210</b>	Well classified: <b>649</b> (99.39%) Badly classified: <b>4</b> (0.61%)	Well classified: <b>49</b> (100%) Badly classified: <b>0</b> (0%)	Found outliers: <b>99</b> (100%) Well-classified images in good cluster: <b>1302</b> (99.69%) Well-classified images in outlier cluster: <b>4</b> (0.31%)
<b>MELF</b>	Well classified: <b>715</b> (98.62%) Badly classified: <b>10</b> (1.38%)	Well classified: <b>60</b> (96.77%) Badly classified: <b>2</b> (3.23%)	Found outliers: <b>121</b> (96.80%) Well-classified images in good cluster: <b>1441</b> (99.31%) Well-classified images in outlier cluster: <b>10</b> (0.69%)
<b>R0805</b>	Well classified: <b>456</b> (90.30%) Badly classified: <b>49</b> (9.70%)	Well classified: <b>37</b> (100%) Badly classified: <b>0</b> (0%)	Found outliers: <b>75</b> (100%) Well-classified images in good cluster: <b>914</b> (90.40%) Well-classified images in outlier cluster: <b>97</b> (9.60%)
<b>SMCTAB</b>	Well classified: <b>557</b> (99.64%) Badly classified: <b>2</b> (0.36%)	Well classified: <b>26</b> (100%) Badly classified: <b>0</b> (0%)	Found outliers: <b>55</b> (100%) Well-classified images in good cluster: <b>1097</b> (98.03%) Well-classified images in outlier cluster: <b>22</b> (1.97%)
<b>SMCTAC</b>	Well classified: <b>633</b> (99.37%) Badly classified: <b>4</b> (0.63%)	Well classified: <b>52</b> (94.55%) Badly classified: <b>3</b> (5.45%)	Found outliers: <b>106</b> (96.36%) Well-classified images in good cluster: <b>1263</b> (99.14%) Well-classified images in outlier cluster: <b>11</b> (0.86%)
<b>SOIC</b>	Well classified: <b>609</b> (99.84%) Badly classified: <b>1</b> (0.16%)	Well classified: <b>54</b> (100%) Badly classified: <b>0</b> (0%)	Found outliers: <b>108</b> (100%) Well-classified images in good cluster: <b>1214</b> (99.43%) Well-classified images in outlier cluster: <b>7</b> (0.57%)

Table 2. Testing results

parameters automatically only by means of actual image database without any external information. Therefore it is hard or impossible to modify the ratio of false positive and false negative errors only by means parameters, so general experiment methods, like ROC curve analysis, were avoided.

One of the most important requirements of the research was to find all real errors or outliers (no or very small false negative error), during the false positive rate stays below a quite low limit. The results mentioned above illustrate, that our general methods based on similarity degree can separate the good and bad pictures with very small error rates, independently from the type of images.

#### 4. Conclusions

This paper presented a novel image similarity measurement method for AOI image databases. The algorithm calculates first a golden template image which represents the average image of good components. Next the image under inspection is compared with this golden template on the level of sub-regions, creating the difference profile. The analysis of this error field (maximum value and standard deviation) results the 2-dimensional similarity degree which symbolize how close the actual image is to the golden template. This means, that the developed similarity degree is a relative measurement value.



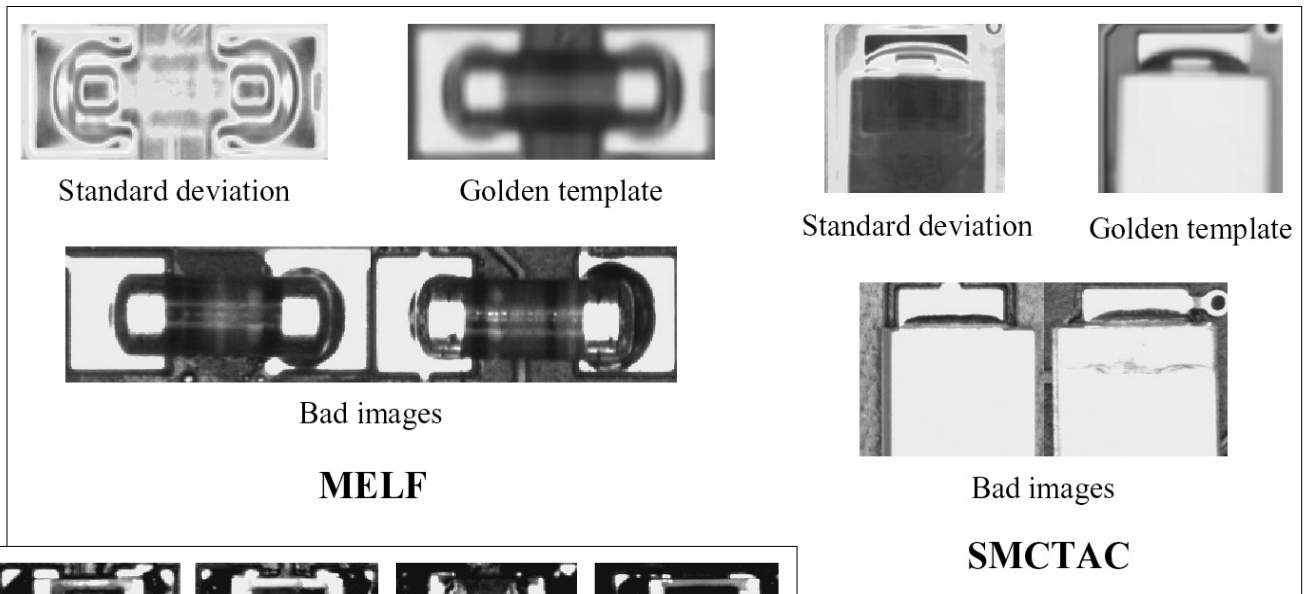


Figure 19.

Illustration of bad classification result

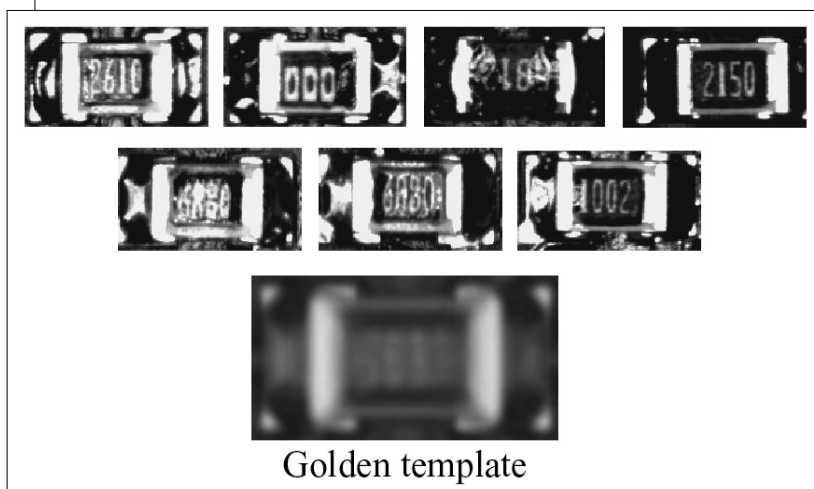


Figure 20.

False-call images in R0805 image base

By means of this measurement method (by analysing the similarity degree vector field), important AOI problem can be solved without serious additional work. This article presented two interesting area: decision support system for human operators at re-inspection process; outlier detection method for filtering the AOI image databases. In the first case, the algorithm uses training images about good and bad components, and calculates their similarity degree vectors.

At re-inspection process, the algorithm determines the similarity degree of the actual error image and determines the appropriate class by means of three different factors (nearest image, smaller average distance, clustering result). The outlier filtering algorithms estimates the golden template as the mean image of all (bad and good) pictures in the database, and by means k-means clustering separates the well-classified and outlier images.

We tested our algorithm on seven different image bases which contain different types of components, like SMT capacitors, resistors, or pins. The experiments show excellent results: in five cases, all real errors or outliers were classified perfectly, in the remaining two image bases the classification rates were 94-97%. The false classification of good images was below 2% in average. This means that our novel decision support and filter-

ing methods based on novel similarity degree calculation are useful also in industrial applications at this stage of our research already. We would like to emphasize again, that our algorithms do not apply outer parameters, they calculate all the stages only by means of actual image database, and they work without any modification on several different types of AOI image bases well.

**Authors**



**TIBOR TAKÁCS** received his Master degree in Computer Science from Budapest University of Technology and Economics in 2007. He is currently working as a Ph.D. candidate at the Department of Control Engineering and Information Technology. His research interest focuses on optimization problems, novel image processing techniques and intelligent solutions in automated optical inspection systems.



**LÁSZLÓ VAJTA** received his Ph.D. degree in Electrical Engineering in 2005 from Budapest University of Technology and Economics, and is currently working as an associate professor at the Department of Control Engineering and Information Technology. His main research activities focus on image processing system and on the area of machine vision solutions.

## References

- [1] Acciani, G., Bruneti, G., Fornarelli, G.,  
A Multiple Neural Network System to Classify  
Solder Joints on Integrated Circuits,  
Int. Journal of Computational Intelligence Research,  
ISSN 0973-1873, Vol. 2, No. 4, pp.337–348, 2006.
- [2] Acciani, G., Bruneti, G., Fornarelli, G.,  
Application of Neural Networks  
in Optical Inspection and Classification of  
Solder Joints in Surface Mount Technology,  
IEEE Transactions on Industrial Informatics,  
Vol. 2., No. 3, pp.200–209, 2006.
- [3] Balakrishnan, N.,  
Handbook of the logistic distribution,  
Marcel Dekker, ISBN 0824785878, 1991.
- [4] Barnett, V., Lewis, T.,  
Outliers in statistical data (3rd ed.).  
John Wiley&Sons, ISBN 0471930946, 1994.
- [5] Belbachir, A.N., Lera, M., Fanni, A., Montisci, A.,  
An automatic optical inspection system  
for the diagnosis of printed circuits based  
on neural networks,  
Industry Applications Conference 2005,  
48th IAS Annual Meeting, Conf. Record of the 2005,  
ISBN 0-7803-9208-6, pp.680–684 (Vol. 1), 2005.
- [6] Gonzalez, R.C., Woods, R.E.,  
Digital Image Processing (3rd ed.), Prentice Hall,  
ISBN 9780131687288, 2008.
- [7] Grinstead, C.M., Snell, J.L.,  
Introduction to Probability,  
ISBN-10 0821807498, ISBN-13 978-0821807491, 1997.
- [8] Grubbs, F. E.,  
Procedures for Detecting Outlying Observations  
in Samples.  
Technometrics (Vol. 11), pp.1–21, 1969.
- [9] Hartigan, J.A.,  
Clustering Algorithms.  
John Wiley&Sons, 1975.
- [10] Hodge, V., Austin, J.,  
A Survey of Outlier Detection Methodologies.  
Artificial Intelligence Review, Springer,  
Vol. 22, No. 2, pp.85–126, 2004.
- [11] Hogg, R.V., Craig, A.,  
Introduction to Mathematical Statistics (5th ed.),  
Prentice Hall,  
ISBN-10 0023557222, ISBN-13 9780023557224, 1994.
- [12] Kong, F., Wang, Y.,  
Reconstruction of Solder Joint Surface Based  
on Shape from Shading,  
3rd Int. Conf. on Natural Computation (ICNC 2007),  
ISBN 9780769528755, pp.58–62, 2007.
- [13] Lee, C.S., Elgammal, A.,  
Dynamic shape outlier detection for human locomotion,  
Computer Vision and Image Understanding,  
Vol. 113, pp.332–344, 2009.
- [14] Lipson, P. R.,  
AOI Systems Simulate Human Brain.  
Test & Measurement World, pp.35–42, 2007.
- [15] Lu, C. T., Kou, Y., Zhao, J., Chen L.,  
Detecting and tracking regional outliers  
in meteorological data.  
Information Sciences,  
No. 177, pp.1609–1632, 2007.
- [16] MacKay, D.,  
Information Theory, Inference and Learning Algorithms,  
Cambridge University Press,  
ISBN-10 0521642981, ISBN-13 9780521642989, 2003.
- [17] Mar, N.S.S., Fookes, C., Yarlagadda, P.K.D.V.,  
Design of automatic vision-based inspection system  
for solder joint segmentation,  
Journal of Achievements in Materials and  
Manufacturing Engineering,  
Vol. 34. Issue 2, pp.145–151, 2009.
- [18] Ong, T.Y., Samad, Z., Ratnam, M.M.,  
Solder Joint Inspection with Multi-angle Imaging  
and Artificial Neural Network,  
The International Journal of Advanced  
Manufacturing Technology, Springer London,  
ISSN 0268-3768 (Print), 1433-3015 (Online),  
Vol. 38, No. 5-6, pp.455–462, 2007.
- [19] Petrakis, E.G.M., Faloutsos, C.,  
Similarity Searching in Medical Image Databases.  
IEEE Trans. on Knowledge and Data Engineering,  
Vol. 9, No. 3, pp.435–447, 1997.
- [20] Prastawa, M., Bullitt, E., Ho, S., Gerig G.,  
A brain tumor segmentation framework based  
on outlier detection.  
Medical Image Analysis, pp.275–283, 2004.
- [21] Vellidoa, A., Lisboa, P.J.G.,  
Handling outliers in brain tumour MRS data analysis  
through robust topographic mapping,  
Computers in Biology and Medicine,  
Vol. 36, Issue 10, pp.1049–1063, 2006.
- [22] Wang, D.Z., Wu, C.H., Ip, A., Chan, C.Y., Wang, D.W.,  
Fast Multi-template Matching Using a Particle  
Swarm Optimization Algorithm for PCB Inspection,  
Lecture Notes in Comp., No. 4974, pp.365–370, 2008.
- [23] Zhan, Y.,  
Advanced image analysis methods for the diagnosis  
of prostate cancer. (PhD dissertation),  
ISBN/ISSN 978 05493 14134, 2008.

# Traffic analysis methods to support decisions at the knowledge plane

PÁL VARGA

BME, Department of Telecommunications and Media Informatics  
 pvarga@tmit.bme.hu

LÁSZLÓ GULYÁS

AITIA International Inc.  
 lgulyas@aitia.ai

Keywords: knowledge plane, monitor plane, traffic mix, traffic matrix

**Traffic analysis of network segments is an effective method to reveal suboptimal configuration, hidden faults and security threats. If the analysis results are promptly acted upon, improvements in service quality are experienced by both the network operator and the end user. The concept of the Knowledge Plane (KPlane), and later the Monitor Plane (MPlane) has been introduced to support Autonomous Networking goals. The tasks of processing the network element, service and traffic information belong to the MPlane. It feeds the KPlane with valuable information, based on which configuration changes are actuated. Although the concept of KPlane is widely used in various levels of network and service management, general traffic analysis is not yet utilized to support decision making procedures. Traffic mix and traffic matrix analysis results are of major interest in the decision making process at the KPlane. In this paper the issues of traffic sensing at the high speed interfaces of the Monitoring Plane are covered, followed by a discussion on traffic mix and traffic matrix analysis methods.**

## 1. Introduction

The optimization of network and service resources and the maximization of end-user experience are not necessarily conflicting terms. The reason for such belief lies in the fact that current network operators and service providers lack of up-to-date, usable information on their traffic. The questions of “how much” of “what” actually are traversed on the various network segments, where is that traffic “originated from” and where is it “distributed toward” are rarely answered.

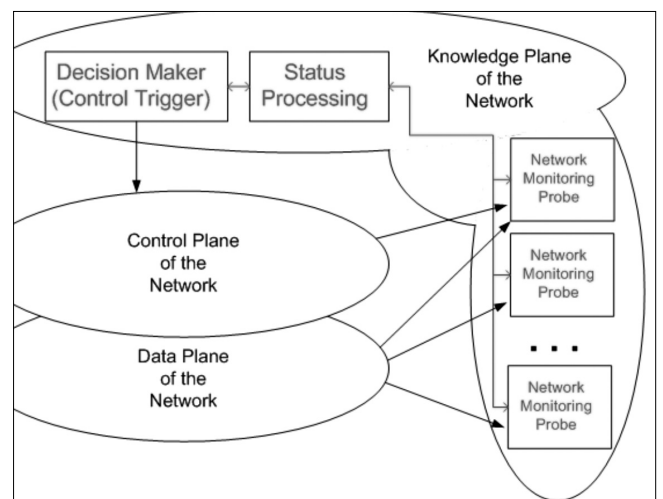
According to the main argument of [1], the users and the operators suffer from the lack of a serious, purposeful optimization effort in the traditional Internet. The transparent core has no knowledge about the data transported, and even if the intelligent edge nodes realize that there is a problem, the core might not be aware of what should be done. The low-level decisions (at the edge) are rarely relate to the higher-level goal (of the core). On the user side this results in meeting the service level agreement only in coarse granularity: it is measured in long periods and more at a network level, rather than on a per-service basis.

The solution for gaining knowledge about network status and traffic characteristics is to gather and process such data, which then provide a basis to trigger corrective actions. The authors of [1] suggest to handle this knowledge in the Knowledge Plane (KP), an abstract entity that completes a triad together with Data Plane and Control Plane (see Fig. 1).

In the original KPlane concept, the input is taken by sensors and the output is given by actuators. A practical variation of this architecture, detailed in [2], splits

the KPlane into *monitoring plane* and *knowledge plane*. The separation of those is an obvious step: the actual “network monitoring units” (sensors) that capture and pre-process traffic data represent the “monitoring plane”, similarly as depicted in Fig. 1. There are further variations and additions to this architecture; we will review these in the section of Related Works, together with a short review of decision making methodologies and practical examples from the field. Fig. 1 depicts the relation between the Knowledge, Control, and Data Planes. The probes/sensors take data from both the control and data planes, and report pre-processed information for the status processing module, where further analysis takes place. The actuator in the model is the de-

Figure 1. Functions of the Knowledge Plane and its connections to Control and Data Planes



cision maker module, which provides triggers for the control plane, completing the self-management cycle.

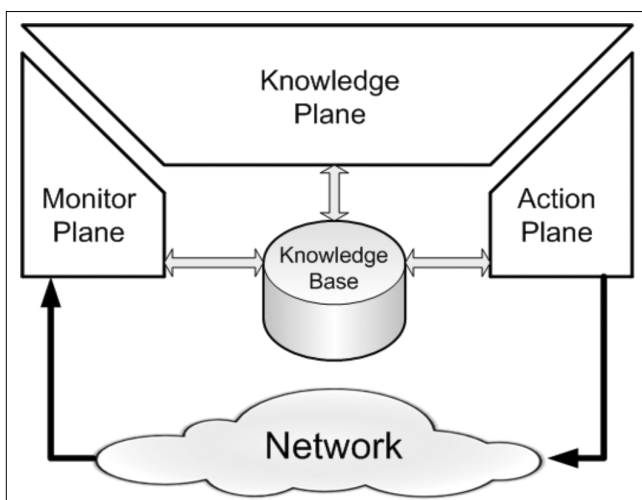
The main source of “knowledge” is the actual traffic of the Control and Data Planes. Although some traffic characteristics can be gathered by analyzing the Control Plane messages, many important applications – such as peer-to-peer (P2P) downloads, video streaming, or interactive voice – hide their control messages, hence their identification is only possible through Deep Packet Inspection (DPI) of the traversed traffic. The aim of Traffic Mix analysis is to determine the distribution of volumes for services and applications utilizing the network. Similarly, Traffic Matrix analysis provides results about traffic volumes – and if possible, further characteristics – broken down by route directions. The second part of this paper discusses our proposed, unique method of Traffic Mix and Traffic Matrix analysis.

## 2. Related work

Kim et al. summarize the research and development ideas and efforts in management of the Future Internet in [4], specifically reviewing the research activity in the EU, USA and Korea. The authors emphasize the common interest and importance of measurements, monitoring, knowledge representation and reasoning. The original idea of introducing a higher level intelligence to the core about its traffic and general status first appeared in [1], where Clark et al. introduced the concept of the Knowledge Plane. Besides providing very clear motivations, this groundbreaking paper suggested to solve networking issues by using methods devised in the field of Artificial Intelligence (AI). Since then, experts of both area – Network Management, and AI – elaborated various versions of the KPlane concept in great depth.

Li described a layered architecture in [3], where NetKP – the network layer – organizes agents to gather and provide valuable information to the higher-level entities, specKPs, which handle and act upon their own interest, i. e., routing optimization or intrusion detection.

Figure 2.  
The original Knowledge Plane concept extended with Monitor and Action Planes



Another variant of splitting is suggested by [6], motivated by the need to get to the kernel of self-functions defined by autonomous networking drives. Hence, the processes in KPlane are based on two loops: a collaborative loop and an adaptation loop. The KPlane itself includes a knowledge base, a reasoning engine, a knowledge sharing process, and a machine learning process. In this model, Monitoring functions remain outside the KPlane.

Dietterich et al. found that the application of distributed, model-based reasoning agents is a feasible and successful approach for certain fault diagnostics tasks that involve the KPlane. In their report (see [5]) one of the main motivations was to involve Machine Learning in KPlane. Although their findings show that these methods can contribute to the KPlane, they do not suggest to have machine learning as a key element of KPlane. Their paper also includes interesting reports on fault detection case studies, including DNS diagnosis, and a scenario where a typo in BGP (Border Gateway Protocol) tables was revealed.

The IST-MUSE project resulted in many ideas and implementations in relation to KPlane. Besides separating the Monitoring Plane from the KPlane in [2,7, and 8], they further introduced the Action Plane (APlane). They also defined a knowledge base that is commonly reachable by KPlane, MPlane and APlane. Fig. 2 depicts their connection and relation to the network. The main motivation in these papers is to eliminate QoS (Quality of Service) and QoE (Quality of Experience) issues in the access network for VoIP, IPTV and other multimedia services. Instead of gathering knowledge from overall data plane traffic, these papers rely on designated protocols (i.e. RTP, Real-time Transport Protocol) and protocol analysis of the control messages.

The Monitor Plane is extensively used in [9] as well, where a complete, “access control list”-based VoIP service management system is described and evaluated. The KPlane in this paper is put in a different context: its functionalities include Call Data Record generation and visualization.

Although KPlane was not mentioned in [11] all of its features appear in the service management framework described in the paper: measurements, monitoring, data processing/mining, decision making, knowledge bases and machine learning. The presented framework has been effectively used for fault detection and elimination for Ethernet services and for VoIP services [11] as well.

A specialized KPlane is suggested in [12] in order to handle current QoS problems with protection routing algorithms in GMPLS over WDM (Generalized Multiprotocol Label Switching over Wavelength Division Multiplexing) networks. This is a clear example of using a variation of the KPlane concept to enhance concrete routing methods’ speed and effectiveness.

It is clear that the concept of Knowledge Plane is widely used in various levels of network and service management. Nevertheless, general traffic analysis is not yet utilized in order to support decision making in

the KPlane. In the following sections we describe the suggested management architecture, traffic analysis concept and two methods to extract valuable information about the traffic mix and the traffic matrix.

### 3. The Monitor Plane

In this paper we follow the architecture suggested in [8] (see Fig. 2), and closely examine the functions and requirements of the Monitor Plane. This function is crystallized at the original definition of autonomous networks, in [13], defining the foursome of “Monitor-Analyze-Plan-Execute” (MAPE) functions. The core function of the MPlane is to provide complete and detailed view of the network and its services. Probes at every element (access nodes, routers, switches, content servers, links, etc.) monitor the element status as well as traffic parameters.

Although built-in probe modules seem convenient, passive probing is more desirable. Active network elements (such as routers or switches) keep their processing priorities to their main job, occasionally leaving the Knowledge Base without information. These occasions of degradation in the status reporting function happen at the worst time from the KPlane’s point-of-view - for practical reasons. It gets degraded at the time when the element is getting overloaded. Coincidentally, such detailed reports of overloading would be the most beneficial for the KPlane. This is why passive probing is more desirable to gather information on these elements.

After capturing the raw data, processed, grouped, and filtered traffic information gets inserted into the Knowledge Base by the probes. Both packet- and flow-level analysis reveal important characteristics on losses, delays, and jitters in the traffic, routing specialties, network structure changes and violations of the SLS (Service-Level Specification).

We are focusing on gathering these characteristics by passive monitoring. In the following subsections we briefly describe the basic requirements and mechanisms enabling this method.

#### 3.1. Basic functions of the probes

The inevitable function of the network monitoring probes is catching, filtering, and preprocessing the traffic. These tasks should be completed for the whole network. Since installing and maintaining such a monitoring network could be an enormous effort for the operator, introducing the MPlane at the highest aggregation parts (i.e. monitoring the fastest links) can be a good decision. Monitoring these relatively few points allows gathering all packets that traverse the network, although some locally looping traffic could be left out of the analysis.

The probes should have the following crucial functionalities:

- *Creating timestamps for the packets.* Time-stamping done by hardware (firmware) facilitates much more pre-

cision than by software, since it avoids possible latencies due to the operating system.

- *Filtering on hardware level.* High-speed traffic (i.e. currently 10 Gbps or above) presently allows no option for on-the-fly filtering in software. Clearly defined, low level filters are very useful: they can dramatically decrease the data to be analyzed.

- *Truncating incoming packets.* For the majority of the network analysis functions, statistics-counting, or fingerprint analysis, it is not necessary to use the whole IP packet, only the first portion of it. A practical example is truncating at 128 bytes, which keeps TCP and IP headers as well as the beginning part of application headers that are helpful for identification, since it contains fingerprints for P2P or video.

- *Traffic processing.* The main traffic processing functionalities are briefed in the next section.

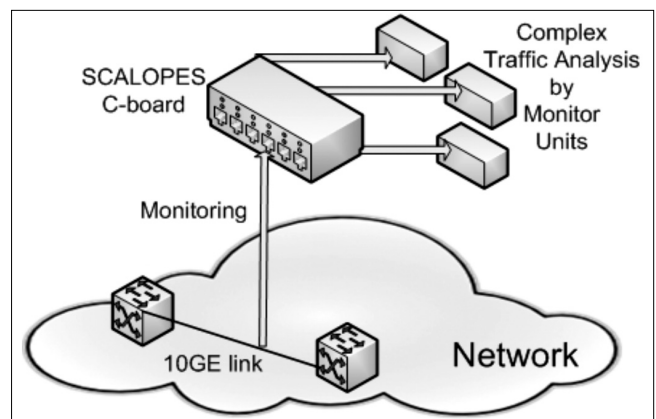
- *Encapsulation and presentation of preprocessed data.* The traffic analysis results must be structured and packed when passed over to the Knowledge Base.

#### 3.2. Traffic processing

The time-stamped, filtered, truncated packets must be processed in order to reveal network and service statuses. Depending on the traffic volume, and the depth of the analysis, this processing can be fed into one or many processors. In order to keep up with the ever increasing traffic and the demand for complex analysis, the processing system must be highly scalable. As discussed earlier, monitoring core links has the advantage of utilizing all through-traffic (that traverses the network), although it requires equipment being able to monitor high-speed (currently 10 Gbps Ethernet) links without frame loss.

For low analysis demand (when one CPU can deal with the challenges), a highly reliable monitoring card, such as SGA10GED can be used to capture the traffic. (This card has been developed as part of the CELTIC TIGER2 project, partially funding our research.) It fits into a PCI slot of an industrial grade PC, where it captures, timestamps, filters, and truncates packets before passing it to the main CPU where Traffic Analysis is performed.

Figure 3. A scalable solution for Traffic Analysis of high-speed network links



In cases where on-the-fly, complex analysis is required on highly utilized links, the SCALOPES C-board is a highly scalable solution. (The C-board has been developed as part of the ARTEMIS SCALOPES project, partially funding our research.). It is a standalone, FPGA-based hardware, equipped with 2x10 Gbps Ethernet interfaces and 16x1 Gbps Ethernet interfaces. When used as part of the Monitor Plane, it is also preprocessing the packets, but rather than passing their data to one CPU, it distributes them among many monitor units through its 1 Gbps Ethernet Interface. The standalone Monitor Units then carry out traffic analysis, and present the results to the knowledge base. *Fig. 3* depicts such a scenario. Detailed description of this system can be found in [14].

The distinct analysis tasks – such as flow separation, application identification, QoS-related parameter calculation per flow/application/route – are managed by separated modules, so the parallel tasks can be run on distinct processors in the same time. Moreover, the inactive modules can be turned off to save power.

The tasks of the monitor units in this architecture are the following:

- collect and decode all the incoming information continuously (in 7/24 manner),
- check filtering rules predefined by the network operator, execute conditional controlled orders/commands (conditional packet saving, alarming),
- structured data storage (raw data, statistics, assays, alerts)
- generation of packet- and flow-level counters on volume, loss, delay, jitter
- generation of specialized traffic reports, such as traffic mix and traffic matrix
- database handling, remote access/query (Remote Capture, Session/Flow Trace)

## 4. Methods for retrieving traffic-specific knowledge

### 4.1. Traffic Matrix calculations

Traffic Matrix is a network planning and development tool. During Traffic Matrix analysis, basic QoS statistics are periodically created on flow-level, and matched to originating and destination routes, network segments, or endpoint pairs (such as IP address(-range) pairs, MPLS tunnel endpoints, etc.). The first step of the analysis is determining the flows by an n-tuple (i.e., “5-tuple”: from-IP, to-IP, from-port, to-port, protocol), and building/refreshing the flow-database. Once the targeted data structure is clarified, the algorithms of Traffic Matrix calculation are of low complexity. Such algorithms are described in [15]. The result of the measurement can be used to display periodical statistics that support network planning or service marketing activities.

The actual Traffic Matrix can easily contain endpoint-pairs in the magnitude of 105. It is challenging to display such huge amount of data in a way that humans

understand. While the raw results should be made available for reference in the Knowledge Base, some kind of data grouping should also be applied for visual presentation. One example of a good solution is to group the matrix elements into network segments, based on their destination addresses. The aim of the grouping algorithm is never to display high, invisible amount of segments (e.g. more than 15). When the operator wishes to peek inside a segment’s statistics, he/she get it displayed as a deeper layer of the matrix.

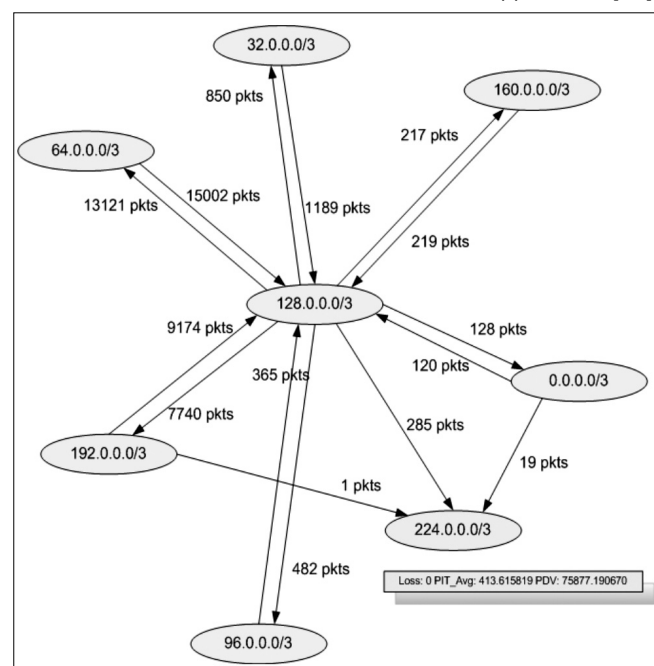
This way the calculated QoS parameters show up in an aggregated manner in the segment-to-segment relation. If the system allows manual definition of segment-creation rules, operators can gather valuable information by grouping their endpoints into various segments. An example screenshot from a solution integrated in our system is shown in *Fig. 4*.

### 4.2. Traffic Mix Statistics

Traffic mix analysis is the classification of traffic flows into application types, and then evaluating these for the service parameters important for the given application type. Flows are classified by means of statistical indicators and, if necessary, behavior heuristics. The most important flow types include video stream, video conference, or simple download of videos, audio stream, VoIP, and peer-2-peer.

An application belonging to a traffic-class can be identified by using static identifiers (e.g. port-based), dynamic identifiers (e.g. changing ports, fingerprints) or by applying packet-level statistics-based evaluation methods (i.e., Naive Bayes). Powerful identification methods for VoIP, video and p2p applications are described in [17-19] respectively. We used these methods successfully during the CELTIC TIGER2 project – see [20] for details.

Figure 4. Screenshot of a Traffic Matrix visualization application [16]



Once a traffic flow is identified (i.e., based on 5-tuple), various metrics are calculated in order to help identifying the traffic-class. These metrics are the following:

- throughput: transferred data bytes per second,
- packet loss: the rate of received packets and total transmitted packets in a given time interval, or during the connection,
- packet delay: depending on the network topology and link load it takes a certain amount of time to receive a packet after it was sent; there is also a gap (delay) between packets on the wire,
- jitter: network load is not always static: as conditions and usage changes over time, packet delay changes as well – this is called jitter,
- round-trip-time: interactive applications require fast replies, which can be characterized with this parameter,
- out of order/duplicated packets.

*Fig. 5* depicts a partial result of one of our measurements at a major ISP. It visualizes the number of parallel VoIP sessions (upper diagram) and the traffic volume (in kbps). The different kind of VoIP traffic are represented with different colors, which are – from bottom to top – a) Skype over UDP, end-to-end; b) Skype over UDP, end-to-office; c) other type of VoIP, d) Skype over TCP.

## 5. Decision Making

Since processing of network status is continuous at the KPlane, and faults/attacks may happen at any time, so decisions on corrective actions have to be made on-the-fly as well. The Action Plane should be notified (instructed) about these actions for execution. Although the accuracy of decisionmaking process is the key, it is limited by the variety of the input information – which is in this case merely traffic-related. Beside the accuracy, speed is also a key factor.

In order to understand the complexity of the decision making problem, a short review the main challenges are necessary. Clark et al. [1] points out three significant issues that need to be addressed by the Knowledge Plane.

1. The KPlane needs to operate in the presence of incomplete and inconsistent information, with the possibility of even misleading or malicious pieces of data.

2. The KPlane needs to be able to handle conflicting or inconsistent high level goals.

3. The KPlane needs to be general and future proof, i.e., the introduction of new technologies and novel applications should be possible. Moreover, the environment in which optimization needs to take place is highly dynamic, where both short and long term changes are possible in the structure and complexity of the network system.

Such challenges are not uncommon in the research and applications of the last decades of Artificial Intelligence (AI) literature. In particular, multi-agent systems (see [21]) are often proposed to handle such challenges.

A multi-agent system (MAS) is a system composed of multiple interacting intelligent agents, where intelligent agents, shortly put, mean autonomous decision making entities with individual information processing capabilities and individual goals. Such agents can naturally incorporate different viewpoints or goals in a system and also provide a natural way to embody components with different levels of data access.

As a consequence, however, the goals and actions of agents in a multi-agent system may partially be aligned or conflicting. Also, even if conflicts are missing or resolvable, information may be unevenly distributed among the agents. Therefore, agents interact and try to resolve conflicts and collaborate according to various protocols and methods. A vast body of the recent AI and MAS literature deals with conflict management, collaboration and cooperation, and distributed optimization in such systems (see [22-24]).

It is worth pointing out that the agent metaphor is a natural abstraction layer to describe conflicting or inconsistent goals – independent of the particular problem at hand. This is also true for matters of trust (cf., malicious information). This way, these issues can be handled by general solution methods and need not be developed for each particular application domain. In other words, these challenges of the Knowledge Plane may be handled by “canned solutions” developed in other research domains.

Multi-agent systems are often said to provide a solution for the introduction of novel applications as well. The idea behind this proposal is that if a new application or requirement appears, a new agent (or bunch of new agents) may be introduced to the system at any point in time. With the general conflict resolution and collaboration protocols in place, the new goals and requirements represented by the new agents will be seamlessly integrated in the system. Similarly, should some of the goals rendered outdated by time, the sets of agents can be gracefully eliminated from the multi-agent system.

Still, in order to proceed towards a decision making solution in the autonomous networking field, further research is required. Although recent AI-related research should be exploited in the area of network management, currently there are no real-time, scalable solutions available. The canned multi-agent solutions have not yet broken into the network management field, and the few prototype systems (e.g. the one described in [25]) remained prototypes up to now.

Due to the aforementioned limiting factors, a scalable, high-performing, yet less accurate solution is suggested for decision making: rule-based reasoning. It is used with success in many areas; see [10] as an example. In connection with the KPlane, we continue future research in the AI-field, and further developments and integration toward a scalable, rule-based reasoning engine that is applied in a distributed manner throughout the KPlane.

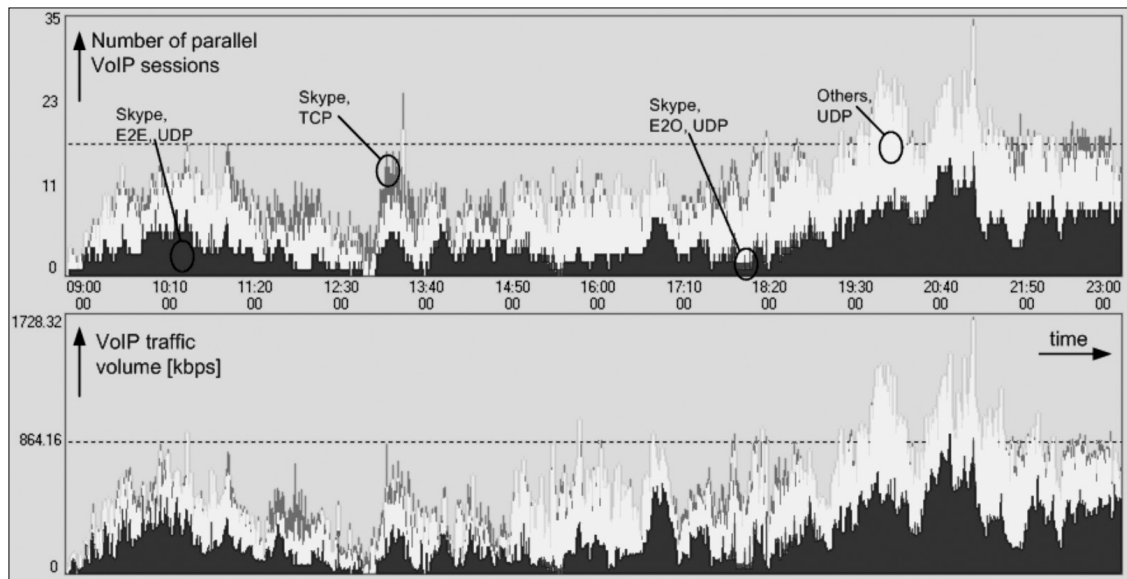


Figure 5.  
VoIP Portion  
visualization of  
a Traffic Mix  
analysis

## 6. Conclusions

Network efficiency and service quality are required to be kept at high standards for both the network operators' and the users' point of view. This can be achieved by keeping the network and service status under continuous monitoring. When inefficiencies become evident, or failures appear, corrective actions should be orchestrated. A recent concept to cover the autonomous loop of "Monitor-Analyze-Plan-Execute" (MAPE) is to utilize a Monitor Plane to gather and process information, introduce a Knowledge Plane to continuously process network and service status according to the requirements, and carry out commands for corrective actions by Action Plane entities.

In this paper we closely examined the tasks of the Monitor Plane, and suggested a scalable architecture to gather and process network traffic in a distributed manner. Since decisions at the Knowledge Plane should be partially made by traffic information, two important traffic analysis methods have been introduced to support decision making. Traffic Mix analysis requires a flow-based approach, where flows get classified into application types based on their characteristics, and then evaluated by related QoS metrics. Traffic Matrix analysis is important for both network and service planning, since it outputs the traffic volumes and characteristics correlated with the traffic endpoints. This information can efficiently support status processing and decision making at the KPlane, since currently these are the most sophisticated traffic-related analysis methods that human experts use during network/ service evaluation and planning.

The brief review of distributed multi-agent systems suggests that based on their problem statement, such systems – when available, – should be able to cover the requirements of the ideal KPlane. Nevertheless, this field requires further applied research, since a scalable, high-performing, tangible MAS – that could serve as a KPlane – is still missing.

## Acknowledgements

This work has been partially funded in the framework of the CELTIC TIGER2 project (CP5-024) as part of the EU-REKA cluster program.

Parts of the research leading to these results has also received funding from the ARTEMIS Joint Undertaking under grant agreement n° 100029 and from the Hungarian National Office for Research and Technology.

The authors would like to thank the contribution of Gyorgy Horvath (BME-TMIT), Laszlo Kovacs (AITIA) and Peter Tatai (AITIA).

## Authors



**PÁL VARGA** is a lecturer at Department of Telecommunications and Media Informatics, at Budapest University of Technology and Economics, Hungary, where he got his MSc from. He is currently proceeding toward his PhD there. Besides, he is the head of the Telecommunications division at AITIA International Inc. His main research interests are service and network management, and traffic analysis. He focuses his research to network performance measurements, fault localization, traffic classification, end-to-end QoS and SLA issues.



**LÁSZLÓ GULYÁS** is a graduate of the Lóránd Eötvös University, Hungary, from where he received his PhD, MSc and BSc degrees, all in Computer Science. He is assistant professor at the Department of History and Philosophy of Science, Lóránd Eötvös University, Budapest. He is also a research partner at AITIA International Inc and a fellow at Collegium Budapest (Institute for Advanced Study). He is a member of the Scientific Advisory Board of the Simulation Center of the Informatics Cooperative Research and Education Center of the Lóránd Eötvös University. Dr. Gulyás has authored several book chapters (5+) and journal articles (5+), and published many conference papers (40+). He participated in two international research consortia under the European Commission's 6th Framework Programme and was project leader or participant in 4 R&D projects funded by the Hungarian Government. His main research interests are computational multi-agent systems where he has worked on 'engineering' desired emergent phenomena. He is currently working on agent-based models of social systems.



## References

- [1] Clark, D.D., Partridge C., and Ramming, J.C., "A knowledge plane for the Internet", In Proc. of the 2003 Conference on Applications, Technologies, architectures, and protocols for computer communications, August 25-29, 2003, Karlsruhe, Germany
- [2] De Vleeschauwer, B., Van de Meerssche, W., Simoens, P., De Turck F, Dhoedt, B., Demeester, P., Gilon E., Struyve, K., Van Caenegem T., "On the Enhancement of QoE for IPTV Services through Knowledge Plane Deployment", In Proc. of Broadband Europe, December 11-14, 2006, Geneva, Switzerland
- [3] Li, J., "Agent Organization in the Knowledge Plane", PhD. Dissertation, MIT, 2008.
- [4] Kim, S., Won, Y.J., Choi, M., Hong, J.W., Strassner, J., "Towards Management of the Future Internet", In Proc. of the 1st IEEE/IFIP Workshop on Management of the Future Internet, Long Island, NY, USA, June 5, 2009.
- [5] Dietterich, T., Harvey, B.T., Miller, D., Jones, D., Gray, A., Fern, A., Tadepalli, P., "Machine Learning for the Knowledge Plane", DARPA Technical Report, Oregon State University, June 2006.
- [6] Mbaye, M., Krief, F., "A Collaborative Knowledge Plane for Autonomic Networks", In Autonomic Communication, Ed. by Vasilakos, A.V., Parashar, M., Karnouskos, S., Pedrycz, W., ISBN 978-0-387-09752-7, Springer, 2010.
- [7] Simoens, P., Vleeschauwer, B.D., Van de Meerssche, W., De Truck, F., Dhoedt, B., Demeester, P., Van Caeneghem, T., Struyve, K., Dequeker, H., Gilon, E., "RTP Connection Monitoring for Enabling Autonomous Access Network QoS Management", In Proc. of 12th European Conference on Networks and Optical Communications, Stockholm, Sweden, June 2007.
- [8] Latre, S., Simoens, P., Vleeschauwer, B.D., Van de Meerssche, W., De Truck, F., Dhoedt, B., Demeester, P., Van Den Berghe, S., Gilon, E., "Design for a Generic Knowledge Base for Autonomic QoS Optimization in Multimedia Access Networks", In Proc. of 2nd IEEE Workshop on Autonomic Communications and Network Management, Salvador, Brazil, April 2008.
- [9] Kobayashi, A., Ishibashi, K., "VoIP Measurement Architecture using Data Mediation", In Proc. of IPOM 2009, Venice, Italy, LNCS 5843.
- [10] Varga, P., Moldovan, I., "Integration of Service-Level Monitoring with Fault Management for End-to-End Multi-Provider Ethernet Services", IEEE Transactions on Network and Service Management, Vol.4 No.1, 2007.
- [11] Varga, P., Moldovan, I., Molnar, G., "Complex Fault Management Solution for VoIP Services," In Infocommunications Journal, Vol. 60, pp.15-21, December 2005.
- [12] Urra, A., Calle, E., Marzo, J.L., "Adding new Components to the Knowledge Plane in GMPLS over WDM Networks", In Proc. of IEEE Workshop on IP Operations and Management, Beijing, China, October 11-13, 2004.
- [13] IBM, "Architectural Blueprint for Autonomic Computing", 2003.
- [14] Plosz, S., Moldovan, S., Varga, P., Kantor, L., "Dependability of a Network Monitoring Hardware", In Proc. of DEPEND 2010, Venice, Italy
- [15] AITIA, BME, "Report on New Architectural Platform and Specification of Example SW Code for Analysis", ARTEMIS SCALOPES Deliverable DA1.3., 2010.
- [16] Szendrei, G., "Calculation and visualization of Traffic Matrices", Technical Report, BME-TMIT, 2010.
- [17] Bonfiglio D., Mellia, M., Meo, M., Rossi, D., Tofanelli, P., "Revealing Skype Traffic: When Randomness Plays with You", In Proc. of SIGCOMM, Japan, 2007.
- [18] Varga, P., Kovacs, L., Moldovan, I., Illes, A.Cs., Kun, G., Sey, G., Turzo, G., "Analysis of Media Communication over the Internet", Technical Report for Hungarian Telecom, 2007.
- [19] Karagiannis, T., Broido, A., Faloutsos, M., Kc claffy, "Transport Layer Identification of P2P Traffic", In Proc. of the 4th ACM SIGCOMM Conference on Internet measurement, Sicily, Italy, 2004.
- [20] Dorgeuille, F., Varga, P., Betoule, C., Thouenon G., Petitdemange, G., Palacios J.F., "Rationales and scenarios for investigations on next generation of access, backhauling and aggregation networks", CELTIC TIGER2 Technical Report, 2009.
- [21] Wooldridge, M., "An Introduction to MultiAgent Systems", John Wiley & Sons Ltd, 2002, ISBN 0-471-49691-X
- [22] Lander, S.E., "Issues in multiagent design systems", IEEE Expert, April 1997.
- [23] Tessier C., Chaudron L., Mueller, H.J., "Conflicting agents: conflict management in multi-agent systems", Kluwer Academic Publishers, 2001.
- [24] Hirayama, K., Yokoo, M., "The distributed breakout algorithms", In Artificial Intelligence, Vol. 161, pp.89-115, 2005.
- [25] Gaiti, D., Pujolle, G., Salaun, M., Zimmermann, H., "Autonomous Network Equipments", In LNCS 3854, Springer, 2006.

## Our reviewers in 2010

*The quality of a research journal depends largely on its reviewing process and, first of all, on the professional service of its reviewers. It is my pleasure to publish the list of our reviewers in 2010 and would like to express my gratitude to them for their devoted work.*

Your Editor-in-Chief

- Boldizsár Bencsáth**, BME, Hungary  
**László Bokor**, BME, Hungary  
**Isabelle Boyer Heard**,  
 Orange Labs, France Telecom, France  
**Aldo Campi**,  
 University of Bologna, Italy  
**Tibor Cinkler**, BME, Hungary  
**Jaume Comellas**,  
 UPC, Spain  
**László Czuni**,  
 University of Pannonia, Hungary  
**Zsolt Félegyházi**,  
 UCB, USA  
**Giuseppe Ferraris**,  
 Telecom Italia, Italy  
**Ádám Földes**, BME, Hungary  
**Péter Fülöp**, BME, Hungary  
**Győző Gódor**, BME, Hungary  
**Ivan Gojmerac**,  
 FTW – Telecom. Res. Center, Austria  
**Oscar Gonzalez de Dios**,  
 Telefonica, Spain  
**Enrico Gregori**,  
 CNR IIT, Pisa, Italy  
**Gábor Gulyás**, BME, Hungary  
**Khairi Hamdi**,  
 University of Manchester, UK  
**Franz Hartleb**,  
 T-Systems, Germany  
**Attila Gábor Hilt**, BME, Hungary  
**Vlasta Hudek**,  
 Croatian Telecom Inc., Croatia  
**Amália Iványi**,  
 University of Pécs, Hungary  
**Gábor Jeney**, BME, Hungary  
**Zoltán Kanizsai**, BME, Hungary  
**Csaba Kántor**,  
 Scientific Assoc. for Infocomm., Hungary  
**Tibor Kolos**,  
 Széchenyi University, Hungary  
**Mario Kusek**,  
 University of Zagreb, Croatia  
**Péter Laborczi**,  
 Zoltán Bay Foundation, Hungary  
**Marc De Leenheer**,  
 Ghent University, Belgium  
**Gábor Lencse**,  
 Széchenyi University, Hungary
- Ignac Lovrek**,  
 University of Zagreb, Croatia  
**Maja Matijasevic**,  
 University of Zagreb, Croatia  
**Oscar Mayora**,  
 CREATE-NET, Italy  
**Miljenko Mikuc**,  
 University of Zagreb, Croatia  
**Annalisa Morea**,  
 Alcatel-Lucent, France  
**Hussein Mouftah**,  
 University of Ottawa, Canada  
**Ralf Nitsch**,  
 T-Systems, Germany  
**Szabolcs Nováczki**, BME, Hungary  
**Géza Paksy**, BME, Hungary  
**László Pap**, BME, Hungary  
**Michał P. Pióro**,  
 Warsaw University of Technology, Poland  
**Ricardo Romeral**,  
 University Carlos III of Madrid, Spain  
**Oana Schnitter**,  
 T-Systems, Germany  
**Robert Schulcz**, BME, Hungary  
**Alexandros Stavidas**,  
 University of Peloponnese, Greece  
**Róbert Szabó**, BME, Hungary  
**Sándor Szabó**, BME, Hungary  
**György Takács**,  
 Péter Pázmány Catholic University, Hungary  
**Ioannis Tomkos**,  
 AIT, Greece  
**Eva Marín-Tordera**,  
 Technical University of Catalonia, Spain  
**Sofie Verbrugge**,  
 Ghent University, Belgium  
**Gyula Veszely**, BME, Hungary  
**Lajos Vonderviszt**,  
 Széchenyi National Library, Hungary  
**Krzysztof Wajda**,  
 AGH University of Technology, Poland  
**Gergely Zaruba**,  
 University of Texas at Arlington, USA  
**Honggang Zhang**,  
 Zhejiang University, China  
**Zoltán Zsóka**, BME, Hungary

(\* BME – Budapest University of Technology and Economics)

---

### Infocommunications Journal

---

**Editorial Office** (Subscription and Advertisements):  
 Scientific Association for Infocommunications  
 H-1055 Budapest, Kossuth Lajos tér 6-8, Room: 422  
 Mail Address: 1372 Budapest Pf. 451. Hungary  
 Phone: +36 1 353 1027, Fax: +36 1 353 0451  
 E-mail: info@hte.hu  
 Web: www.hte.hu

**Articles can be sent also to the following address:**  
 Budapest University of Technology and Economics  
 Department of Telecommunications  
 Tel.: +36 1 463 3261, Fax: +36 1 463 3263  
 E-mail: szabo@hit.bme.hu

**Subscription rates for foreign subscribers:**  
 4 issues 50 USD, single copies 15 USD + postage

Publisher: PÉTER NAGY • Manager: ANDRÁS DANKÓ

HU ISSN 2061-2079 • Layout: MATT DTP Bt. • Printed by: Regisztr Kft.