JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

4/2018

Editorial Board

JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

Preface

The advent of 5G networks – what do you need to know?

Undoubtedly, the advent of 5G networks will permanently change the future of telecommunications, affecting the professional work of every engineer. In light of the above, a brief outline of the standard's details is presented herein to emphasize the significance of the new concept.

Fifth-generation networks have been, until recently, an unclear future prospect, as no specific details concerning their operation were available. Answers to the "what is 5G?" question were merely speculations.

Although many technical aspects of the 5G architecture and even of the 5G technologies still remain unclear, some technical solutions have been assuming a more specific shape over the past few years. The most probable paths along which the 5G technology will be developing (with some of them not confirmed and subject to potential change) are described below.

From the technical point of view, the type of modulation used is of primary importance, and a lot of attention has been devoted to this specific aspect. Selection of the modulation type was preceded by a detailed analysis of the pros and cons of various techniques. The standardization body finally decided that cyclic prefix orthogonal frequency division multiplexing (CPTD) will be used in fifth generation networks, with QPSK, 16-, 64- or 256-QAM sub-carrier modulation. Such a choice was based on the numerous advantages of this particular solution.

The OFDM technique has been known since the 1960s and started to evoke some serious interest in the late 1980s, when research concerning GSM networks was conducted, and later, when the UMTS standard was developed. However, OFDM failed to be implemented in any of the abovementioned cases, as the computing power of processors available at that time was insufficient.

The technical progress in this area has enabled the popularization of OFDM in, inter alia, the following: digital audio and video broadcasting systems (DAB, DVB), asymmetric digital subscriber line data transmission technology (ADSL) and data transmission via power

lines (PLC). The popularity of OFDM results from the fact that it provides high data rates, simultaneously eliminating signal distortion caused by the phenomenon of multipath inter-symbol interference (ISI). Doing away with that type of interference is a serious issue in other types of modulation.

In OFDM, a data stream with a high bitrate is split into several slower substreams that are transmitted simultaneously using multiple carriers, which extends the duration of one symbol. As a result, signals that are reflected by multipath phenomena and reach the receiver with a delay, exert a smaller impact on the quality of received signal. In addition, in the CP-OFDM modulation variant, inter-symbol interference is prevented by the interval between consecutive symbols, the so-called cyclic prefix (CP). CP is a copied symbol end inserted at its beginning. Thanks to its application, if the delay between the original signal and the reflected signal does not exceed its length, the transmitted information can be easily retrieved at the receiver.

Thanks to the fact that OFDM is not a new method, solutions have been developed for LTE networks that will now reduce the cost of its adaptation to the needs of the 5G standard. The possibility of implementing multi-antenna transmission systems (MIMO-OFDM) is a considerable advantage as well. Furthermore, the plans assume that the multi-user (MU) version of the multi-antenna MIMO transmission protocol will be implemented in 5G networks. New beamforming techniques that will be introduced in 5G NR networks will additionally maximize the strength of the signal sent towards the receiver, simultaneously reducing, to the minimum, the strength of signals disseminated in other directions. Future NR systems are also expected to rely on the Massive MIMO technology, where the number of antennas is increased to several hundred.

It is worth noting that OFDM has a few drawbacks, too. One its limitations is the high peak-average power ratio (PAPR). At higher values, the operating point of the RF power amplifier shifts to the linear part of the characteristic to avoid signal distortion. This causes a reduction in efficiency. In the case of base stations, higher power consumption is not a big issue, but in the mobile, battery-powered user devices, this is an unfavorable phenomenon. Therefore, in the case of 5G uplink, the DFT-S-OFDM technique is used with the following modulations: $\pi/2$-BPSK (a novelty), as well as 16-, 64- or 256-QAM. The advantage of DFT-S-OFDM is its ability to achieve a low PAPR coefficient, while simultaneously ensuring multipath immunity.

The importance of mmWave

5G networks will operate on numerous frequencies, both higher and lower. The range of below 1 GHz will be used, e.g. 600, 700 and 800 MHz bands. As far as the range between 1 and 6 GHz used in China, Europe, South Korea and Japan is concerned, the 3.3–3.8 GHz section is of particular interest. Great hopes are also related to the millimeter wave (mmWave) band, especially 52.6 GHz, although research is already conducted on relying on 64–71 GHz and 71–76 GHz bands in wireless communication. Examples of bands that have been pre-reserved for 5G NR networks in different parts of the world include: 24.25–27.5 GHz and 40.5–43.5 GHz (Europe), 27.5–28.35 GHz, 37–38.6 GHz and 38.6–40 GHz (USA), 24.75–27.5 GHz and 37–43.5 GHz (China).

The transition to higher operational frequencies will have its consequences. The higher the carrier frequency, the greater the phase noise. The narrower the spacing between subcarriers (subcarrier spacing, SCS), due to phase noise, the higher is error vector magnitude (EVM) which characterizes the quality of modulation. Also, the higher the carrier frequency, the higher the Doppler shift. With the user's speed of several km/h and the frequency of several dozen GHz, it can reach up to several percent of the space between subcarriers.

On the other hand, the larger the SCS, the shorter the OFDM symbol, and, therefore, the CP prefix. This, in turn, results in a greater impact of the multipath phenomenon on the quality of transmission. Therefore, the choice of the width of the gap between the subcarriers requires a compromise solution. That is why a decision has been made that it will be scalable in 5H NR networks. The multiples of 15 kHz are allowed: 15, 30, 60, 120 and 240 kHz. This basic value has been chosen to allow the coexistence of 5G and 4G networks, where SCS equals 15 kHz. With the change of the subcarrier gap, the cyclic CP prefix is also shortened or extended, respectively. This will enable to adjust its length to propagation conditions prevailing in a given communication channel. The scalability of the frame is also a consequence of SCS' flexibility. This feature will ensure compatibility of 5G NR networks.

5G networks will also introduce a completely new concept of dividing the channel into parts, the so-called Bandwidth Parts (BWP). Each BWP will consist of a group of adjacent PRB blocks and will have an SCS and a CP assigned thereto. The individual BWPs will be allocated to different users according to their needs and capabilities, which will ensure energy savings.

Undoubtedly, the 5G technology will be a key factor in creating a commonly available ecosystem of the future. Due to its breakthrough functionality, this subject will certainly be the focal point of numerous research papers, and the findings thereof will also be presented in our Journal.

I would like to take this opportunity to declare the era of next generation telecommunication networks to have officially begun.

<div align="right">
Robert Magdziak, Ph.D.
Managing Editor
</div>

A Comparative Study of PSO and CMA-ES Algorithms on Black-box Optimization Benchmarks

Paweł Szynkiewicz

Research and Academic Computer Network (NASK), Warsaw, Poland

Abstract—Numerous practical engineering applications can be formulated as non-convex, non-smooth, multi-modal and ill-conditioned optimization problems. Classical, deterministic algorithms require an enormous computational effort, which tends to fail as the problem size and its complexity increase, which is often the case. On the other hand, stochastic, biologically-inspired techniques, designed for global optimum calculation, frequently prove successful when applied to real life computational problems. While the area of bio-inspired algorithms (BIAs) is still relatively young, it is undergoing continuous, rapid development. Selection and tuning of the appropriate optimization solver for a particular task can be challenging and requires expert knowledge of the methods to be considered. Comparing the performance of viable candidates against a defined test bed environment can help in solving such dilemmas. This paper presents the benchmark results of two biologically inspired algorithms: covariance matrix adaptation evolution strategy (CMA-ES) and two variants of particle swarm optimization (PSO). COCO (COmparing Continuous Optimizers) – a platform for systematic and sound comparisons of real-parameter global optimization solvers was used to evaluate the performance of CMA-ES and PSO methods. Particular attention was paid to the efficiency and scalability of both techniques.

Keywords—benchmarking, black-box optimization, CMA-ES, global optimization, PSO, stochastic optimization.

1. Introduction

Many issues related to real-life problems require that the optimal solution be calculated. Traditionally, optimization problems are solved using deterministic solvers which normally assume that the objective function and the set of admissible solutions are convex and known in an analytical form. In practice, however, there are many problems for which an algebraic model is missing. Either, we lack the insight into the system to be optimized and the model is entirely unavailable, or its analytical form is too complicated and intractable to conventional optimization solvers. In the latter case the load of mathematical and practical knowledge fed into the optimization model usually results in numerous formulas the solution of which can be obtained only numerically. In general, problems for which algebraic conventional optimization solver models are un-

suitable or entirely unavailable, are referred to as black-box problems. Thereby we assume that the black-box can be queried through a simulation, experimental measurements or the so-called *surrogate model* to provide crucial performance characteristics [1]–[5] for specified values of system inputs. Practical applications in network system design, cybersecurity, large scale systems modeling, optimization and control, etc. are discussed in [6]–[11].

Let us consider the following black-box optimization problem:

$$\min_{x \in \mathfrak{R}^{dim}} f(x)$$
$$\text{subject to: } x_i^{min} \leq x_i \leq x_i^{max}, \quad i = 1, \dots, dim, \quad (1)$$

where f is the real-valued, *dim*-dimensional function. In general, f can be a non-convex, non-smooth, ill-conditioned or multi-modal. It is assumed that in the above problem the function values of the evaluated search points x are the only accessible information. The search points to be evaluated can be freely chosen. Hence, the search cost is equal to the number of function evaluations executed in order to reach the target solution.

This paper addresses issues associated with the black-box optimization benchmarking (1) for the comparison of two biologically-inspired global optimum calculation algorithms, i.e., the covariance matrix adaptation evolution strategy (CMA-ES) [12], [13] and particle swarm optimization (PSO) [14]. CMA-ES and PSO have already proved successful in solving various black-box problems. The aim of the research was to examine how well CMA-ES and PSO perform on both well- and ill-conditioned optimization problems, and how strongly the efficiency of both algorithms depends on the complexity of the problem and on the prescribed accuracy of the solution.

All experiments were conducted with the use of the black box optimization benchmarking (BBOB) test bed [15] that provides numerous testing problems with various characteristics, dimensions and degrees of complexity. These problems are divided into groups, each with specific characteristics of the objective function, i.e., separable, moderate, ill-conditioned, multi-modal and weakly structured multi-modal functions.

The remainder of this paper is organized as follows. A short survey of black-box techniques is presented in Section 2. A brief description of two bio-inspired algorithms, CMA-ES and PSO in their local and global versions, is presented in Section 3. The overview of implementation, organization and usage of the benchmark COCO and BBOB test bed platform is presented in Section 4. The experimental procedure and the performance measures are presented in Section 5. The results of performance evaluation of both optimization algorithms, conducted with the use of various benchmarks, are presented and discussed in Section 6. Finally, conclusions are drawn in Section 7.

2. Related Work

Many optimization techniques that could be employed for solving the black-box optimization problem (1) have been reported in literature. The overview of various techniques is presented in [2]–[3], [16]–[18]. The algorithms can be classified into the following groups: stochastic approximation techniques (gradient-based approaches), sample path optimization, response surface methodology, deterministic search methods, random search methods, heuristics and metaheuristics.

A stochastic approximation is the well-known gradient search method that is similar to the steepest descent gradient algorithm. The procedure requires a gradient estimation. A computer simulation is performed to obtain estimates of the gradient. It seems that the simulation outcome has to contain gradient evaluations. The advantages and disadvantages of this technique are discussed in detail in [5].

Stochastic gradient algorithms need a simulation run for every iteration in order to calculate the gradient value (K iterations require at least K experiments). In the sample path method [5], the original problem is converted into an approximated deterministic problem. The approximation of the objective function f in (1) is calculated based on simulations performed for a randomly generated set of independent observations. Then, standard optimization algorithms are applied to locate the optimal solution.

Response surface methodology [5], [16] is a sequential strategy based on local approximation $F(x, \alpha^{(k)})$ of the performance f in the neighborhood of x, where the parameters α are calculated using simulations performed every k-th iteration. Next, the minimal value of $F(x, \alpha^{(k)})$ is calculated. The process is repeated until the acceptable solution is found.

Standard deterministic search techniques [18], [19], i.e. algorithms developed by Hook and Jeeves, Rosenbrock or nonlinear simplex (as Nelder and Mead) or complex methods can be applied to solve non-differentiable simulation optimization problems.

As the capabilities of modern computers increase, we can observe a growing interest in the development of the global optimization methods. Global optimization algorithms are linked with the computation of a global solution of non-convex, non-smooth, ill-conditioned and multi-extreme problems. The greatest challenge in the process of designing such algorithms consists in deciding whether the local optimum is a global one in the absence of any local criteria. Over the past decades, numerous theoretical and computational contributions, whose results are described broadly in literature, helped solve such problems. Many of the developed and widely recommended techniques are based on random searches [5], [17], [18], [20]. Pure random search, multi-start local search and controlled random search methods belong to this category. Many algorithms utilize random search and are biologically- or heuristics-inspired by biology or physics. Genetic algorithms, evolutionary strategies, simulated annealing, swarm optimization are all of the heuristic nature [13], [14], [21]–[23]. Randomized search algorithms are regarded to be flexible, robust and less demanding in terms of problem properties than deterministic methods. Unfortunately, efficient stochastic algorithms require a large number of iterations and objective function evaluations, especially for high dimension problems. Therefore, the efficiency and scalability of solvers is often a key issue.

3. Description of Optimization Algorithms

3.1. CMA-ES

The Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [12] is an evolutionary algorithm based on Gaussian mutation and deterministic selection. Evolutionary strategies are stochastic search methods inspired by the principles of biological evolution typically using a multivariate normal mutation distribution. They operate on a set of search points. CMA-ES is considered to be one of the best choices against ill-conditioned, non-convex black-box optimization problems in the continuous domain. The general algorithm to solve the black-box problem (1) is to sample a numerous independent points from a given distribution P, evaluate these points based on their performance measures f and update the distribution parameters. All these steps are executed until the termination criterion is met. In the CMA-ES algorithm, P is a multivariate normal distribution that is a generalization of the one-dimensional (univariate) normal distribution to higher dimensions. Hence, a random vector is said to be n-variate normally distributed, if every linear combination of its n components has a univariate normal distribution. Normal distribution is a good candidate for randomized searches – its entropy for the mean values, variances and covariances given is the largest of all distributions in $\Re^n$ and the coordinate directions are not distinguished in any way. Therefore, in CMA-ES, a population of new search points (set of individuals) is generated by sampling a multivariate normal distribution. In every iteration k new individuals $x_i^k \in \Re^n$ are calculated as:

$$x_i^{k+1} = m^k + \sigma^k \times \mathcal{N}_i^k(0, C^k), \quad i = 1, \ldots, I, \quad (2)$$

where m^k and C^k denote the approximated mean value and the $n \times n$ covariance matrix of the search distribution at iteration k, $\sigma^k > 0$ is the standard deviation – step-size at the k-th iteration, $\mathcal{N}_i^k(0, C^k)$ is a normal distribution with the mean 0 and I is a population size. Hence, the mutation is performed by a perturbation with a covariance matrix which is iteratively updated to guide the search towards areas with expected lower objective values.

After generation of the population of individuals, they are evaluated on f (1), sorted and transformed according to (2). Upon every iteration, all distribution parameters (m^k, C^k, σ^k) are updated.

3.2. PSO

The particle swarm optimization (PSO) algorithm [14] is a simple population-based stochastic optimization technique that emerged from the simulations of the behavior of bird flocks and fish schools. At each iteration a set of points (a swarm of particles) evolve their position in the search space with a velocity vector associated with each particle to find the global optimum. A new population of particles is generated from the previous swarm using randomly generated weights and parameters specific to the algorithm. Hence, both positions and velocities of the i-th particle $x_i^k \in \mathfrak{R}^n$ ($i = 1, \ldots, I$) are updated for every k iteration ($k = 1, \ldots, K$) according to the following rules:

$$x_i^{k+1} = x_i^k + v_i^{k+1} , \qquad (3)$$

where x_i^{k+1} and v_i^{k+1} denote the position and the velocity of the i-th particle at the $k+1$ iteration, respectively. The velocity is updated as:

$$v_i^{k+1} = w v_i^k + \varphi_1 U_1^k (x_{i_{opt}}^k - x_i^k) + \varphi_2 U_2^k (x_{i_{nopt}}^k - x_i^k) , \qquad (4)$$

where w is the weighting parameter, φ_1 and φ_2 denote weights of global and local information, U_1^k and U_2^k are $n \times n$ diagonal matrices with elements randomly drawn from a uniform distribution $[0,1]$, $x_{i_{opt}}^k$ is the best position of the i-th particle found so far and $x_{i_{nopt}}^k$ is the best position of all particles within the neighborhood of x_i^k or in the special case, within the swarm.

The first component in (4) makes a particle move in the previous direction, the second one makes a particle return to the best position calculated so far and the last component makes a particle follow the best position in the current iteration (within neighborhood or whole swarm). The main disadvantage of the PSO algorithm is its sensitivity to a velocity – if the velocity is too low the convergence speed is low, if it is too high, the algorithm falls into the local minimum.

In the research presented two versions of PSO with different concepts of $x_{i_{nopt}}^k$ in (4) were considered:

- Global-best PSO algorithm (PSO-glob) using a star topology. Every particle compares itself with the best-performing particle in the swarm.
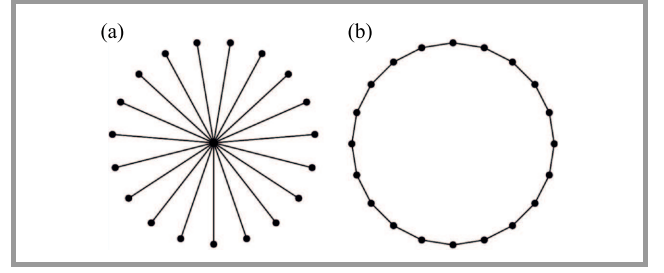
Fig. 1. PSO variants and topologies: (a) star topology (PSO-glob) and (b) ring topology (PSO-loc).

- Local-best PSO algorithm (PSO-loc) using a ring topology. Every particle compares itself only with its nearest-neighbors computed by applying the chosen distance metric.

4. Implementation Overview

The COmparing Continuous Optimizers (COCO) [24] benchmarking platform version 2.2.1 was used to evaluate the performance of PSO and CMA-ES methods. COCO is an integrated software environment that can be successfully used for systematic and sound comparisons of global optimization solvers. It provides a set of benchmark functions and tools for processing and visualizing results of calculations. The COCO source code is available at https://github.com/numbbo/coco. A detailed description of the experimental procedure for conducting the experiments using the COCO system is presented in [25]. The measures of performance assessment implemented in COCO are described in [26].

The COCO platform offers Python language support to provide the implementation of optimization solvers. Therefore, both CMA-ES and PSO algorithms implemented in Python were adopted and incorporated in COCO. The CMA-ES solver was taken from the *pycma* library [27], while the PSO solver was taken from the *pyswarms* library [28].

The benchmark optimization problems were taken from the black box optimization benchmarking (BBOB) [15] suite. It provides numerous testing problems with various characteristics, dimensions and complexities. All benchmark functions to be minimized are divided into groups, each with specific characteristics of the objective function, i.e., separable, moderate, ill-conditioned, multi-modal, weakly structured multi-modal.

5. Experimental Procedure and Performance Measures

Numerous experiments whose results are presented in this paper were conducted. The experimental procedure was executed within the COCO benchmarking platform, following the best practices for the assessment of performance

of optimization algorithms executed in a black-box scenario [25]. The performance of the algorithms was measured based on the number of run times, i.e. objective function evaluations, needed to reach one or several quality indicator target values. The details of the experimental setup are presented below.

5.1. Experimental Procedure

The goal of each experiment was to find the global solution of the problem (1) with the prescribed accuracy, i.e. the point $\{x \in \Re^{dim} : f(x) = f_{target}\}$ with f_{target} defined as:

$$f_{target} = f_{opt} + \varepsilon_{min} , \qquad (5)$$

where $f(x)$ denotes the objective function, $f_{opt} = f(x_{opt})$, where x_{opt} is the solution of (1), and $\varepsilon_{min} = 10^{-8}$ is the assumed precision. As mentioned above, the performance assessment of each experiment was based on the number of the objective function evaluations. The evaluations conducted by the algorithm were recorded for each target $f_{target}^{\varepsilon} = f_{opt} + \varepsilon$ that was reached for various precisions ε. Precisions were chosen uniformly on the logarithmic scale from the interval $\varepsilon \in [10^{-8}, 10^2]$.

All benchmark problems, i.e. 24 noiseless unconstrained optimization problems concerned with the minimization of objective functions from the BBOB test suite $\Omega_f = \{f1, f2, \ldots, f24\}$ were taken into consideration. Problems were solved with the increasing number of decision variables (dimension), i.e. $dim \in \{2, 3, 5, 10, 20, 40\}$.

Every single setup, i.e. the problem of optimization of a given benchmark function and the dimension of the problem, was executed over $N_{trial} = 15$ independent trial runs. The performance was evaluated over all trials.

Three solvers were used and compared: CMA-ES, global-best PSO and local-best PSO. Additionally, the performance of the considered methods was confronted with that of the artificial algorithm labeled *best2009*. The scores of the *best2009* algorithm are based on the results reached by solvers submitted to the COCO 2009 competition. For each of the setups the best performing algorithm's results were incorporated.

The calculations were stopped after reaching the target value of the objective function f_{target} defined in (5) or depleting the budget Max_{iter} for the number of objective function evaluations equal to $3000 \cdot (dim + 2)$, where dim denoted the problem dimension for a given setup (trial).

5.2. Performance Measures

Two measures were used to evaluate the performance and to compare the efficiency of CMA-ES and two variants of PSO algorithms. The aim of the first measure – the average running time (aRT) – is to show the successful performance of the tested techniques. Average runtime is an estimation of the algorithm's expected runtime expressed

in the number of objective function evaluations. The smaller the value, the better [25].

$$aRT = \frac{1}{N_{success}} \sum_{t=1}^{N_{trial}} fevals_t, \quad N_{success} \leq N_{trial} , \quad (6)$$

where aRT denotes the average runtime for N_{trial} relevant trials for a given setup. $fevals_t$ is the number of objective function evaluations performed in the t-th trial until the target value f_{target} was reached. $N_{success}$ is the number of successful trials, i.e., trials where f_{target} was reached.

The second measure – an empirical cumulative distribution function (ECDF) [29] – was used to display the proportion of problems solved within a given limit of objective function evaluations. In general, the purpose of the measure is to show and compare the speed of convergence of specific optimization algorithms [25].

Let us consider a subset of the available benchmarking functions $\omega_f \subseteq \Omega_f$ with a fixed number of dimensions dim. We define S as the set of bootstrap executions, where each element $s \in S$ is denoted by a triplet $s = (f_m, f_{m,target}^{\varepsilon}, t)$, $f_m \in \omega_t$. $f_{m,target}^{\varepsilon}$ is the target solution with the required precision ε. In this experiment, the solutions were calculated for 51 precisions chosen from $[10^{-8}, 10^2]$. t denotes a trial index, $t = 1, \ldots, N_{trial}$.

Every triplet s corresponds to the number of function f_m evaluations performed by the optimization algorithm to reach the target solution with $f_{m,target}^{\varepsilon}$. This number is denoted by R_s.

We calculate ECDF as follows:

$$\hat{F}_S(r) = \frac{|\{s \mid R_s \leq r\}|}{|S|}, \qquad (7)$$

where the variable r denotes the number of function evaluations, $|S|$ is a number of all triplets from the set S. Hence, $\hat{F}_S(r)$ is the fraction of bootstrap executions for which the target value of the objective function was reached after R_s function evaluations, such that $R_s \leq r$.

To provide a compact assessment of all tested algorithms, taking into account both quality of the solution obtained and efficiency of the optimization solvers, ECDF was computed over subsets of multiple benchmark functions (ω_f). The functions were grouped based on their characteristics. The proposed groups are described in the Section 6.

6. PSO and CMA-ES Performance Evaluation

Multiple experiments for 24 benchmark functions given in [15], [30] were conducted. The results are presented and discussed in this section.

6.1. CMA-ES and PSO Configuration Setup

The following values of parameters typical of the tested methods were determined.

Table 1
Number of particles in PSO for various problem dimensions

Problem dimension (dim)	Swarm size
$dim \leq 10$	40
$10 < dim \leq 25$	60
$25 < dim \leq 40$	90

- The CMA-ES method was executed with the step size of $\sigma = 0.2$.

- The hyperparameters of PSO solvers were consistent across all tests. The following values were used: inertial coefficient $w = 0.9$, acceleration coefficients $c_1 = 0.5$, $c_2 = 0.3$. For the local-best variation, the Euclidean norm was used to measure the distance between neighbors. The number of nearest neighbors considered was equal to $k = 2$. The swarm size was adapted to the size and complexity of the problem.

Hence, the number of particles was different for different tests (see Table 1).

6.2. Experimental Results

The global minima of 24 benchmark functions in the search space $[-5, 5]^{dim}$ were calculated with the use of CMA-ES and PSO methods. Test results were compared with the reference solutions of *best2009*.

All experiments were performed on a unified hardware platform: Intel Core i7-2640M CPU @ 2.80 GHz with 1 processor and 4 cores.
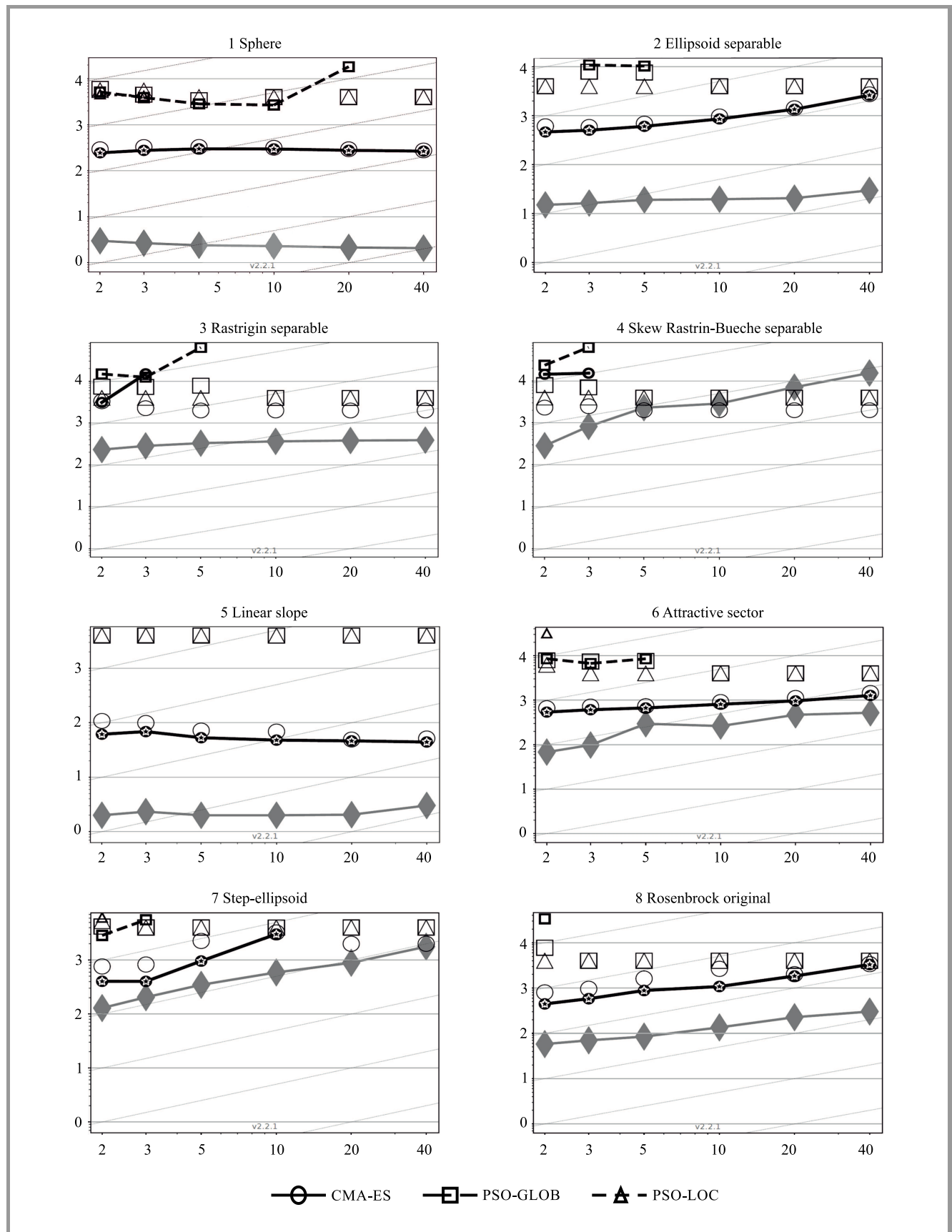
The results, i.e. efficiency of all tested algorithms tested for various test functions and problem dimensions are presented in Table 2 and Figs. 2, 3, 4. Different markers used in all figures correspond to different algorithms:

- a circle – CMA-ES results,

- a square – PSO global-best results,

- a triangle – PSO local-best results.

Table 2
The average running time divided by the best *aRT* obtained for *best2009*; 24 benchmark functions ($dim = 5$), $N_{trial} = 15$, $Max_{iter} = 21000$. "–" denotes that the target solution was not reached.

Function	CMA-ES			PSO-glob			PSO-loc		
	$\varepsilon = 10^0$	$\varepsilon = 10^{-2}$	$\varepsilon = 10^{-5}$	$\varepsilon = 10^0$	$\varepsilon = 10^{-2}$	$\varepsilon = 10^{-5}$	$\varepsilon = 10^0$	$\varepsilon = 10^{-2}$	$\varepsilon = 10^{-5}$
f1	20	44	86	83	274	668	846	4361	–
f2	18	23	28	239	443	524	–	–	–
f3	21	–	–	88	188	190	–	–	–
f4	–	–	–	–	–	–	–	–	–
f5	26	27	27	–	–	–	–	–	–
f6	3.4	3.5	2.3	41	35	26	1343	–	–
f7	3.1	2.4	2.5	60	–	–	–	–	–
f8	8.3	9.1	10	210	–	–	–	–	–
f9	3.9	5.7	6.2	131	40 5	–	710	–	–
f10	3.3	3.5	3.2	–	–	–	–	–	–
f11	6.6	1.7	1.5	–	–	–	–	–	–
f12	12	14	6	1079	–	–	1099	–	–
f13	7.7	7.3	2.1	451	–	–	–	–	–
f14	5.3	7.6	6.8	23	44	–	295	–	–
f15	7.9	–	–	32	–	–	–	–	–
f16	2.6	1.1	1.3	26	–	–	46	–	–
f17	18	2.6	24	25	14	–	503	–	–
f18	8.6	3	–	30	–	–	374	–	–
f19	741	0.4	1.2	9820	–	–	–	–	–
f20	19	2.9	2.7	18	–	–	75	–	–
f21	7.8	8.2	8.2	23	20	23	77	–	–
f22	23	17	16	65	89	142	92	292	–
f23	7.2	1.1	0.99	39	–	–	16	–	–
f24	–	–	–	–	–	–	–	–	–

Fig. 2. PSO and CMA-ES efficiency. The average running time divided by dimension (aRT/dim); 24 benchmark functions, $\varepsilon = 10^{-8}$, number of trials $N_{trial} = 15$. Legend: ◯ – covariance matrix adaptation evolution strategy, ☐ – global particle swarm optimization, △ – local particle swarm optimization. Slanted grid lines indicate quadratic scaling with the dimension.
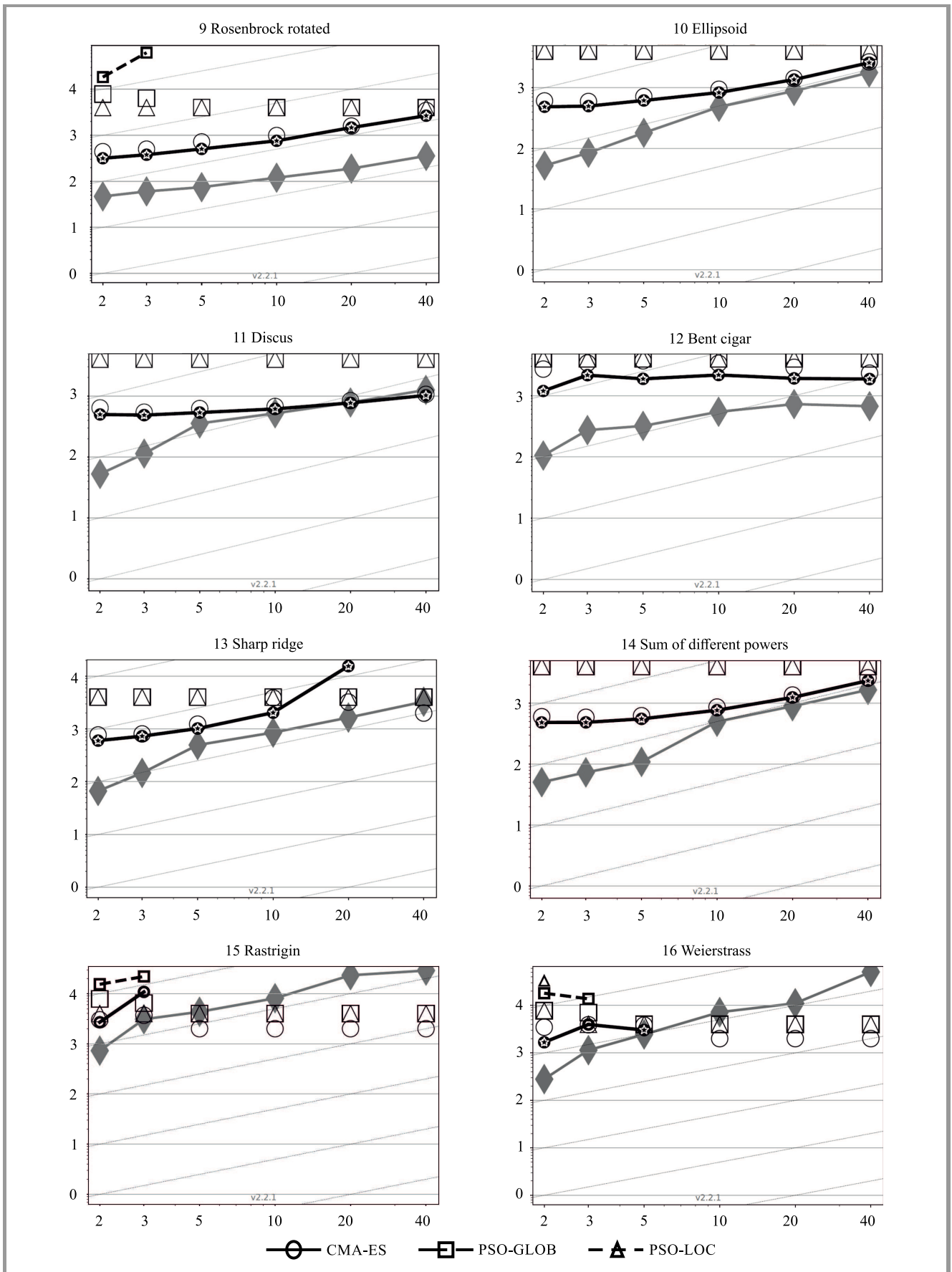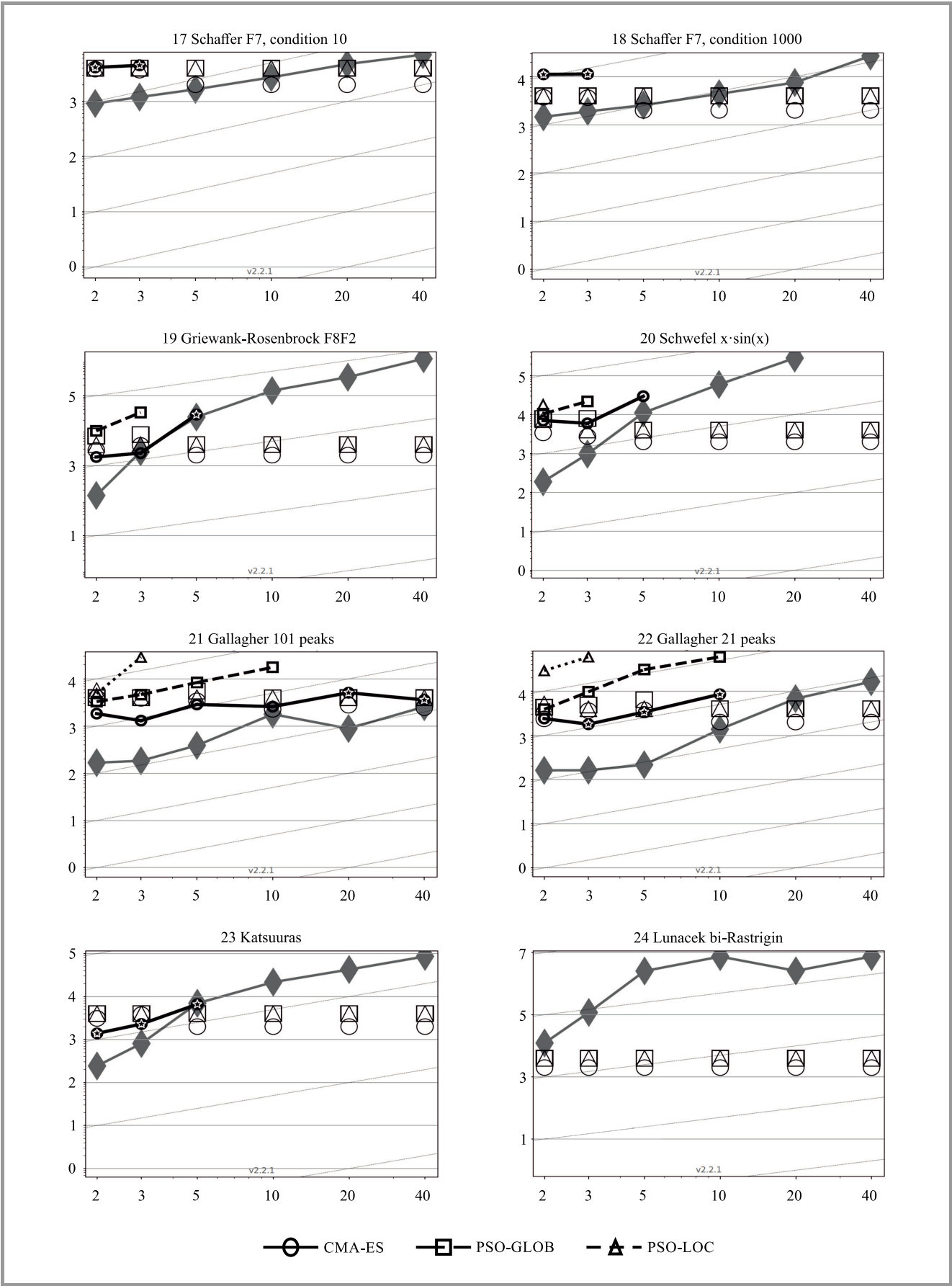
Fig. 2 – continued.

Fig. 2 – continued.

The table shows the average running time *aRT* (6) divided by the respective best *aRT* obtained for *best2009*, each for one of 24 benchmark functions and the problem dimension of *dim* = 5. The plots in Fig. 2 show the average running time *aRT* (6) as $\log_{10}$ value divided by the problem dimension, i.e. $\log(aRT)/dim$, each for one of 24 benchmark functions and various problem dimensions. The experiments were conducted for the following number of decision variables $dim \in \{2, 3, 5, 10, 20, 40\}$ for each function and the prescribed accuracy of $\varepsilon = 10^{-8}$. All tests were executed $N_{trial} = 15$ times. The meaning of the additional symbols in all 24 plots are as follows:

- light symbols (circle, square, triangle) give the maximum number of function evaluations from the largest trial divided by the problem dimension,

- black stars indicate a statistically better result compared to all other algorithms.

The results obtained, as shown in Table 2 and Fig. 2, indicate that PSO was unable to reach the target function value f_{target} for the problems with $dim > 3$ for most of the benchmark functions. The best results were recorded for the *Sphere* function, Fig. 2-1 – global-best PSO (PSO-glob) was able to find the solution for a 20-dimensional problem. In this case, the number of function evaluations decreased until the number of dimensions reached 10, which suggest that a better choice of the algorithm's hyperparameters could be made. For the *Gallagher* functions, Fig. 2-21, we can observe that *aRT* grows on a nearly quadratic basis until the number of dimensions is equal to 10. Local-best PSO (PSO-loc) performed very poorly, reaching f_{target} only for *Sphere* and *Gallagher* functions with a small number of dimensions.

As for CMA-ES, the results were much better. The method succeeded in finding the target solution of 11 functions from the benchmark set, even for $dim = 40$. In other cases the results of CMA-ES were similar to PSO, e.g. they failed completely for *Lunacek bi-Rastrigin*, Fig. 2-24. In some cases like the *Elipsoid* function, Fig. 2-10, or *Discus*, Fig. 2-11, the *aRT* was similar to the referential solutions from *best2009* (for bigger dimensions). However, in general, CMA-ES did considerably better then both PSO versions.

The goal of the second series of experiments was to test the statistical significance of results. All benchmark functions (f1–f24) listed in Fig. 2 were divided into 6 groups with different characteristics:

1. Separable functions (f1–f5) – optimal value of a given coordinate of the decision variable does not depend on the choice of the remaining coordinates.

2. Moderate functions (f6–f9) – moderate dimension of the decision variable vector.

3. Ill-conditioned functions (f10–f14) – different variables, or different directions in search space, show a largely different sensitivity in their contribution to the objective function value.

4. Multi-modal functions (f15–f19) – multiple minima and maxima.

5. Weakly structured multi-modal functions (f20–f24) – many solutions with similar values of the performance measure.
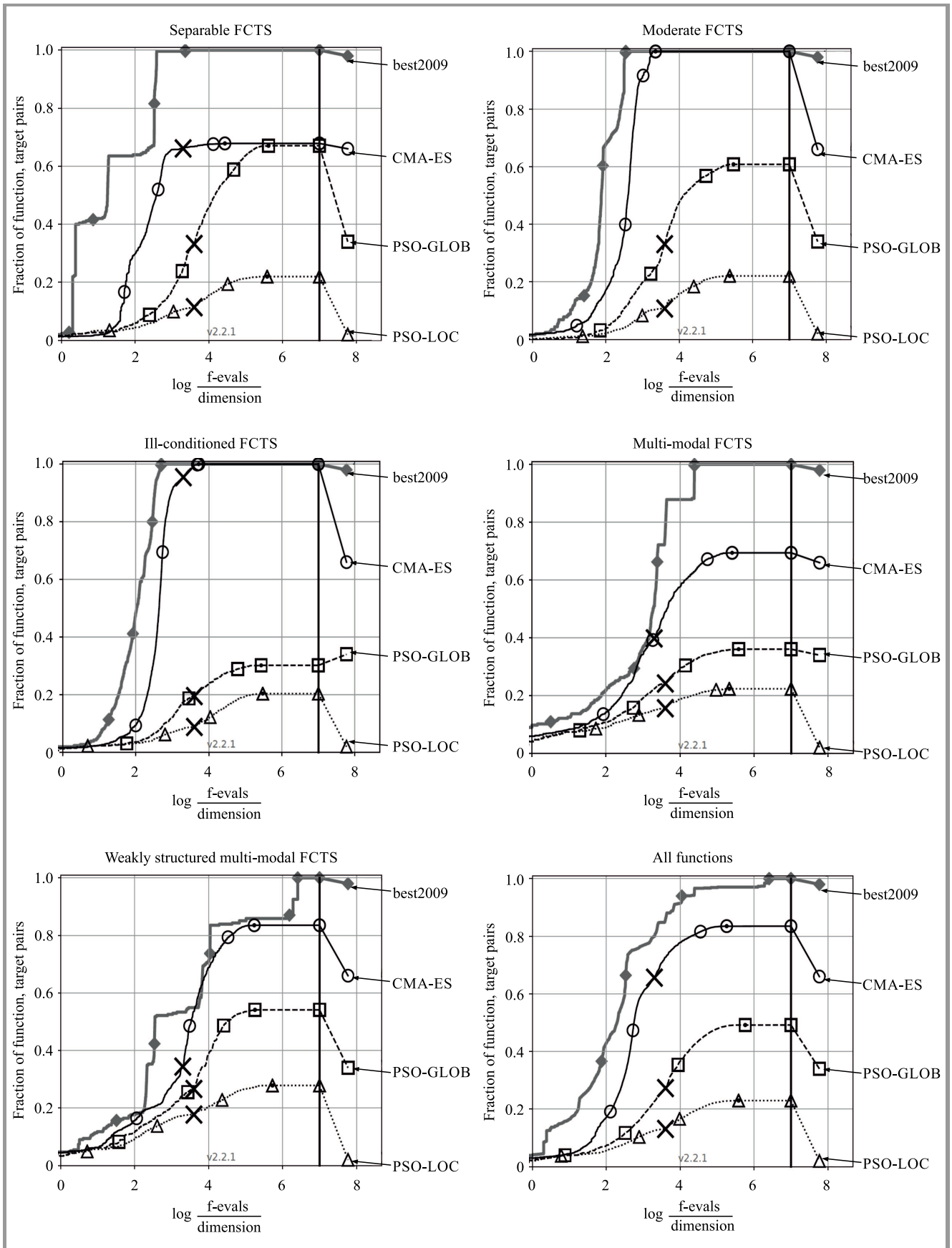
6. All functions.

Multiple experiments for all functions, as well as PSO and CMA-ES optimization methods were performed. The number of trials $N_{trial} = 15$. Calculations were conducted for 51 values of f_{target} with various precisions $\varepsilon \in [10^{-8}, 10^2]$. Rank-sum test for a given target f_{target} using, for each trial, either the number of needed function evaluations needed to reach f_{target} (inverted and multiplied by -1), or, if the target was not reached, the best precision – ε-value achieved, measured only up to the smallest number of overall function evaluations for any unsuccessful trial under consideration. Problems with two dimensions were tested: $dim = 5$ (Fig. 3) and $dim = 20$ (Fig. 4). Both figures present the cumulative distribution of the measure $\hat{F}_S$ (7). The results obtained for PSO and CMA-ES are compared with the reference solutions from *best2009* (shown as a thick line with diamond markers).

Figures depicting empirical cumulative distribution functions (ECDFs) confirm that overall CMA-ES performs better then both versions of the PSO method. For dimensions smaller then 5 (Fig. 3), the differences in the optimization of separable functions are not so significant, PSO-glob performs rather well, compared to CMA-ES. On the other hand, in the case of ill-conditioned functions, CMA-ES hugely outranks PSO, with its performance nearly matching the reference solution from *best2009*. A difference is also noticeable for the larger dimensions, Fig. 4, especially in the case of ill-conditioned and moderate functions. However, for the more demanding, multi-modal problems, the results of all tested algorithms fall short compared to the reference solution. PSO-loc clearly stands out as the most ineffective of all tested methods.

As the final observation, it is worth mentioning that in solving a given black-box problem, the choice of the proper optimization algorithm and proper tuning of its parameters are of crucial significance. In the presented experiments, the *best2009* solutions outclass, in all cases, the results calculated by CMA-ES and PSO. This is to be expected, as *best2009* consists of the solutions of multiple algorithms, each adjusted for a given set of problems.

6.3. Possible Improvements of CMA-ES

CMA-ES is a population-based stochastic technique. The population size plays a big factor in the algorithm's efficiency, depending on the use case. With a default (small) population size, CMA-ES is a robust and fast method intended mainly for local search optimization. By increasing the size of the population, the algorithm can be successfully employed for more global search problems. Taking into account both of those principles, a modified version of this

Fig. 3. Bootstrapped empirical cumulative distribution of the objective function evaluations divided by dimension (*fevals/dim*), *dim* = 5, $\varepsilon \in [10^{-8}, 10^{2}]$.
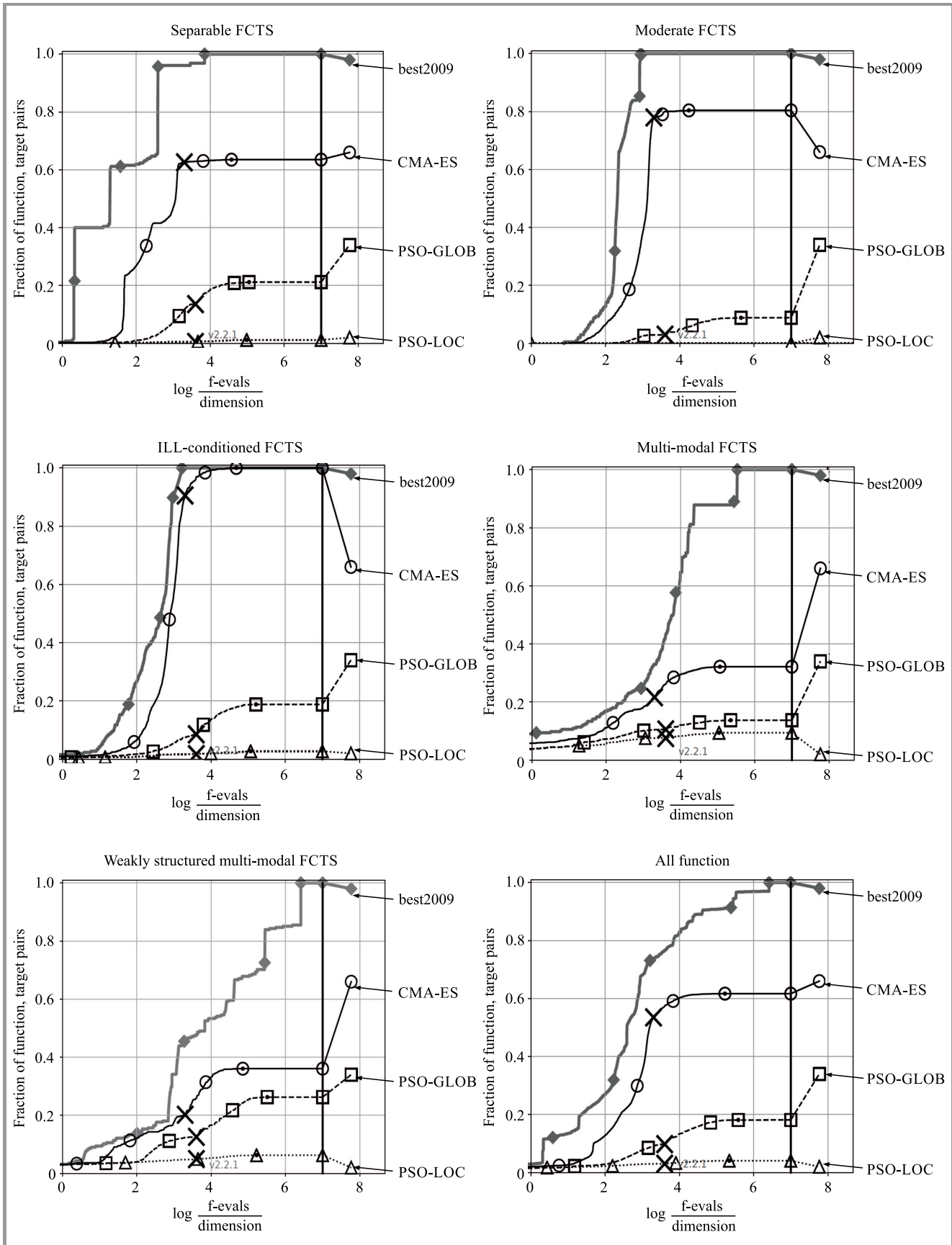
Fig. 4. Bootstrapped empirical cumulative distribution of the objective function evaluations divided by dimension (*fevals/dim*), *dim* = 20, $\varepsilon \in [10^{-8}, 10^2]$.

method, i.e. BI-POP CMA-ES was developed. It applies two interlaced multistart regimes altering the size of the population. One of them increases the population size by the factor of two and starts with a higher initial step, while the other decreases the population size and uses smaller steps equal to σ in 2. The promising results of BI-POP CMA-ES in global optimization are reported in literature, i.e. [31].

7. Summary and Conclusion

The paper provides a short report on the efficiency and availability of two biologically-inspired CMA-ES and PSO methods that are designed to tackle non-convex and ill-conditioned black-box optimization problems.

The worst performance was observed for the local version of the PSO method. The global version was better in all of the cases, while CMA-ES outranked both PSO methods.

This should not come as a surprise, as various modifications of the CMA-ES algorithm are currently considered to be state-of-the-art in the field of black-box optimization. The bare CMA-ES algorithm performs well, although the numerous experiments confirmed that the reference algorithm outclasses both PSO and the classic variant of CMA-ES.

The final conclusion is that PSO techniques are very sensitive to hyperparameters of the algorithms and tuning of these parameters is a challenging task. A better choice of the algorithm's hyperparameters adapted to each function and dimension can seriously influence the final result. Since the CMA-ES method does not require tedious parameter tuning, the choice of the strategy to be adopted while setting the internal parameters is not left to the user. Therefore, it is much more convenient then algorithms such as PSO.

With the standard version of CMA-ES, there is room for improvement. Modification of the original algorithm and its adaptation to the optimization problem to be solved can lead to better performance overall, making it a more reliable and versatile method. Therefore, we plan to compare the performance of the original CMA-ES and BI-POP CMA-ES mentioned in Subsection 6.3 in the future.

Acknowledgments

References

[1] M. J. Asher, B. F. W. Croke, A. J. Jakeman, and L. J. M. Peeter, "A review of surrogate models and their application to groundwater modeling", *Water Resources Res.*, vol. 51, no. 8, pp. 5957–5973, 2013 (doi: 10.1002/2015WR016967).

[2] T. Bartz-Beielstein, "A survey of model-based methods for global optimization", in *Proc. of 7th Int. Conf. on Bioinspired Optimiz. Methods and their Applications BIOMA 2016*, Bled, Slovenia, 2016, pp. 3–20, 2016.

[3] A. T. Nguyen *et al.*, "A review on simulation-based optimization methods applied to building performance analysis", *Applied Energy*, vol. 113, pp. 1043–1058, 2014 (doi: 10.1016/j.apenergy.2013.08.061).

[4] E. Niewiadomska-Szynkiewicz and J. Błaszczyk, "Simulation-based optimization methods applied to large scale water systems control", in *Proc. of 16th IEEE Int. Conf. on Scalable Comput. and Commun. ScalCom2016*, Toulouse, France, 2016, pp. 649–656 (doi: 10.1109/UIC-ATC-ScalCom-CBDCom-IoP-SmartWorld. 2016.0108).

[5] J. C. Spall, *Introduction to Stochastic Search and Optimization*. Wiley, 2003 (doi: 10.1002/0471722138, ISBN 9780471330523).

[6] S. Gil, A. Kott, and A. L. Barabási, "A genetic epidemiology approach to cyber-security", *Scientific Rep.*, vol. 4, pp. 1–7, 2014 (doi: 10.1038/srep05659).

[7] G. Kumar, K. Kumar, and M. Sachdeva, "The use of artificial intelligence-based techniques for intrusion detection: a review", *Artif. Intell. Rev.*, vol. 34, no. 4, pp. 369–387, 2010 (doi: 10.1007/s10462-010-9179-5).

[8] E. Niewiadomska-Szynkiewicz, A. Sikora, and J. Kołodziej, "Modeling mobility in cooperative ad hoc networks", *Mobile Netw. and Appl.*, vol. 18, no. 5, pp. 610–621, 2013 (doi: 10.1007/s11036-013-0450-2).

[9] P. Szynkiewicz, "A novel GPU-enabled simulator for large scale spiking neural networks", *J. of Telecommun. and Inform. Technology*, no. 2, pp. 34–42, 2016.

[10] M. Marks, E. Niewiadomska-Szynkiewicz, and J. Kołodziej, "High performance wireless sensor network localisation system", *Int. J. of Ad Hoc and Ubiquitous Comput.*, vol. 17, no. 2–3, pp. 122–133, 2014 (doi: 10.1504/IJAHUC.2014.065776).

[11] P. Szynkiewicz and A. Kozakiewicz, "Design and evaluation of a system for network threat signatures generation", *J. of Comput. Sci.*, vol. 22, pp. 187–197, 2017 (doi: 10.1016/j.jocs.2017.05.006).

[12] N. Hansen, "The CMA evolution strategy: A tutorial", Tech. Rep., INRIA, 2016, arXiv: 1604.00772.

[13] J. A. Lozano, P. Larranaga, I. Inza, and E. Bengoetxea, Eds., *Towards a New Evolutionary Computation. Advances on Estimation of Distribution Algorithms*. Berlin: Springer, 2006 (ISBN 9783540324942).

[14] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory", in *Proc. of the 6th Int. Symp. on Micro Machine and Human Sci. MHS'95*, Nagoya, Japan, 1995, pp. 39–43 (doi: 10.1109/MHS.1995.494215).

[15] N. Hansen, S. Finck, R. Ros, and A. Auger, "Real-parameter black-box optimization benchmarking 2009: Noiseless functions definitions", Tech. Rep. RR-6829, INRIA, 2009 [Online]. Available: https://hal.inria.fr/inria-00362633/document

[16] S. Amaran, N. V. Sahinidis, B. Sharda, and S. J. Bury, "Simulation optimization: a review of algorithms and applications", *Annals of Operations Res.*, vol. 240, no. 1, pp. 351–380, 2016 (doi: 10.1007/s10479-015-2019-x).

[17] J. April, F. Glover, J. P. Kelley, and M. Laguna, "Practical introduction to simulation optimization", in *Proc. of the 2003 Conf. on Winter Simul. WSC 2003*, New Orleans, LA, USA, 2003, pp. 71–78 (doi: 10.1109/WSC.2003.1261410).

[18] L. M. Rios and N. V. Sahinidis, "Derivative-free optimization: a review of algorithms and comparison of software implementations", *J. of Global Optimiz.*, vol. 56, no. 3, pp. 1247–1293, 2013 (doi: 10.1007/s10898-012-9951-y).

[19] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C. The Art of Scientific Computing*. Cambridge University Press, Cambridge, 1992 (ISBN 9780521431088).

[20] M. M. Ali, A. Törn, and S. Vittanen, "A numerical comparison of some modified controlled random search algorithms", *J. of Global Optimiz.*, vol. 11, no. 4, pp. 377–385, 1997 (doi: 10.1023/A:1008236920512).

[21] A. Dekkers and E. Aarts, "Global optimization and simulated annealing", *Mathem. Programming*, vol. 50, no 3, pp. 367–393, 1999 (doi: 10.1007/BF01594945).

[22] J. Kołodziej, *Evolutionary Hierarchical Multi-Criteria Metaheuristics for Scheduling in Large-Scale Grid Systems. Studies in Computational Intelligence*, vol. 419. Springer, 2012 (ISBN 9783642289705).

[23] Z. Michalewicz and D. B. Fogel, *How to Solve It: Modern Heuristics*. Berlin Heidelberg: Springer, 2002 (ISBN 3540660615).

[24] N. Hansen, A. Auger, O. Mersmann, T. Tušar, and D. Brockhoff, "COCO: A platform for comparing continuous optimizers in a black-box setting", *ArXiv e-prints*, 2016, arXiv: 1603.08785.

[25] N. Hansen, A. Auger, D. Brockhoff, D. Tušar, and T. Tušar, "COCO: Performance assessment", *ArXiv e-prints*, 2016, arXiv: 1605.03560.

[26] N. Hansen, T. Tušar, O. Mersmann, A. Auger, and D. Brockhoff. "COCO: The experimental procedure", *ArXiv e-prints*, 2016, arXiv: 1603.08776.

[27] N. Hansen, "Python implementation of CMA-ES", 2016 [Online]. Available: https://github.com/CMA-ES/pycma

[28] L. J. V. Miranda, "PySwarms: a research toolkit for particle swarm optimization in Python", *The J. of Open Source Softw.*, vol. 3, no. 21, pp. 1–2, 2018 (doi: 10.21105/joss.00433).

[29] A. W. van der Vaart, *Asymptotic Statistics*. Cambridge University Press, 2000 (doi: 10.1017/CBO9780511802256, ISBN 9780511802256).

[30] S. Finck, N. Hansen, R. Ros, and A. Auger, "Real-parameter black-box optimization benchmarking 2009: Presentation of the noiseless functions", Tech. Rep. 20, Research Center PPE, 2009.

[31] H. Beyer and B. Sendhoff, "Simplify your covariance matrix adaptation evolution strategy", *IEEE Trans. on Evolut. Comput.*, vol. 21, no. 5, pp. 746–759, 2017 (doi: 10.1109/TEVC.2017.2680320).

Paweł Szynkiewicz received his M.Sc. in Computer Science from the Warsaw University of Technology, Poland, in 2015. Currently he is a Ph.D. student at the Systems Research Institute, Polish Academy of Science. Between 2015–2016 he was employed at Comarch, and between 2016–2017 at Bright-Solutions IT. He has been working for NASK since 2017. His research interest focuses on software technologies, computer networks security, global optimization, machine learning, neural networks and evolutionary algorithms.

https://orcid.org/0000-0001-8872-5505
E-mail: pawel.szynkiewicz@nask.pl
Research and Academic Computer Network (NASK)
Kolska 12
01-045 Warsaw, Poland

Examples of Objectified Multiple Criteria Ranking in the Selection of Infrastructural Projects

Edward Klimasara and Andrzej P. Wierzbicki

The National Institute of Telecommunications, Warsaw, Poland

Abstract—The paper addresses the issue of multiple criteria rankings of infrastructural projects (buildings, roads, etc.). Although the amount of literature devoted to this subject is considerable, all methods proposed produce subjective rankings, dependent on a direct or indirect definition of weighting coefficients applicable to subsequent evaluation criteria. Infrastructural projects are usually selected and approved collegially, however, by a group of decision makers with preferences that may potentially differ significantly. Therefore, an objectified ranking, independent from subjectively defined weighting coefficients, is needed for infrastructural projects. Such a ranking is proposed, analyzed and applied by the authors of this paper. This ranking depends originally only on the multiple objective evaluation data, i.e. the values of evaluation criteria related to decision scenarios or alternatives. Such an approach does not render a fully objective ranking, since one of this kind does not exist at all. Even the choice of the ranking method is a subjective decision, but it is objectified to the extent possible. The paper presents several examples of multiple criteria evaluation of infrastructural projects, derived from literature, and compares subjective rankings published in literature with objectified rankings that are independent of weighting coefficients.

Keywords—multiple criteria rankings, objectified multiple criteria ranking, objectified ranking method, subjective rankings.

1. Introduction – Korhonen Paradox

We show, first, that a linear aggregation of criteria values leads essentially to wrong results, thus such methods as the Analytic Hierarchy Process (AHP) [1], [2] should never be applied in serious problems. This results from the Korhonen paradox [3]: a young man considers three candidates for marriage, evaluated based on such criteria as "intelligence"[1] and "sex appeal". The first candidate received 10 points for sex appeal and 0 points for intelligence. The second one is evaluated at 0 points for sex appeal and 10 points for intelligence. The third one is evaluated at 5 points for sex appeal and 4 points for intelligence.

---

[1]Pekka Korhonen used here the "ability to cook" criterion, but we changed its name to "intelligence", since we do not wish to be accused of sexism.

Actually, any evaluation that is inside the convex cover of two first evaluations can be used. The paradox is that *when applying linear aggregation of criteria, one of the first two points would be selected only – never the third one*. This is illustrated in Fig. 1, where in addition to the linear criteria
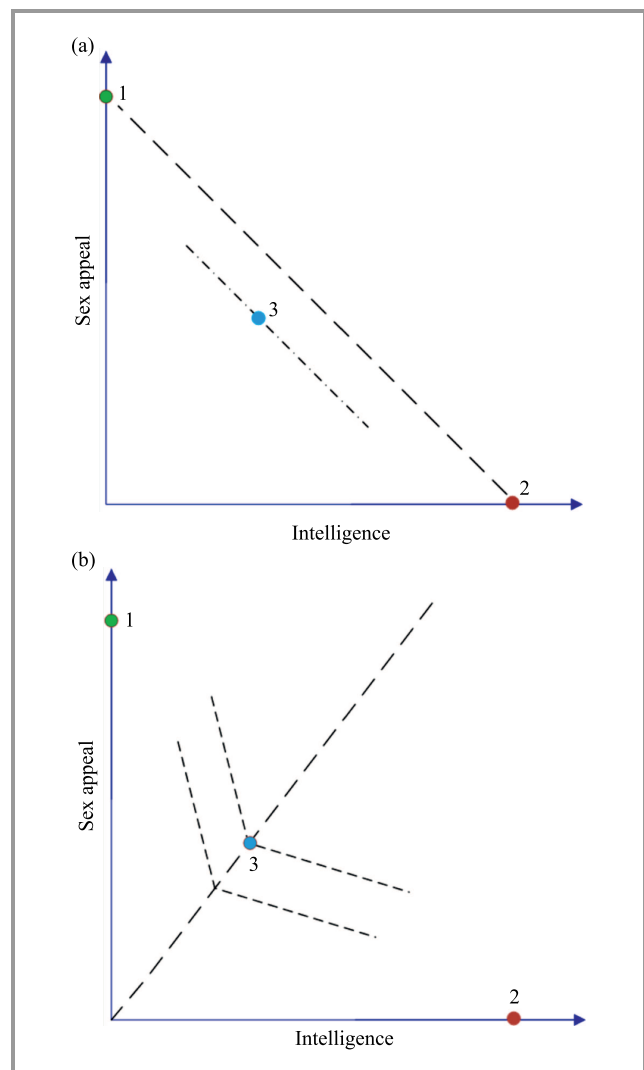
Fig. 1. Graphical illustration of the Korhonen paradox and the manner of overcoming it: (a) linear aggregation of criteria, (b) nonlinear aggregation of criteria.

aggregation method, its nonlinear counterpart is presented as well (as defined in one of following sections).

Pekka Korhonen, the then chairman of the international society of multiple criteria analysis, presented this paradox to show the inadequacy of methods relying on linear aggregation of criteria, in particular the AHP method. But in literature concerned with evaluation of infrastructural projects, no awareness of this difficulty is evident, and the AHP method is widely used.

2. Collegial Selection and Objectified Ranking

Another serious problem is the fact that infrastructural projects are usually evaluated collegially – by groups of decision makers usually representing quite varied and different preferences. Therefore, it is wrong to ask them at the beginning of the evaluation, as required by the AHP method, to define relative importance of subsequent criteria. They much more prefer to be presented a ranking – a list of preferred projects possibly with an objective justification of their selection – that they can later discuss and evaluate.

In fact, the method of objectified ranking was devised precisely because of this particular reason. At a Japanese research institute (actually, a university offering doctoral studies only), a questionnaire concerning diverse aspects of creative work and study organization was filled out by the majority of students, see [4]. The question was which aspects are most important and the dean supervising this event – a specialist in multiple criteria decision making himself – refused to specify any weighting coefficients or aspiration levels, because he wanted to present, to fellow professors, results that were as objective as possible. Therefore, we observed that the data from the questionnaire – the values of criteria related to subsequent issues – suffice to define a nonlinear achievement function that can be used to rank these issues.

Specifically, we assume that we have a decision problem with K criteria, indexed by $k = 1, \ldots, K$ (also denoted by $k \in \mathbf{K}$), and J decision options also referred to as alternatives or variants, indexed by $j = 1, \ldots, J$ or $j = A, B, \ldots, J$ (also denoted by $j \in \mathbf{J} = \{1, \ldots, J\}$). The corresponding criteria values are denoted by q_{kj}. We assume that all are maximized or converted to maximized variables. The maximal values $\max_{j \in J} q_{kj} = q_k^{\text{up}}$ are called upper bounds for criteria and are often equivalent to the components of the so called ideal or utopia point $\mathbf{q}^{\text{uto}} = \mathbf{q}^{\text{up}} = (q_1^{\text{up}}, \ldots, q_k^{\text{up}}, \ldots, q_K^{\text{up}})$ – except for cases when they were established a priori as a measurement scale. The minimal values $\min_{j \in J} q_{kj} = q_k^{\text{lo}}$ are called lower bounds and, generally, are not equivalent to the components of the so called nadir point $\mathbf{q}^{\text{nad}} \geq \mathbf{q}^{\text{lo}} = (q_1^{\text{lo}}, \ldots q_k^{\text{lo}}, \ldots q_K^{\text{lo}})$; the nadir point qnad is defined similarly as the lower bound point $\mathbf{q}^{\text{lo}}$, but with minimization restricted to Pareto optimal or

efficient or nondominated alternatives. We recall that an alternative $j^* \in \mathbf{J}$ is Pareto optimal (Pareto-nondominated or shortly nondominated, also called efficient), if there is no other alternative $j \in \mathbf{J}$ that dominates j^*, that is, if we denote $\mathbf{q}_j = (q_{1j}, \ldots, q_{kj}, \ldots q_{Kj})$, there is no $j \in \mathbf{J}$ such that $\mathbf{q}_j \geq \mathbf{q}_{j^*}, \mathbf{q}_j \neq \mathbf{q}_{j^*}$.

In the objectified ranking method, see [5], [6] for more details, reference point approach is used while aspiration and reservation levels a_k and r_k are determined by the following equation. The assumption is made that the corresponding criterion k is maximized. If it is minimized, we just reverse the places of aspiration and reservation levels:

$$m_k = \sum_{j \in \mathbf{J}} \frac{q_{jk}}{J},$$
$$r_k = 0.5(q_k^{\text{lo}} + m_k),$$
$$a_k = 0.5(q_k^{\text{up}} + m_k), \ k = 1, \ldots, K. \quad (1)$$

After determining these reference levels for all criteria, values of partial achievement functions for a given criterion and variant are determined as follows. For maximized criteria:

$$\sigma_{kj}(q_{kj}, a_k, r_k) = \begin{cases} \frac{\alpha(q_{kj} - q_k^{\text{lo}})}{r_k - q_k^{\text{lo}}} & \text{for } q_k^{\text{lo}} \leq q_{kj} \leq r_k \\ \alpha + \frac{(\beta - \alpha)(q_{kj} - r_k)}{a_k - r_k} & \text{for } r_k < q_{kj} \leq a_k \\ \beta + \frac{(10 - \beta)(q_{kj} - a_k)}{q_k^{\text{up}} - a_k} & \text{for } a_k < q_{kj} \leq q_k^{\text{up}} \end{cases} \quad (2)$$

where $0 < \alpha < \beta < 10$, α is a parameter denoting the value of the partial achievement function for $q_{kj} = r_k$, and β is a parameter denoting the value of the partial achievement function for $q_{kj} = a_k$. For minimized criteria (where the roles of r_k and a_k are exchanged):

$$\sigma_{kj}(q_{kj}, a_k, r_k) = \begin{cases} \beta + \frac{(10 - \beta)(a_k - q_{kj})}{a_k - q_k^{\text{lo}}} & \text{for } q_k^{\text{lo}} \leq q_{kj} \leq a_k \\ \alpha + \frac{(\beta - \alpha)(r_k - q_{kj})}{r_k - a_k} & \text{for } a_k < q_{kj} \leq r_k \\ \frac{\alpha(q_k^{\text{up}} - q_{kj})}{q_k^{\text{up}} - a_k} & \text{for } r_k < q_{kj} \leq q_k^{\text{up}} \end{cases} \quad (3)$$

The overall achievement function for the alternative k, whose values are used to determine objectified ranking lists, is computed on the basis of partial achievement functions, as:

$$\sigma(\mathbf{q}_j, \mathbf{a}, \mathbf{r}) = \min_{k \in \mathbf{K}} \sigma_{kj}(q_{kj}, a_k, r_k)$$
$$+ \frac{\varepsilon}{J} \sum_{k \in \mathbf{K}} \sigma_{kj}(q_{kj}, a_k, r_k), \ j = 1, \ldots, J, \quad (4)$$

where ε is a small parameter (e.g. $\varepsilon = \frac{0.1}{J}$). The ranking obtained on the basis of values $\sigma(\mathbf{q}_j, \mathbf{a}, \mathbf{r})$ for consequent variants $j \in \mathbf{J}$ is referred to as objective or objectified, rather, since it is based only on the data of the problem (the values of q_{kj} for $j \in \mathbf{J}$ and $k \in \mathbf{K}$). Such a ranking

Table 1
Data for calculations

Criterion number	Criterion	Variants				
		S	T	Z	U	P
1	PM [m^2]	2826	2920	2689.1	2689.1	3044
2	PZ	3.71	4.07	3.43	3.43	3.78
3	LM	60	78	67	62	78
4	KR [PLN·1000]	7714	8450	7350	7430	8490
5	CR [months]	12	18	14	15	11
6	PU	1	1	0.8	0.4	0.8
7	EE	1	0.6	0.8	1	1
8	DT	0.6	1	0.4	0.8	0.2

can also be subjectified, but in collegial decision making it is always better to start by presenting the group of decision makers with the objectified ranking first, before asking them whether they would also like to take into account the importance of criteria. This is because it happens very seldom that all decision makers agree on the importance of criteria and a specially devised voting procedure is necessary to achieve a consensus in this respect. However, if the group of decision makers agrees to subdivide the criteria into two sets $\mathbf{K}_1 \cup \mathbf{K}_2 = \mathbf{K}$ ($\mathbf{K}_1$ more important, $\mathbf{K}_2$ less important), then by assuming weighting coefficients $\alpha_1 > \alpha_2$ (e.g. $\alpha_1 = 1$, $\alpha_2 = 0.1$) for these two subsets of criteria we can modify Eq. (4), e.g. to the following form:

$$\sigma(\mathbf{q}_j, \mathbf{a}, \mathbf{r}) = \alpha_1 \min_{k \in \mathbf{K}_1} \sigma_{kj}(q_{kj}, a_k, r_k) + \alpha_2 \min_{k \in \mathbf{K}_2} \sigma_{kj}(q_{kj}, a_k, r_k)$$
$$+ \frac{\varepsilon}{J} \sum_{k \in \mathbf{K}} \sigma_{kj}(q_{kj}, a_k, r_k), \ j = 1, \dots, J. \quad (5)$$

In the following text, we apply the objectified ranking method to several examples of multiple criteria ranking of infrastructural projects found in literature, compare it with subjective ranking presented in literature and with a subjectification of the objectified ranking.

3. Example Case

We use the case published in [7], concerning the investors selection of the best architectural and technological variant of a multi-apartment residential building. The following variants of architectural, technological and material solutions are considered:

- S – four buildings with a single staircase, four storeys and a steel frame,

- T – four buildings with two segments each, built with the use of traditional technology,

- Z – one building of four segments with a monolith reinforced concrete constructure,

- U – one building with four segments and four storeys, traditional technology,

- P – one building with four segments and four storeys, new raised decking technology.

These variants are evaluated according to the following criteria:

- PM – living space (max),

- PZ – utilization of building space (max),

- LM – number of apartments (max),

- KR – full cost of realization (min),

- CR – planned realization time (min),

- PU – urban planning compliance (max),

- EE – esthetics of façade (max),

- DT – availability of technology (max).

The evaluation of criteria for subsequent variants, given in [7], is contained in Table 1 (source: Table 10.75 in [7]). When we apply the objectified ranking method, i.e. Eqs. (1)–(4) with $\alpha = 3$, $\beta = 7$, $\frac{\varepsilon}{J} = 0.1$, we obtain the following values of the overall achievement function, see Table 2.

Table 2
Values of achievement function for Table 1 data

Variants				
1 – S	2 – T	3 – Z	4 – U	5 – P
0.97396	0.93426	0.64526	0.63164	1.02908

Table 3
Data for calculations

Criterion no.	Criterion	Variants		
		1	2	3
Transport				
1	Length of road [km]	3.88	3.82	4.794
2	Number of junctions	1	1	2
Environmental				
3	Interference with "Nature 2000" area [ha]	8.1	2.1	1.9
4	Length of route leading through a forest [km]	2.5	1	2.1
Economic				
5	Construction cost [millions of PLN]	104.76	103.14	129.44
6	Economic net present value [millions of PLN]	32.4	27.5	19.5
Social				
7	Number of houses to demolish	6	18	4
8	Number of people exposed to excessive noise	45	120	20

Thus, the ranking list based on the objectified method is:

$$5 - P \blacktriangleright 1 - S \blacktriangleright 2 - T \blacktriangleright 3 - Z \blacktriangleright 4 - U, \qquad (6)$$

where $\blacktriangleright$ denotes *is better than*. In [7], five different subjective ranking methods are used, each resulting in a variant of the ranking list ([7], Table 10.90):

AHP/ANP add.: $5 - P \blacktriangleright 2 - T \blacktriangleright 1 - S \blacktriangleright 3 - Z \blacktriangleright 4 - U$

AHP/ANP mult.: $5 - P \blacktriangleright 1 - S \blacktriangleright 2 - T \blacktriangleright 3 - Z \blacktriangleright 4 - U$

DEMATEL: $5 - P \blacktriangleright 1 - S \blacktriangleright 2 - T \blacktriangleright 3 - Z \blacktriangleright 4 - U$

MUZ: $5 - P \blacktriangleright 2 - T \blacktriangleright 1 - S \blacktriangleright 3 - Z \blacktriangleright 4 - U$

TW: $5 - P \blacktriangleright 1 - S \blacktriangleright 3 - Z \blacktriangleright 2 - T \blacktriangleright 4 - U$

We note that the rankings obtained by all methods - both objectified and subjective – are similar, and the objectified method renders the same results as multiplicative AHP/ANP and DEMATEL [8]. However, it is not a particular ranking list that is the advantage of the objectified method, but the fact that it is not necessary to agree collegially on the relative importance of subsequent criteria.

4. Road Infrastructure

This example follows the data from [9]. The choice concerns the best variant of the selection of a route of a road. The data for calculations are given in Table 3 (source: Table 2 in [9]).

Maximized criterion is 6 – economic net present value. Minimized criteria are 1 – length of road, 2 – number of junctions, 3 – collision with "Nature 2000" area, 4 – length of route through forest, 5 – construction cost, 7 – number of houses to demolish, 8 – number of people exposed to excessive noise. When applying the method of objectified ranking as above, we obtain the following values of the overall achievement function shown in Table 4.

This example follows the data from [9]. The choice concerns the best variant of a path along which a road is to be constructed. The data for calculations are given in Table 3 (source: Table 2 in [9]).

The maximized criterion is 6 – economic net present value. Minimized criteria are 1 – length of road, 2 – number of junctions, 3 – interference with "Nature 2000" sites, 4 – length of route leading through a forest, 5 – construction cost, 7 – number of houses to demolish, 8 – number of people exposed to excessive noise. When applying the objectified ranking method, as presented above, we obtain the following values of the overall achievement function shown in Table 4.

Thus, the resulting ranking is V2 ▶ V1 ▶ V3. In the paper [9] six subjective ranking methods and five preference scenarios are considered, resulting in 30 different rankings, in which V2 was ranked first 16 times, while V1 only 5 times, and V3 9 times. Therefore, objectified ranking renders results that are not inconsistent with subjective rankings, but all that is done in a much simpler way and without the necessity to collegially agree on the scenario of preferences.

Table 4
Values of achievement function for the data in Table 3

Variants		
1	2	3
1.74425	1.84128	1.11481

Table 5
Data for calculations

No.	Criterion	Variants							
		1	2	3	4	5	6	7	8
		Ferro-concrete	Teriva	Cerambet	Ackerman	Muro-therm	Ytong	Filigran	Strop Smart
1	Cost [PLN]	1943.20	1123.80	1566.10	1528.80	1485.30	2105.20	1249.40	1346.80
2	Time [w-h]	34.50	14.20	20.10	27.30	14.90	1.80	4.80	2.50
3	Thermal insulation [U]	11.33	2.70	2.22	4.00	1.02	0.53	11.33	5.88
4	Acoustic insulation	5.00	4.00	4.00	4.00	3.00	2.00	5.00	5.00
5	Fire integrity	5.00	3.00	3.00	3.00	2.00	4.00	3.00	3.00
6	Simplicity of execution	1.00	3.00	3.00	2.00	3.00	5.00	4.00	5.00
7	Comfort of usage	4.00	3.00	5.00	5.00	3.00	2.00	4.00	4.00

Table 6
Values of the achievement function for Table 5 data

Variants							
1	2	3	4	5	6	7	8
0.45787	3.44978	2.78297	1.23442	0.286690	0.333333	4.669905	4.649105

5. Type of Ceiling Choice Case

This example is based on data given in [10] and concerns the selection of the best technology for constructing a ceiling. The data for calculations are presented in Table 5 (source Table 1 in [10]:

Minimized are criteria 1 – cost, 2 – time of construction. Maximized are all other criteria: 3 – thermal insulation, 4 – acoustic insulation, 5 – fire integrity, 6 – simplicity of execution, 7 – comfort of usage. The values of the overall achievement function calculated according to Eqs. (1)–(4) are given in Table 6.

Thus, the objectified ranking in this case is: V7 ▶ V8 ▶ V2 ▶ V3 ▶ V4 ▶ V1 ▶ V6 ▶ V5. In paper [10], calculations using ideal point method resulted in a slightly different ranking: V8 ▶ V7 ▶ V2 ▶ V5 ▶ V3 ▶ V4 ▶ V6 ▶ V1. Variants V7 and V8 are ranked best in both methods, although their placing is reversed.

6. Choosing Best Logistics Center Variant Case

This example is related to the choice of the best variant of a building for a logistics center, or actually any industrial building located in a big city. The data given in [11], [12] express the knowledge of experts in this field by evaluating numerous criteria with the use of values from the range of $<0, 1>$ and by applying weighting coefficients to these criteria. We summarize this data in Table 7 (source: [11], [12]).

Criteria in group A were minimized, in groups B, C, D, E they were maximized. While using the entropy method [11], [13] and the ideal point method [12], [14], the following two rankings were obtained from the data presented in Table 7.

Ideal point method: V2 ▶ V1 ▶ V3 ▶ V4 ▶ V5.

Entropy method: V2 ▶ V5 ▶ V1 ▶ V3 ▶ V4.

The objectified method results in the overall achievement values presented in Table 8.

Thus, the resulting ranking is:

Objectified method: V2 ▶ V3 ▶ V4 ▶ V5 ▶ V1.

While the best variant in all three methods is V2, the remaining ones are ordered differently. The question is whether this is the effect of taking into account the weighting coefficients in the first two methods. Knowing that the criteria from groups D, E are less important, we can apply Eq. (5) with $\mathbf{K}_1 = (A, B, C)$, $\mathbf{K}_2 = (D, E)$, $\alpha_1 = 1$, $\alpha_2 = 0.1$,

Table 7
Evaluation of five logistics center variants

Type and name of criterion	Weight	Evaluation of five variants				
Basic criteria	**0.89**	**V1**	**V2**	**V3**	**V4**	**V5**
A. Technology of construction	**0.31**	**0.92**	**0.83**	**0.81**	**0.89**	**0.80**
Time of construction	0.10	0.77	0.92	1.00	0.81	0.95
Cost of construction	0.10	1.00	0.80	0.70	0.90	0.70
Durability	0.03	1.00	0.75	1.00	1.00	1.00
Technological complexity	0.07	1.00	0.80	0.60	1.00	0.60
Ecological value	0.01	0.80	0.60	1.00	0.60	1.00
B. Adaptability of construction	**0.36**	**0.80**	**0.87**	**0.68**	**0.62**	**0.85**
Height utilized	0.12	1.00	0.80	0.80	1.00	0.80
Space between structural elements	0.18	1.00	1.00	0.50	0.50	0.90
Technology and construction of curtain walls	0.06	0.40	0.80	1.00	0.20	0.80
C. Functionality and utility value	**0.22**	**1.00**	**1.00**	**1.00**	**1.00**	**0.90**
Degree of fulfilment of qualitive requirements	0.11	1.00	1.00	1.00	1.00	0.80
Degree of fulfilment of safety requirements	0.06	1.00	1.00	1.00	1.00	1.00
Possibility of repeated usage	0.05	1.00	1.00	1.00	1.00	0.80
Additional criteria	0.11	**V1**	**V2**	**V3**	**V4**	**V5**
D. Placement of structure	**0.07**	**0.83**	**0.86**	**0.93**	**0.94**	**0.79**
Proximity of city railway lines	0.04	0.80	0.80	1.00	1.00	0.75
Proximity of trams and buses	0.01	1.00	1.00	1.00	1.00	0.80
Proximity of a throughway	0.01	0.60	0.80	1,00	0.60	0.80
Proximity of another industrial site	0.01	1.00	1.00	0.50	1.00	0.80
E. Standard of veneer	**0.04**	**0.79**	**0.90**	**0.94**	**0.95**	**0.83**
External walls	0.01	0.75	1.00	0.75	1.00	0.80
Internal walls	0.01	0.80	0.80	1.00	0.80	1.00
Administrative rooms	0.01	0.80	0.80	1.00	1.00	0.60
Social rooms	0.01	0.80	1.00	1.00	1.00	1.00

which leads to the following modification of Table 8. It is given as Table 9.

Table 8
Values of the achievement function for Table 7 data

Variants				
1	2	3	4	5
0.33912	5.34933	3.29121	0.61894	0.42953

Table 9
Values of the achievement function
according to Eq. (5) for Table 7 data

Variants				
1	2	3	4	5
0.33512	7.6816	3.87428	1.25894	0.38257

The resulting ranking is the same as one without weights:

Objectified method with weights: V2 ▶ V3 ▶ V4 ▶ V5 ▶ V1.

We conclude that the objectified ranking is quite robust and is not strongly influenced by criteria with small weighting coefficients.

7. Summary and Conclusions

This paper presented a comparison between the objectified ranking method and diverse subjective ranking methods used in various infrastructural applications. While the results achieved (rankings) are very similar, the objectified method is simpler and does not require the specification of weighting coefficients. Therefore, we can conclude as follows:

- The objectified ranking method renders good results when compared with other methods, while it is simpler in application and, most importantly, does not require any collegial agreement on weighting coefficients or on the importance of specific criteria.

- If the weighting coefficients are known, they might be used in a subjectified variant of the objectified method. However, the presence of many criteria of lesser importance does not necessarily impact the outcome ranking.

- Each method has its advantages and disadvantages, but subjective methods used in infrastructural problems are usually more complicated than the objectified ranking method. They also utilize weighting coefficients and often assume a linear combination of criteria, which can lead to erroneous results.

The ranking of potential solutions is not always decisive when selecting the variant of the investment project to be followed, but other factors beyond the merits, such as polit-

ical factors, are taken into account as well. As a result, financial resources, often originating from state budget, may be wasted.

References

[1] T. L. Saaty, *The Analytic Hierarchy Process: Planning Setting Priorities, Resource Allocation*. New York: McGraw-Hill Int. Book Co., 1980 (ISBN: 9780070543713).

[2] D. Sabaei, J. Erkoyuncu, and R. Roy, "A review of multi-criteria decision making methods for enhanced maintenance delivery", *Procedia CIRP*, vol. 3, pp. 30–35, 2015 (doi: 10.1016/j.procir.2015.08.086).

[3] T. Kuszewski and A. Sielska, "The useful art of social-economic rankings", *Współczesna Ekonomia*, vol. 13, no. 1, pp. 143–162, 2010 [in Polish].

[4] J. Tian, A. P. Wierzbicki, H. Ren, and Y. Nakamori, "A study on knowledge creation support in a Japanese Research Institute", in *Proc. of First Int. Conf. on Knowledge Science, Engineer. and Manag. KSEM06*, Guilin, Kuangsi, China 2006, pp. 405–417.

[5] A. P. Wierzbicki, "The problem of objective ranking: foundations, approaches and applications", *J. of Telecommun. and Infor. Technol.*, no. 3, pp. 15–23, 2008.

[6] E. Klimasara and A. P. Wierzbicki, "Examples of multiple criteria ranking in the design of cellular networks of mobile telecommunication", *Computer Science*, vol. 19, no. 1, pp. 101–114, 2018 (doi: 10.7494/csci.2018.19.1.2637).

[7] M. Dytczak, *Selected Methods of Solving Multi-criterial Decision Analysis in Civil Engineering*. Opole: Oficyna Wydawnicza Politechniki Opolskiej, 2010 (ISBN: 9788360691946) [in Polish].

[8] M. Dytczak and G. Ginda, "DEMATEL-based ranking approaches", *The Central Europ. Rev. of Econom. and Manag.*, vol. 16, no. 3 pp. 191–201, 2016 (doi: 10.29015/cerem.226).

[9] P. Żabicki and W. Gardziejczyk, "Issues of criteria normalization in the multicriteria analyzes in the design of roads", *Budownictwo i Architektura*, vol. 13, no. 4, pp. 325–333, 2014 [in Polish].

[10] M. Książek, P. Nowak, and J. Rosłoń, "Multi-criteria evaluation of selected slabs' design solutions", *Logistyka*, no. 6, pp. 6241–6250, 2014 [in Polish].

[11] M. Krzemiński and M. Książek, "Multicriteria assessment of logistic centres' structure with use of ideal entropy method", *Theor. Foundat. of Civil Engineer., Polish-Ukrainian Transac.*, vol. 21, pp. 419–428, 2013 [in Polish].

[12] M. Krzemiński and M. Książek, "Multicriteria assessment of logistic centres' structure with use of ideal point method", *Autobusy: Technika, Eksploatacja, Systemy Transportowe*, no. 3, pp. 741–748, 2013, [in Polish].

[13] F. Hosseinzadeh Lotfi and R. Fallahnejad, "Imprecise Shannon's entropy and multi attribute decision making", *Entropy*, vol. 12, no. 1, pp. 53–62, 2010 (doi: 10.3390/e12010053).

[14] H. Zamani-Sabzi, J. P. King, C. C. Gard, and S. Abudu, "Statistical and analytical comparison of multi-criteria decision-making techniques under fuzzy environment", *Operat. Research Perspec.*, no. 3 pp. 92–117, 2016 (doi: 10.1016/j.orp.2016.11.001).

Edward Klimasara received his M.Sc. from the Faculty of Mathematics and Mechanics, University of Warsaw, in 1977. Since 1984 he has been employed at the National Institute of Telecommunications in Warsaw, currently as the main specialist at the Department of Advanced Information Technology. He is an author and

co-author of publications in the field of computer science and telecommunications. His professional interests include knowledge management, multi-criteria optimization, Big Data, application of information techniques in telecommunications, medicine, transport, administration and education.

https://orcid.org/0000-0002-9761-0088

E-mail: E.Klimasara@itl.waw.pl

National Institute of Telecommunications

Szachowa 1

04-894 Warsaw, Poland

Andrzej Piotr Wierzbicki received his M.Sc. in telecommunications and control engineering in 1960, his Ph.D. in nonlinear dynamics in control in 1964, and his D.Sc. in optimization and decision science in 1968. He worked as the Dean of the Faculty of Electronics, Warsaw University of Technology (WUT), Poland (1975–1978); the Chairman of Systems and Decision Sciences Program of the International Institute for Applied Systems Analysis in Laxenburg n. Vienna, Austria (1979–1984). He was elected a member of the State Committee for Scientific Research of the Republic of Poland and the Chairman of its Applied Research Committee (1991–1994). He was the Director General of the National Institute of Telecommunications in Warsaw (1996–2004). He worked as a research Professor at the Japan Advanced Institute of Science and Technology (JAIST), Nomi, Ishikawa, Japan (2004–2007). Beside teaching and lecturing for over 45 years and promoting over 100 master's theses and 20 doctoral dissertations at WUT, he also lectured at doctoral studies at many Polish and international universities. Professor Wierzbicki is an author of over 200 publications, including 14 books, over 80 articles in scientific journals, over 100 conference papers; He is the author of 3 industrially applied patents. His current interests include vector optimization, multiple criteria and game theory approaches, negotiation and decision support, information society and knowledge civilization, rational evolutionary theory of intuition, theories of knowledge creation and management.

https://orcid.org/0000-0001-5817-3906

E-mail: A.Wierzbicki@itl.waw.pl

National Institute of Telecommunications

Szachowa 1

04-894 Warsaw, Poland

Energy Aware Data Centers and Networks: a Survey

Piotr Arabas

Institute of Control and Computation Engineering, Warsaw University of Technology, Warsaw, Poland

Abstract—The past years have brought about a great variety of clusters and clouds. This, combined with their increasing size and complexity, has resulted in an obvious need for power-saving control mechanisms. Upon presenting a basis on which such solutions - namely low-level power control interfaces, CPU governors and network topologies – are constructed, the paper summarizes network and cluster resources control algorithms. Finally, the need for integrated, hierarchical control is expressed, and specific examples are provided.

Keywords—*energy efficiency, green networks, resource allocation, HPC.*

1. Introduction

Distributed computer systems – clusters and clouds – have gained in popularity over the past years due to their versatility and easily scalable processing power. The important drivers of this are of economic nature: leased infrastructure is usually cheaper than owned hardware, thanks to better utilization and greater efficiency of high-end equipment. On the other hand, the costs of running large installations are tremendous. Therefore any improvement, negligible in a small-scale scenario, is worth considering.

One of the major cost factors is energy consumption [1]–[3]. It is important to note that energy consumed by IT equipment is nearly fully transformed into heat. Therefore, efficient cooling systems must be built – a task which becomes more and more difficult and costly, as the packing density increases.

The electronics industry has made a great effort to lower energy consumption of hardware: processors, memory, etc., but the demand for computing power grows so quickly that the activities undertaken are insufficient to solve the problem [4]. The most viable strategy consists in the application of power-aware control algorithms enabling to reduce power consumption during periods of limited load. This is possible, as lower level mechanisms, namely measurement and power control interfaces of specific IT system components, are readily available. Furthermore, the most popular operating systems offer some power control functionalities as well. The main challenge now is to orchestrate all these mechanisms to build efficient and flexible power control systems encompassing all elements of cloud infrastructure. The remaining part of this paper is organized as follows.

Section 2 briefly summarizes hardware-related, low level technologies, namely power-scaling, operating system level controllers and power-aware functionalities and the design of networks. Section 3 reviews a solution devoted to maximizing power efficiency of networks used to connect computing nodes of clusters and clouds. Section 4 describes selected resource allocation algorithms, and Section 5 proposes some integrated systems controlling all aspects of cloud operation. Section 6 concludes the survey.

2. Available Technologies

This section briefly presents technologies developed in recent years, enabling to control power consumption of computers and network equipment, and thus serving as a basis for designing more complex solutions.

2.1. Power-scaling Techniques

With the growing computing power of processors[1], the problem of energy dissipation and efficient cooling becomes important. So, the first attempts to reduce power consumption were more concerned with preventing the generation of excessive heat, rather than with energy savings. However, both targets were addressed by ACPI specification propositions. Now the ACPI standard [5] defines a relatively large number of energy-aware states for the operation of processors. *Performance states P*, associated with higher consumption levels, make it possible to choose the computing power required for processing tasks by means of decreasing processor clock frequency and supply voltage. This technique, known as dynamic voltage and frequency scaling (DVFS), is used by CMOS circuits which are controlled by the potential present at transistor gates.

As energy is required only when this potential is to be changed, lowering the clock frequency and voltage enables to reduce the power consumed. Additional reduction in power consumption is possible in the so-called *C states*, where the processor may enter a temporary sleep mode – i.e. may stop its operation and switch off some of its circuits. Figure 1 shows variations in computing power and energy consumption for specific energy-aware states. The

---

[1] Introduction of Intel 486 may be considered a time when the problem became broadly recognized, as it was the first consumer-use processor which needed fan cooling.

higher the number of processor subsystems switched off, the less energy is consumed, at the expense of a longer transition time, however. Common adoption of the ACPI standard has made it possible to implement processor power control modules in the majority of operating systems (e.g. Linux power governors) – see Subsection 2.2.

The adoption of similar techniques in network hardware is a little bit slower. However, IEEE802.3az [6] may be considered to be the most important standard, formerly known
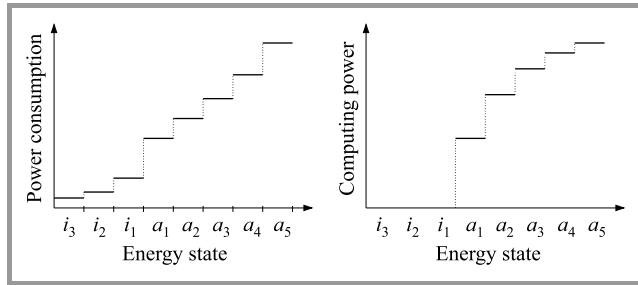
Fig. 1. Power consumption (a) and computation power (b) in subsequent energy-aware states. States i_3 to i_1 are idle (C states). States a_1 to a_5 are active (P states). For more detailed characteristics, see e.g. [4], [7], [8].

as Energy Efficient Ethernet employing Low Power Idle (LPI). It operates in a manner similar to C processor states. Implementation of LPI for highly redundant links present in typical cluster topologies may be the most natural way of reducing power consumed by the network. It must be also noted that some network devices are composed of hierarchically connected subsystems (e.g. Mellanox InfiniBand switches [9]) allowing to switch off some components, reducing capacity but preserving connectivity). Furthermore, modern routers and switches tend to use operating systems stemming from the main line of universal OSes (FreeBSD or Linux clones) running on PC-like equipment providing their control plane functions. This allows to accompany data plane-based techniques (e.g. 802.3az) with these inherited from baseline OS and data plane processors (e.g. ACPI and power governors).

Successful and broad exploitation of power saving mechanisms requires standardized interfaces. The ACPI specification referred to above unifies control of the processor and some other devices. A number of other standards, e.g. RAPL for Intel processors [10] or PAPI specification [11], and, to some extent, the IPMI interface [12], allow to access power measurements related to some components of a computer system[2]. Unified interfaces, such as the green abstraction layer (GAL) [13], [14], are the next generation solution having the form of a generalized interface providing the functionality of setting energy states for all computer system components and also, which is another in-

---

[2]Running average power limit (RAPL) provides access to various processor registers containing, inter alia, measurements of power consumed by processor components, cache memory, etc. While the original aim was primarily to provide power capping, these measurements are precise, short term averages relevant for designing dynamic processor frequency controllers [7].

novation, of querying energy-aware capabilities of the said components.

2.2. Local CPU Power Control

The availability of power consumption statistics and frequency scaling interfaces along with system load measurements makes the operating system kernel an ideal location to implement power saving processor control. In Linux, the most popular controller – on-demand [15] power governor – applies a simple yet robust mechanism responding to load fluctuations by selecting CPU frequency.

The operation relies on two thresholds. When the processor load grows above the higher threshold, the controller sets the maximum frequency value. If the load decreases below the lower threshold, the frequency is lowered by one step. As processor load is relative to frequency, this algorithm leads to selecting the clock rate which should be sufficient to handle the current tasks without causing excessive processor loads and delaying task execution. The resulting power characteristic, i.e. the ratio between power consumption and processor load, usually has the form of a nonlinear function – see Fig. 2. It must be noted, how-
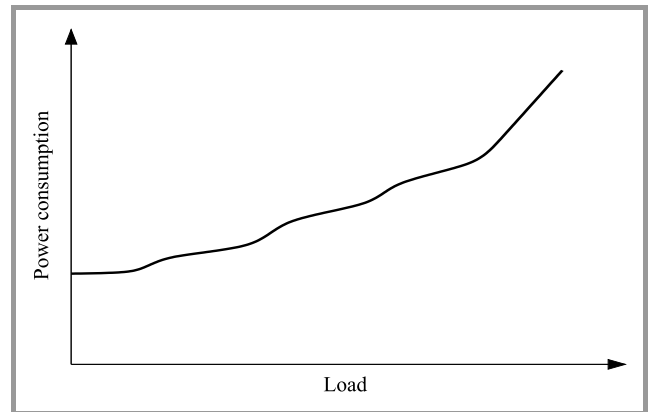

Fig. 2. Example of a power consumption curve for a processor with an on-demand governor. For more precise measurement data, see e.g. [4], [17].

ever, that even under the no-load condition, the processor still consumes power. Intel *p-state* [16] may be an example of a more complex governor, where the PI controller is relied upon to keep processor frequency close to the value ensuring optimized efficiency. Although these controllers offer significant energy savings, there is still some area for improvement, especially when the load characteristics are known. In such a case it is possible to construct specialized power governors suitable for specialized usage scenarios, e.g. a web server, large scale computations or network traffic filtering [7], [8], [17]. The savings achieved by the algorithm presented in [7] are attained mostly by exploiting identified dynamics of applications running, and thus designing a control law that makes it possible to adequately react to load changes. A discussion concerned with the possibility of designing PI and PID processor frequency

controllers may be also found in [18]. Similar mechanisms may be applied to control network devices, either software (Linux)-based routers implemented on general grade PC-class machines [19], [20] or specialized network devices – see, for instance, [9].

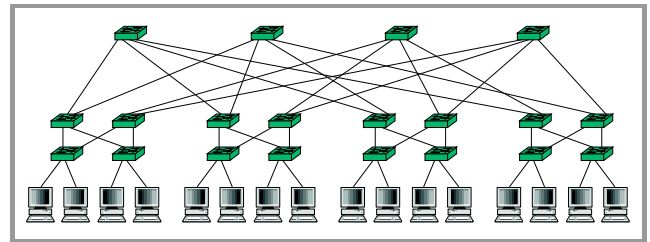### 2.3. Data Center Interconnect Network

Networks connecting machines that operate in clusters are usually designed to maximize throughput between any two components while providing extremely low latency and high availability. As a result, typical topologies are highly regular, usually hierarchical. Dominant technologies for machine-to-machine networks are Ethernet 1 Gbps, 10 Gbps and higher speeds or InfiniBand [21]. The use of a consistent technology across the data center enables to build a switched network limiting delays and complexity. Traditional topologies include two or three layers of switches with the lowest level switches installed on top of the rack (ToR) in the cabinets housing the servers (Fig. 3). To attain high reliability and to multiply bandwidth, the computers may use more than one network interface, connected to different switches. Similarly, ToR switches are interconnected with upper level devices [23]. It must be noted that providing full bandwidth between any pair of hosts requires that links connecting the individual levels offer capacity being at least equal to the sum of lower level links connected to the node. Such a topology requires using costly high-end switches to provide appropriate switching capacity and the number of ports needed, e.g. a 24-port 10 Gbps switch requires six 40 Gbps uplinks terminated at the upper level switch, to provide full bisection bandwidth. The example demonstrates the high cost involved, which grows exponentially as the network expands. In practice, based on the available size (i.e. number of ports) and the switching capacity of devices, it is possible to build clusters of up to several thousands of nodes using two levels of switches: ToR and core [22]. For larger installations, it is necessary to use at least three levels: ToR, aggregation, and core. The cost may be reduced by oversubscription, i.e. by connecting more nodes to ToR switches than their aggregated uplink bandwidth allows in the base, fully provisioned scenario. However, it also reduces the bandwidth available to hosts[3].
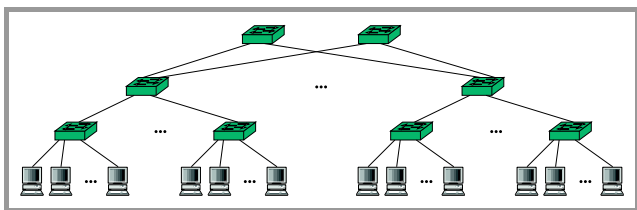
Fig. 3. Traditional topology of a cluster interconnect network [22], [23].

While oversubscription makes it possible to shift the limit, it does not allow to scale the network economically. To lower the cost and provide scalability, a number of topologies using baseline switches have been proposed (see e.g. [22]). They all exploit, to some point, the concept of a Clos network [24] to build a mesh network of a large number of simple devices having the same number of inputs, i.e. links connecting to the lower level of the tree, and outputs (uplinks). As the cost of higher class devices rises rapidly, it is possible to build a cheaper network, at the expense of more complex wiring[4]. A number of similar topologies attempting to solve some of the complexity-related issues have been proposed, with fat tree [25], [26] being the most common of them (along with some variations – see, e.g. [27] – flattened fat trees or [28], [29] – numerous variants of butterfly networks).

Fig. 4. Example of fat tree topology of a cluster interconnect network [22].

Regardless of which class of equipment is used, the resulting multiple tree topologies are substantially redundant, as they tend to provide full mesh connectivity and an equal level of service to all nodes. However, as the traffic pattern depends on the type of tasks performed by the cluster, it is often possible that the load is not spread equally across the network, so a substantial amount of energy may be saved if some of the devices are switched off or can operate with reduced performance. Provided they are equipped with some of the mechanisms mentioned in Subsection 2.1, it is possible to incorporate power saving functions into the network management system. Furthermore, as the entire network is managed by the same institution, application of the centralized controller is possible and in many aspects favorable. On the other hand, the large number of nodes and links makes the task of mathematical programming (e.g. known from [30]) formidable. Section 3 contains some proposals on how to solve this problem.

3. Energy Efficient Network Control

The power saving mechanism described in Subsection 2.1 may be helpful for constructing a local control mechanism that is relevant for a single device. However, especially when a whole network is considered, it may be more beneficial to rely on global, network-wide control. Such a mechanism, apart from acting reactively by setting the speed of

---

[3]At least in the worst-case scenario, when all nodes communicate simultaneously. In scenarios in which statistical multiplexing is possible, oversubscription is viable, provided its level is properly calculated.

[4]Which, in turn, may also be relatively cheap, as simpler copper cables may be used.

a single device in response to load, may reconfigure traffic paths to optimize energy levels of all devices within the network. Results of recent studies considering general purpose networks suggest that the overall energy consumption may be considerably reduced by the application of appropriate control mechanisms [31], [32]. In short, the proposed solutions may be divided into two major groups:

- distributed algorithms,
- centralized controllers.

The first group of algorithms is typically built on the base of existing routing protocols. Their main advantage is good scalability and relative robustness achieved by decomposition. Therefore, they are supposed to be correct (if not the sole feasible) approach in the case of large, general purpose networks. However, networks connecting machines within a cluster are very specific. Not only is their topology specialized (see Subsection 2.3), but they often use homogenic equipment and, more importantly, they are usually controlled by a single organization with the help of dedicated software. All these characteristics make it more acceptable to build a centralized controller in a scenario involving a free topology network.

Techniques used in this case are mostly related to traffic engineering and provisioning and require solving complex optimization tasks. However, as most large installations use some kind of a network management system, implementation of control may be relatively easy.

Both approaches are based on the obvious observation that redundant links, so abundant in cluster topologies, could be switched off or could operate at a limited rate during periods of lower loads. An additional source of redundancy present in cluster networks has the form of bundled links used to multiply their bandwidth. Typically, distributed power saving algorithms rely also on distributed sensing of the network load and react to it by increasing or decreasing the set of active paths. Centralized algorithms may rely on the monitoring system (if present) and request load statistics or even try to forecast traffic patterns based on historic data. The problem that most centralized algorithms suffer from is the availability of a traffic matrix necessary to define target demands between network nodes and to assess QoS. A situation in which the traffic matrix is provided by upper control level (e.g. tasks scheduling, see Section 5) and reflects the real needs of end users is beneficial, but not common. Deriving the traffic matrix from on-line measurements – an activity performed by many distributed controllers – may lead to an oscillatory behavior of flows due to interference with low level flow control mechanisms, e.g. TCP. Decoupling control by using different time scales (much longer for traffic engineering then for flow control) may be a solution here. A more detailed discussion is presented, for instance, in [33]–[36].

3.1. Centralized Power Save Network Control

Before offering specific, cluster-related formulations, some general proposals will be presented. Typically, they tend to formulate an optimization problem similar to the network design problem [30] or to the QoS provisioning task [34], [37], but with a cost function defined as total energy consumed by all components of the network. Such mathematical programming tasks are presented in [2], [33], [38], [39] and [40]. In [38], Chiaraviglio *et al.* uses integer linear programming to identify network nodes and links that can be switched off. Chabarek *et al.* proposes, in [2], reducing power consumption by finding links and line cards that can be switched off when a large mixed-integer linear problem is solved for a given traffic matrix. Similarly, [41] tries to scale link rates by selectively switching the fibers they are made of on and off. Optimization of a two-level structure with an IP network built with the use of optical equipment is covered in [42].

Optimization of an energy-aware network involves more difficulties than in the case of typical shortest path calculations, as paths are not independent but should be typically aggregated on a subset of links to allow switching off the remaining ones. Furthermore, energy consumption of devices often depends, in a non-convex way, on their load due to intrinsic nonlinearities and must be modeled using binary variables. All this makes the fully formulated problem NP-complete and large. Relaxing some constraints to simplify the solution may result in suboptimality or instability [43] of the system. In [44], the following variants of the mathematical programming task were presented, from an exact mixed integer programming (MIP) formulation, including complete routing and energy-state decisions, to simplified, continuous formulations:

1. complete network management problem with full routing and energy state control – MIP task,

2. energy state control with predefined paths – simplification of above MIP task,

3. continuous relaxation of formulation 1,

4. continuous relaxation of formulation 2.

Unfortunately, formulation 1 is a complex, NP-complete problem, so finding the solution for larger networks is usually impossible. Avoiding routing, like in the case of formulation 2, makes the computation easier, at the cost of the earlier path generation procedure referred to earlier[5]. Application of continuous relaxation variants yields multipath solutions, which are typically avoided due to the unacceptable jitter level and packet reordering.

Application of heuristics to eliminate some paths and to consolidate flows may be the solution here, at the expense of power efficiency or QoS. However, in some data center network topologies (e.g. fat tree), multipath routing is a must to balance the load of multiplied links, so the continuous solution may be easier to apply (although

---

[5]The resulting mechanism is usually suboptimal and, depending on network complexity, path generation may be a difficult and a time consuming task. On the other hand, for highly regular topologies, preparation of some predefined sets of paths seems viable.

some rounding off to meet equipment capacities may be necessary).

To demonstrate its complexity, the definition of the full MIP task [45]–[47] is provided below:

$$\min_{x_c, y_{ek}, z_r, u_{ed}} \left[\sum_{r=1}^{R} T_r z_r + \sum_{c=1}^{C} W_c x_c + \sum_{e=1}^{E} \sum_{k=1}^{K} \xi_{ek} y_{ek} \right] , \quad (1)$$

subject to the following constraints:

$$\forall_{e=1,\dots,E} \quad \sum_{k=1}^{K} y_{ek} \leq 1 , \quad (2)$$

$$\forall_{\substack{d=1,\dots,D, \\ c=1,\dots,C}} \quad \sum_{p=1}^{P} l_{cp} \sum_{e=1}^{E} a_{ep} u_{ed} \leq x_c , \quad (3)$$

$$\forall_{\substack{d=1,\dots,D, \\ c=1,\dots,C}} \quad \sum_{p=1}^{P} l_{cp} \sum_{e=1}^{E} b_{ep} u_{ed} \leq x_c , \quad (4)$$

$$\forall_{\substack{r=1,\dots,R, \\ c=1,\dots,C}} \quad g_{rc} x_c \leq z_r , \quad (5)$$

$$\forall_{\substack{d=1,\dots,D, \\ r=1,\dots,R, \\ p=s_d}} \sum_{c=1}^{C} g_{rc} l_{cp} \sum_{e=1}^{E} a_{ep} u_{ed} - \sum_{c=1}^{C} g_{rc} l_{cp} \sum_{e=1}^{E} b_{ep} u_{ed} = 1 , \quad (6)$$

$$\forall_{\substack{d=1,\dots,D, \\ r=1,\dots,R \\ p \neq t_d, p \neq s_d}} \sum_{c=1}^{C} g_{rc} \sum_{p=1}^{P} l_{cp} \sum_{e=1}^{E} a_{ep} u_{ed}$$
$$- \sum_{c=1}^{C} g_{rc} \sum_{p=1}^{P} l_{cp} \sum_{e=1}^{E} b_{ep} u_{ed} = 0 , \quad (7)$$

$$\forall_{\substack{d=1,\dots,D, \\ r=1,\dots,R, \\ p=t_d}} \sum_{c=1}^{C} g_{rc} l_{cp} \sum_{e=1}^{E} a_{ep} u_{ed}$$
$$- \sum_{c=1}^{C} g_{rc} l_{cp} \sum_{e=1}^{E} b_{ep} u_{ed} = -1 , \quad (8)$$

$$\forall_{e=1,\dots,E} \quad \sum_{d=1}^{D} V_d u_{ed} \leq \sum_{k=1}^{K} M_{ek} y_{ek} , \quad (9)$$

where: W_c and T_r are power consumption values of the card c and the router r, respectively, M_{ek} is throughput and ξ_{ek} power consumption of link e in the state k, $y_{ek} = 1$ if the energy state of link e is set to k (0 otherwise), $z_r = 1$ if router r transmits any flow (0 otherwise), $x_c = 1$ if card c transmits any flow (0 otherwise), $l_{cp} = 1$ if port (one of link endpoints) p is on the card c (0 otherwise), $u_{ed} = 1$ if path d leads through the link e (0 otherwise), binary constants a_{ep} and b_{ep} are used to define ingress and egress links (e) of port p, g_{rc} is set to 1 if card c belongs to the router r.

The complexity of the problem results from the fact that authors have combined the routing task - see flow continuity constraints (6)–(8) and link capacity constraint (9) – with hierarchic layout of the network node (router) – constraints (3)–(5) and the multiple energy state model of a link – constraint (2).

An efficient solution of such a complex task is possible by relying on heuristics, usually built as repetitive solving of simpler (usually relaxed) mathematic programming tasks. Relevant examples may be found in [41], [48], where the algorithm is run for a predefined set of links or in [49], [50], where the solution of a full routing task is replaced by the selection of paths from a set provided. Similarly, Garroppo *et al.* [51] solves a relaxed task to determine the use of links, and then runs heuristics to find out which links within the bundle may be switched off. Aggregation of nodes and demands may provide a precise solution by mathematical programming. However, one must remember that additional operations are needed to de-aggregate the results [38].

As it was explained in Subsection 2.3, the topologies used in cluster networks are highly regular and redundant. Although they provide full mesh connectivity, active links form a relatively sparse tree during periods of limited load. Such behavior allows to aggregate link loads, easily leading to reduced power consumption. The design of heuristics may be greatly simplified as well. The most popular strategies are greedy algorithms, attempting to pack as many flows as possible within a limited number of switches, moving from the lower layer up, to reduce power consumption. In typical two-layer fat tree topologies, the upper layer (i.e. aggregation) switches form a full mesh. Therefore, after consolidating flows to reduce the number of active lower layer switches, upper layer switches may be chosen easily, as the selection task boils down to powering up an appropriate number of such switches. In topologies with more than two layers, the procedure described should be repeated recursively, just like in HERO [52] (although authors discussed only two-layer case).

ElasticTree [53] may serve as another example of such an approach, where flows are assigned starting from leftmost switch, based on their declared peak rate. While the algorithm is time efficient, its main drawback consists in the fact that it needs a correct traffic matrix and allocates flows based on peak rate, which leads to a relatively low network utilization level. To achieve higher loads on selected links and to enable better consolidation, the algorithm must be aware, in some way, of the traffic level and must act adaptively. In [54], Wang *et al.* proposed CARPO – Correlation-Aware Power Optimization, based on observations showing that in the usual cluster operation not only is the network underutilized, but also many if not most of flows are negatively correlated. This observation implies that a much higher number of flows may be served by a single link than when it is computed based solely on the declared peak rate. Instead, the authors propose to use 90 percentile rates and to construct a greedy flow consolidation algorithm using flow correlation to calculate traffic mixing coefficients. The algorithm is repeated regularly and starts with calculating the flow correlation matrix. Then, it packs flows on switches, left to right, and taking into account the computed coefficients.

The packing flows from left to right may result in (possibly short-term) overload of the leftmost switches and in lowering QoS. To prevent this, authors of AgreeFlow [55]

propose to balance the load of active switches, and propose another interesting concept known as Flow-set routing and Lazy rerouting. The former is based on the presence of many flows which can be routed along the same path. To simplify the solution and speed-up flow consolidation, they are assigned to the common flow set based on a hash function. The algorithm's calculations are meant to be regularly repeated to react to flow fluctuations that may lead to a change in a large number of routing paths. To limit the stress of the underlying control plane protocols (OpenFlow), they apply Lazy rerouting, i.e. delay the setting paths until transmitting the first packet of a new flow.

PowerFCT [56] attempts to deal with QoS by very precise modeling of transmit queues in the switches. Thanks to the application of an ECN derived protocol (namely DCTCP), the queue length may be calculated in a simple manner. To facilitate the algorithm's operation, different classes of service flows are divided into two groups: long lasting noncritical flows and the ones with a defined completion time. The flows are routed by a heuristic algorithm which sets the energy states of network devices to meet QoS requirements.

3.2. Distributed Power Save Network Control

In general networks, the distributed control scheme is preferred over the centralized one due to its reliability and scalability. Some solutions of this class may be suitable for cluster networks as well, especially for larger ones that rely on IP routing protocols to simplify management, because distributed energy-aware mechanisms are typically built as extensions of existing routing protocols – e.g., OSPF [57], [58], [59] or MPLS. Extension of the signaling infrastructure with green functions allows to partially overcome the absence of the traffic matrix [48] – the past state of the network can be used to compute flows [58]. Apart from traditional traffic engineering, the Software Defined Network (SDN) concept may be exploited here as well. Although SDN usually uses a central controller, decisions may be implemented locally with the help of existing routing protocols.

In *GRiDA* [58], Bianzino *et al.* relied on information inferred from augmented OSPF LSA messages to build the topology and find out the congested links. With this data, nodes may decide to reduce their power consumption by switching links off. The decision is taken based not only on the current network state, but also on past observations. The node tries to set the configuration of links minimizing its cost function being a sum of power consumption and penalty. Penalty is used to accumulate the knowledge about the node's role in the network. If the decision leads to congestion, penalty is increased additively, while for beneficial decisions, it is decreased multiplicatively. The most important feature of GRiDA is that it does not need the traffic matrix. Instead, it learns the network's topology and state from LSA. As sending an augmented LSA may be spawned by topology change or congestion, it allows to construct a reactive algorithm suitable for networks of medium dynamics (i.e. measured in seconds). A similar

algorithm was proposed in [60]. The main difference is the use of a rule-based mechanism for switching the links leading to the router on and off. Relatively good results (at least in low-load periods) were obtained for the rollback last (RL) activation strategy and the least loaded link (LLL) deactivation strategy. The former switches on the link whose deactivation has caused the congestion, the latter selects the link carrying the least (possibly no) load to be switched off.

SENATOR [61] may serve as an example of the algorithm exploiting OSPF and SDN infrastructure, in which the SDN controller is used to switch off the unnecessary links, while OSPF is used to propagate topology changes. To ensure a smooth topology change, temporary tunnels are used to redirect traffic along the envisaged paths. All this makes the algorithm is a hybrid solution – the decision to switch off links is computed globally and is executed using an SDN controller. The new routing paths are computed in a distributed manner, however.

4. Power-saving Resource Allocation

As most of energy consumed by any cluster is used by computing servers, it is crucial to manage resource usage in a manner allowing to limit their power needs without deteriorating QoS significantly. This implies cooperation between the resource allocation system and low level power monitoring and power-saving mechanisms described in Subsection 2.1. To achieve this, one can formulate a power-aware resource allocation task similar to the network control task described in Subsection 3.1. Defining this task requires the knowledge of power characteristics of the servers. As the servers are not fully power-proportional, its solution typically results in some kind of task and server consolidation. The solution of this task may be either precise, determined with the help of mathematic programming, or approximated, relying on various heuristics, and serves as a basis for designing many algorithms.

Another group of purely heuristic algorithms is based solely on the server consolidation postulate – i.e. they are policy-based mechanisms assuming that assigning tasks to a minimum number of servers allowing to maintain the required QoS level is the optimal or close-to-optimal approach. It may be easily shown by a computational experiment that when long term average power consumption and QoS metrics are considered and the equipment is homogenic, such a solution is optimal [62]. In a more dynamic case, however, especially when task characteristics are uncertain or varying, a more complicated mechanism should be applied. Following the classification proposed in [63], it is possible to distinguish three types of resource allocation mechanisms:

- predictive allocation algorithms,
- reactive allocation strategies,
- a hybrid of the above.

The predictive approach takes into account historical observations of the load imposed on servers by tasks to forecast its future evolution and to assign resources to meet power and QoS criteria. On the contrary, the reactive mechanism relies on system observations and tries to execute some previously prepared actions to bring the system close to the defined working point. Typically, predictive algorithms operate based on longer time frames of hours, while reactive algorithms need to be much faster, with repetition occurring every few minutes. The shortcoming of the reactive approach is that large workload changes are managed with difficulty, while they can be easily handled by predictive mechanisms, provided that forecasts are available. On the other hand, predictive mechanisms cannot react to small workload fluctuations between repetitions, which may lead to unsatisfactory performance.

The proposal of hybrid mechanisms is a simple consequence of these observations. The predictive mechanism may be applied in the long term to facilitate long calculations needed to find the solution, while the reactive one acts between repetitions to accommodate workload fluctuations.

4.1. Predictive Allocation Algorithms

The graph coloring algorithm [64] may serve as an example of the predictive approach. It uses graph coloring to assign, to the servers, resource units demanded by tasks represented by graph nodes. Links in the graph describe time dependencies among tasks. The resulting problem may be solved by mathematic programming. However, a more efficient heuristic is proposed. The pMapper mechanism described in [62] is a type of a predictive algorithm aiming to migrate virtual machines between servers and, hence, packing tasks in a power efficient manner. Importantly, the mechanism takes into account not only current power consumption and QoS requirements of the individual applications, but also the migration history to prevent excessive overheads.

The allocation algorithm proposed in [65] is much more detailed, as it computes task allocations and fine-grained energy configuration of servers (namely P-states), simultaneously modeling heat generation and air conditioning costs. Two versions of mathematical programming tasks are proposed. One tends to minimize the aggregated tasks' completion time based on the total power consumption constraint, while the other minimizes power consumption with limited completion time. The allocations are found by solving relaxed continuous problems and then applying a heuristic algorithm to get a feasible integer solution.

4.2. Reactive Allocation Strategies

Most virtual machine migration algorithms may be considered reactive, they are typically spawned by the loss of QoS or power efficiency and do not analyze any historical data (pMapper [62] seems to be a special, very complex example). One of the typical strategies used to allocate virtual machines to servers is best fit decreasing (BFD) [66], which starts with VMs sorted in the descending order of the demanded resources and assigns them to the server having the minimum computing capacity. Power and computation capacity BFD (PCA-BFD) is a mutation of this method, and it computes the servers' power consumption to computing capacity ratio to take into account power efficiency as well. Similarly, low perturbation bin packing (LPBP) attempts to migrate VMs from the most power consuming server to the one with the lowest power demand. The variation of the method proposed in [66] differs in terms of the sorting order: it starts from the least demanding VM and allocates it to the server with the lowest power consumption. Similar strategies are described in [67], the most trivial being first fit (FF), in which the task is assigned to the first machine with sufficient capacity. Next fit (NF) tries to assign the task to the last allocated machine, as the easiest way of consolidation. Max rest on requirements (MRR) is also called the worst fit method, as it tends to allocate tasks to servers with the maximum remaining capacity. The idea of delayed allocation is the most important contribution presented in [67]. Task are allocated in groups – such an approach utilizes server capacity much better than in the case of allocating single tasks. The authors of [68] not only allocate tasks to servers based on the best fit strategy, but also control processor frequency to attain best efficiency.

4.3. Hybrid Allocation Algorithms

The hybrid mechanism proposed in [69] comprises two complementary tasks. The first one is a predictive algorithm based on Fourier analysis of historic workload traces. Such an approach makes it possible to find out regular changes of the workload profile and to prepare base resource allocation. On the other hand, it cannot account for sudden workload surges, which are controlled by the reactive mechanism adding resources to the tasks violating QoS thresholds.

The algorithm presented in [70] may be considered of the hybrid variety, as it uses predictions for the allocation of virtual machines, while implementing a reactive migration algorithm as well. Apart from a rather sophisticated framework incorporating some network monitoring functions, a proposition of novel statistics to estimate thresholds used for host overload and VM selection is another important contribution of this particular solution. The allocation mechanism presented in [71] may be considered a trivial case of a hybrid algorithm. It uses an auto regressive model to predict the future load for the task relied upon by the machine allocation mechanism, and enters unused machines into idle state. The reactive mechanism switches off those machines that have been idle past the declared time-out. By applying such a delayed power-off strategy, the authors attempt to refrain from switching off all unused servers, at least during periods of high load variations.

5. Hierarchical Control

While network control algorithms may help reduce the power consumed by network devices, and while task allocation or scheduling may economize power necessary for computers, the overall result may be greatly improved by coordinating operation of both components. It is obvious that consolidation of tasks on the part of a cluster may lead to reducing power consumption of both servers and the network. To build a manageable control structure, it is usually necessary to divide responsibility between a number of controllers with task allocation performed on the upper levels. Then it is possible to predict the traffic matrix and to identify transmission paths reducing power consumption.
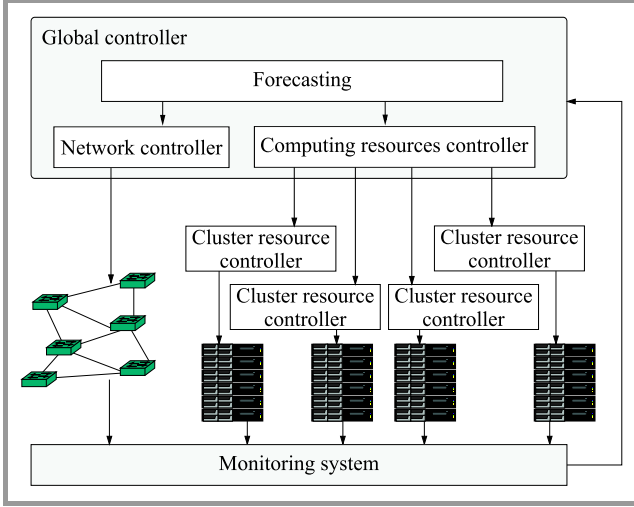
Fig. 5. Hierarchical control system [72].

The control framework proposed in [72] comprises two levels: the global level with separate network and computing resources controllers, and the local level with resource managers assigned to subsequent clusters (see Fig. 5). The global controller is responsible for network configuration (including selecting power states of links) and for assignment of tasks to the clusters. The allocation of tasks to processors is carried out by cluster resource controllers. The proposed mathematical programming task resembles that of the network control case [46], Eqs. (1)–(9) with their performance index augmented by the addition of the total power consumed by computing servers:

$$E_F(\mathbf{x_f}) = \sum_{f=1}^{F} \sum_{k=1}^{K_f} \bar{P}_f^k x_f^k , \qquad (10)$$

where $x_f^k = 1$ if the cluster f is in the state k (0 otherwise), and $\bar{P}_f^k$ is power consumed by the cluster f in state k. Constraints related to network function are defined in a manner similar to (2)–(9), and are accompanied by a set of inequalities describing the limitations of clusters:

$$\forall_{f=1,\ldots,F} \quad \sum_{k=1}^{K_f} x_f^k \leq 1 , \qquad (11)$$

$$\forall_{f=1,\ldots,F} \quad \sum_{j=1}^{J} W_j \vartheta_{fj} \leq \sum_{k=1}^{K_f} \bar{\Theta}_f^k x_f^k , \qquad (12)$$

$$\forall_{f=1,\ldots,F} \quad \sum_{j=1}^{J} M_j \vartheta_{fj} \leq \Psi_f , \qquad (13)$$

$$\forall_{j=1,\ldots,J} \quad \frac{W_j}{\sum_{f=1}^{F} \vartheta_{fj}(\sum_{k=1}^{K_f} \bar{\Theta}_f^k x_f^k - \sum_{i=1,i\neq j}^{J} W_i \vartheta_{fi})} \\ \leq T_j , \quad (14)$$

where: $\vartheta_{fj} = 1$ is assignment of task j to cluster f, $\bar{\Theta}_k^f$ is the computing capacity of cluster f in energy state k, W_j, M_j and T_j are workload (in MIPS), data size (in MB) and completion time T_j of task j.

The constraint (11) forces the cluster to work in a single energy-aware state, (12) and (13) are processing and memory capacity constraints. The most complicated, non-linear constraint (14) is the makespan limitation. Although the constraint (14) may be easily linearized, the resulting task complexity resembles that of the network control case. It must be also noted that hierarchical decomposition allows to use a simplified model of the cluster – it is described by a single processing element of the aggregated capacity and multilevel energy state. The authors suggest to solve the problem using a heuristic method based on successive continuous relaxation.

An example of a relatively complex, hierarchical structure is presented in [73]. It involves three levels: local controllers find the power-saving paths in their domain (part of cluster network), resource allocation controllers allocate tasks to processors, while the global controller coordinates actions concerning the network and the servers by controlling paths leading to top level switches. In general, both the task allocation algorithm and the path selection algorithm tend to consolidate task on servers and paths on links, respectively. The algorithms used are relatively simple heuristics (reuse of link/server till capacity limit). The authors demonstrated some power saving capabilities of their approach which may be fine-tuned by selecting a proper version of the algorithm. However, some trade-offs between efficiency and quality of service (e.g. longer delay) may be observed. The path selection algorithms draw heavily on the fat-tree structure of the network. In fact, it is the adoption of a regular topology that allows to construct a simple but effective heuristic.

6. Conclusions

This survey shows that many energy aware algorithms have been worked out for particular purposes: controlling utilization of processors by choosing the operating frequency, managing networks by routing and traffic engineering methods, allocating resources of clusters and grids. The need for this mechanism is obvious: energy should be used carefully, as it is one of the major components of costs involved, and, even more importantly, it is converted into harmful

heat. The environmental factors should also be considered, as we all should take care to limit emissions. The last issue may be to some extent mitigated by the application of heat recuperation systems (relied upon, for instance, to heat offices) or by incorporating green energy sources. In order to take advantage of all these particular solutions, an integrated hierarchical control system is needed. As presented in Section 5, designing such a system is possible, as most necessary interfaces (to control the network and the computing nodes, to provide power and utilization measurements, etc.) have been already standardized and implemented.

Acknowledgements

References

[1] S. Nedevschi, L. Popa, G. Iannacone, D. Wetherall, and S. Ratnasamy, "Reducing network energy consumption via sleeping and rate adaptation", in *Proc. 5th USENIX Symp. on Netw. Syst. Design and Implement. NSDI 2008*, San Francisco, CA, USA, 2008, pp. 323–336 [Online]. Available: https://www.usenix.org/legacy/events/nsdi08/tech/full_papers/nedevschi/nedevschi.pdf

[2] J. Chabarek *et al.*, "Power awareness in network design and routing", in *Proc. 27th Conf. on Comp. Commun. INFOCOM 2008*, Phoenix, AZ, USA, 2008, pp. 457–465 (doi: 10.1109/INFOCOM.2008.93).

[3] A. Shehabi *et al.*, "United States data center energy usage report", 06/2016, 2016 [Online]. Available: https://eta.lbl.gov/publications/united-states-data-center-energy

[4] R. Bolla, R. Bruschi, A. Carrega, and F. Davoli, "Theoretical and technological limitations of power scaling in network devices", in *Proc. Australasian Telecommun. Netw. and Appl. Conf. ATNAC 2010*, Auckland, New Zealand, 2010, pp. 37–42 (doi: 10.1109/ATNAC.2010.5680253).

[5] "Advanced Configuration and Power Interface Specification, Revision 5.0", Hewlett-Packard, Intel, Microsoft, Phoenix Technologies, and Toshiba, 2011 [Online]. Available: http://www.acpi.info/DOWNLOADS/ACPIspec50.pdf

[6] "IEEE 802.3az Energy Efficient Ethernet Task Force", IEEE, 2012 [Online]. Available: http://grouper.ieee.org/groups/802/3/az/public/index.html

[7] M. Karpowicz, "Energy-efficient CPU frequency control for the Linux system", *Concur. and Comput.: Pract. and Exper.*, vol. 28, no. 2, pp. 420–437, 2015 (doi: 10.1002/cpe.3476).

[8] M. Karpowicz, P. Arabas, and E. Niewiadomska-Szynkiewicz, "Design and implementation of energy-aware application-specific CPU frequency governors for the heterogeneous distributed computing systems", *Future Gener. Comp. Syst.*, vol. 78, pp. 302–315, 2018 (doi: 10.1016/j.future.2016.05.011).

[9] "Power saving features in Mellanox products", Melanox Technologies, 2013 [Online]. Available: http://www.mellanox.com/related-docs/whitepapers/WP_ECONET.pdf

[10] "Intel 64 and IA-32 Architectures Software Developer's Manual", 2015 [Online]. Available: http://www.intel.com/content/dam/www/public/us/en/documents/manuals/64-ia-32-architectures-software-developer-manual-325462.pdf

[11] S. Terpstra, H. Jagode, H. You, and J. Dongarra, "Collecting performance data with PAPI-C", in *Tools for High Performance Computing 2009. Proceedings of the 3rd International Workshop on Parallel Tools for High Performance Computing, September 2009, ZIH, Dresden*, M. S. Müller, M. M. Resch, A. Schulz, and W. E. Nagel, Eds. Berlin, Heidelberg: Springer, 2010, pp. 157–173.

[12] "IPMI – Intelligent Platform Management Interface Specification, Second Generation", Intel, Hewlett-Packard, NEC, and Dell, 2015 [Online]. Available: https://www.intel.com/content/www/us/en/servers/ipmi/ipmi-intelligent-platform-mgt-interface-spec-2nd-gen-v2-0-spec-update.html

[13] R. Bolla *et al.*, "The green abstraction layer: A standard power-management interface for next-generation network devices", *IEEE Internet Comput.*, vol. 17, no. 2, pp. 82–86, 2013 (doi: 10.1109/MIC.2013.39).

[14] R. Bolla *et al.*, "Large-scale validation and benchmarking of a network of power-conservative systems using ETSI's green abstraction layer", *Trans. on Emerg. Telecommun. Technol.*, vol. 27, no. 3, pp. 451–468, 2016 (doi: 10.1002/ett.3006).

[15] V. Pallipadi and A. Starikovskiy, "The ondemand governor: past, present and future", in *Proc. Linux Symp.*, Ottawa, Ontario, Canada, 2006, vol. 2, pp. 223–238 [Online]. Available: https://www.kernel.org/doc/ols/2006/ols2006v2-pages-223-238.pdf

[16] K. Accardi, "Balancing power and performance in the Linux kernel", 2015 [Online]. Available: https://events.static.linuxfound.org/sites/events/files/slides/LinuxConEurope_2015.pdf

[17] P. Arabas and M. Karpowicz, "Server power consumption: measurements and modeling with measurements", in *Challenges in Automation, Robotics and Measurement Techniques. Proceedings of AUTOMATION-2016, March 2-4, 2016, Warsaw, Poland*, R. Szewczyk, C. Zieliński, and M. Kaliczyńska, Eds. Springer, 2016, pp. 233–244 (ISBN 9783319293578).

[18] J. Gong and Ch. Xu, "A gray-box feedback control approach for system-level peak power management", in *Proc. 39th Int. Conf. on Parallel Process. ICPP-2010*, San Diego, CA, USA, 2010, pp. 555–564 (doi: 10.1109/ICPP.2010.63).

[19] R. Bolla, R. Bruschi, and A. Ranieri, "Green support for PC-based software router: Performance evaluation and modeling", in *IEEE Int. Conf. on Commun.*, Dresden, Germany, 2009, pp. 1–6 (doi: 10.1109/ICC.2009.5199050).

[20] R. Bolla and R. Bruschi, "Energy-aware load balancing for parallel packet processing engines", in *Online Conf. on Green Commun. GreenCom 2011*, New York, NY, USA, 2011, pp. 105–112 (doi: 10.1109/GreenCom.2011.6082516).

[21] M. Benito, E. Vallejo, and R. Beivide, "On the use of commodity Ethernet technology in exascale HPC systems", in *Proc. IEEE 22nd Int. Conf. on High Perform. Comput. HiPC 2015*, Bangalore, India, 2015, pp. 254–263 (doi: 10.1109/HiPC.2015.32).

[22] M. Al-Fares, A. Loukissas, and A. Vahdat, "A scalable, commodity data center network architecture", in *Proc. ACM SIGCOMM 2008 Conf. on Data Commun.*, Seattle, WA, USA, 2008, pp. 63–74 (doi: 10.1145/1402958.1402967).

[23] "Cisco Data Center Infrastructure 2.5 Design Guide", Cisco Systems, Inc., 2011 [Online]. Available: https://www.cisco.com/c/en/us/td/docs/solutions/Enterprise/Data_Center/DC_Infra2_5/DCI_SRND_2_5a_book.html

[24] C. Clos, "A study of non-blocking switching networks", *The Bell Syst. Tech. J.*, vol. 32, no. 2, pp. 406–424, 1953 (doi: 10.1002/j.1538-7305.1953.tb01433.x).

[25] Ch. E. Leiserson, "Fat-trees: Universal networks for hardware-efficient supercomputing", *IEEE Trans. Comput.*, vol. 34, no. 10, pp. 892–901, 1985 (doi: 10.1109/TC.1985.6312192).

[26] "Introduction to cloud design four design principals for IaaS", Melanox Technologies, 2012 [Online]. Available: http://www.mellanox.com/pdf/whitepapers/WP_Cloud_Computing.pdf

[27] Y. Xia and T. S. E. Ng, "Flat-tree: A convertible data center network architecture from CLOS to random graph", in *Proc. 15th ACM Worksh. on Hot Topics in Netw. HotNets'16*, Atlanta, GA, USA, 2016, pp. 71–77 (doi: 10.1145/3005745.3005763).

[28] J. Kim, W. J. Dally, and D. Abts, "Flattened butterfly: A cost-efficient topology for high-radix networks", in *Proc. of the 34th Ann. Int. Symp. on Comp. Architec. ISCA'07*, San Diego, CA, USA, 2007, pp. 126–137 (doi: 10.1145/1273440.1250679).

[29] A. Shpiner, Z. Haramaty, S. Eliad, V. Zdornov, B. Gafni, and E. Za-havi, "Dragonfly+: Low cost topology for scaling datacenters", in *Proc. IEEE 3rd Int. Worksh. on High-Perform. Interconnec. Netw. in the Exascale and Big-Data Era HiPINEB 2017*, 2017, Austin, TX, USA, pp. 1–8 (doi: 10.1109/HiPINEB.2017.11).

[30] M. Pióro, M. Mysłek, A. Juttner, J. Harmatos, and A. Szentesi, "Topological Design of MPLS Networks", in *Proc. Global Telecom-mun. Conf. GLOBECOM'2001*, San Antonio, TX, USA, 2001 (doi: 10.1109/GLOCOM.2001.965071).

[31] F. Bianco, G. Cucchietti, and G. Griffa, "Energy consumption trends in the next generation access network – a telco perspective", in *Proc. 29th Inter. Telecommun. Energy Conf. INTELEC 2007)*, Rome, Italy, 2007, pp. 737–742 (doi: 10.1109/INTLEC.2007.4448879).

[32] S. N. Roy, "Energy logic: a road map to reducing energy consump-tion in telecommunications networks", in *Proc. 30th Int. Telecom-mun. Energy Conf. INTELEC 2008*, San Diego, CA, USA, 2008 (doi: 10.1109/INTLEC.2008.4664025).

[33] A. Karbowski and P. Jaskóła, "Two approaches to dynamic power management in energy-aware computer networks – methodological considerations", in *Proc. Federated Conf. on Comp. Sci. and Inform. Syst. FedCSIS*, Łódź, Poland, 2015 (doi: 10.15439/2015F228).

[34] P. Jaskóła and K. Malinowski, "Two methods of optimal band-width allocation in TCP/IP networks with QoS differentiation", in *Proc. Symp. on Perform. Eva. of Comp. and Telecommun. Systems SPECTS'04*, San Jose, CA, USA, 2004, pp. 373–378, 2004.

[35] A. Kozakiewicz and K. Malinowski, "Network traffic routing using effective bandwidth theory", *Eur. Trans. on Telecommun.*, vol. 20, no. 7, pp. 660–667, 2009 (doi: 10.1002/ett.1383).

[36] M. Kamola and P. Arabas, "Dynamically established transmission paths in the future internet – proposal of a framework", *Bull. of the Polish Acad. of Sciences: Tech. Sciences*, vol. 59, no. 3, pp. 357–366, 2011 (doi: 10.2478/v10175-011-0043-9).

[37] K. Malinowski, E. Niewiadomska-Szynkiewicz, and P. Jaskóła, "Price method and network congestion control", *J. of Telecommun. and Inform. Technol.*, no. 2, pp. 73–77, 2010.

[38] L. Chiaraviglio, M. Mellia, and F. Neri, "Minimizing ISP network energy cost: formulation and solutions", *IEEE/ACM Trans. on Netw.*, vol. 20, no. 2, pp. 463–476, 2011 (doi: 10.1109/TNET.2011.2161487).

[39] E. Niewiadomska-Szynkiewicz, A. Sikora, P. Arabas, and J. Kołodziej, "Control framework for high performance energy aware backbone network", in *Proc. of 26th Eur. Conf. on Modell. and Simul. ECMS 2012*, Koblenz, Germany, 2012, pp. 490–496 (doi: 10.7148/2012-0490-0496).

[40] J. Restrepo, C. Gruber, and C. Machuca, "Energy profile aware rout-ing", in *Proc. IEEE Int. Conf. on Commun. Workshops ICC 2009*, Dresden, Germany, 2009, pp. 1–5, 2009 (doi: 10.1109/ICCW.2009.5208041).

[41] W. Fisher, M. Suchara, and J. Rexford, "Greening backbone net-works: reducing energy consumption by shutting off cables in bun-dled links", in *Proc. 1st ACM SIGCOMM Worksh. on Green Net-working Green Networking'10*, New Delhi, India, 2010, pp. 29–34 (doi: 10.1145/1851290.1851297).

[42] F. Idzikowski, S. Orlowski, Ch. Raack, H. Rasner, and A. Wolisz, "Saving energy in IP-over-WDM networks by switching off line cards in low-demand scenarios", in *Proc. 14th Conf. on Opt. Netw. Design and Model. ONDM'10*, Kyoto, Japan, 2010 (doi: 10.1109/ONDM.2010.5431569).

[43] N. Vasić and D. Kostić, "Energy-aware traffic engineering", in *Proc. 1st Int. Conf. on Energy-Efficient Comput. and Netw. E-ENERGY 2010*, Passau, Germany, 2010 (doi: 10.1145/1791314.1791341).

[44] P. Arabas, K. Malinowski, and A. Sikora, "On formulation of a net-work energy saving optimization problem", in *Proc. of 4th Int. Conf. on Commun. and Electron. ICCE 2012*, Hue, Vietnam, 2012, pp. 122–129 (doi: 10.1109/CCE.2012.6315903).

[45] E. Niewiadomska-Szynkiewicz *et al.*, "Network-wide power manage-ment in computer networks", in *Proc. 22nd ITC Special. Seminar on Energy Effic. and Green Netw. SSEEGN 2013*, Riccarton, New Zealand, 2013, pp. 25–30 (doi: 10.1109/SSEEGN.2013.6705398).

[46] E. Niewiadomska-Szynkiewicz *et al.*, "Dynamic power management in energy-aware computer networks and data intensive systems", *Fu-ture Gener. Comp. Syst.*, vol. 37, pp. 284–296, 2014 (doi: 10.1016/j.future.2013.10.002).

[47] M. Karpowicz, P. Arabas, and E. Niewiadomska-Szynkiewicz, "Energy-aware multilevel control system for a network of Linux software routers: design and implementation", *IEEE Syst. J.*, vol. 12, no. 1, pp. 571–582, 2018 (doi: 10.1109/JSYST.2015.2489244).

[48] L. Chiaraviglio, M. Mellia, and F. Neri, "Energy-aware backbone networks: a case study", in *Proc. IEEE Int. Conf. on Commun. Worksh. ICC2009*, Dresden, Germany, 2009, pp. 1–5 (doi: 10.1109/ICCW.2009.5208038).

[49] M. Zhang, Ch. Yi, B. Liu, and B. Zhang, "GreenTE: power-aware traffic engineering", in *Proc. IEEE Inter. Conf. on Netw. Protoc. ICNP'2010*, Kyoto, Japan, 2010 (doi: 10.1109/ICNP.2010.5762751).

[50] G. Shen and R. S. Tucker, "Energy-minimized desig for IP over WDM networks", *J. of Optical Commun. and Netw.*, vol. 1, no. 1, pp. 176–186, 2009 (doi: 10.1364/JOCN.1.000176).

[51] R. G. Garroppo, S. Giordano, G. Nencioni, and M. G. Scutella, "Power-aware routing and network design with bundled links: So-lutions and analysis", *J. of Comp. Netw. and Commun.*, vol. 2013, Article ID 154953, 2013 (doi: 10.1155/2013/154953).

[52] Y. Zhang and N. Ansari, "Hero: Hierarchical energy optimization for data center networks", *IEEE Systems J.*, vol. 9, no. 2, pp. 406–415, 2015 (doi: 10.1109/JSYST.2013.2285606).

[53] B. Heller *et al.*, "ElasticTree: Saving energy in data center net-works", in *Proc. 7th USENIX Conf. on Network. Syst. Design and Implemen. NSDI'10*, San Jose, CA, USA, 2010, p. 17.

[54] X. Wang, X. Wang, K. Zheng, Y. Yao, and Q. Cao, "Correlation-aware traffic consolidation for power optimization of data center net-works", *IEEE Trans. Parallel Distrib. Syst.*, vol. 27, no. 4, pp. 992–1006, 2016 (doi: 10.1109/TPDS.2015.2421492).

[55] Z. Guo, Sh. Hui, Y. Xu, and H. J. Chao, "Dynamic flow scheduling for power-efficient data center networks", in *Proc. IEEE/ACM 24th Int. Symp. on Quality of Service IWQoS 2016*, Beijing, China, 2016 (doi: 10.1109/IWQoS.2016.7590399).

[56] K. Zheng, X. Wang, and X. Wang, "PowerFCT: Power optimization of data center network with flow completion time constraints", in *Proc. IEEE Int. Paral. and Distrib. Proces. Symp.*, Hyderabad, India, 2015 (doi: 10.1109/IPDPS.2015.22).

[57] A. Cianfrani, V. Eramo, M. Listani, M. Marazza, and E. Vittorini, "An energy saving routing algorithm for a green OSPF protocol", in *Proc. IEEE Conf. on Comp. Commun. INFOCOM 2010*, San Diego, CA, USA, 2010, pp. 1 – 5 (doi: 10.1109/INFCOMW.2010.5466646).

[58] A. P. Bianzino, L. Chiaraviglio, and M. Mellia, GRiDA: a green dis-tributed algorithm for backbone networks", in *Online Conf. on Green Commun. GreenCom 2011*, New York, NY, USA, 2011, pp. 113–119 (doi: 10.1109/GreenCom.2011.6082517).

[59] F. Cuomo, A. Abbagnale, A. Cianfrani, and M. Polverini, "Keeping the connectivity and saving the energy in the Internet", in *Proc. IEEE INFOCOM 2011 Workshop on Green Communications and Networking*, Shanghai, China, 2011, pp. 319–324 (doi: 10.1109/INFCOMW.2011.5928831).

[60] M. Kamola and P. Arabas, "Shortest path green routing and the importance of traffic matrix knowledge", in *Proc. 24th Tyrrhenian Int. Worksh. on Digit. Commun. – Green ICT (TIWDC)*, Genoa, Italy, 2013 (doi: 10.1109/TIWDC.2013.6664215).

[61] H. Huin *et al.*, "Bringing energy aware routing closer to reality with SDN hybrid networks", in *Proc. Global Commun. Conf. GLOBE-COM'2017*, Singapore, 2017, pp. 1101–1107 (doi: 10.1109/GLO-COM.2017.8254456).

[62] A. Verma, P. Ahuja, and A. Neogi, "pMapper: Power and migration cost aware application placement in virtualized systems", in *Middle-ware 2008. ACM/IFIP/USENIX 9th International Middleware Con-ference Leuven, Belgium, December 1-5, 2008 Proceedings*, V. Is-sarny and R. Schantz, Eds. LNCS, vol. 5346, pp. 243–264. Springer, 2008 (doi: 10.1007/978-3-540-89856-6_13).

[63] A. Hameed *et al.*, "A survey and taxonomy on energy efficient resource allocation techniques for cloud computing systems", *Computing*, vol. 98, no. 7, pp. 751–774, 2016 (doi: 10.1007/s00607-014-0407-8).

[64] Ch. Ghribi and D. Zeghlache, "Exact and heuristic graph-coloring for energy efficient advance cloud resource reservation", in *Proc. 7th IEEE Int. Conf. on Cloud Comput. CLOUD 2014*, Anchorage, AK, USA, 2014, pp. 112–119 (doi: 10.1109/CLOUD.2014.25).

[65] A. M. Al-Qawasmeh, S. Pasricha, A. A. Maciejewski, and H. J. Siegel, "Power and thermal-aware workload allocation in heterogeneous data centers", *IEEE Trans. on Comp.*, vol. 64, no. 2, pp. 477–491, 2015 (doi: 10.1109/TC.2013.116).

[66] K. Gupta and V. Katiyar, "Energy aware virtual machine migration techniques for cloud environment", *Int. J. of Comp. Appl.*, vol. 141, no. 2, pp. 11–16, 2016 (doi: 10.5120/ijca2016909551).

[67] V. Armant, M. De Cauwer, K. N. Brown, and B. O'Sullivan, "Semi-online task assignment policies for workload consolidation in cloud computing systems", *Future Gener. Comp. Syst.*, vol. 82, pp. 89–103, 2018 (doi: 10.1016/j.future.2017.12.035).

[68] M. Zhang, S. Wang, G. Yuan, Y. Li, and Z. Qian, "Energy-efficient real-time task allocation in a data center", in *Proc. IEEE Int. Conf. on Internet of Things and Green Comput. and Commun. and Cyber, Phys. and Soc. Comput. and Smart Data iThings-GreenCom-CPSCom-SmartData 2016*, Chengdu, China, 2016, pp. 680–687 (doi: 10.1109/iThings-GreenCom-CPSCom-SmartData.2016.147).

[69] A. Gandhi, Yuan Ch., D. Gmach, M. Arlitt, and M. Marwah, "Minimizing data center SLA violations and power consumption via hybrid resource provisioning", in *Proc. Int. Green Comput. Conf. and Worksh. IGCC'11*, Orlando, FL, USA, 2011, pp. 1–8 (doi: 10.1109/IGCC.2011.6008611).

[70] S. Hasan and E. Huh, "Heuristic based energy-aware resource allocation by dynamic consolidation of virtual machines in cloud data center", *TIIS*, vol. 7, no. 8, pp. 1825–1842, 2013 (doi: 10.3837/tiis.2013.08.005).

[71] Y. Shao, C. Li, W. Dong, and Y. Liu, "Energy-aware dynamic resource allocation on hadoop YARN cluster", in *Proc. 18th Int. Conf. on High Perform. Comput. and Commun., 14th Int. Conf. on Smart City, 2nd Int. Conf. on Data Sci. and Syst. HPCC/SmartCity/DSS 2016*, Sydney, NSW, Australia, 2016, pp. 364–371 (doi: 10.1109/HPCC-SmartCity-DSS.2016.0059).

[72] E. Niewiadomska-Szynkiewicz and P. Arabas, "Resource management system for HPC computing", in *Automation 2018. Advances in Automation, Robotics and Measurement Techniques, Warsaw, Poland*, R. Szewczyk, C. Zieliński, and M. Kaliczyńska, Eds. *Advances in Intelligent Systems and Computing*, vol. 743, pp. 52–61. Springer, 2018 (doi: 10.1007/978-3-319-77179-3_5).

[73] M. Ray, S. Sondur, J. Biswas, A. Pal, and K. Kant, "Opportunistic power savings with coordinated control in data center networks", in *Proc. 19th ICDCN Int. Conf. on Distrib. Comput. and Netw.*, Varanasi, India, 2018, Article no. 48, pp. 48:1–48:10 (doi: 10.1145/3154273.3154328).

Piotr Arabas received his Ph.D. in Computer Science from the Warsaw University of Technology, Poland, in 2004. Currently he is an Assistant Professor at the Institute of Control and Computation Engineering at the Warsaw University of Technology. He has been with the Research and Academic Computer Network (NASK) since 2002. His research interest focuses on energy-efficient control of computer systems and networks, analysis of social networks, predictive control and hierarchical systems.

https://orcid.org/0000-0002-4173-3249
E-mail: parabas@ia.pw.edu.pl
Institute of Control and Computation Engineering
Warsaw University of Technology
Nowowiejska 15/19
00-665 Warsaw, Poland

Performance Modeling
of Database Systems: a Survey

Antonina Krajewska

Institute of Control and Computation Engineering, Warsaw University of Technology, Warsaw, Poland

Abstract—This paper presents a systematic survey of the existing database system performance evaluation models based on the queueing theory. The continuous evolution of the methodologies developed is classified according to the mathematical modeling language used. This survey covers formal models – from queueing systems and queueing networks to queueing Petri nets. Some fundamentals of the queueing system theory are presented and queueing system models are classified according to service time distribution. The paper introduces queueing networks and considers several classification criteria applicable to such models. This survey distinguishes methodologies, which evaluate database performance at the integrated system level. Finally, queueing Petri nets are introduced, which combine modeling power of queueing networks and Petri nets. Two performance models within this formalism are investigated. We find that an insufficient amount of research effort is directed into the area of NoSQL data stores. Vast majority of models developed focus on traditional relational models. These models should be adapted to evaluate performance of non-relational data stores.

Keywords—database systems, NoSQL data stores, performance evaluation, queueing networks, queueing Petri nets.

1. Introduction

Database servers play a crucial role in information system infrastructures. With the rapid expansion of Big Data analytics, NoSQL data stores keep gaining strategic significance and supplement traditional relational databases. Extensive research has been conducted in the area of performance of relational databases, which has been reviewed thoroughly [1], [2]. The vast majority of existing models is based on the queueing theory. Still, the performance of NoSQL data stores remains unexplored. Therefore, existing performance models should be revisited to identify their potential in capturing the dynamics of non-relational systems. This paper presents a survey of the existing database performance models with their underlying queueing networks and queueing Petri nets.

Several classification criteria may be considered in the analysis of the performance models constructed. Existing models may be categorized according to the field of study, i.e. concurrency control, replication mechanism or database architecture. Depending on the phenomena investigated, researchers may choose different performance criteria, such as request response time or transaction through-put. Finally, analytical and simulation studies may be distinguished.

In this paper, existing models are classified according to their mathematical modeling formalism. Firstly, models which are based on the queueing system are considered, then a review of the queueing network models is presented. Lastly, the paper presents queueing Petri nets as a powerful tool for qualitative and quantitative analysis of the performance of database systems. With this approach adopted, the article demonstrates the evolution of performance evaluation models. Section 2 presents fundamentals of the queueing systems theory and a review of queueing system-based representation of database dynamics. In Section 3, queueing network models are considered and various aspects of such models, including the open and closed character of a network or service time distribution, are recorded. Following Osman and Knottenbelt, we also classify them according to the granularity of transaction modeling and describe two performance evaluation models which enable the mapping of database design specification at the integrated system level to the queueing network structure [2]. Finally, in Section 4, queueing Petri nets are introduced which combine the expressiveness of modeling of queueing networks and Petri nets.

2. Queueing Systems

This section presents database performance evaluation models with a single underlying queueing system node. A queueing system is defined by the following: (i) stochastic process describing the customer arrival stream $A(t)$, (ii) probability distribution of service time $B(t)$ per customer, (iii) number of service channels K and (iv) scheduling discipline in the queue, i.e. first-come-first-serve (FCFS), last-come-first-serve (LCFS), processor sharing (PS), round robin (RR), etc. [3]. In this paper, the simple Kendall's notation is used: $A/B/K$ [3]. If no other scheduling discipline is denoted, the FCFS variety is adopted. The queueing theory enables a probabilistic analysis of such systems considering the average length, probability that the queue has a given length, or average service time. The relation between the average number of customers in a system is given by applying Little's law. It claims that the average number of customers in a queueing system is equal to the average arrival rate of customers to that system, times

the average time spent in that system. Queueing system models are capable of system-level workload evaluation for both centralized and distributed database systems [3]. This section focuses on Markovian models and classification according to the service time distribution.

2.1. Markovian Models with Exponentially Distributed Service Times

Firstly, we will discuss $M/M/m$ models in which customers arrive at rate λ, according to the Poisson process, and are served by m servers. The service time for each customer has an exponential distribution with parameter μ. The $M/M/1$ system with one server (Fig. 1) is a special kind of a Markovian model with exponentially distributed service time.
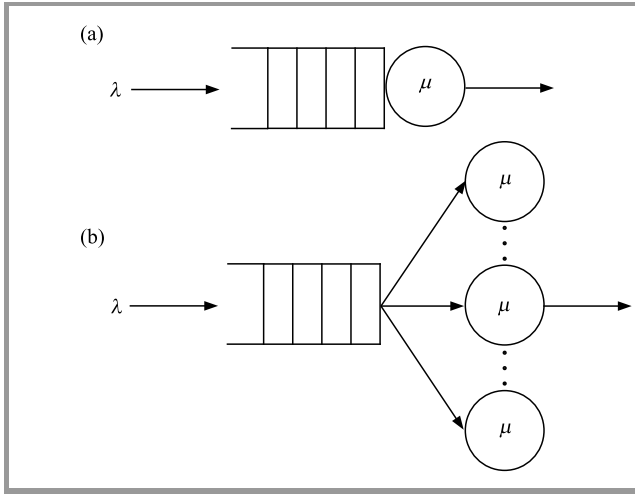
Fig. 1. Queueing system: (a) $M/M/1$, (b) $M/M/m$.

Let $N(t)$ denote the number of customers in a system at time t, then $N(t)$ is a continuous time Markov chain shown in Fig. 2.
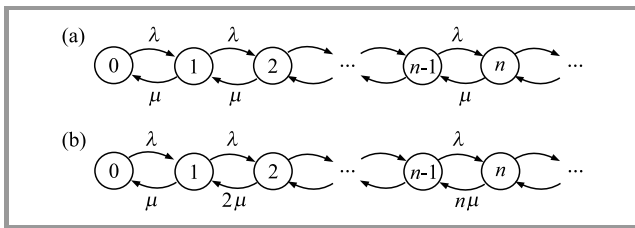
Fig. 2. State space diagram for the Markov chain for: (a) $M/M/1$, (b) $M/M/m$ systems.

Nicola and Jarke in [1] review the performance models of distributed and replicated database systems. They record that first queueing models of distributed databases were constructed by Coffmann *et al.* [4], Bacelli and Coffman [5], as well as Nelson and Iyer [6]. Nelson and Iyer [6] compare the performance of synchronous and non-synchronous updating policies in the $M/M/m$ system with non-preemptive processing of write operations. By contrast, Bacelli and Coffmann *et al.* [5] analyze different repli-

cation policies in a system with preemptive priority for write requests over read requests. Interruptions dynamics was captured by the $M/M/1$ system in the following way: read transactions could occupy m servers concurrently, until write requests arrived with the rate λ, according to the Poisson process, and were processed by all of the m servers with exponentially distributed service time.

2.2. Markovian Models with Generally Distributed Service Times

In a $M/G/1$, queueing system customers arrive according to the Poisson distribution process and have generally distributed service times. In contrast to $M/M/1$ queues, general models do not a have closed form describing the number of jobs in the system in a stationary state. However, according to the Pollaczek-Khinchine formula, the average $M/G/1$ queue length is given by Eq. 1 [3]:

$$L = \rho + \frac{\rho^2 + \lambda^2 \mathrm{Var}(S)}{2(1-\rho)} \ , \qquad (1)$$

where λ is the arrival rate of the Poisson process, $\frac{1}{\mu}$ is the mean of the service time distribution S, $\rho = \frac{\lambda}{\mu}$ is the utilization and $\mathrm{Var}(S)$ is the variance of the service time distribution S.

Arzauga and Kaeli [7] construct $M/G/1$ model of a storage area network (SAN) system that manages multiple applications stored in the same volume. Their experiments included various types of workloads running in the same system.

2.3. Load Dependent Systems M/M/m-LDS

Kihl *et al.* [8] have shown that the $M/M/1$ model does not capture the high load dynamics of a database for the write operations workload. Paper [9] proposes a model which adds load dependency to the service time. Its authors construct the $M/M/m$-LDS model in which service time depends on the number of concurrent requests.
Let $x_t(n)$ be the service time at time t, and n be the number of concurrent requests in the system. Then:

$$x_t(n) = x_{base}(1+f)^{n-1} \ , \qquad (2)$$

where x_{base} is exponentially distributed base service time for a system containing one job and $f \in [0,1]$ is a dependency factor. The model has been validated with experimental data and its accuracy in predicting response time for a write-heavy workload has been proved. Figure 3 presents
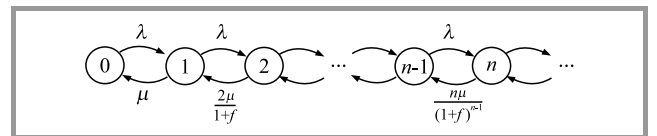

Fig. 3. State space diagram for Markov chain for the $M/M/m$-LDS system.

Markov chain for $N(t)$ process of the number of customers in the $M/M/m$-LDS system.

3. Queueing Networks

Database models based on single queueing systems can capture only particular aspects of database system dynamics. The authors of [1] have noted disadvantages of this approach in modeling replicated and distributed database systems, such as using the same queue by all database sites or neglecting inter-site communication. Furthermore, in the case of replicated database systems, full replication must be assumed. These limitations do not have to apply to queueing network models [3]. Queueing networks consist of several queueing nodes. Customers are routed between the nodes probabilistically. After being serviced at one queue node, the customer may join another node to receive additional service or may depart the network [3].

3.1. Open and Closed Queueing Networks

Open queueing networks allow external arrivals of customers. The number of customers in the network is variable and the arrival rate does not depend on the number of currently processed requests. To the contrary, *closed queueing networks* consider s fixed number of customers in the system. Once a request has been served, it is replaced with another one.

Open queueing networks have been widely used in performance modeling of database systems [10]–[13]. In [14], Mei *et al.* represent a database with an open two-node queueing network with a central processor-sharing node and multiple multi-server nodes. Parallel access to the backend database is modeled as a multi-service FCFS queue with exponential service times.

Since the number of requests arriving at the database system is not fixed, open queueing networks are generally more adequate. However, in specific problems, closed networks simplify performance evaluation. Nicola and Jarke record their significance in an analysis of concurrency control mechanisms. Performance of such methods depends on the multiprogramming level and thus its evaluation becomes challenging in networks with a variable number of jobs [1].

Liang and Tripathi in [15] introduce saga transactions – a special case of long-lived transactions (LLTs) – which release their locks as soon as possible. To evaluate the performance of saga systems, authors define an analytical model with a closed queueing network underlying. Carey and Livini [16] deploy a closed queueing network to analyze the performance of different concurrency control mechanisms in distributed database systems.

3.2. Service Time Distribution

Exponentially distributed service time. The simplest networks consist of $M/M/1$ queues, thus service time of all transactions is distributed according to the same exponential distribution. Ciciani *et al.* [17], [18] develop an analytical model to compare concurrency control in a replicated, distributed environments. In the queueing network, each database site is represented as an $M/M/1$ queue.

Generally distributed service time. More general models use $M/G/1$ queues. Banerjee *et al.* [19] use networks of $M/G/1$ queue nodes to benchmark the concurrency control protocol developed in a distributed environment against existing concurrency control schemes. Hwang *et al.* [20] compare the performance of three replication schemes using a model in which each database site is represented by an $M/G/1/RR$ node with the round robin scheduling discipline.

$M/H_n/*$ models. The above-mentioned models do not distinguish transactions according to service time distribution. As a remedy for this unrealistic assumption, networks with $M/H_n/*$ queue nodes might be used. In such models, arriving requests are divided into n categories with different service times. Since the service time in each class has exponential distribution, service time for the combined arrival process follows an n-phase hyper-exponential distribution. Leung [21] assigns different exponentially distributed service times for read and update requests. He uses a network with $M/H_2/1$ queueing nodes with 2-phase hyper-exponentially distributed service times.
Nicola and Jarke [1] introduce an analytical model which emphasizes the mutual influence between replication and inter-site communication. Each database site is represented by an $M/H_n/1$ queueing node.

Deterministic service time. Born in [22] uses an $M/D/1$ queueing network to investigate trade-offs of different implementation strategies for distributed lock management. The algorithms compared differ in the management of the lock database, in its optimization and in the communication protocol.

3.3. Granularity of Transaction Modeling

Osman and Knottenbelt [2] classify queueing database performance models according to the granularity of transaction modeling. In their exhaustive survey, the authors distinguish four types of performance evaluation models: *black box, transaction processing, transaction size and transaction phase*.

Black box models. Such models categorize transactions according to their service demands at the system level.
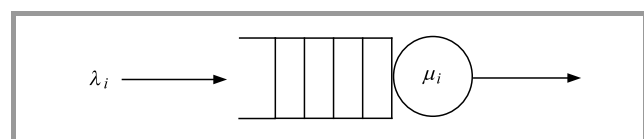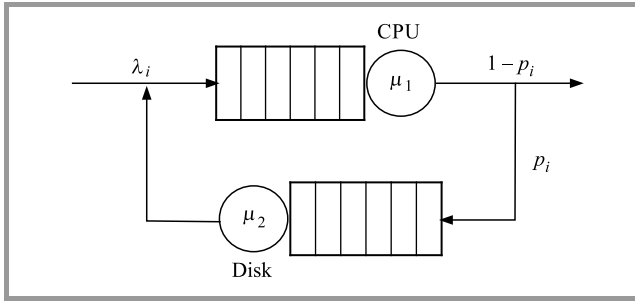

Fig. 4. Black box model: transaction class i is defined by arrival rate λ_i and service rate μ_i.

Transaction T_i from class i arrives to the database with the rate λ_i and a service rate μ_i (Fig. 4).

Black box models may represent both centralized and distributed database systems. The former design type is represented by a queueing system, whereas the latter one can be modeled both with a queueing system with multiple servers [5], [6] or as a queueing network with multiple nodes [17], [18]. Such a model can vary in service time distribution as discussed in the previous subsection.

Transaction processing models. In transaction processing models, the queueing network represents hardware architecture which lies below the database system. Transactions arriving to the system are classified according to their service demands on particular components of the hardware architecture, for example CPU or disk resources (Fig. 5). For each transaction class T_i arrival rate λ_i is defined. Transactions are routed throughout the network probabilistically.


Fig. 5. Transaction processing model. Transaction class i is defined by arrival rate λ_i and resource demands: CPU service rate μ_1 and disk service rate μ_2. Transactions are routed between service stations probabilistically.

In the transaction processing approach, the main variables to be optimized are capacity or quantity of physical resources [2].

Menasce *et al.* [23] emphasize the impact of considering both software and hardware contention. Software contention is caused by a limited number of threads that may process arriving transactions. When all threads in the system are busy, requests are placed in a queue. The number of active threads is defined by a birth-death process. Hardware contention is captured by a closed queueing network with nodes representing CPU and disk resources. Authors study both single- and multi-class requests categorization.

Gijsen *et al.* [24] study sojourn times in an open queueing network with a single processor sharing node and an arbitrary number of multi-server FCFS nodes. PS-node represents the front-end application server, while multi-server FCFS nodes stand for distributed database sites. The Poisson process defines the processor sharing node's arrival stream. After being processed by the front-end node, the request is routed to one of a database site nodes or departs the system. After customer has been serviced in one of FCFS nodes, it jumps back to a processor sharing node.

Transaction size model. In this approach, transaction class T_i is defined by the number n_i of data objects it accesses. Osman and Knottenbelt in [2] mention various object types including table rows, data pages or locks. Resource demands on hardware resources depend on n (see Fig. 6).
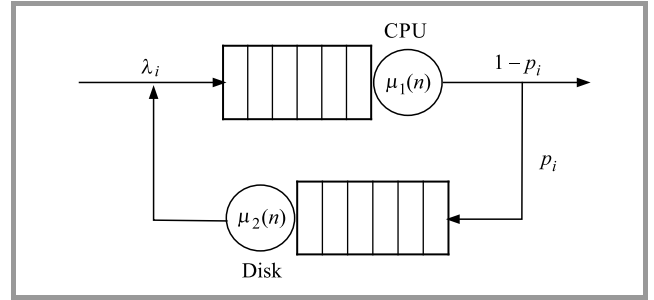
Fig. 6. Transaction size model. Transaction class i is defined by arrival rate λ_i and number of objects n it accesses. Resource demands depends on n: CPU service rate $\mu_1(n)$ and disk service rate $\mu_2(n)$.

Thomasian and Ryu [25] develop a model predicting the maximum throughput of both locking and optimistic concurrency control algorithms in a centralized database environment. Each transaction class accesses a fixed number of randomly chosen data objects. This set is then used to determine locking conflict probability. A closed queueing network represents the hardware architecture of a system with processor sharing and disks queues. Authors define the lock scheduling overhead as a function of the number of idle and active transactions. The same approach has been used by Morris and Wong [26]. However, the authors neglected the concurrency overhead.

Transaction phase. In the transaction phase model, transactions are categorized according to execution phases they consist of. Phases might be parallel or sequential. Queue nodes in a network represent these phases and transactions are routed between them probabilistically (see Fig. 7).
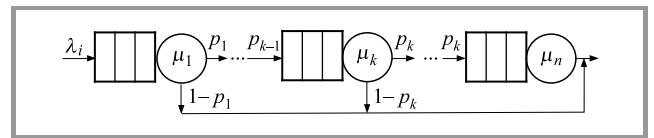
Fig. 7. Transaction phase model. Transaction class i is defined by arrival rate λ_i and number of phases n (p_i – probability of moving from phase i to phase $i+1$ and μ_i service rate in phase i).

Yu *et al.* [27] distinguish three processing phases: *front-end*, *application* and *database request*. In their study, the authors investigate dynamic transaction routing in locally distributed databases.

3.4. Database System Performance Evaluation Models

Following Osman and Knottenbelt [2], we would like to describe performance analysis methodologies which do not focus on particular database management system (DBMS)

components or constructs, but evaluate DBMS performance of the integrated system. Such an approach allows mapping database system specifications onto queueing network models. In their survey shown in paper [2], the authors present eight methodologies. In this paper, attention is focused on the two most recent of them.

Parallel relational database system performance evaluation. Tomov *et al.* [28] describe an analytical methodology for response time estimation. The queries analyzed execute within a shared-nothing parallel DBMS. The proposed approach consists of three steps: *preparation*, *mean resource time estimation* and *mean query response time estimation*. In the first step, the query is transformed into a *query resource profile*. Via an execution plan transaction, it is reduced to a set of low-level resource usage specifications which determines its demands for hardware components. In the second stage, response time of particular resources is estimated. Each hardware resource is a $M/M/1$ or $M/M/G$ node in the queueing network. Synchronization between query execution phases, including pipelined execution or partitioned parallelism, is not taken into consideration during this stage. At the last stage, the mean time of a query is estimated by accommodating hardware resource times for the entire query usage profile. In this stage, intra-operator parallelism, such as pipelined or partitioned execution, determines the way usage time is accumulated.

QuePED model. Osman *et al.* [29] define a database design as a set of tables and transaction types accessing these tables. Authors introduce QuePED – a queueing network performance evaluation model for database designs. Each table in the database is represented as a system of queues in a network. It is noted that partitioned or replicated tables are treated as separate nodes. Each table is characterized by the following qualities: the attribute data types and selectivity, the expected number of rows and row length, as well as index type and structure. Customers arriving to the network correspond to transactions and are categorized according to their service demands in terms of the number of I/O pages required to process the transaction. To obtain this value, a query optimizer is used to return an optimal execution plan comprising SQL statements. Then, for each SQL statement, I/O cost is calculated with the approach described by Ramakrishnan and Gehrke [30]. Thus, the cost depends on the file structure of the table, for example heap file with no index, sorted file, cluster B+ tree file, and the type of SQL operation, such as scan, equality or range search, insert, update or delete. QuePED was validated with a TPC-C benchmark. Since the model assumes the mean time for read/write page requests, the *strace* utility was used to measure the time in which the kernel fulfilled a database page request.

4. Queueing Petri Nets

Queueing Petri Net (QPN) models combine queueing networks and Petri nets formalism and were introduced by

Bause [31]. They extend Petri nets with queueing places which consist of two parts: the queue and the depository. Tokens arriving to the queueing place are firstly placed in a queue according to the scheduling strategy of the queue server. After being serviced, each token is placed in a depository for future transitions. Bause [31] distinguishes two types of queueing places. In *timed queueing*, service time distribution is given, while in *immediate queueing places*, the scheduling strategy is implemented without any delay. The motivation behind QPN is to fuse the modeling power of quantitative and qualitative analysis provided by queueing networks and Petri nets, respectively. Several other advantages were listed by Kounev *et al.* [32], such as ability to model simultaneous resource possession, synchronization, asynchronous processing and software contention. These models may capture hardware- and software-related aspects of system behavior. Moreover, the graphical representation of QPN is intuitive.

4.1. Notation and Definitions

Following Kounev *et al.* [32], we will now provide a formal definition of QPN.

Definition 1: A queueing Petri net is an 8-tuple $QPN = (P, T, C, I^+, I^-, M_0, Q, W)$ where:

1. $P = p_1, \ldots, p_n$ is a finite and non-empty set of places;

2. $T = t_1, \ldots, t_m$ is a finite and non-empty set of transitions, $P \cap T = \emptyset$;

3. C is a color function that assigns a finite and non-empty set of colors to each place and a finite and non-empty set of modes to each transitions;

4. I^+ and I^- and forward and backward incidence functions defined on $P \times T$, such that $I^-(p,t), I^+(p,t) \in [C(t) \rightarrow C(p)_{MS}], \forall (p,t) \in P \times T$, where $C(p)_{MS}$ denotes the set of all finite multisets of $C(p)$;

5. M_0 is a function on P describing *initial marking* such that $M_0(p) \in C(p)_{MS}$;

6. $Q = (Q_1, Q_2, (q_1, \ldots, q_{|P|}))$ where:

   - $Q_1 \subseteq P$ is the set of timed queueing places,
   - $Q_2 \subseteq P$ is the set of immediate queueing places,
   - $Q_1 \cap Q_2 = \emptyset$,
   - q_i denotes the description of a queue taking all colors of $C(p_i)$ into consideration if p_i is a queueing place or equals keyword 'null', if p_i is an ordinary place;

7. $W = (W_1, W_2, (w_1, \ldots, w_{|T|}))$ where:

   - $W_1 \subseteq T$ is the set of timed transitions,
   - $W_2 \subseteq T$ is the set of immediate transitions,
   - $T = W_1 \cup W_2, W_1 \cap W_2 = \emptyset$ and

- $w_i \in [C(t_i) \to \mathbb{R}^+]$, $\forall t_i \in T$, $c \in C(t_i)$ $w_i(c) \in R^+$ is interpreted as a rate of a negative exponential distribution specifying the firing delay due to color c, if $t_i \in W_1$ or a firing weight specifying the relative firing frequency due to color c, if $t_i \in W_2$.

A more detailed information about QPN may be found in [31].

4.2. Cassandra Replication Modeling

Osman and Piazzola [33] adapt the queueing Petri nets formalism to model asynchronous replication in a Cassandra data store. Cassandra is a column-based, scalable data store which offers following multi-master replication and data distribution [34]. The *key-value* space is mapped onto a ring. The ring is then split into ranges and each of the cluster members is assigned to one or more key subsets. Database clients can contact any of the cluster nodes, which then becomes a coordinator node. Each of the cluster nodes has the knowledge about the mapping between other nodes and data ranges. Cassandra allows to configure a *replication factor* (RF) variable which determines number of nodes that store replicated data. The replication process is asynchronous and thus offers lower response times. It may result, however, in the lack of data consistency. The trade-off between those two behaviors is controlled by the user-defined *consistency level* (CL), denoting the number of nodes in which a given operation must succeed before the coordinator node responds to the client.

Osman and Piazzola [33] point out the significant role of QPNs in capturing crucial aspects of asynchronous replication: scheduling, synchronization and blocking request in a particular node. The model was validated for read operation workload.

In the QPN (see Fig. 8), the database client is represented with a queueing place with exponential think time. The client sends *read* tokens and *enter-cluster* immediate transition fires. This results in random distribution of the incoming requests. Let i be the index of the node chosen to be the coordinator. A read token is placed in the *enter-node_i* – a timeless queueing place with the FIFO scheduling strategy. The number of parallel requests which may be processed by the node is controlled with thread tokens. The model was initialized with 32 tokens placed in *threads_i* ordinary place. When at least one thread token is in the threads_i place, at cluster node i is modeled by the timed queueing place *node_i* with exponentially distributed service time, single server and processor sharing discipline. The mechanism caused by process-request_i differs depending on whether node i stores data for the requested key or not.

Local and Non-Local Requests Processing. Since each of the nodes stores only a subset of data, coordinator node_i might not hold data for the requested key. The QPN model described differentiates processing local and non-local requests. For local operations, an immediate tran-

sition process-request_i relocates the read token from the enter-node_i place to the node_i timed queueing place. It is noted that the remote request does not affect the nodes' ability to process local operations, as it does not use thread tokens. Coordinator node i routes the request and blocks it until it receives responses from CL other nodes. This is done by firing the immediate transition distribute-request_i. As a result, one read token is placed in *blocked_i* ordinary place and each of CL contacted servers receives *read_i*. Let j be the index of one of the contacted members, the token of color read_i is put in its enter-node_j. The replica nodes contacted are chosen according to the consistency level and replication factor.

When read_i token leaves the *node_j* place, the *distribute-request_j* immediate transition returns the thread token to *threads_j* place and deposits response token in the *replica-response_i* place of *node_i*. When the number of tokens in the replica-response_i place equals CL, the immediate transition unblock_i fires. One read token is passed from the blocked_i place to the the exit-cluster place and CL tokens are removed from the replica-response_i place. The read request is then transferred back to the client.

In contrast to remote requests, local requests address keys that are stored by node_i. When the immediate transition process-request_i fires, it removes both the read token from the enter-node_i place and the thread token from the threads_i place. Additionally, it deposits a *local-read* token in the timed queueing place node_i. In order to synchronize replies, a read token is placed in blocked_i. After node i processes the local request, it fires the distribute-request_i transition. Thread token is returned and one token is placed in replica-response_i. If CL > 1, the distribute-request_i transition will deposit a read_i token in the *enter-node* places of CL − 1 other nodes and the request is processed as a remote request. When the number of tokens in the replica-response_i place reaches a consistency level, unblock_i transition fires. One read token is replaced from the blocked_i place to the exit-cluster place and CL tokens are removed from the replica-response_i place. The read request is then transferred back to the client.

4.3. MongoDB Replication Modeling

The model described can be adapted to evaluate the performance of other NoSQL data stores which have implemented asynchronous replication. Although MongoDB [35] is a document-oriented database and uses different high availability and scalability mechanisms, the request response time may be obtained with an analogical QPN. To avoid a single point of failure and to ensure the balancing of load among data clusters, MongoDB has developed replica sets [35]. Replica set is a group of mongod processes which hold the same set of data. In the group, one of the nodes acts as a master node and receives all write operations. Such a node is called the primary replica set member. Secondary replica set members are slave nodes to which all write operations are asynchronously replicated from the primary member. Read operations may be routed
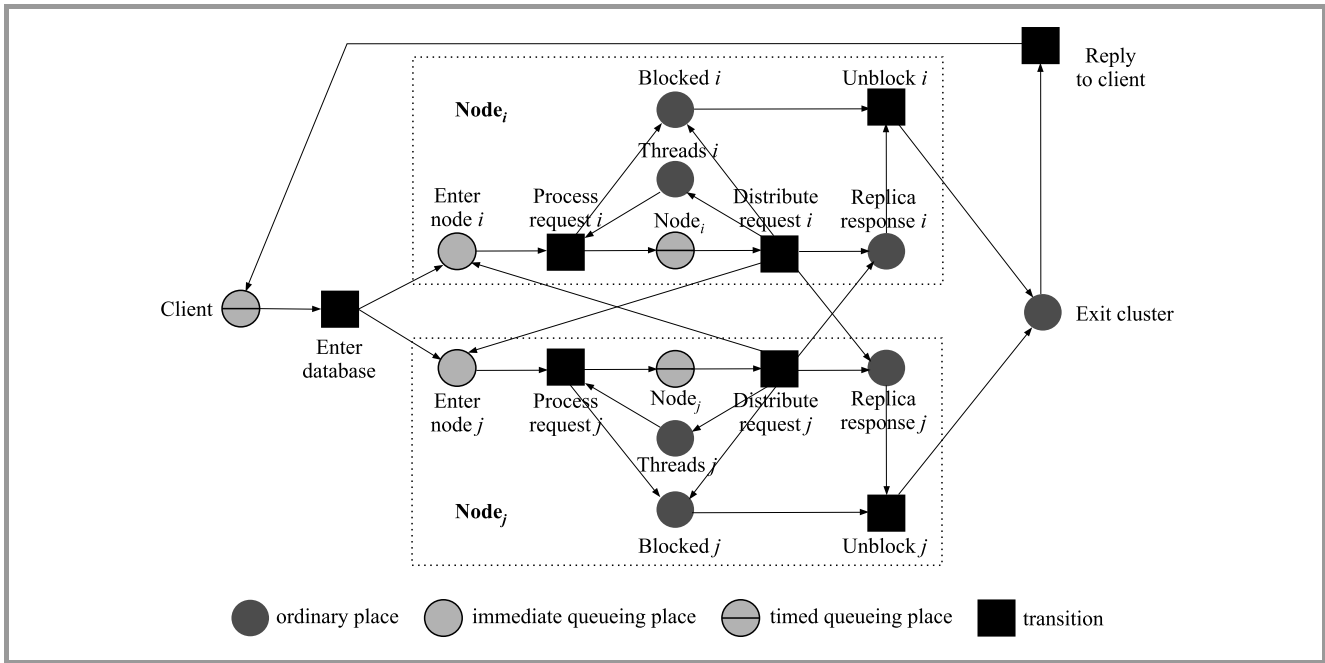
Fig. 8. QPN model of replication in a Cassandra cluster.

both to primary or secondary nodes, depending on the user-defined read preference variable. In the case of failure of the primary node, one of the running servers is chosen to be the new primary node. To allow horizontal scalability, MongoDB introduces *sharding* [35]. Sharding is automated data partitioning at the collection level which is based on the arbitrary shard key values. A shard key has to be an immutable field which exists in every document in the sharded collection. A shard is a subset of the collection data and is advised to be deployed as a replica set. Another component of a sharded cluster is the config database, storing meta data of the cluster. In particular, it contains mapping between data subsets and shards. A client sends a query to the router process *mongos*, which contacts the config server to retrieve information necessary for correct routing of the query to the shards. Depending on the requested shard key values, queries might be performed in one shard or distributed to several shards.

Applying the Osman and Piazzola [33] model for a replica set of a non-sharded database is straightforward. The number of the nodes holding the data in the replica set is equal to the replication factor RF. Similarly to the Cassandra data store, MongoDB allows the user to configure consistency level CL by setting read and write concerns. Depending on the read preference, the primary node or one of the secondary nodes receives the read request. Then, the read operation is processed similarly to a local request in the coordinator node, as described in the previous model.

The QPN formalism described sheds a light on the way future sharded cluster models may be constructed. Once a request arrives to the mongos process, it might be evaluated in a similar manner as remote requests processed by the coordinator node. However, communication between

the router and the config database should be incorporated. Based on the knowledge from the config server, the router distributes the request to one or more shards. The number of target shards plays a role that is analogical to the consistency level of the previous model.

The above consideration may lead to the construction of a QPN model of replication and data distribution in the MongoDB cluster. However, such a work requires additional research. The complexity of QPN for a cluster composed of shards deployed as replica sets should be examined.

4.4. Database Contention Performance Modeling

Coulden *et al*. [36] construct a QPN model of a table-level database concurrency control through a strict two-phase locking protocol (Strict 2PL). Write operations hold exclusive locks on data objects, while read requests hold
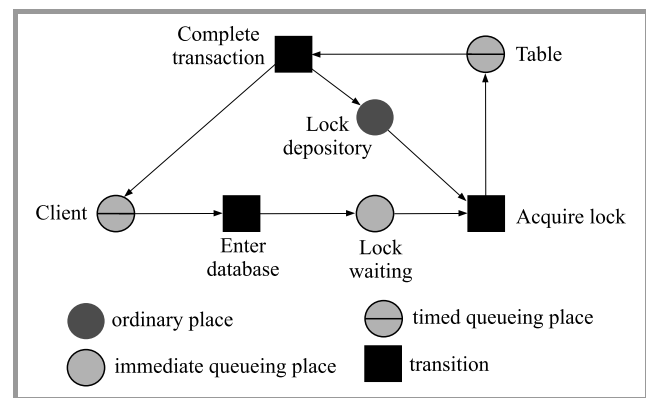

Fig. 9. QPN model of table level locking.

shared locks. A shared lock on a given data object may be acquired only when the object is not covered with an exclusive lock. An exclusive lock may be acquired on an object only when it is not covered with any other lock. This phenomenon can be captured with an ordinary repository place containing lock tokens. A shared transaction uses one lock token, while exclusive request requires the maximum number of tokens in the system. The database client is represented as a timed queueing place. The table accessed is represented as a timed $G/M/\infty$ queueing place with an infinite server queue. Figure 9 presents QPN of a table-level contention. The model was validated with the use of the *PostgreSQL* database management system [36].

5. Conclusion

In this paper a review of existing database performance models developed within the queueing theory approach was presented. The vast majority of conducted studies investigate the performance of relational databases. Only one of the models presented deals with the replication mechanism in the NoSQL Cassandra data store. The above-mentioned approach can be used to model performance of other replicated and distributed non-relational data stores, such as MongoDB. We claim that future studies should be pursued to rearrange the presented methodologies in order to evaluate the performance of NoSQL data stores. Moreover, while extensive research focuses on particular database components or constructs, the number of methodologies capable of mapping database system specifications onto queueing network models continues to be low. We believe that such studies should be pursued to develop methodologies for industrial applications.

Acknowledgements

References

[1] M. Nicola and M. Jarke, "Performance modeling of distributed and replicated databases", *IEEE Trans. on Knowl. and Data Engin.*, vol. 12, pp. 645–672, no. 4, 2000 (doi: 10.1109/69.868912).

[2] R. Osman and W. J. Knottenbelt, "Database system performance evaluation models: A survey", *Perform. Eval.*, vol. 69, pp. 471–493, no. 10, 2012 (doi: 10.1016/j.peva.2012.05.006).

[3] L. Kleinrock, *Queueing Systems. Volume 1: Theory*. New York: Wiley-Interscience, 1975 (ISBN: 9780471491101).

[4] E. G. Coffmann, E. Gelenbe, and B. Plateau, "Optimization of the number of copies in a distributed data base", *IEEE Trans. on Software Engin.*, vol. SE-7, no. 1, pp. 78–84, 1981 (doi: 10.1109/TSE.1981.234510).

[5] F. Bacelli and E. Coffmann, "A database replication analysis using an M/M/m queue with service interruptions", in *Proc. of the 1982 ACM SIGMETRICS Conf. on Measur. and Model. of Comp. Syst. SIGMETRICS'82*, Seattle, VA, USA, 1982, pp. 102–107 (doi: 10.1145/1035332.1035309).

[6] R. Nelson and R. Iyer, "Analysis of a replicated data base", *Perform. Eval.*, vol. 5, no. 3, pp. 133–148, 1985 (doi: 10.1016/0166-5316(85)90008-2).

[7] E. Arzuaga and D. S. Kaeli, "An M/G/1 queue model for multiple applications on storage area networks", in *Proc. for the 11th Worksh. on Comp. Architec. Eval. using Commercial Workloads CAECW-11*, Salt Lake City, UT, USA, 2008, pp. 25–32.

[8] M. Dellkrantz, M. Kihl, and A. Robertsson, "Performance modeling and analysis of a database server with write-heavy workload", in *Proc. 1st Eur. Conf. on Serv.-Orient. and Cloud Comput. ESOCC 2012*, Bertinoro, Italy, 2012, pp. 184–191, 2012 (doi: 10.1007/978-3-642-33427-6_13).

[9] M. Kihl, P. Amani, A. Robertsson, G. Radu, M. Dellkrantz, and B. Aspernas, "Performance modeling of database servers in a telecommunication service management system", in *Proc. 7th Int. Conf. on Digit. Telecommun. ICDT 2012*, Chamonix, France, 2012, pp. 123–129.

[10] J. Mc Dermott and R. Mukkamala, "Performance analysis of transaction management algorithms for the SINTRA replicated architecture database systems", in *Proc. of the IFIP WG11.3 Working Conf. on Datab. Secur. VII*, Lake Guntersville, Alabama, USA pp. 215–234, 1993.

[11] B.-C. Jenq, B. C. Twichell, and T. W. Keller, "Locking performance in a shared nothing parallel database machine", *IEEE Trans. on Knowl. and Data Engin.*, vol. 1, no. 4, pp. 530–543, 1989 (doi: 10.1109/69.43427).

[12] H. Garcia-Molina, "Performance of the update algorithms for replicated data in a distributed database", Ph.D. Thesis, Stanford University, Stanford, CA, USA, 1979 [Online]. Available: http://www.dtic.mil/dtic/tr/fulltext/u2/a075268.pdf

[13] H. Garcia-Molina and G. Wiederhold, "Read-only transactions in a distributed database", *ACM Trans. on Datab. Syst.*, vol. 7, no. 2, pp. 209–234, 1982 (doi: 10.1145/319702.319704).

[14] R. D. van der Mei, A. R. de Wilde, and S. Bhulai, "A method for approximating the variance of the sojourn times in star-shaped queueing networks", *Stochastic Models*, vol. 24, no. 3, pp. 487–501, 2008 (doi: 10.1080/15326340802232327).

[15] D. Liang and S. Tripathi, "Performance analysis of long-lived transaction processing systems with rollbacks and aborts," *IEEE Trans. on Knowl. and Data Engin.*, vol. 8, no. 5, pp. 802–815, 1996 (doi: 10.1109/69.542031).

[16] M. J. Carey and M. Liviny, "Distributed concurrency control performance: A study of algorithms, distribution, and replication", in *Proc. of the 14th Int. Conf. on Very Large Datab. VLDB'88*, Los Angeles, CA, USA, 1988, pp. 13–25.

[17] B. Ciciani, D. M. Dias, and P. S. Yu, "Analysis of replication in distributed database systems", *IEEE Trans. on Knowl. and Data Engin.*, vol. 2, no. 2, pp. 247–261, 1990 (doi: 10.1109/69.54723).

[18] B. Ciciani, D. Dias, M., and P. S. Yu, "Analysis of concurrency-coherency control protocols for distributed transaction processing systems with regional locality", *IEEE Trans. on Knowl. and Data Engin.*, vol. 18, no. 10, pp. 899–914, 1992 (doi: 10.1109/32.163606).

[19] S. Banerjee, V. O. Li, and C. Wang, "Performance analysis of the send-on-demand: A distributed database concurrency control protocol for high-speed networks", *Comp. Commun.*, vol. 17, no. 3, pp. 189–204, 1994 (doi: 10.1016/0140-3664(94)90005-1).

[20] S. Y. Hwang, K. Lee, and Y. H. Chin, "Data replication in a distributed system: A performance study", in *Proc. 7th Int. Conf. Database and Expert Systems Applications*, Zurich, Switzerland, 1996, pp. 708–717 (doi: 10.1007/BFb0034724).

[21] K. K. Leung, "An update algorithm for replicated signaling databases in wireless and advanced intelligent networks", *IEEE Trans. on Comp. – Special issue on mobile computing archive*, vol. 46, no. 3, pp. 362–367, 1997 (doi: 10.1109/12.580431).

[22] E. Born, "Analytical performance modelling of lock management in distributed systems", *Distributed Systems Engineering*, vol. 3, no. 1, pp. 68–76, 1996 (doi: 10.1088/0967-1846/3/1/008).

[23] D. A. Menascé and M. N. Bennani, "Analytic performance models for single class and multiple class multithreaded software servers", in *Proc. 32nd Int. Computer Measur. Group Conf.*, Reno, NV, USA, 2006, pp. 475–482.

[24] B. M. M. Gijsen, R. D. van der Mei, P. Engelberts, J. L. van den Berg, and K. M. C. van Wingerden, "Sojourn time approximations in queueing networks with feedback", *Perform. Eval.*, vol. 63, no. 8, pp. 743–758, 2006 (doi: 10.1016/j.peva.2005.08.002).

[25] A. Thomasian and K. Ryu, "A decomposition solution to the queueing network model of the centralized DBMS with static locking", in *Proc. of the Int. Conf. on Measur. and Model. of Comp. Syst. SIGMETRICS 1983*, Minneapolis, MN, USA, 1983, pp. 82–92, 1983 (doi: 10.1145/800040.801397).

[26] R. J. T. Morris and W. S. Wong, "Performance analysis of locking and optimistic concurrency control algorithms", *Perform. Eval.*, vol. 5, no. 2, pp. 105–118, 1985 (doi: 10.1016/0166-5316(85)90043-4).

[27] P. S. Yu, S. Balsamo, and Y. Lee, "Dynamic transaction routing in distributed database systems", *IEEE Trans. on Software Engin.*, vol. 14, no. 9, pp. 1307–1318, 1988 (doi: 10.1109/32.6174).

[28] N. Tomov *et al.*, "Analytical response time estimation in parallel relation database systems", *Parallel Computing*, vol. 30, no. 2, pp. 249–283, 2004 (doi: 10.1016/j.parco.2003.11.003).

[29] R. Osman, I. Awan, and M. E. Woodward, "QuePED: Revisiting queueing networks for the performance evaluation of database designs", *Simulation Modelling Practice and Theory*, vol. 19, no. 1, pp. 251–270, 2011 (doi: 10.1016/j.simpat.2010.06.010).

[30] R. Ramakrishnan and J. Gehrke, *Database Management Systems*. Boston: McGraw-Hill, 2002 (ISBN: 9780072465631).

[31] F. Bause, "Queueing Petri nets – a formalism for the combined qualitative and quantitative analysis of systems", in *Proc. of 5th Int. Worksh on Petri Nets and Perf. Models*, Toulouse, France, 1993 (doi: 10.1109/PNPM.1993.393439).

[32] S. Kounev, S. Spinner, and P. Meier, "Introduction to queueing Petri nets: Modeling formalism, tool support and case studies", in *Proc. of the 3rd ACM/SPEC Int. Conf. on Perform. Engin. ICPE 2012*, Boston, MA, USA, 2012, pp. 9–18 (doi: 10.1145/2188286.2188290).

[33] R. Osman and P. Piazzolla, "Modeling replication in noSQL datastore", in *Proc. 11th Int. Conf. Quantitat. Eval. of Syst. QEST 2014*, Florence, Italy, 2014, pp. 194–209 (doi: 10.1007/978-3-319-10696-0_16).

[34] R. Cattell, "Scalable SQL and noSQL data stores", *ACM SIGMOD Record*, vol. 39, no. 4, pp. 12–27, 2010 (doi: 10.1145/1978915.1978919).

[35] K. Chodorov, *MongoDB: The Definitive Guide*, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, Inc., 2013 (ISBN 9781449344795).

[36] D. Coulden, R. Osman, and W. J. Knottenbelt, "Performance modelling of database contention using queueing Petri nets", in *Proc. of the 4th ACM/SPEC Int. Conf. on Perform. Engin. ICPE 2018*, Prague, Czech Republic, 2013 (doi: 10.1145/2479871.2479919).
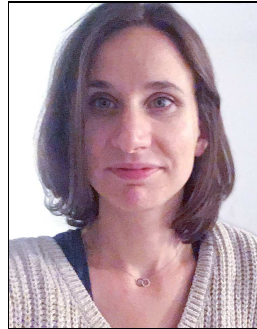
Antonina Krajewska is a second-year Ph.D. student at the Systems Research Institute of the Polish Academy of Sciences and a fellow working on a research project titled "An energy-aware computer system for HPC computing" at the Faculty of Electronics and Information Technology of the Warsaw University of Technology. She has been with the Research and Academic Computer Network (NASK) since 2016. She holds an M.Sc. in Mathematics and an M.Sc. in Psychology from Warsaw University. Her research areas focus on modeling of performance of database systems, Markov chains and NoSQL data stores.

iD https://orcid.org/0000-0001-6626-5667
E-mail: antonina.krajewska@nask.pl
Research and Academic Computer Network (NASK)
Kolska 12
01-045 Warsaw, Poland

Optimization of Direct Direction Finding Method with Two-Dimensional Correlative Processing of Spatial Signal

Vitaliy V. Tsyporenko and Valentyn G. Tsyporenko

Zhytomyr State Technological University, Zhytomyr, Ukraine

Abstract—In this article, the main parameter of the correlative-interferometric direction finding method with two-dimensional correlative processing of spatial signal in the aperture of a linear antenna array (AA) is determined as the value of spatial shift within the AA aperture. The corresponding objective function is also formed. Analytical optimization of this parameter is presented and a comparative analysis of analytical calculations based on simulation results is conducted. In the simulation, a range of dependencies of the middle square deviation of estimation of direction on the value of the spatial shift for a signal-to-noise ratio of 0 dB, for minimum 3-sample and 4-sample Blackman-Harris windows of the spectral analysis, is received. The value of the middle square deviation of estimation of direction will be minimal and will equal 0.02 degrees using a minimum 3-sample Blackman-Harris window with the −67 dB level of side lobes. It offers high noise immunity and high accuracy of direction finding.

Keywords—*antenna array, correlative interferometric, direct method, radio direction finding.*

1. Introduction

Modern radio electronic telecommunications and radio monitoring systems must function in a difficult and rapidly changing electromagnetic situation (EMS). The said situation is characterized by multipath radio propagation, frequency coverage of the desired signal, as well as by intrusive noises and an a priori uncertainty. It is to because of a large number of mobile network devices operating in dense urban areas. With these conditions taken into consideration, the use of an adaptive digital antenna array (AA) is a promising area of improvement for telecommunications and radio monitoring systems. Its application allows to increase functional efficiency of telecommunications and radio monitoring systems by significantly optimizing AA parameters and operating modes. Optimization of the configuration of AA, the number and location of elements within its aperture, their number, methods for processing the weight of AA element signals, the form and the relationship between parameters of the main and the side lobes of the AA radiation pattern, are widely used methods. Dif-

ferent criteria of optimization are presented in this paper. Also, at the expense of preliminary estimation of directions of interference sources, zeros in the radiation pattern (RP) of a digital adaptive AA can be formed. Optimization of the means of estimation of EMS allows to reduce the cost of telecommunications and radio monitoring equipment and significantly increases the speed of such systems.

The adaptation speed of a digital AA and its interference immunity depend significantly on the time-intensity and computational costs of the current EMS estimation methods, on the correlation interference matrix and on identification directions towards its source. Therefore, the task of increasing the speed of EMS analysis and estimating directions towards the sources of noise in a digital AA is relevant. The promising direction for the implementation of direction finding (DF) for these conditions is the use of digital wide-band correlative-interferometric direction finders coupled with AA, and the digital synthesis of its radiation pattern [1], [2].

Correlation-interferometric DF techniques offer a wide frequency range, resistance to interference caused by multipath reception, high sensitivity and good accuracy. However, the most plausible and unbiased assessment of estimation of directions towards radio emission sources (RES) is based on the search of a sequential correlational analysis and on the review of space. This greatly limits performance and requires a lot of hardware-related expenditure for data processing systems, therefore reducing the effectiveness of their application in dynamic EMS conditions [1]. Another disadvantage of this method is the low DF precision of RES, with the spectra being completely overlapped by frequency.

Effectiveness of DF depends on such parameters as speed, accuracy, noise immunity and hardware quality. Therefore, research on and optimization of digital direct correlative-interferometric direction finders, as well as their adaptation to specific EMS conditions, are an important scientific challenge.

In [3], the direct method of DF with two-dimensional correlative processing of a spatial signal reconstructed in the aperture of a linear antenna array which provides a combination of high speed and precision direction finding, was

presented. However, this method has not been researched and optimized under real life conditions yet.

In [4], [5] correlative-interferometric methods of estimating directions towards RES, using AA, which are effectively implemented in a digital domain, were studied. However, they rely on the search compensatory method of DF, which is slow in wideband DF of RES.

In [6], [7] the results of optimizing radio signal processing algorithms and basic parameters of digital correlation direction finders that rely on different AA configurations were obtained. The high level of efficiency in monitoring noise-like radio signals in a complex EMS was shown.

In [8]–[10] nonlinear spectral methods of DF that provide a high level of spatial resolution of the received signal were studied. However, they have a number of significant drawbacks for radio monitoring systems, such as:

- high computational complexity, being several times higher than that of the searching correlative DF method,

- requirement of a priori information on the exact amount of emission received in the mix,

- biased direction estimates which reduce DF accuracy,

- loss of operational stability at low signal-to-noise levels (about 10 dB).

As a result, the efficiency of spectral DF methods which is determined, in the first place, by the speed-to-accuracy ratio, is insufficient in radio monitoring systems. This is confirmed by the absence of their application in modern digital DF of radio monitoring systems [1], [2].

In [11]–[14], the effectiveness and optimization of primary methods used for estimating direction towards RES, using AA, is researched. The optimal algorithms used for estimation of spatial radio emission parameters, as well as the optimal relationships between radiation pattern (RP) parameters ensuring effective spatial selection were determined. However, in these articles, optimization of the direct methods of correlation-interferometric DF, including the use of two-dimensional reconstructed signal processing, have not been investigated. Therefore, the results obtained there cannot be directly used to solve the research problem in question.

In light of the above, optimization, development and analysis of digital methods of correlation-interferometric DF to be applied in automated radio monitoring systems continue to be task with a high degree of urgency.

The aim of the article is to research and optimize the direct digital method of correlative-interferometric radio DF with two-dimensional correlative processing of spatial signal.

2. Analytical Approach

First, optimization of the direct digital method of correlative-interferometric DF with two-dimensional correlative processing of spatial signal reconstructed in the linear

AA aperture [3] to provide a minimum DF variance error is performed. Let a horizontal linear equidistant AA, consisting of Z identical DF radio channels, receive a useful random Gaussian stationary signal $S(t)$ from an unknown direction θ. It has a uniform energy spectrum $S^2(\omega)$, the width of which does not exceed the width Δf_k of the band of DF radio channels. Radio channels of the AA have their own additive stationary normal noise $n_z(t)$, with a zero mathematical expectation and an equal two-sided spectral density N of power, constant within the passband $[\omega_{S,L}; \omega_{S,,H}]$ of the radio direction finder. We assume that the own noises of radio channels of the AA are not characterized by inter-channel correlation and correlation with the signal. Also, we assume that the phase fluctuations at the signal paths are absent. Thus, the initial research conditions are:

$$U_z(t) = S_z(t - \tau_z) + n_z(t), \qquad (1)$$

where:

- $U_z(t)$ is the mixture taken by the z-th DF channel,

- $S_z(t - \tau_z)$ is the useful signal received by the z-th DF channel,

- τ_z is the delay of reception of the desired signal by the z-th DF channel, relative to the reference channel, depending on the direction towards RES,

- $n_z(t)$ is the own additive Gaussian noise within the band of the simultaneous analysis of the z-th DF channel.

DF should be performed within time T_a of the analysis process, and within one correlation processing cycle, to ensure that the conditions correspond to the DF method analyzed, estimation $\hat{\theta}$ of direction towards RES is [3]:

$$\hat{\theta} = \arccos \frac{\hat{\Omega}}{\omega_{S,L}}, \qquad (2)$$

where $\hat{\Omega} = \dfrac{1}{\Delta z} \arctan \dfrac{\displaystyle\sum_{k=k_l}^{k_H} U_{vk}(\Omega_p) \sin[\Delta \hat{\Psi}_{Ak}(\Delta z, \ \Omega_p)]}{\displaystyle\sum_{k=k_l}^{k_H} U_{vk}(\Omega_p) \cos[\Delta \hat{\Psi}_{Ak}(\Delta z, \ \Omega_p)]}$

is the estimation of the spatial frequency of the signal group for received signal $S(t)$ using dispersion-correlation processing:

- c is the propagation speed of electromagnetic emission in space,

- Δz is the spatial shift within the AA aperture.

The $U_{vk}(\Omega_p)$ is defined, in above equation, by:

$$U_{vk}(\Omega_p) = \left\{ \left[\sum_{p=p_L}^{p_H} U_k^2(\Omega_p) \sin(\Omega_p \Delta z) \right]^2 \right.$$
$$\left. + \left[\sum_{p=p_L}^{p_H} U_k^2(\Omega_p) \cos(\Omega_p \Delta z) \right]^2 \right\}^{\frac{1}{2}}.$$

$U_{vk}(\Omega_p)$ is the module of the complex mutual space (obtained by implementing spatial processing) spectrum of the k-th frequency component of emission $S(t)$ received by the linear AA with multi-lobe RP. The other designations are:

- $\Omega_p = \frac{2\pi p}{dZ}$ is the value of spatial frequency, which determines the direction of the p-th lobe of the multilobe RP, $p = 0, 1, \ldots, Z-1$,

- p_L, p_H are the numbers of lower and upper spatial frequencies of the selected signal group,

- d is the distance between AA elements,

- $\Delta\hat{\Psi}_{Ak}(\Delta z, \Omega_p) = \arctan \dfrac{\displaystyle\sum_{p=p_L}^{p_H} U_k^2(\Omega_p)\sin(\Omega_p\Delta z)}{\displaystyle\sum_{p=p_L}^{p_H} U_k^2(\Omega_p)\cos(\Omega_p\Delta z)}$

  is the estimation of the phase shift spatial of analytic signal in the distance,

- $\omega_{S.L}$, $\omega_{S.k}$ are the lower and the k-th frequencies of spectrum $U_z(j\omega_{S.k})$ of the received mixture $U_z(t)$, respectively,

- k_L, k_H are the numbers of lower and upper frequencies of the spectrum $U_z(j\omega_{S.k})$ of the received mixture $U_z(t)$, respectively.

Analysis of Eq. (2) shows that the estimation $\hat{\theta}$ of the direction towards RES is determined on the basis of two-dimensional space-time correlation processing of a complex mutual spatial spectrum $U_{vk}(\Omega_p)$ of the emission $S(t)$ received from the linear AA, with multi-lobe RP within the time T_a of the analysis. Counts of the mutual spatial spectrum $U_{vk}(\Omega_p)$ are determined by spatial correlation processing of the selected signal group. Thus, the k-th signal group is a complex array $\{U_k(j\Omega_p)\}_{p \in [p_L; p_H]}$ of spatial spectrum components from the AA, with the multi-lobe RP on action of the k-th time spectral component with frequency $\omega_{S.k}$ for emission, received by overlapping main lobes of adjacent partial RP.

The use of parallel frequency and spatial analysis with the following dispersion-correlation estimation of direction allows to obtain the direct estimation $\hat{\theta}$ of direction towards RES.

The variant of the direction finder structure, implementing the analyzed method is shown in Fig. 1.

The direction finder shown in Fig. 1 is characterized by the following features. The mixture of $U_z(t)$ of radio emission is received by the Z-channel linear AA. The first channel of the AA is constantly connected to the first input of a two-channel radioreceiver with a common heterodyne. The AA channels from 2 to Z are connected to the second input of the radio receiver, using an antenna switch, at the time T_a of the analysis. From the two intermediate frequency outputs of the radio receiver, the pairs of signals are formed by a two-channel analog-to-digital converter (ADC) and are converted into a digital form with a sampling frequency

$F_d = 2\Delta f_k$. As a result, $Z-1$ pairs of arrays on $N_s = T_a F_d$ samples of the received mixture $U_z(t)$ of radio emission are accumulated and transmitted to the computing unit. Then, using the fast Fourier transform (FFT) algorithm, $Z-1$ pairs of signal spectra are obtained. From the received pairs of spectra the reciprocal spectra are formed by means of a complex conjugation of the spectrum of the first channel AA and its multiplication with signals from 2 to Z channel AA that are equivalent to the spectra of signals obtained by using the Z channel radio receiver.
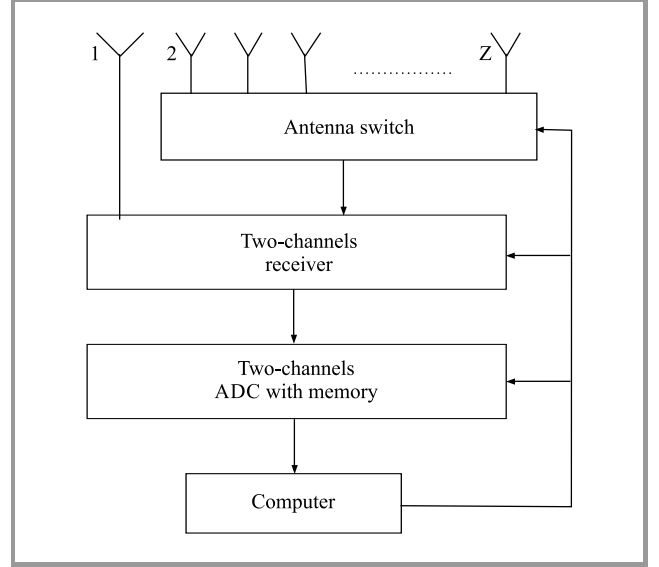
Fig. 1. Structure scheme of correlative-interferometric DF with two-dimensional correlative processing of spatial signal.

According to the proposed DF method [3], subsequent processing of the received mutual spectra is performed on a computer, based on Eq. (2). The use of a two-channel radio receiver, compared to the -channel minimizes hardware costs, but leads to an increase in the total time of radio emission analysis by $Z-1$ times.

The operation of the antenna switch, of the radio receiver and of the ADC is controlled by a computer.

The main indicator of the accuracy of DF is dispersion σ_θ^2 of the direction estimation error. It is determined, in the DF method analyzed, according to [2], [11]:

$$\sigma_\theta^2 = \frac{2K_{W\theta}c^2 m_s}{\omega^2 T_a \mu_I \Delta f_k Z K_\tau^2(\Delta z)}$$
$$\times \frac{1}{(m_s - 1 + K_{W\theta})(\Delta z)^2 d^2 \sin^2\theta}, \quad (3)$$

where:

- $K_{W\theta}$ is the noise band coefficient of the weight function $W_\theta(z)$ of the digital spatial spectral analysis,

- m_s is the number of spectral-spatial counts of the signal group,

- μ_I is the signal to noise ratio at the input of DF channels of the AA,

- $K_\tau(\Delta z)$ is the value of the normalized autocorrelation function for the weight function $W_\theta(z)$ of digital RP forming for the shift given by:

$$K_\tau(\Delta z) = \frac{\sum\limits_{z=0}^{Z-1} W_\theta(z) W_\theta(z - \Delta z)}{\sum\limits_{z=0}^{Z-1} W_\theta^2(z)},$$

- $\omega_0 \in [\omega_{S.L}; \omega_{S.H}]$ is the average frequency of passband of radio channels or the carrier frequency of the emission.

In view of Eq. (3) the general optimization conditions are defined as follows:

$$\sigma_\theta^2 \to \min. \qquad (4)$$

While optimizing the researched DF method we have to define the type of the objective function and communication functions [15]. In order to do this based on Eqs. (2) and (3), we perform an analysis of the implementation of the optimized DF method. The analysis of Eq. (3) has shown that the dispersion σ_θ^2 of direction estimation depends on the procedure of reconstructing the complex analytical signal within the AA aperture, and on correlation processing. In addition, the direction finder parameters, such as the number Z of DF radio channels, the width Δf_k of the passband of the radio channels and the sensitivity μ_{BX}, have significant limitations in the optimization process, with the capabilities of modern technology and implementation of direction finder's compactness and price-related requirements taken into consideration.

However, such parameters as $K_{W\theta}$ and m_S are determined by the requirements for noise immunity of DF in a difficult EMS and by the provision of effective spatial selection of station interference and multipath interference. Therefore, their variation is also severely limited in order to optimize the accuracy of DF.

Therefore, the emission parameters of DF RES, such as the average or the carrier frequency ω_0 of the time spectrum (obtained from the processing time) and the direction θ of arrival of radio waves are not affected in the algorithm of DF and have only a global limit concerning the range operating frequency range and the width of DF sector, respectively. The time T_a during which radio emissions which are received at the same time within a passband Δf_k of radio channels are analyzed, significantly affects interference immunity and the speed of DF's operation. However, analysis of time T_a of the radio emission is not affected in the case of the general algorithm implementation and also in the reconstruction of a complex analytic signal.

Therefore, analysis of Eq. (3) shows that a significant impact on the dispersion σ_θ^2 of direction estimation in constant devices and time expense is exerted by the spatial shift value Δz, as well as by the correlation coefficient $K_{W\theta}(\Delta z)$ of the weight function, which depends on the spatial shift Δz. Therefore, it is advisable to determine

the objective function $F(\Delta z)$ for further optimization as a functional of variables Δz and $K_{W\theta}(\Delta z)$:

$$F(\Delta z) = f[\Delta z, K_{W\theta}(\Delta z)] = \Delta z^2 K_{W\theta}^2(\Delta z). \qquad (5)$$

We perform a qualitative assessment of the features of the objective function $F(\Delta z)$. Analysis of Eq. (3) indicates that the spatial shift Δz and the correlation coefficient $K_{W\theta}(\Delta z)$ of the weight function $W_\theta(z)$ are antagonistic parameters. The increase of the shift Δz defines the essential decrease in the dispersion σ_θ^2 of estimation of direction for back-quadratic dependence. In turn, the correlation coefficient $K_{W\theta}(\Delta z)$ decreases by increasing the value of the spatial shift Δz. The second variable of the objective function $K_{W\theta}(\Delta z)$ is the coefficient of correlation $K_{W\theta}(\Delta z)$, that is defined as [3]:

$$K_\tau(\Delta z) = \frac{\sum\limits_{z=0}^{Z-1} W_\theta(z) W_\theta(z - \Delta z)}{\sum\limits_{z=0}^{Z-1} W_\theta^2(z)}. \qquad (6)$$

Analysis of Eq. (6) shows that the maximum value of the correlation coefficient $K_{W\theta}(\Delta z)$ is achieved at zero spatial shift $\Delta z = 0$: $\max[K_{W\theta}(\Delta z)] = 1|_{\Delta z=0}$. By increasing the spatial shift Δz, the value of the correlation coefficient $K_{W\theta}(\Delta z)$ is reduced, which, in turn, determines an increase in the dispersion σ_θ^2 of estimation of direction. Therefore, the task consists in finding the global extremum of the objective function $F(\Delta z)$. Taking into account the dependence between the value of the correlation coefficient $K_{W\theta}(\Delta z)$ and the spatial shift Δz, it is advisable to perform subsequent optimization, based on the scalar method, with an analysis of the relevant set of communication functions $f_l(X)$ [15].

3. Optimization

Considering Eqs. (3)–(6), the optimization equation relying on the objective function $F(\Delta z)$ and the optimum criterion, may be determined as:

$$F(\Delta z) = \Delta z^2 \frac{\sum\limits_{z=0}^{Z-1} W_\theta(z) W_\theta(z - \Delta z)}{\sum\limits_{z=0}^{Z-1} W_\theta^2(z)} \to \max. \qquad (7)$$

The solution of Eq. 7 is the optimal value Δz_{opt} of the spatial shift Δz. Hence, the optimization of Eq. (7) takes the form of:

$$\Delta z_{\text{opt}} = \arg\max F(\Delta z). \qquad (8)$$

We define the type of communication functions $f_l(X)$, the number L and the type of the desired extremum of the objective function $F(\Delta z)$ for optimization.

At first, general spatial correlation processing must be carried out for one correlation processing cycle, using a single-channel correlator. Taking into consideration the

first $f_1(N_C)$ and the second $f_2(N_k)$ the communication functions take the following form:

$$f_1(N_C) = N_C = 1\,, \\ f_2(N_k) = N_k = 1\,, \qquad (9)$$

where N_C and N_k are the numbers of cycles and apparatus channels used in correlation processing respectively.

Secondly, a reconstructed spatial analytic signal $U_{k.S}(\mathrm{j}z)$ is used as a limited sum of spatial harmonic components. Its spatial frequencies Ω_p are the result of FFT properties. They are multiples of the base frequency Ω_1, whose period is equal to the value dz of AA aperture [16]:

$$U_{k.S}(\mathrm{j}z) = \sum_{p=p_L}^{p_H} U_{k.S}(\Omega_p) e^{\mathrm{j}\Omega_p z + \Psi_p}\,, \qquad (10)$$

where $p_H = p_L = m_S$ is the dimension of the signal group and $U_{k.S}(\Omega_p)$, Ψ_p are the module and the argument of the p-th component of the signal group. Equation 10 determines the periodicity of distribution of counts of the spatial analytical signal $U_{k.S}(\mathrm{j}z)$ with a period equal to the value dz of the AA aperture:

$$U_{k.S}((\mathrm{j}z)) = U_{k.S}[\mathrm{j}(z \pm dZm)]\,, \qquad (11)$$

where $m = 1, 2, \ldots, Z-1$.

Let's also note that the weight function $W_\theta(z)$ of the digital spatial spectral analysis has the mirror-symmetric distribution of counts within the AA aperture, concerning its center coordinates $\frac{dZ}{2}$.

Therefore, an unambiguous value of the correlation coefficient $K_{W\theta}(\Delta z)$ will be formed by using a spatial shift Δz, that does not exceed in terms of value $\frac{dZ}{2}$, the AA aperture. The domain of the correlation coefficient $K_{W\theta}(\Delta z)$ and the objective function $F(\Delta z)$ will be $[-0.5dZ; 1.5dZ]$. As a result, the third $f_3(\Delta z)$ and the fourth $f_4(z)$ communication functions have the form:

$$f_3(\Delta z) = \frac{dZ}{2} > \Delta z > -\frac{dZ}{2}\,, \\ f_4(z) = d(Z-1) > z \geq 0\,. \qquad (12)$$

In the third step, as spatial reconstruction of the analytical signal $U_{k.S}(\mathrm{j}z)$ is realized based on of the FFT algorithm, its spatial counts will coincide with the spatial counts $U_z(t_k)$ of received mixtures within the AA aperture only at md locations of the AA antenna elements. In other locations of AA, the approximation error may be significant [16]. Considering this, it is appropriate to use only the discrete spatial shift which is a multiple of the AA step d. As a result, the corresponding value of the fifth communication function $f_5(\Delta z)$ is as follows:

$$f_5(\Delta z) = \Delta z = md\,. \qquad (13)$$

Let's perform a probabilistic analysis of the features of the spatial distribution of counts of the analytical signal $U_{k.S}(\mathrm{j}z)$

that should be presented as an additive mixture of signal $S_{aS}(\mathrm{j}z)$ and noise $S_{aN}(\mathrm{j}z)$ components within the AA aperture:

$$U_{k.S}(\mathrm{j}z) = S_{aS}(\mathrm{j}z) + S_{aN}(\mathrm{j}z)\,. \qquad (14)$$

The noise component $S_{aN}(\mathrm{j}z)$ is formed as the sum of m_S harmonic noise components $S_{aN.m}(\mathrm{j}z)$ of the signal group, that is:

$$S_{aN}(\mathrm{j}z) = \sum_{m=1}^{m_S} S_{aN.m}(\mathrm{j}z)\,. \qquad (15)$$

Every harmonic noise component $S_{aN.m}(\mathrm{j}z)$ is a narrowband random process with a normal distribution of probability density of counts. As a result, the noise component $S_{aN}(\mathrm{j}z)$ of the spatial analytical signal $U_{k.S}(\mathrm{j}z)$ is also a normal narrowband random process, and its power P_aN will be equal to the sum of powers $P_{aN.m}$ of m_S constituent components:

$$P_{aN} = \sum_{m=1}^{m_S} P_{aN.m}\,. \qquad (16)$$

In addition, the noise component $S_{aN}(\mathrm{j}z)$ of the spatial analytical signal $U_{k.S}(\mathrm{j}z$ has a correlation interval Δz_{cor}, which is determined by its bandwidth $\frac{m_S-1}{dz}$ as [11]:

$$\Delta z_{\mathrm{cor}} = \frac{dz}{m_S - 1}\,. \qquad (17)$$

The analysis of Eq. (17) shows that it is necessary for the spatial shift Δz not to be less than the correlation interval Δz_{cor} to ensure effective statistical estimation of the difference $\Delta\hat{\Psi}_{Ak}(\Delta z, \Omega_p)$ between the arguments of the signal component $S_aS(\mathrm{j}z)$ of the spatial analytical $U_{k.S}(\mathrm{j}z)$ [11]. Taking this into account, the sixth communication function $f_6(\Delta z)$ takes the form:

$$f_6(\Delta z) = \Delta z \geq \frac{dz}{m_S - 1}\,. \qquad (18)$$

It can be concluded from Eqs. (9)–(18) that optimization should be carried out taking into account the six $L = 6$ functions of communication and the search for a global constrained extremum type maximum.

Considering the optimum value of the spatial shift Δz_{opt}, Eq. (8) is defined as the solution of the corresponding differential equation:

$$\Delta z_{\mathrm{opt}} = \arg\frac{\partial F(\Delta z)}{\partial \Delta z} = 0\,. \qquad (19)$$

We take into account that, in general, the weight function $W_\theta(z)$ is nonlinear. Therefore, it is advisable to use nonlinear programming methods to solve the optimization problem [15].

The Mathcad software package was used under the following conditions. The number of DF channels $Z = 64$, the types of weight functions are minimum 3-sample Blackman-Harris and minimum 4-sample Blackman-Harris – for the solution of Eq. (19).

The advantage of such an approach is that they are optimized by the minimum of the area of all side lobes and have

a sufficiently low level of maximum side lobes (-67 dB and -92 dB, accordingly). Additionally, the windows show a decrease in the level of side lobes -6 dB/oct, as opposed to the Dolph-Chebyshev window, for which the decrease in the level of side lobes is -6 dB/oct [17]. Consequently,
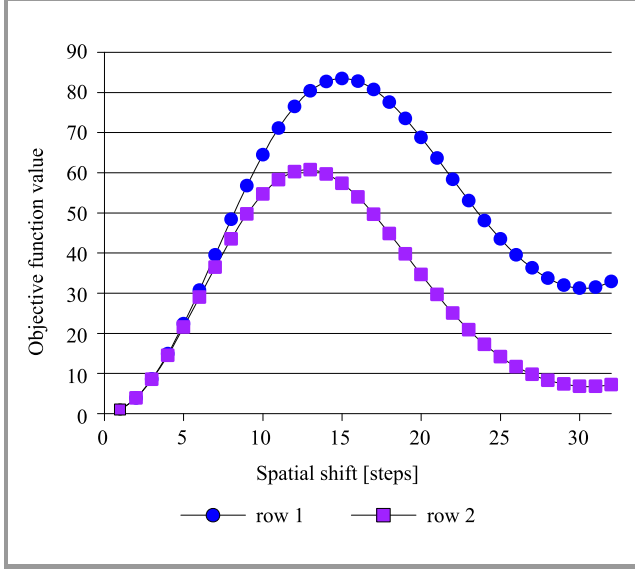

Fig. 2. Spatial shift optimization results.

the use of these windows provides an efficient selection of powerful random interference that has the uniform and the Gaussian distribution of the probability density in the direction of arrival. Such a version of EMS is typical for conducting of radio monitoring in the urban conditions in the presence of networks of comprised of moving sources of radio emission sources. The minimum 3-sample Blackman-Harris weight function $W_\theta(z)$ has a -67 dB level of side lobes and is defined by [16]:

$$W_\theta(z) = 0.42329 + 0.49755\cos\frac{2\pi z}{Z} + 0.07922\cos\frac{4\pi z}{Z}. \quad (20)$$

The minimum 4-sample Blackman-Harris weight function has 92 dB level of side lobes and is defined by [16]:

$$W_\theta(z) = 0.35875 - 0.48829\cos\frac{2\pi z}{Z} + 0.14128\cos\frac{4\pi z}{Z} - 0.01168\cos\frac{2\pi z}{Z}. \quad (21)$$

Optimization results for the minimum 3-sample Blackman-Harris window $W_\theta(z)$ (row 1) and for the minimum 4-sample Blackman-Harris window $W_\theta(z)$ (row 2) are shown in Fig. 2.

Analysis of Fig. 2 shows that the objective function $F(\Delta z)$ for both types of windows has a bell-shaped form with one global maximum. The maximum value of the objective function $F(\Delta z)$ will be achieved when $\Delta z_{opt} = 15$ for the minimum 3-sample Blackman-Harris window, and

$\Delta z_{opt} = 13$ for the minimum 4-sample Blackman-Harris window.

Therefore, the purpose of this article that is concerned with the optimization problem of direct digital method of correlative-interferometric radio DF with two-dimensional correlative processing of a reconstructed complex spatial signal and with determination of the optimal spatial shift has been achieved.

4. Simulation Results

A simulation of the direction finder was conducted [3]. A software model was developed in Mathcad for the following initial conditions:

- signal type – continuous with linear frequency modulation,

- width of signal $\Delta f_S = 0.6$ MHz,

- passband of DF radio channels $\Delta f_k = 10$ MHz,

- working frequency of the signal $f_S = 2$ GHz,

- sampling frequency $f_d = 2\Delta f_k = 20$ MHz,

- numbers of signal time counts analyzed $N_S = 2048$,

- duration of analysis $T_a = 0.1$ ms,

- AA type – linear AA with the number of DF reception channels $Z = 64$,

- given direction towards RES $\theta = 60°$ in relation to the AA aperture line,

- AA step size $d = 0.5$ m,

- number of tests to assess the middle square deviation (MSD) of estimation of direction $= 50$,

- number of counts of signal groups $m_S = 6$ for the minimum 3-sample Blackman-Harris window $W_\theta(z)$,

- number of counts of signal groups $m_S = 8$ for the minimum 4-sample Blackman-Harris window $W_\theta(z)$.

As a result, the simulations performed for the analyzed DF method we receive a range of dependencies concerning MSD of direction estimation, obtained based on the Δz spatial shift values (in AA steps of d) for the signal-to-noise ratio of 0 dB for the minimum 3-sample and 4-sample Blackman-Harris windows, as shown in Fig. 3.

Figure 3 presents two curves: row 1 – for the minimum 3-sample Blackman-Harris window $W_\theta(z)$, and row 2 – for the minimum 4-sample Blackman-Harris window $W_\theta(z)$.

Figure 3 shows that the value of MSD is a smooth line with one global minimum. The minimum MSD value is equal to $\sigma_\theta = 0.02$, when $\Delta z = 15$, row 1 and $\sigma_\theta = 0.025$, when $\sigma_\theta = 0.025$, row 2. The lower value of MSD of direction estimation is obtained by using the Blackman-Harris 3-sample window (row 1). This is consistent with

the theory, as this window has a smaller equivalent noise bandwidth [17].
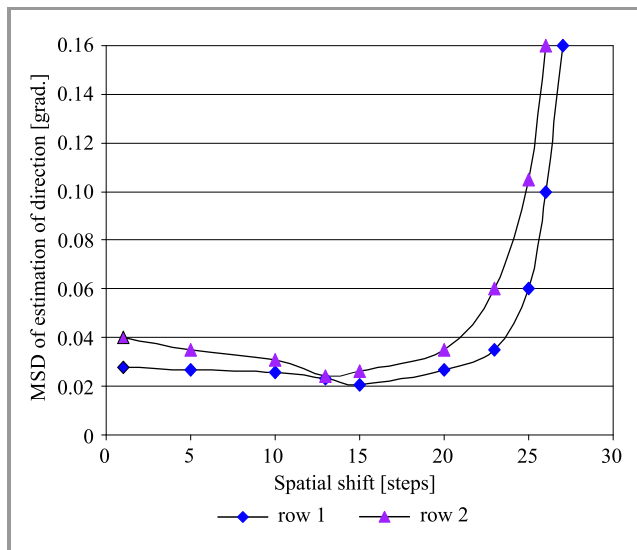
Fig. 3. Range of dependencies between MSD of direction estimation and spatial shift values.

Let's evaluate the effect of optimization on the accuracy of DF. The experiments have shown that within the range of Δz changes from 1 to 32, the value of MSD of direction estimation varies within $[0.02; 1]$ degrees, that is by 50 times. This proves that a significant impact is exerted on the accuracy of the optimization process. With the use of certain optimal spatial shift values of $\Delta z = 13$ and $\Delta z = 15$, it is possible to improve the accuracy of direction estimation of RES or to increase the speed of the system by significantly reducing the duration of radio frequency analysis.

The received dependence shown in Fig. 3 can be conventionally divided into three ranges of Δz values. For example, for row 2 in the first range $[1; 13]$, the value of MSD of direction estimation decreases gradually by almost 2 times, from $\sigma_\theta = 0.04$ to $\sigma_\theta = 0.025$. In the second range of $\Delta z = [14; 22]$, the value of MSD of direction estimation increases gradually to $\sigma_\theta = 0.04$. In the third range of values $\Delta z = [23; 32]$, the MSD of direction estimation increases rapidly with the quadratic dependence. The obtained results confirm the high level of accuracy of the DF method investigated, at a low signal-to-noise ratio of 0 dB.

The comparative analysis of Figs. 2, 3 shows that the analytical results of optimization and the simulation results are the same for spatial shift values Δz for which the MSD of direction estimation is minimal. Therefore, the aim of the article is achieved.

5. Conclusions and Future Work

In this article, the main optimized parameter of the correlative-interferometric method of direction finding with two-dimensional correlative processing of spatial signal in the aperture of a linear antenna array is determined. It is the Δz value of the spatial shift within the AA aperture for which the phase shift and the frequency of spatial analytical signal are estimated.

Analytical optimization of the parameters of the studied method has shown that the maximum value of the objective function $F(\Delta z)$ with the optimal spatial shift $F(\Delta z)$ for the minimum 3-sample Blackman-Harris window $W_\theta(z)$ and $W_\theta(z)$ for the minimum 4-sample Blackman-Harris window $W_\theta(z)$ is achieved.

Analysis of the simulation results has shown that the value of MSD of direction estimation is monotonically dependent on the spatial shift Δz with one minimum, and equals $\sigma_\theta = 0.02$ for the spatial shift of $\Delta z = 15$ for the minimum 3-sample Blackman-Harris window $W_\theta(z)$, and equals $W_\theta(z)$ for the spatial shift of $W_\theta(z)$ for the minimum 4-sample Blackman-Harris window $W_\theta(z)$. The lower value of MSD of direction estimation is obtained by using the Blackman-Harris 3-sample window. This is consistent with theory, because this window has a smaller equivalent noise bandwidth.

The analytical results of optimization and the simulation results are the same for both windows. The value of MSD will be minimal and is equal $\sigma_\theta = 0.02$ for the spatial shift of $\Delta z = 15$ while using the minimum 3-sample Blackman-Harris window the side lobes level of -67 dB, which offers a high degree of interference immunity of DF.

The results obtained are suitable for application in the operation of radio monitoring, telecommunications and radio navigation systems which function in a complex dynamic EMS.

References

[1] G. Kratschmer (Ed.), *Introduction into Theory of Direction Finding. Radiomonitoring and Radiolocation 2010/2011*. Munich: Rohde & Schwarz GmbH & Co. HW – UKD, 2011, pp. 85-101.

[2] A. M. Rembovskiy, A. V. Ashyhmin, and V. A. Kuzmin, *Radiomonitoring – tasks, methods, devices – 2nd ed*. Hotline – Telecom, 2010.

[3] V. V. Tsyporenko and V. G. Tsyporenko, "Development of direct method of direction finding with two-dimensional correlative processing of spatial signal", *Eastern-Europ. J. of Enterpr. Technologies* vol. 6, no. 84, pp. 63–70, 2016 (doi: 10.15587/1729-4061.2016.85599).

[4] J.-H. Lee and J.-M. Woo, "Interferometer direction finding system with improved DF accuracy using two different array configurations", *IEEE Anten. and Wirel. Propag. Let.*, vol. 14, pp. 719–722, 2014 (doi: 10.1109/LAWP.2014.2377291).

[5] J. Yang, W. Chen, L. Li, and X. Ni, "Long baseline direction finding and localization algorithms for noise radiation source", in *Proc. 12th Int. Conf. on Signal Process. ICSP 2014*, Hangzhou, Zhejiang, China, 2014, pp. 52–57 (doi: 10.1109/ICOSP.2014.7014968).

[6] H. Gazzah, J. P. Delmas, and S. M. J. Larsys, "Direction-finding arrays of directional sensors for randomly located sources", *IEEE Transac. on Aerospace and Elec. Sys.*, vol. 52, no. 4, pp. 1995–2003, 2016 (doi: 10.1109/TAES.2016.150655).

[7] S. Van Doan, J. Vesely, P. Janu, P. Hubacek, and X. L. Tran, "Optimized algorithm for solving phase interferometer ambiguity", in *Proc. 17th Int. Radar Symp. IRS 2016*, Kraków, Poland, 2016, pp. 1–6 (doi: 10.1109/IRS.2016.7497353).

[8] X. Fu, N. D. Sidiropoulos, W.-K. Ma, and J. Tranter, "Blind spectra separation and direction finding for cognitive radio using temporal correlation-domain ESPRIT", in *Proc. 39th Int. Conf. on Acoustics, Speech and Signal Proces. ICASSP 2014*, Florence, Tuscany, Italy, 2014, pp. 7749–7753 (doi: 10.1109/ICASSP.2014.6855108).

[9] K. V. Rangarao and S. Venkatanarasimhan, "Gold-MUSIC: A variation on MUSIC to accurately determine peaks of the spectrum", *IEEE Transac. on Anten. and Propag.*, vol. 61, no. 4, pp. 2263–2268, 2013 (doi: 10.1109/TAP.2012.2232893).

[10] W. Chen, X. Xu, S. Wen, and Z. Cao, "Super-resolution direction finding with far-separated subarrays using virtual array elements", *IET Radar, Sonar & Navigation*, vol. 5, no. 8, pp. 824–834, 2011 (doi: 10.1049/iet-rsn.2010.0289).

[11] V. V. Karavajev and V. V. Sazonov, *Statistic theory of passive location*. Moscow: Radio and Communication, 1987 [in Russian].

[12] Q. Cui, J. Wang, J. Han, and X. Tao, "Pattern optimisation algorithm for vertical cell splitting system with active antenna arrays", *IET Elec. Let.*, vol. 50, no. 20, pp. 1416–1417, 2014 (doi: 10.1049/el.2014.2023).

[13] J.-H., Lee, J.-H. Lee, and J.-M. Woo, "Method for obtaining three- and four-element array spacing for interferometer direction-finding system", *IEEE Anten. and Wirel. Propag. Let.*, vol. 15, pp. 897–900, 2016 (doi: 10.1109/LAWP.2015.2479224).

[14] T. Sánchez, C. Gómez, G. I. Angel Castillo, and M. Patiño Bernal, "Radio direction finding system for spectrum management activities in developing countries", in *Proc. 17th Int. Symp. on Anten. and Propag. APSURSI 2016*, Fajardo, Puerto Rico, 2016, pp. 1663–1664 (doi: 10.1109/APS.2016.7696538).

[15] G. Fornarelli and L. Mescia, *Swarm Intelligence for Electric and Electronic Engineering*. Hershey, PA, USA: IGI Global, 2012 (ISBN: 9781466626669).

[16] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing, Principles, Algorithms, and Applications*, 4th Edition. Upper Saddle River, NJ, USA: Prentice-Hall Inc., 2006 (ISBN: 9780131873742).

[17] F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform", *Proc. of the IEEE*, vol. 66, no. 1, pp. 51–83, 1978 (doi: 10.1109/PROC.1978.10837).

Vitaliy Tsyporenko obtained his Ph.D. in radio-technical and television systems in 2011, with the thesis focusing on researching new direct digital correlation-interferometric DF methods. At present, he works as an associate professor at the Zhytomyr State Technological University (ZSTU), Ukraine. His current research interests focus on correlation-interferometric direction finding.

https://orcid.org/0000-0001-8559-006X
E-mail: tsyporenko.1985@gmail.com
Zhytomyr State Technological University
Zhytomyr, Ukraine

Valentyn Tsyporenko obtained his Ph.D. in radio-technical systems in 1995, with the thesis focusing on researching new digital signal processing methods. At present, he holds the position of an associate professor at ZSTU, Ukraine. His current research activities focus on direction finding.

https://orcid.org/0000-0002-6843-8960
E-mail: tsyporenkovg@gmail.com
Zhytomyr State Technological University
Zhytomyr, Ukraine

Radio Vision Systems Ensuring Movement Safety for Ground, Airborne and Sea Vehicles

Andrey Ananenkov, Anton Konovaltsev, Vladimir Nuzhdin, Vladimir Rastorguev, and Pavel Sokolov

Moscow Aviation Institute (National Research University), Moscow, Russia

Abstract—This article presents the features of an all-weather radio vision system (RVS) ensuring the safety of movement of ground, airborne and sea vehicles and automation of vehicle traffic control under limited or non-existent visibility conditions. New and promising RVS applications in the aviation and rail transport sectors are presented. The potential use of RVS based on an interferometric radar with aperture synthesis, capable of estimating the position of ice fields and the height of icebergs is considered as well.

Keywords—automotive radio vision system, full-scale tests, interferometric radar with aperture synthesis, traffic control automation, traffic safety.

1. Introduction

One of the urgent problems that needs to be faced in the development of modern vehicles is to ensure the safety and automation of controlling the traffic of both manned and autonomous (unmanned) vehicles under limited or no optical visibility conditions. This problem is equally relevant for road, air, water (sea and river) and rail transport. However, the specific nature of each of those modes of transport is linked with additional requirements that need to be met by the safety sensors.

In road transport, as evidenced by the results of research conducted by NTHSA (USA) [1], in addition to vehicle malfunction, 68% of road accidents are caused by inattention and inadequate awareness of the driver about the traffic situation. These are directly dependent on the visibility of the roadway, on the presence of markings and signs, on the degree of contamination of the cabin glazing, on the presence of fog, snow, rain, dust, smoke and other disturbing factors.

Attempts to solve the driver awareness problem using sensors that operate based on various physical principles (ultrasonic, optical and infrared) are known to have been made, but they failed to provide any major breakthroughs. The drawback of these solutions is obvious – of all negative factors that reduce optical visibility are linked to insufficient illumination. Thus, the main way to solve the actual problem of improving the safety of road transport, under limited or non-existent optical visibility conditions, is to use a radar sensor whose operation does not depend on the time of the day, on weather conditions (snow, rain, fog),

or on the presence of smoke or dust [2]–[9]. The range of such a sensor can be a multiple of the expected length of the stopping distance.

The task of creating a compact radar that can be used in a car, enabling the driver to continue driving under insufficient visibility conditions (dense fog) has been successfully solved by a team of researchers from the Department of Radio Receivers of the Moscow Aviation Institute (MAI). In 2001, the first model of the automotive radio vision system (ARVS), relying on the W band ($\lambda = 4$ mm) and approved for this class of radar by the EU and USA, was created. The use of a new holographic antenna, digital signal processors and an LCD indicator made it possible to reduce the weight and the dimensions of ARVS, so that it closely resembles infrared (IR) systems with respect to these parameters. The system can be easily installed on almost any car [10].

The market of rapidly developing automobile safety systems is not the only one where demand for small-size radar sensors exists. In order to enter the segment of safety systems used on small aircraft and/or in rail transport, it is necessary to significantly increase the operational range of ARVS. To ensure the safety of sea/river transport in icing conditions, it is also necessary to ensure that the radar sensor has the capability to estimate the height of ice formations.

As shown by the results of research focusing on ARVS [10]–[12], it is possible to increase the system's operational range to 1500 m and even to 3000 m without any major changes involving technology. The increase in operational range of ARVS will expand the scope of its application. This version of RVS may be used:

- for ensuring the safety of take-off, landing and cruise phases of light and ultra-light aircraft, which do not currently have any radar equipment, in conditions of limited optical visibility;

- for improving the safety of railway traffic in under limited or non-existent optical visibility conditions, including at level crossings with roads;

- for solving ship navigation-related problems:

  – berthing, including in automated mode,

  – passage under bridges or through other "bottle necks",

– passage through locks,

– high-precision navigation within ports premises,

– detection of semi-submerged objects,

– solution of the problem of the "blind" zone of regular navigational radars,

– detection of ice sheets and icebergs.

In this article, we consider the technical specifications of such a prospective sensor, and present a corresponding model – created at MAI (NIU) – a technology demonstrator, describing the research and field tests conducted. The successful creation of a sample radio vision system (RVS) having the form of a panoramic short-range radar (SRR) that meets the requirements applicable to road transport use, has led to the introduction of such RVSs in aviation, water and rail transport. Further development of radio vision systems used in water transport is linked to the creation of an interferometric radar with the synthesis of the antenna aperture (SAR). This will make it possible to estimate the height of ice surfaces, and detect dilution and icebergs, thus solving the problem of safe navigation of vessels in icing conditions. In this article, estimates of the potential characteristics of such a system, based on ARVS, are given.

2. Aviation RVS for Flight Safety

One of the actual problems faced while developing modern small aircraft is to ensure the safety of flight by preventing mid-air collisions, as well as by enabling to perform safe low-altitude flights, taxiing, take-offs and landings under limited optical visibility conditions. However, due to their low take-off weight, small aircraft are not equipped with radar. The increase in the working range of radio vision systems, with the acceptable weight and size characteristics maintained, has made it possible to propose to develop a radar sensor suitable for small aircraft.

The result of the project implemented jointly by JSC "Techaviakompleks" (Zhukovsky, Moscow region) and the research team of MAI's Department of Radio Receiving Devices in 2016 was a model of an onboard safety system that is mounted on small aircraft and used to prevent mid-air collisions and to ensure safe performance of a low-altitude flights, with the use of a small-size airborne radar complex (RVS) [13]–[14].

The RVS solution considered (Fig. 1) is of the navigational two-coordinate SRR variety, which determines a number of radar information processing and presentation-related features.

The key parameters of the experimental RVS model intended for the aviation industry are listed in Table 1.

Table 1
Key specifications of aviation-grade RVS

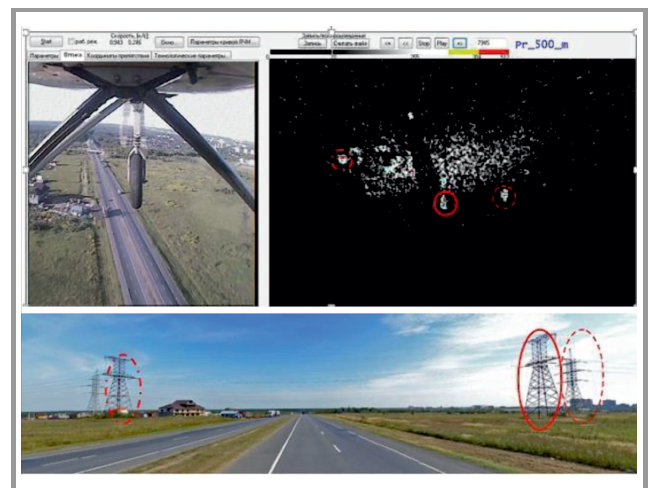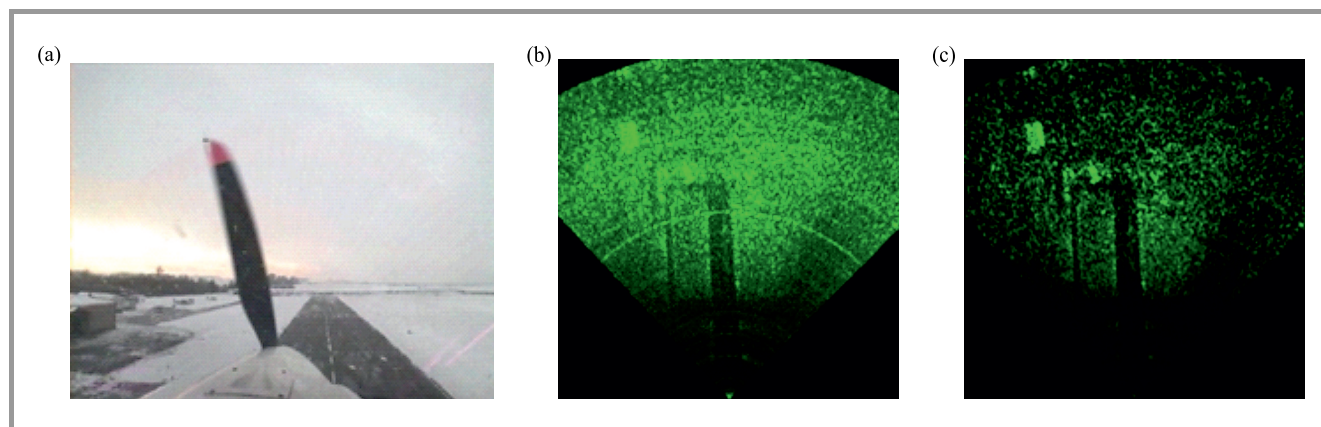| Instrumental range	20–2000 m
Range resolution	1.5 m
Resolution azimuth	$\geq 1°$
Operating frequency	39 GHz
Radiation power	30–50 mW
Antenna	waveguide-slot
Antenna-view in the azimuth plane	$\pm 60°$ (360°)
Antenna-view in elevation plane	from $+10°$ (up) to $-40°$ (down)
The number of information points of the radar image	256×1024 for shot
Energy consumption	< 100 W (12–24 V power supply)
Information update speed	$5 \div 10$ Hz
Radar weight	< 10 kg

Fig. 2. Radar and video images of a location with power transmission lines (highlighted in red) and low-rise buildings. (For color pictures visit www.nit.eu/publications/journal-jtit)

The technical characteristics and functional capabilities of the modernized RVS have confirmed, during the test flights, that the following tasks can be performed undelimited

Fig. 1. IKARUS C42 plane with aviation-grade RVS, during testing.

Andrey Ananenkov, Anton Konovaltsev, Vladimir Nuzhdin, Vladimir Rastorguev, and Pavel Sokolov

Fig. 3. Images of the runway (during take-off): (a) video, (b) radar, and (c) radar with noise filtration.

optical visibility conditions (twilight, fog, drizzle, snow, smoke, etc.):

- airspace control for the presence of airborne objects (helicopters, airplanes, gliders, etc.) and high-rise ground-based objects (relay towers, factory chimneys, tall buildings, etc.), creation of the outline of those objects (Fig. 2);

- formation of highly informative and detailed radar images of the Earth's surface (mapping), enabling the aircraft's navigation systems to rely on the radar map, to search for landing sites, to "see" the runway and taxiways (Fig. 3);

- detection of fixed obstacles and moving vehicles present on runways and taxiways, while landing and taxiing.

The aviation-grade RVS that has been developed provides radar data information that can be presented on the multifunctional display in real time [13]–[14].

3. Application of RVS in Rail Transport

One of the actual problems faced in modern rail transport is related to the improvement of railway traffic safety under limited or zero visibility conditions. With quick development of high-speed and automated railway transport systems observed today, the hazards posed by potentially dangerous situations increase substantially:

- unauthorized and uncontrolled presence on crossing of tracks by: people, animals, foreign objects,

- emergency situations related to the presence of wagons, locomotives, rail cars and road vehicles on the tracks, within the area of unguarded level crossings,

- consequences of flooding and blurring the roadway.

One of the most effective methods of improving the safety of rail transport is the use of a radar-optical complex (ROC)

located on the locomotive and including a short-range RVS radar and an optical sensor. Both devices use a shared display and enable remote detection of obstacles on the tracks:

- under limited zero visibility conditions,

- regardless of weather conditions and of time of day,

- at railway stations – against a background of intense reflections from buildings, structures and the underlying surface,

- in a programmable controlled/monitoring zone,

- in the presence of numerous objects.

IDS-76 – a radar-based range and speed measurement device used in rail transport and developed by DOC (St. Petersburg, Russia) [15] is an example of a system relied upon to improve rail transport safety. The IDS-76 radar is installed permanently on the railway track, for example, at the end of the track (dead-end) and measures the movement parameters of the approaching rolling stock (speed, distance) and transfers this data to the PRISMA-K integrated traffic safety system. It is used at dead-end tracks of passenger stations and protects high-speed passenger suburban traffic and shunting works. The PRISMA-K system determines that a given approach is dangerous and transfers the data and relevant commands to the driver, demanding speed reduction or forcing the locomotive or the head car of the approaching train to perform an automatic stop (Fig. 4).

The IDS-76 radar is based on a 76 GHz car radar and its detection range equals 1.5 to 500 m, with the speed range of 1–100 km/h. Due to the short operating range of this type of radar, millimeter waves are only slightly attenuated by fine-dispersed interference in the form of smoke, dust and precipitation. In other words, the IDS-76 radar suffers from practically no weather – related distortions under the range of 500 m.

The diameter of the antenna, including the radio-transparent cap, is 600 mm. The radar has, with its wavelength considered, a very narrow beam width of $0.42°$. This makes

Fig. 4. IDS-76 radar.

it possible to control only those objects that are within the line of sight on a given track, without reacting trains arriving on adjacent tracks, or passengers and vehicles present on the platform.

Obviously, this radar station, positioned on a permanently fixed barrier, increases safety at braking and stopping stage of the journey, but not while traveling at speed. However, positive experience concerning the IDS-76 radar confirms the relevance of using radar in railway applications.

To ensure a higher level of safety along the route, the following systems should be incorporated into rail transport ROC solutions:

- panoramic front-end short-range radar,

- optoelectronic system with television and/or infrared channels,

- sound and visual warning system for the driver,

- channels transmitting information and communications to traffic supervisors,

- server for data processing, registration and storage,

- optionally – an emergency braking system.

Expert assessment of the Russian Railways (JSC "RZD") specialists enabled RVS – MAI designers to formulate the following requirements related to the main ROC characteristics:

- size of the controlled security zone: person on the tracks detection range of 400 m, vehicle on the tracks detection range of 800 m, train detection range of 1000 m,

- size of the scanned area, within the angle of $\pm 30°$, and with the elevation angle of $\pm 5°$,

- information update rate of 5–10 Hz,

- number of objects to be detected – not less than 3,

- priority of objects according to the criterion of shortest distance,

- RVS is interfaced with optoelectronic facilities,

- output of information about the object and the presentation of a visual/audible signal when obstacles are detected on the railway tracks within the safety zone,

- RVS positioned on the roof of the locomotive, under a radio-transparent cover, with an indicator mounted in the driver's cab,

- environmental safety.

Obviously, the given technical requirements make it feasible to design a similar RVS based on automotive-grade RVS (ARVS), which is discussed in detail in [12].

At the end of 2016, tests demonstrating RVS developed by MAI (NRU) for railway transport application were carried out. Standard electric trains made by the Gorky branch of "RZD" were used. The tests were performed both in stationary mode (within the depot area), and with the electric train moving at a purpose-designed pattern.

In static tests, two variants of RVS placement were tested: on the coupling and on the roof of the electric train's locomotive. During dynamic tests in motion, the placement that was possible in such a scenario was used - on the roof of the electric train's locomotive. Figure 5 shows a photograph of the control module and the RVS display located in the driver's cab. The results of testing the railway version of RVS are shown in Figs. 6–11.

Figure 6 shows a synchronous optical (OI) and radar image (RI) of railway tracks, with a person who is near the tracks, at different distances. In this situation, the radar was placed on the roof of the locomotive. The mark indicating the person on the tracks (highlighted) in OI and RI images proves that the solution is stable at the range of 120 m. Figures 7 and 8 show OI and RI images of railway tracks, with a person who is near the tracks (highlighted), as well as near the locomotive (highlighted).

It has been established, based on the tests results obtained with the electric train moving (Fig. 9), that the experimental RVS system:

- generates a highly detailed RI image in real-time,

- enables observation and detection of people walking along and standing on the tracks,

Fig. 5. Control and display modules in the driver's cab, with the radar placed on the roof.
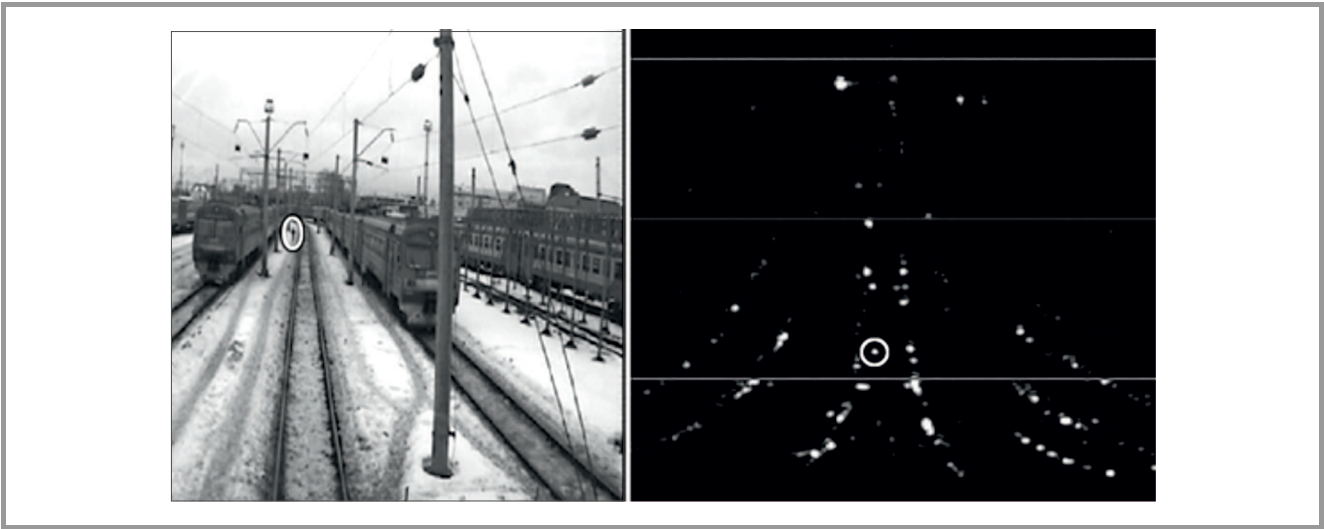

Fig. 6. OI and RI images (vertical coordinate – azimuth, horizontal – range) of railway tracks with a man 120 m away (highlighted).
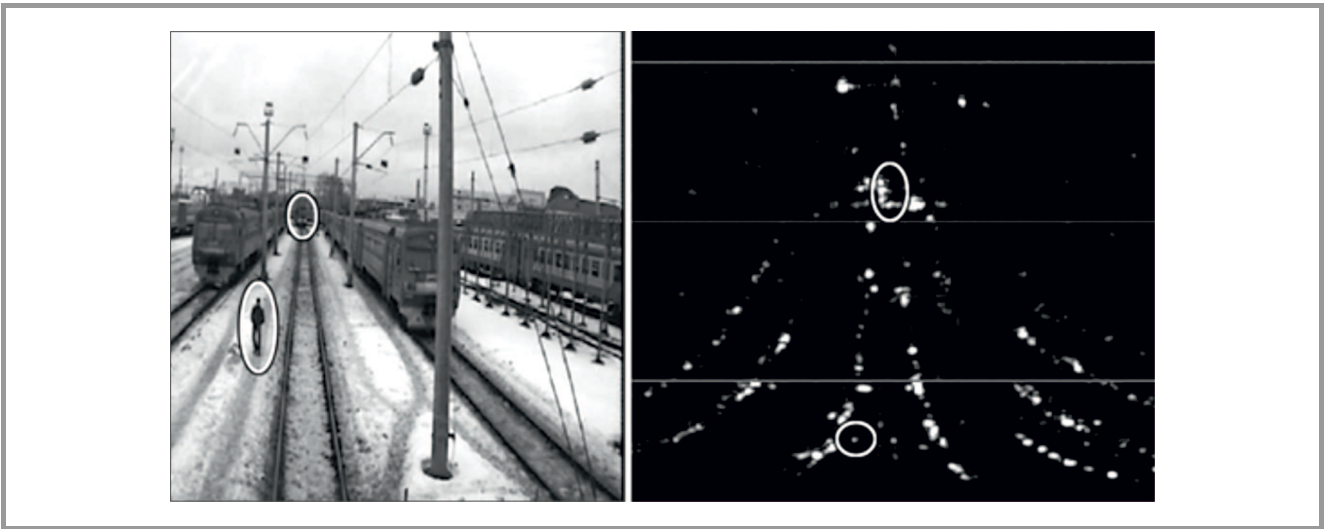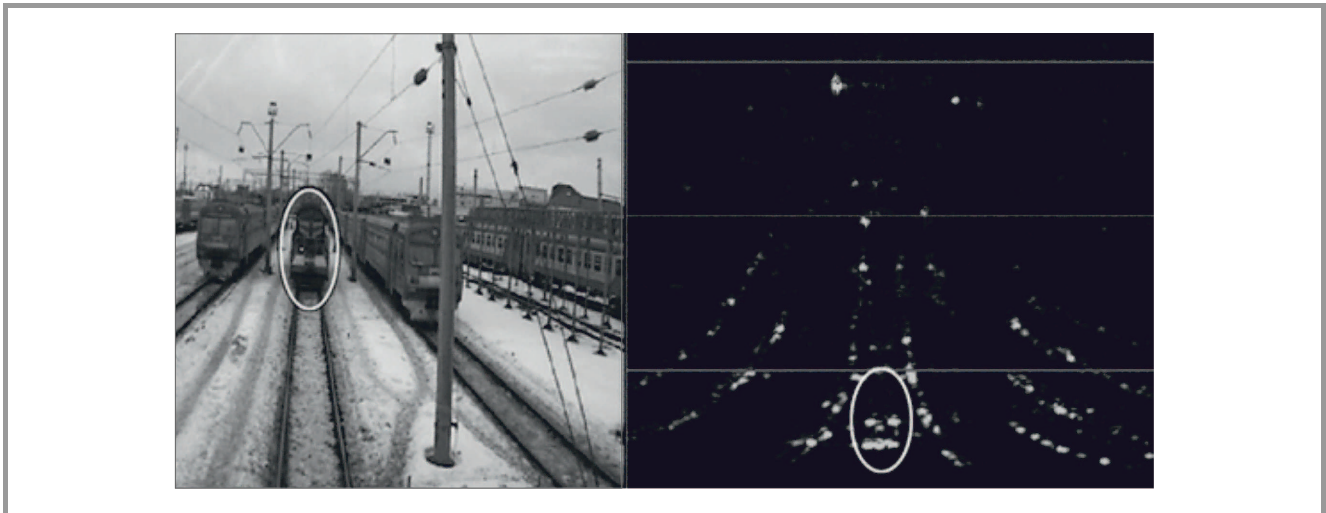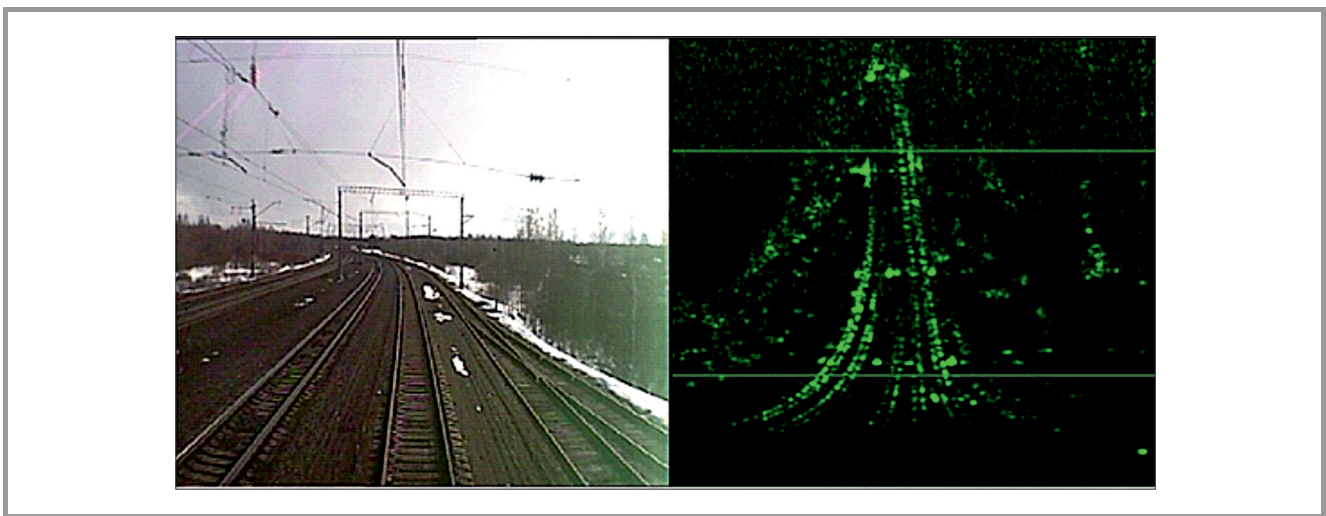

Fig. 7. OI and RI images of railway tracks with a man and a locomotive (highlighted).

Fig. 8. OI and RI images of railway tracks with a Diesel locomotive (highlighted).

Fig. 9. OI and RI images of railway tracks, at speed.

- enable detection and observation of obstacles such as locomotives, wagons and other objects.

The tests carried out with the use of the experimental RVS model installed on an electric train, both in stationary mode and moving modes, confirmed ability to ensure remote detection of obstacles on the railway tracks.

4. Building Interferometric SAR based on RVS

The experience obtained while working with highly detailed radar images, with the view to ensure higher level of safety in transport, enabled us to consider the next step in the development of the radio vision technology. Further development of RVS should focus on small-size radars with antenna aperture synthesis (SAR) capabilities, which will soon find wide practical use to ensure the safety of ships in icing conditions prevailing in northern regions of the planet, due to their low cost and low expenses associated with the operation of their carriers – small-sized unmanned aerial vehicles (UAVs) [16].

Expanding the range of applications of such SAR systems may offer an additional advantage consisting in measuring, in one pass, the height of the underlying surface, with the use of interferometry methods [17]. The mostobvious areas for application of interferometric small-size SAR are:

- all-weather operational monitoring of the ice situation, to identify dilution and icebergs that threaten navigation, to boost both the safety of naval navigation and to ensure the safety of oil platforms in northern oceans,

- round-the-clock, all-weather operational monitoring of flood zones, ice congestion on rivers, avalanche zones, landslides and mudflows,

- operational monitoring of oil pipelines to identify fistulas (leaks of oil products) and critical deformations caused by subsidence of the ground.

Such an instrument will be especially effective when obtaining periodical, comparative images of the same section of terrain, in small time intervals, to assess the dynamic character of any emergency situation. Here, small-size interferometric SAR will offer some unquestionable advantages, such as its independence from illumination, meteorological conditions, and the presence of smoke and dust in the atmosphere. The low per-hour cost of the carrier – a small UAV, is important as well.

To create such SAR, RVS equipment must be supplemented by a second, identical receiving path and a second antenna. The receiving path antenna must be vertically spaced in relation to the base, which ensures unambiguous measurements of altitude at the selected range. The SAR design should generate radar images of the underlying surface with a range resolution and azimuth of the order of 1 m within a 3 km survey strip, from an altitude of about 1 km. In this case, the angles of incidence of the electromagnetic waves (EMW) on the surface (in the approximation of a plane surface) will equal from 15.5 to 60°. With the antenna's elevation plane beam width of about 15°, the capture bandwidth will range from 430 to 1900 m. The width of the signal spectrum necessary to provide the required range resolution (at the near edge of the capture band) is 300 MHz. The estimation of SAR's power while mapping the underlying surface, performed based on the radiolocation formula [17], shows that at the required signal-to-noise ratio $q = 20$ dB, at the maximum operating range, at the gain of the antenna $G = 33.4$ dB, at the specific radar cross section (RCS) of the underlying surface $\sigma = -20$ dB, the average continuous radiation power of SAR with the selected modulation will be $P = 5$ W.

The requirements for the SAR antenna system are contradictory. On the one hand, it is necessary to provide sufficient amplification of the useful signal (which requires an increase in the antenna size) and sufficient suppression of the side-lobe maxima of the synthesized aperture. On the other hand, if possible, the real antenna beam must be wide enough to minimize losses from the random bank angle and fluctuations of the course angle during the movement of the air carrier in air streams at low altitude.

Let us define the minimum size of antenna diagram in azimuthal plane $d_{a\min}$:

$$d_{a\min} = 4VT_m = 0.12 \text{ m} , \qquad (1)$$

where V – carrier speed (30 m/s), and T_m – probing signals period (1000 samples per second).

To increase SAR energy, we will choose the azimuth aperture of 0.5 m. The ray beam azimuth width will be about 0.9°. Such a wide azimuth beam allows the use of non-coherent accumulation of forth radar image frames to reduce the level of speckle noise. In this case, the requirement for the accuracy of maintaining the heading angle will be less than ± 0.5°. To ensure the required gripping strip, the antenna beam width in the elevation plane should be less than 15.2°, which is provided in the Ka band by the size of the antenna aperture of approximately 3 cm.

The use of a homodyne receiver (substantially reducing the bandwidth of received signals) and frequency expansion of the received signals relying on the RVS technology, as described in [18], can bypass the problem of creating a broadband amplifier with a wide dynamic range. This result allows for a significant narrowing of the dynamic range of signals at the output of the last cascades of the amplifier.

Let us estimate the band of the beat signal F_σ to be amplified, digitized, and processed:

$$F_\sigma = \frac{2\Delta F F_{\text{M}} R}{c} , \qquad (2)$$

where F – the HF signal band providing the necessary delay resolution (300 MHz), F_{M} – signal modulation frequency (1 kHz), R – maximum range to the objects being observed (3740 m), c – speed of light.

In consideration of the above, the necessary bandwidth F_σ for amplification and processing of the received signals will be 3.7 MHz, which enables the use of cheap, light and low-energy demand equipment for amplification, digitization and processing of radar signals.

Let us consider the altitude estimation algorithm, assuming that there are two coupled matrices of the complex amplitudes of the signals received at the antennas $A_1(i, j)$ and $A_2(i, j)$, in the on-board computer. Indices i and j correspond to discrete range and azimuth readings, respectively. The procedure for multiplying the conjugate complex amplitudes must be realized in the computer. As a result, we obtain the following matrix:

$$A_1(i, j) \cdot A^*(i, j) = |a(i, j)|^2 \cdot e^{j[\varphi_1(i,j) - \varphi_2(i,j)]} , \qquad (3)$$

where $|a(i, j)|^2$ – power multiplier proportional to the intensity of the signals received on two antennas, or the radio brightness field, and $\Delta\varphi(i, j) = \varphi_1(i,j) - \varphi_2(i,j)$ – field of phase distribution or phase difference.
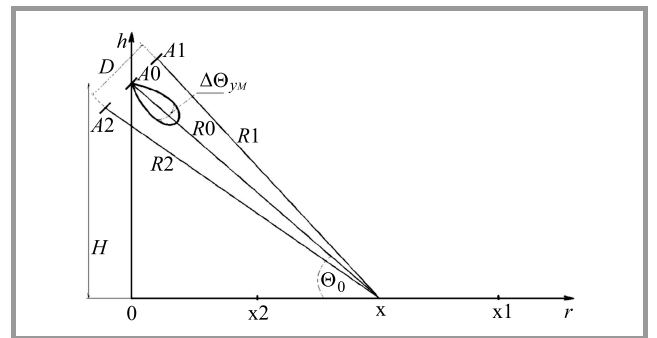

Fig. 10. Geometric relationships in the interferometer.

Consider the dependence of the phase difference of the signals, for a fixed value of the azimuthal coordinate $j = \text{constant}$, and a variable of the inclined range $R = i \cdot \Delta R$. The analysis of this dependence is carried out for different slope angles of the base of the separation of the receiving

antennas – Θ_0, relative to the direction of the local horizontal. The geometrical relationships used are explained in Fig. 10.

It is easy to show that the phase difference $\Delta\varphi(i, j = \text{const})$ will be determined by the expression:

$$\Delta\varphi(r, j = \text{const})$$

$$= kD \cdot \left[\frac{H}{R} \cdot \cos\Theta_0 - \sqrt{1 - \left(\frac{H}{R}\right)^2} \cdot \sin\Theta_0 \right] , \quad (4)$$

where: $k = \frac{2\pi}{\lambda}$ – wave factor, D – receiving antennas spacing base, H – height of the antenna system phase center location over the reflecting surface. The graphs of the normalized phase difference, for various angles – the inclination of the base – Θ_0, are shown in Fig. 11. These graphs are obtained in the absence of unevenness of the relief (flat horizontal reflecting surface) and the constant altitude of the flight of the air carrier ($H = 1000$ m).
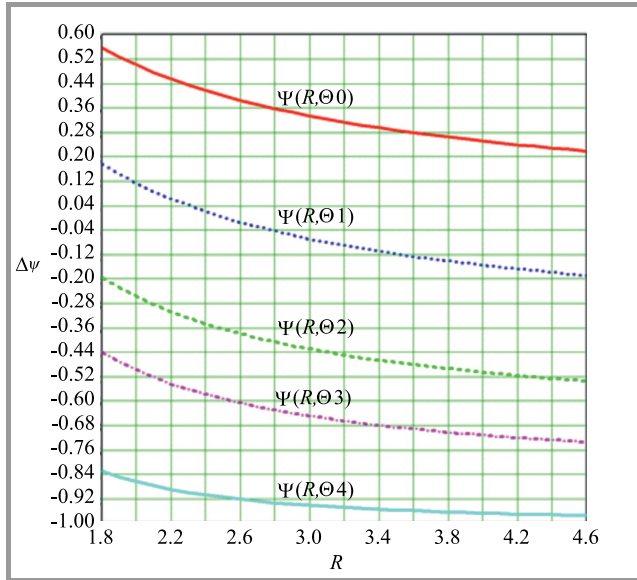
Fig. 11. Normalized dependence of the phase difference (in radians) at a constant altitude H and a variable inclined distance R in km (for setting angles $\Theta_0 = 0$, 23.5, 45, 60 and 90°).

To identify the relationship between the phase difference and the height of the unevenness h (discriminating characteristic), we differentiate the phase difference function with respect to the variable H and write the product of the derivative with variable h, which will give us the required characteristic:

$$\Delta\psi(h) = kD \cdot \frac{H}{R} \left[\cos\Theta_0 + \frac{\frac{H}{R}}{\sqrt{1 - \left(\frac{H}{R}\right)^2}} \cdot \sin\Theta_0 \right] \cdot \frac{h(R)}{H} .$$

$$(5)$$

Analysis of the expression shows that the maximum slope of the discrimination characteristic is realized at $\Theta_0 = 0$, increasing the slope angle to $\Theta_0 = 90°$ leads to a decrease

in the slope and, accordingly, to a decrease in the sensitivity of the interferometer. The slope decreases inversely with the range at which the estimated elevation of the relief is above the average level. An example of a discriminatory characteristic for the setting angles $\Theta_0 = 0$, 23.5, 45, 60 and 90°, $\frac{D}{\lambda} = 20$ for medium range $R = 3600$ m, and $H = 1000$ m, is shown in Fig. 12.
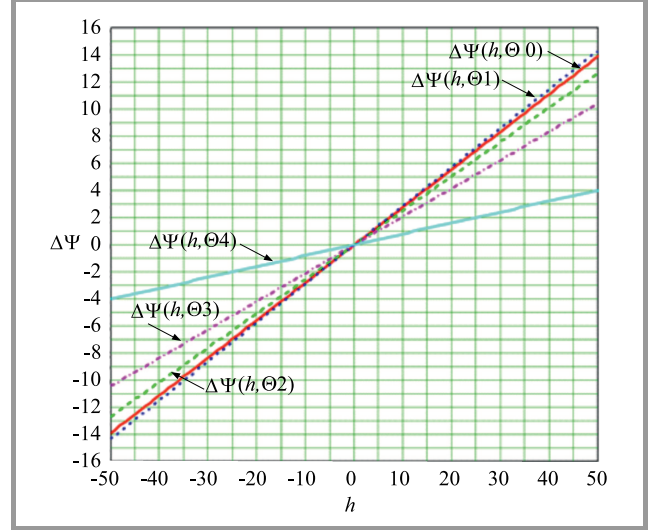
Fig. 12. Discriminatory characteristic (the dependence of the phase difference in radians from the height of the surface in m) for the angles 0, 23.5, 45, 60 and 90°.
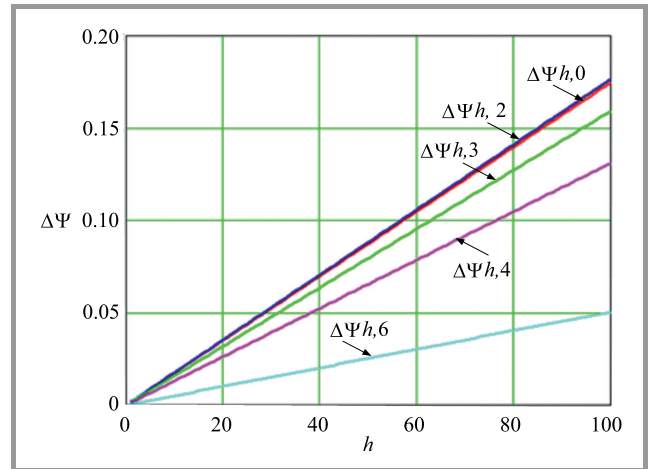
Fig. 13. Normalized (to $\frac{D}{\lambda}$) discriminatory characteristics.

Figure 13 shows that normalized discrimination characteristics are constructed as functions of the inclined distances for a fixed value $\frac{h}{H} = 0.1$, where the inclination angle of the interferometer base Θ_0 is used as the variable parameter.

The results of discriminatory characteristics analysis show that with vertical spacing between the receiving antennas, the maximum sensitivity of the interferometer changes along with the height of irregularities. However, at large bases $\frac{D}{\lambda} = 20 \div 30$, the ambiguity of estimates of the current phases is encountered. The use of the horizontal lo-

Andrey Ananenkov, Anton Konovaltsev, Vladimir Nuzhdin, Vladimir Rastorguev, and Pavel Sokolov

cation $\Theta_0 = 90°$, leads to a significant decrease in the sensitivity of the interferometer, which can be increased by applying significant separation bases $\frac{D}{\lambda} > 30$.

Changing the angle of inclination by $\Theta_0 = 0 \div 60°$ does not reduce the sensitivity of the interferometer. Hence, we can consider the angle of $\Theta_0 = 60°$ to be the operating inclination angle of the antenna base.

5. Conclusion

The article deals with the application of a radio vision system (RVS), which is the optimal all-weather sensor improving the safety of road traffic, aviation, as well as rail and water transport. An example presenting the use of RVS to improve the safety of a modern small aircraft is given. The aircraft-grade RVS will prevent mid-air collisions and will enable to safely perform low-altitude flights, takeoffs and landings in difficult weather conditions, with limited visibility.

Directions in which long-term development of the RVS technology used in railway transport should be heading are formulated. The form and the main technical characteristics proposed for the use of RVS in water transport are presented – enabling all-weather operational monitoring of ice conditions, identifying dilution and icebergs that threaten navigation, both to increase the safety of maritime navigation and that of oil rigs in the world's northern oceans.

Acknowledgements

References

[1] Traffic Safety Facts Research Note NHTSA HS 812101, "2013 Motor Crashes: Overview", Dec. 2014 [Online]. Available: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812101

[2] Y. Asano, S. Ohshima, M. Ogawa, and K. Nishikawa, "Proposal of millimeter-wave holographic radar with antenna switching", in *Proc. IEEE MTT-S Int. Microw. Symp. Digest*, Phoenix, AZ, USA, 2001 (doi: 10.1109/MWSYM.2001.967086).

[3] M. Chiaberge (Ed.), *New Trends and Developments in Automotive System Engineering*. Intech, 2011 (doi: 10.5772/552, ISBN: 9789533075174)

[4] Maria S. Greco, "Automotive radar", in *IEEE Radar Conf. RADAR 2012*, Atlanta, USA, 2012 [Online]. Available: http://www.iet.unipi.it/m.greco/esami_lab/Radar/automotive_radar.pdf

[5] K. Shirakawa, S. Kobashi, Y. Kurono, M. Shono, and O. Osamuisaj, "3D-scan millimeter-wave radar for automotive application", *Fujitsu TEN Tech. J.*, no. 38, pp. 3–7, 2013.

[6] L. Stanislas and T. Peynot, "Characterization of the Delphi Electronically Scanning Radar for robotics applications", in *Proc. of Australasian Conf. on Robot. and Autom. ACRA 2015*, Canberra, Australia, 2015, pp. 434–443.

[7] M. Kishida, K. Ohguchi, and M. Shono, "79 GHz band high-resolutions millimeter-wave radar", *Fujitsu Scient. and Tech. J.*, vol. 51, no. 4, pp. 55–59, 2015.

[8] "Implementing digital processing for automotive radar using SoC FPGAs", White Paper, no. WP-01183-1.2, Altera Corp., 2013.

[9] C. Waldschmitt and H. Meinel, "Future trends and directions in radar concerning the application for autonomous driving", in *Proc. of the 11th Eur. Radar Conf. EuRAD 2014*, Rome, Italy, 2014, pp. 416–419 (doi: 10.1109/EuRAD.2014.6991296).

[10] A. E. Ananenkov *et al.*, "MM range ARVS as a means of investigating into the objects of natural and and artificial origin", *Elektronika i Informatika*, no. 4, 2002 [in Russian].

[11] A. E. Ananenkov, A. V. Karpyshev, V. M. Nuzhdin, V. V. Rastorguev, V. B. Shnaider, and G. A. Morozov, "Microwave distance sensor of the helicopter fire fighting system", *Russian Aeronautics*, vol. 57, no. 4, pp. 406–411, 2014 (doi: 103103/S1068799814040151).

[12] A. E. Ananenkov, A. V. Konovaltsev, V. M. Nuzhdin, V. V. Rastorguev, and P. V. Sokolov, "Optical data-bus and microwave systems for automotive application in vehicles, airplanes and ships", in *Optical and Microwave Technologies for Telecommunication Networks*, O. Strobel, Ed. Wiley, 2016 (ISBN: 9781119971900).

[13] Y. V. Likharev, V. M. Nuzhdin, P. V. Sokolov, V. I. Akhrameev, and M. C. Shelagurova, "Determining actual aircraft positions relative to the runway using the airborne radar and approach navigation", *Int. J. of Control Theory and Appl.*, vol. 9, no. 30, pp. 83–93, 2016.

[14] A. Ananenkov, Y. Likharev, V. Rastorguev, and P. Sokolov, "Research of opportunities of short-range radar to prevent flight accidents", in *Proc. of 18th Int. Conf. on Transparent Opt. Netw. ICTON 2016*, Trento, Italy, 2016 (doi: 10.1109/ICTON.2016.7550247).

[15] "All weather radar IDS-76", DOK Company, 2002–2018 [Online]. Available: http://dokltd.ru/products/a20192

[16] A. E. Ananenkov, A. I. Kanashchenkov, V. M. Nuzhdin, V. V. Rastorguev, and A. M. Smolyar, "Estimation of potential characteristics of onboard radar for ice surface monitoring", in *Proc. of 19th Int. Conf. on Transparent Opt. Netw. ICTON 2017*, Girona, Spain, 2017 (doi: 10.1109/ICTON.2017.8025027).

[17] M. I. Skolnik (Ed.), *Radar Handbook*, 3rd Ed. McGraw-Hill, 2008 (ISBN: 9780071485470).

[18] A. E. Ananenkov, A. V. Konovaltsev, V. M. Nuzhdin, V. V. Rastorguev, and P. V. Sokolov, "Homodyne radar", patent no. RU2626405, Russia, 27/07/2017.

Andrey E. Ananenkov is a senior scientific employee of the Scientific Center for Special Radio-Electronic Systems and Management of the Moscow Aviation Institute – National Research University. He obtained his Ph.D. from MAI in 1999. He is the author of over 30 scientific publications focusing on signal processing in radar systems.

E-mail: pan_angej@rambler.ru

Moscow Aviation Institute (National Research University)

4, Volokolamskoye shosse

125993, Moscow, Russia

Anton V. Konovaltsev is the leading engineer of the Scientific Center for Special Radio-Electronic Systems and Management of the Moscow Aviation Institute – National Research University. He is the author of 27 scientific publications focusing on hardware and software used for processing signals in radio-electronic systems.

E-mail: anton_ko@mail.ru
Moscow Aviation Institute (National Research University)
4, Volokolamskoye shosse
125993, Moscow, Russia

Vladimir V. Rastorguev was the Deputy Director and Professor at the Scientific Center for Special Radio-Electronic Systems and Management of the Moscow Aviation Institute – National Research University. He obtained his Ph.D. from MAI in 1978. He was an expert in the area of optimal signal processing in radar and radio-navigation systems. He was also the author of more than 170 scientific publications.
Professor Vladimir V. Rastorguev passed away on December 10, 2018.

Vladimir M. Nuzhdin, D.Sc., works at the Scientific Center for Special Radio-Electronic Systems and Management of the Moscow Aviation Institute – National Research University. He obtained his Ph.D. from MAI in 1983. He is an expert in the area of radar and radio-navigation systems and the author of more than 75 scientific publications.
E-mail: kaf407@mai.ru
Moscow Aviation Institute (National Research University)
4, Volokolamskoye shosse
125993, Moscow, Russia

Pavel V. Sokolov is a researcher at the Scientific Center for Special Radio-Electronic Systems and Management of the Moscow Aviation Institute – National Research University. He is an expert in the area of signal processing software used in radio-electronic systems. He is also the author of 25 scientific papers.
E-mail: VerbludSlon@list.ru
Moscow Aviation Institute (National Research University)
4, Volokolamskoye shosse
125993, Moscow, Russia

Network Selection in Wireless Heterogeneous Networks: a Survey

Fayssal Bendaoud[1], Marwen Abdennebi[2], and Fedoua Didi[3]

[1] *LabRI-SBA Lab., Ecole supèrieure en Informatique, Sidi Bel Abbes, Algeria*
[2] *L2TI Lab, University of Paris 13, France*
[3] *LRIT Lab, University of Tlemcen, Algeria*

Abstract—Heterogeneous wireless networks is a term referring to networks combining different radio access technologies with the aim of establishing the best connection possible. In this case, users with multi-mode terminals can connect via different wireless technologies, such as 802.16, 802.11, UMTS, HSPA and LTE, all at the same time. The problem consists in the selection of the most suitable from all radio access technologies available. The decision process is called network selection, and depends on several parameters, such as quality of service, mobility, cost, energy, battery life, etc. Several methods and approaches have been proposed in this context, with their objective being to offer the best QoS to the users, and/or to maximize re-usability of the networks. This paper represents a survey of the network selection methods used. Multiple attribute-dependent decision-making methods are presented. Furthermore, the game theory concept is illustrated, the use of the fuzzy logic is presented, and the utility functions defining the network selection process are discussed.

Keywords—always best connected, fuzzy logic, game theory, multiple attribute decision making, quality of service, radio access technologies.

1. Introduction

Two decades ago, the IEEE dominated the Internet market with the IEEE 802.11 standard. It was a big innovation in terms of cost and high data throughputs. At that time, 3GPP's cellular technology was the Global System of Mobile Telecommunication (GSM). It is considered to be an excellent technology for calling and sending text messages, but it can't provide Internet access. This has pushed 3GPP developers to introduce the Global Packet Radio Service (GPRS). Unfortunately, Internet access with a higher data throughput was not possible either. In the next update, 3GPP changed the circuit-switched functionality to the packet-switched variety, and they called this release the Universal Mobile Telecommunications System (UMTS) or the third generation network (3G).

Before 3G networks arrived, radio access networks were mainly homogeneous. Later, development of network technologies has led to an impressive growth of Internet applications and services, as well as to the development of the mobile user industry. Presently, people are equipped with smartphones and seek "always the best connectivity"

(ABC). It is obvious that no single Radio Access Technology (RAT) can offer the ABC. Therefore, it was necessary to move from homogeneous systems to their heterogeneous counterparts. The aim of the fourth generation (4G) networks is to satisfy the ABC concept by offering mobile users the ability to take advantage of those networks which offer different architectures and performance levels. Nowadays, we have a variety of RATs: WLAN, IEEE 802.11, UMTS, HSPA and LTE. All of them make up a heterogeneous environment (Fig. 1). A heterogeneous system allows mobile users to choose multiple RATs based on several criteria. This choice is known as network selection (NS) and is the very area this paper is concerned with.
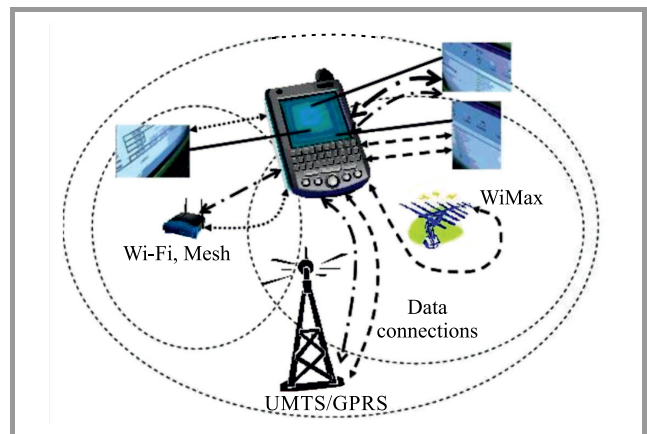
Fig. 1. Heterogeneous wireless environment.

The network selection procedure consists of selecting the best network from all those that are available. However, due to many parameters, such as cost, QoS and the amount of energy consumed, the decision is complex. This allows us to say that the NS problem may be defined, in basic terms, as a dynamic and automatic choice of the best wireless access network, taking into account a number of parameters. In the case of older cellular technologies, NS was based on physical layer parameters, and the mobile terminal was often assigned to the best received base station. Such a selection policy is obviously not suitable for heterogeneous wireless access technologies, because the user may favor to connect to a RAT with a lower load, located at a greater

distance and offering a lower level of the received signal strength (RSS), rather than to a more loaded cell with high RSS.

NS in a heterogeneous environment can be described as a multiple attribute decision making (MADM) problem, because of the number of parameters and criteria involved [1]–[4]. Other methods, such as fuzzy logic, game theory and utility functions have been proposed to solve the NS problem in [5], [6] and will be surveyed in this paper. Other methods pertaining to multi-criteria optimization have been used to deal with the NS problem as well. These include artificial intelligence, neural networks and genetic algorithms, and will not be in this paper because of their limited popularity.

This paper is organized as follows. After introduction, Section 2 is devoted to the NS procedure, while in Section 3 we focus on the approaches and methods used to solve the NS problem, as presented in literature. In Section 4, a summary of the discussed methods and approaches is presented, along with a recap table.

2. Network Selection Process

The NS process consists of switching between different RATs, to be always best served. So, when a multi-mode user discovers the existence of various RATs within its area, it should be able to select the best network in terms of delay, jitter, throughput and packet loss rate (Fig. 2). The NS procedure is the general case of the handover process (HO), which can be either centralized (network-centric) or decentralized (user-centric).

For the network-centric approach, the operator controls the whole process and makes decisions. It is a good strategy to avoid problems, such as selfish behavior of users who try to get the best RAT at the same time, which results in congestion. On the other hand, this strategy cannot be used in the case of multiple operators. For the user-centric approach, users make decisions by themselves. This approach is known as decentralized and can easily generate congestion because of the selfish nature of users. Nowadays, almost all operators offer 3G and 4G radio access and also Wi-Fi connections, so the first approach is more suitable for regular use.

Many parameters influence the process of selecting the best RAT: battery level, energy required to get the services requested, RSS received, cost, bandwidth acquired, user preferences, QoS required, etc. These parameters can be categorized into four groups:

- network conditions parameters – information about network conditions, such as network load, coverage area, network connection time, available bandwidth,

- application requirements parameters – information about the threshold needed by the service application to be in the normal state, as well as required throughput, delay, jitter, packet loss rate, and energy needed for the application,

- user preference parameters – information relative to end users, i.e. user acceptable cost, preference between cost and service quality,

- mobile equipment parameters – information about the user's device, i.e. battery level status and mobility.

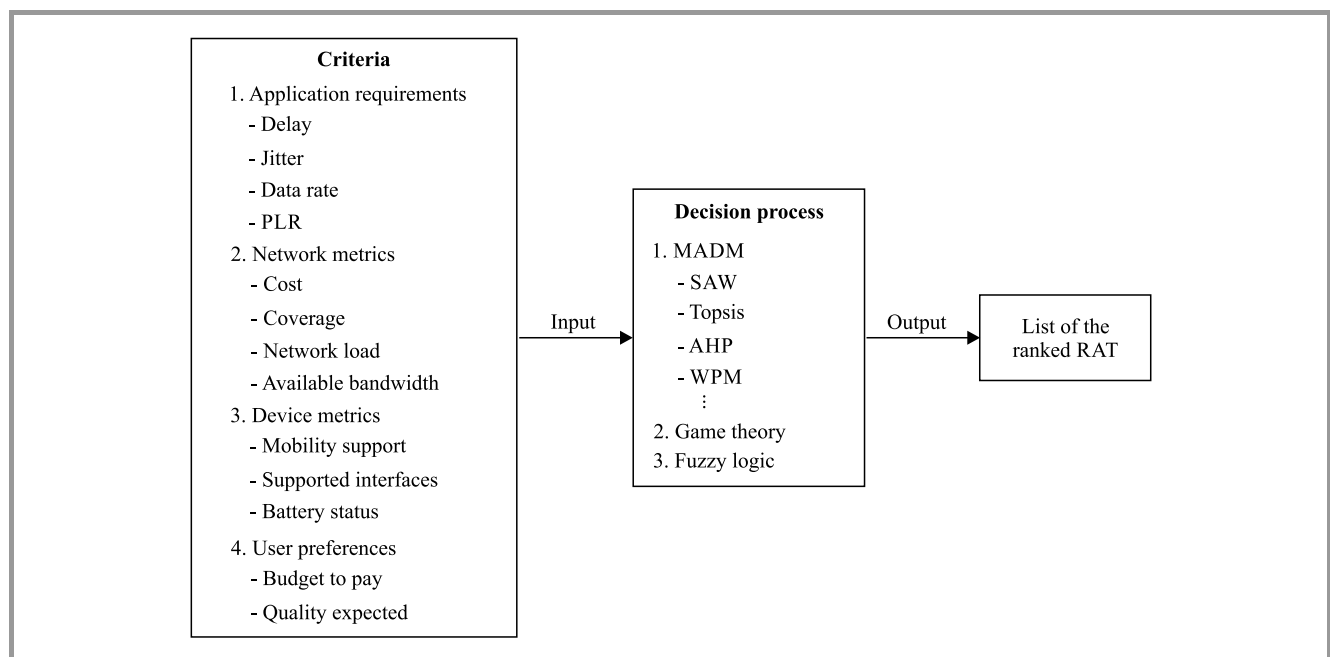The parameters may also be dynamic or static, and may require to be maximized or minimized. For example, the

Fig. 2. Network selection process.

Table 1
Network selection inputs and classification of parameters

Parameters group	Parameters	Type	Expected as
Network conditions	• network load	Dynamic	Minimized
	• network coverage	Static	Fixed
	• network connection time	Dynamic	Minimized
	• available bandwidth		Maximized
Application requirements	• throughput	Dynamic	Maximized
	• delay		
	• jitter		
	• PLR		
	• energy consumption		
User preferences	• budget	Static	Fixed
	• cost		
Mobile equipment	• battery level	Dynamic	Fixed
	• mobility		

delay parameter is related to network conditions, network load and user's RSS – it is a dynamic parameter that must be minimized (Table 1).

The network selection process consists mainly of the following actions:

- Monitoring step. It consists in identifying the available RAT, collecting the network's radio parameters and other RAT characteristics. In this stage, some of the parameters are estimated and others are calculated.

- Decision step. It initiates the NS decision. The choice of the best network is based on the monitoring process and on other parameters provided by the mobile device, such as user's preferences. In this stage, the decision process is applied to rank the RATs.

- Execution step. It consists in connecting to the target RAT.

The NS procedure is started when a new service is requested, such as a video/VoIP call or a data transfer service, also when the received RSS drops below the threshold value and after the user's radio connection worsens, for instance when the user is mobile. As far as the application requirements are concerned, NS depends on the type of service desired. For VoIP, delay and packet loss rate are important parameters. For a video service – bandwidth and delay, while for a best effort service – the bandwidth acquired.

3. Network Selection Solutions

Many authors have modeled the solution of the NS problem, presenting different propositions in order to find the most appropriate one. In this section, we present a survey of the methods used to solve the NS problem.

3.1. MADM Methods

MADM is an analytical approach focusing on preferential decisions. It treats problems with numerous decision-related criteria, and is widely used in various areas of expertise, such as economy [7]–[9]. The basics of this approach are divided into three groups:

- Alternatives – a set of the actors who will be ranked. In the NS scenario, the alternatives have the form of RAT lists.

- Set of attributes – it represents the parameters or the criteria used in the decision-making process. For the NS scenario, the parameters are the throughput achieved, jitter, packet loss and delay.

- Weights – the importance of a given parameter or the criteria relied upon in the decision process.

By using such a taxonomy, we get a decision matrix representing the system, where the columns are the criteria and the lines are the alternatives.

Several decision-making methods have been proposed in MADM context, such as: simple additive weight (SAW), technique to order preference by similarity to ideal solution (TOPSIS), weighted product model (WPM) and analytical hierarchy process (AHP) [8].

SAW, TOPSIS and WPM are also qualified as ranking methods that need other methods to weigh the criteria, while AHP relies on a process that generates the weights for the criteria.

It is important to note that these methods are applicable only when all data of the input matrix are expressed with

the use of the same unit. Therefore, data must be normalized, which is an important step in the network selection procedure. Table 2 represents a non-exhaustive list of common normalization methods.

Table 2
Normalization methods

Method	Value
Max-min	$e_{ij} = \dfrac{x_{ij} - \min(x_{ij})}{\max(x_{ij}) - \min(x_{ij})}$
Sum	$e_{ij} = \dfrac{x_{ij}}{\sum\limits_{i=1}^{n} x_{ij}}$
Square root	$e_{ij} = \dfrac{x_{ij}}{\sqrt{\sum\limits_{i=1}^{n} x_{ij}}}$

The weight that a given parameter has for the use is another important point, e.g. QoS-based, cost-based and/or energy-based. Weights are related to the user profile and can be subjective or objective. Subjective weights are empirical values based on experience. For example, in the case of QoS-based users, while initiating a VoIP session, such parameters as delay and packet loss ratio have 60 to 70% of importance and bandwidth is not as important. In the case of a video session, bandwidth is more important than other parameters (50% of importance). Objective weights are given by the formulas shown in Table 3.

Table 3
Weighting methods

Method	Value
Entropy	$w_j = 1 - \dfrac{1}{N} \sum\limits_{i=1}^{n} \left[x_{ij} \ln(x_{ij}) \right]$
Variance	$w_j = \dfrac{1}{N} \sum\limits_{i=1}^{n} x_{ij}$
Eigenvector	$w(B - \lambda I) = 0$

The AHP method contains an auto-creative system to generate the weight vector using the eigenvectors and eigenvalues of the input matrix.

3.1.1. Simple Additive Weight (SAW)

SAW is a method for the case of multiple-criteria systems [8], [10], [11]. In SAW, the first data is normalized, then the candidate having the highest/lowest value is selected:

$$R_{SAW} = \sum_{i=1}^{n} \left(w_j \times r_{ij} \right) , \qquad (1)$$

where R_{SAW} is the value of each candidate, w_j is the weight value of the parameter j and r_{ij} is the normalized value of parameter j and network i.

The SAW method has been widely used in the context of network selection. In [11] and [12], authors have used the SAW method to get a ranked list of networks, while in [8], authors made a mix between the game theory and the SAW method. When the NS problem is approached by using the SAW method and other variants, the main benefit of the SAW method resides in its simplicity and low complexity. However, its drawback is that one parameter can be outweighed by another one.

3.1.2. Technique for Order Preference by Similarity to Ideal Solution (TOPSIS)

TOPSIS is an aggregating compensatory method based on the concept that the chosen solution should have the shortest geometric distance from the positive ideal solution [13] and the longest geometric distance from the negative ideal solution. The normalized data for each parameter are weighted and therefore the geometric distance between each alternative and the ideal alternative is computed. The TOPSIS process is carried out as follows.

- First, an evaluation matrix consisting of m alternatives and n criteria is created, with the intersection of each alternative and criterion given as x_{ij}. It results in a matrix $(x_{ij})_{m \times n}$.

- The matrix $(x_{ij})_{m \times n}$ is then normalized to get $(r_{ij})_{m \times n}$ using one of the methods from Table 2.

- Calculate the weighted normalized decision matrix where $t_{ij} = w_j \times (r_{ij})_{m \times n}$, $w_j = \frac{W_j}{\sum_{i=1}^{n}(W_j)}$, $\sum_{i=1}^{n}(W_j) = 1$ and $j = 1, \ldots, m$, $i = 1, \ldots, n$.

- Determine the best and worst alternative A_b and A_w, respectively.

$$A_b = b_j^P = \max(t_{ij}) , \; j \in J_+ ,$$

$$A_w = w_j^P = \min(t_{ij}) , \; j \in J_- ,$$

where J_+ and J_- contain the criteria with positive and negative impact respectively.

- Calculate the separation measure for each alternative:

$$D_P = \sqrt{\sum_{i=1}^{n} \left(w_j^2 \times (r_{ij} - b_j^P)^2 \right)} ,$$

$$D_N = \sqrt{\sum_{i=1}^{n} \left(w_j^2 \times (r_{ij} - a_j^N)^2 \right)} .$$

- Calculate the relative closeness to the ideal solution

$$R_{TOPSIS} = \frac{D_P}{D_P + D_N} . \qquad (2)$$

TOPSIS has been applied to the network selection problem in several works, e.g. [9], [11], [14]. In [11], authors compare the performance of vertical handover using SAW

and TOPSIS. They concluded that TOPSIS outperforms the SAW method. In general, TOPSIS and other compensatory methods managed to avoid the problem that a parameter can be outperformed by another one by allowing a trade-off between criteria. This means that a poor value of one criterion can be neglected by a good value in another. This offers a huge benefit and is more sensible than non-compensatory methods, which use a threshold system.

3.1.3. Weighted Product Model (WPM)

WPM, also known as multiplicative exponential weighting (MEW), is a method similar to SAW [15]. The difference consists in the replacement of the addition operation used in the SAW method, with multiplication. Each alternative decision is compared with the remaining ones by multiplying a number of ratios, one for each decision criterion. Each ratio is raised to the power equivalent to the relative weight of the corresponding criterion:

$$R_{WPM} = \prod_{i=1}^{n} (r_{ij})_j^w .$$ (3)

Authors in [16] made a comparison between SAW and WPM methods in the context of vertical handover. They used the relative standard deviation as a metric of comparison and they arrived at a conclusion that WPM is better than SAW. In [17], the WPM method was used in the context of heterogeneous systems. Their conclusion is that the WPM method is a more robust approach for dynamic decision making and it penalizes the attributes with poor quality.

3.1.4. Analytical Hierarchy Process (AHP) and Grey Relational Analysis (GRA)

AHP assumes that one complicated problem is decomposed into a multiple-hierarchy simple sub-problems. AHP steps are:

- decomposing the problem into a hierarchy of sub-problems, where the top node is the final goal and where alternatives are listed for each criterion,

- pair-wise comparison of attributes and translating them into numerical values from 1 to 9,

- calculating the weights of each level of the hierarchy,

- synthesizing weights and getting overall weights.

As far as the GRA method is concerned, it is used to rank the candidate networks and it involves the following steps:

- normalization of data is performed considering three situations: higher is better through lower is better, and nominal is the desired,

- definition of the ideal sequence in the three situations considered is set: the ideal sequence contains the higher bound, lower bound and moderate bound,

- computing the grey relational coefficient (GRC): the sequence in which GRC is larger is more favorable.

The AHP is usually coupled with the GRA method: AHP for weighting, and GRA for ranking alternatives. Authors in [18] used a modified version of AHP and compared it with normal AHP using the QoE criterion. Their numerical results show that the proposed scheme outperforms the conventional AHP scheme, resulting in a good load balance. In [19], authors relied on AHP to rank various criteria used to compare the desirability of different Internet advertising networks. The proposed model provides an objective and effective decision model to be used by advertisers in selecting an Internet advertising network.

To recapitulate: MADM methods are widely used to solve the network selection problem, this is due to the fact that network selection involves the same problems as are solved by MADM. Moreover, these methods are known for their ease of use, clarity and low complexity of computation. The disadvantages of these methods are listed below.

Firstly, these methods do not offer the same level of performance with respect to different services (VoIP, video calls and web browsing) (Table 4). Secondly, they suffer from the problem of ranking abnormality, i.e. the phenomenon occurs in the MADM methods when an exact replica or a copy of an alternative is introduced or eliminated.

Authors in [20] have shown that the rank reversal problem occurs in the majority of well-known MADM methods. This problem has been addressed in other works [21], [22], by introducing specific modifications, but the original versions of MADM methods suffer from the rank reversal problem. Additionally, the AHP method is very complicated and requires complex computations when calculating the vector of weights. Due to all these reasons, we can say that MADM are a good solution, but the lack of a general method that would be suitable for all kinds of services is a problem.

Table 4
MADM method pros and cons

Advantages	Disadvantages
Easy to understand	Rank reversal phenomena
Easy to implement	High complexity such as AHP
Good results in some cases	Good performance with some services and bad results for others

3.2. Game Theory

Game theory allows to model competitive situations, which implies an interaction between rational decision makers and mutual, and possibly conflicting interests [9], [11], [12]. It provides an analytical tool to predict the outcome of complicated interactions between rational conflicting entities.

In this paper, we focus on the use of game theory for modeling the network selection problem [23].

An example of a conflicting scenario in which a game theory solves the problem is described below. As a rule, the fundamental characteristic of a game is that the gain of a player depends on his choices and also on the choices made by the other players. Games of this type are known as strategic games.

A game is represented mainly by three sets: a set of players that contains the rational actor competing to get a bigger pay-off. A set of actions or strategies which depends on the information available in the system. Obviously, each player seeks the action which maximizes their revenue. The pay-off is the objective function representing the player's revenue when choosing a specific strategy. The pay-off for each player can be represented as the actual or expected benefit the player receives by playing the current strategy. The game is played until the player is capable of obtaining more gains. When the pay-off cannot be enhanced any further with any other strategy combinations, an equilibrium known as the Nash equilibrium is reached.

The Nash equilibrium occurs basically when no player can obtain more gain by changing their strategy, with the strategies of the other players remaining unchanged. The Nash equilibrium is a combination of the best strategies for all players. A detailed representation of a strategic game is:

- $Game = \{P, A, G\}$;

- $P = \{1, n\}$ the set of players;

- $A = \{a_1, a_n\}$ the set of actions, denoting the set of strategies available to the player i, $1 \leq i \leq n$;

- $G : \{payoff\}$ represents the reward achieved when choosing a strategy. Here, to simplify things, the pay-off function is a linear summation of local gains, with the weights of each parameter applied.

Different types of games are used to model various cooperative or competitive situations between rational decision makers. Some of the most widely used game theory models are outlined below.

3.2.1. Cooperative and Non-cooperative Games

A cooperative game is a situation in which players claim to display cooperative behaviors. In this situation, the players plan, in groups, to choose their actions. In a non-cooperative game, also known as a competitive game, all options available to the players are specified, while contracts underlying the coalitions in a cooperative game are not described. Each player tries to reach his goal without regarding the other players. Here, the players are called rational [24]. Generally, non-cooperative games admit a solution called the Nash equilibrium, while for the cooperative games the solution is a total surplus generated by the coalition of players.

3.2.2. Games with Complete/Perfect Information

Complete information is a term used in game theory to describe games in which knowledge about other players is available to all participants. Every player knows the pay-off and the strategies available to other players. Games with complete and perfect information are significantly different. In a game with complete information, the structure of the game and the pay-off functions of the players are commonly known, but players may not see all of the moves made by other players. For the perfect information games, each player is perfectly informed of all the events that have previously occurred, but may lack some information about the pay-off of others players or on the structure of the game. Inversely, in games with incomplete information, some players do not know the information of other participants, like other players' pay-off. In a game with imperfect information, players are simply unaware of the actions chosen by others. However, they know who the other players are, their possible strategies and their preferences. Hence, the data about the other players is imperfect, but complete [23], [24].

3.2.3. Repeated Games

In strategic games, players make their decisions simultaneously at the beginning of the game. On the contrary, the extensive game model defines the possible orders of events. In this case, players can make decisions during the game and they can react to others decisions. Extensive games can be finite or infinite. Repeated games are a variety of extensive games in which a game is played for a number times and the players can observe the outcome of the previous game before attending the next iteration [24], [25].

3.2.4. Zero-sum Games

A situation in which one person's gain is equivalent to another's loss, so the net change in wealth or benefit is zero. A zero-sum game may have two or more players. They are usually called strictly competitive games in the game theory discipline, but are less widespread than non-zero sum games [25].

3.3. Game Theory on Network Selection

In this subsection, we will discuss works that use the game theory to solve the network selection problem. As we mentioned above, a game is defined by three sets: players, actions and payoff. In the network case, players can be users, networks or both. A distinction between those categories will be made below.

3.3.1. Game between Users

In [5], authors modeled a competition between users for one access point as an evolutionary game. Users represent players that compete to maximize the transmission rate. The latter represents the strategy, hence the pay-off is modeled by an objective function. This function takes the delay

and packet loss rate and determines the mean opinion score MOS, which is a measure for voice quality. In [5], the notion of free users is used. VoIP is the only service used and the authors claim that the equilibrium reached is optimal.

In [26], the problem of the least congested access point selection is emphasized. The set of strategies contains different access points available and the pay-off is a trade-off between bandwidth gained and the effort generated when moving to the new access point. The authors demonstrate that the result of this game is the distribution of users on the access points.

In [27] authors consider the scenario with a single Wi-Fi network and multiple access points. In this case study, users can choose one access point to connect to, the pay-off function depends on the congestion level of the access point and cost. The authors affirm that the Nash equilibrium is reached.

The works mentioned above suggest that a game between users means that players seek, in general, non-cooperative behavior and because of the selfish nature of users, these games lead to the situation of congestion, and also to the monopolization of resources by certain users.

3.3.2. Game between Networks

When players represent the networks, the latter compete to get the attention of users and maximize the number of users connected, to achieve more revenue. In [28], authors present a non-cooperative game mechanism between networks in which the players compete for a service requirement and try to gain the maximum of an access point. This represents user satisfaction. The problem of this scheme is that the preferences toward players are the same. In [29] authors have introduced the strategy space and quality point concepts, the players are the networks and the pay-off function determines which access network will provide the service requested by the user, which corresponds to the distribution of service requests amongst the networks.

In [30] researchers have investigated the admission control problem by modeling a multi-round game between two Wi-Fi networks. Here, the players in the game are the two networks, and the strategy set is the user's service requests. The pay-off of the game is the distribution of the service requests between the competitive networks.

The papers referred to above prove that games between networks are characterized by one well-known strategy, i.e. by seeking the maximum revenue or maximizing the number of users connected to the network. This approach guides users to think about their corresponding network selection schemes under this network in a competitive environment.

3.3.3. Game between Users and Networks

In this case, players act as mobile users and/or networks. In [31] authors propose a reputation-based network selection mechanism by modeling the interaction between users and networks using a repeated prisoner's dilemma game. To reinforce the cooperation between users and networks,

authors combine the reputation-based systems and game theory. The network reputation factor represents the network's past behavior in the network selection decision. Researchers show that using reputation is essential in the case of cooperation and that repeated interaction maintains cooperation.

In [32] the network selection problem having the form of a non-cooperative auction game is modeled, in which buyers represent users, sellers/bidders are the available network operators and the auction component is the requested bandwidth with its associated attributes. The auction that maximizes the user utility is the winning bid.

In [33] a non-cooperative resource allocation based on the Cournot game between a provider and his customers is presented, where users are classified into three classes: premium, gold and silver. The strategies for the provider and the customer are as follows. The provider seeks customers who bring high revenue, while if the customers are not satisfied from the received QoS, they can decide to leave the network. Users are accepted into the network if the new benefit computed when a new customer arrives is less than the provider's benefit value. Finally, the authors identify the equilibrium for resource distribution. This game type can be summarized in the following manner: users compete against networks, each seeking to maximize their own benefits. On the one hand, users try to maximize their benefits (cost- and/or QoS-related). On the other hand, the networks try to maximize the profit for the services provided.

3.4. Fuzzy Logic

In fuzzy logic, there are few degrees of satisfaction of a condition [34]. Unlike in Boolean algebra, where a proposal is considered to be true or false, fuzzy logic adds a degree of truth to choose from the $0\dots1$ range. It is a tool of artificial intelligence used in various fields [35]. The concept is based on the theory of fuzzy sets with an extension of the classical theory. Fuzzy logic brings the concept of partial truth, where the truth value may vary from completely true to completely false.

3.4.1. Fuzzy Logic in Network Selection

A few studies have addressed the network selection problem using fuzzy logic as a core of the ranking scheme. Basically, authors use fuzzy logic in network selection in two ways: as a combination of MADM and fuzzy logic, or they use it as the selection scheme.

In [6] authors propose a general scheme to solve the multi-criteria network selection problem. In the proposed scheme, the multi-criteria network selection solution is obtained by considering the users' requirements and QoS. The proposed scheme is scalable and is capable of handling any number of RATs with a large set of criteria. The simulation results show that the proposed solution has a better and more robust performance over the reference solutions. In [36] researchers describe two novel, fuzzy logic based ranking schemes. These schemes enable users to evaluate the correctness of different combinations of P2P-based grid

networks. A fixed set of commonly used attributes is used, such as cost, capacity and reliability. The proposed ranking algorithm is based on an intuitive rule optimization design applying Boolean logic to capture input combinations.

In [37] authors propose a fusion method-based fuzzy logic approach for different network schemes. The main advantage is the consideration of the relative importance of different networks. The authors show that the proposed scheme significantly improves the generalization capability.

Most of the recent works using the fuzzy logic are combined with MADM methods [38], [39]. Generally, in the field of network selection, the use of fuzzy logic as a core of the ranking scheme is not widely adopted. Instead, fuzzy logic has always been combined with MADM.

3.5. Utility Functions

Utility is the "satisfaction" we get from using, owning or doing something. It is what allows us to choose between options.

A preference function utility assigns values to the ranking of a set of choices. This is useful in analyzing consumer behavior in the maximization problem. Faced with a set of options and a budget constraint, we will choose what satisfies them to the highest degree. Utility functions are often expressed as $U(x1, x2, x3)$, which means that U (utility) is a function of the quantities of $x1, x2$ and so on. In the case of monotonic functions, if A is a set of goods, and $A > B$, then $U(A) > U(B)$. That is, if A is preferred to B.

For making a decision, utility refers to the level of satisfaction that goods or a service provide to the decision maker [40]. Utility function is an associated term which relates to the utility derived by a consumer from goods or a service. Different consumers with various user preferences will have different utility values for the same product. Thus, the individual preferences should be taken into account in the utility evaluation.

In the paper [41] authors show that many of the commonly used MADM algorithms, such as SAW, WPM and TOP-SIS, in their standard form, are not best suited because of their assumptions concerned with a monotonous increase or decrease of the attributes' utilities. They affirm that both monotonic and non-monotonic utilities can be taken into consideration, and are therefore better suited for achieving this type of optimization objectives.

In [42] researchers proposed a user-centric RAN selection strategy based on maximizing consumer surplus, subject to meeting user-defined constraints in terms of transfer completion time. An exploration of a number of possible utility functions based on different user's attitudes to risk is presented. They affirm that simulations produced results that correspond to the user utility descriptions input. The risk taker ends up paying more, but enjoying less delay.

In [43] researchers proposed a user-centric RAN selection strategy based on maximizing consumer surplus subject to meeting user-defined constraints in terms of transfer completion time. An exploration of a number of possible utility functions based on different user's attitudes to risk is presented. They affirm that simulations produced results that correspond to the in-putted user utility descriptions. The risk taker ends up paying more, but enjoying less delay.

In [44] the researchers have proposed a method called SU-TIL, which is a mechanism for network selection in the context of next generation networks. It prioritizes networks with higher relevance to the application and lower energy consumption, enabling full and seamless connectivity with mobile devices and applications. They even propose some future works, such as conducting an investigation to address the interaction between multiple instances of SUTIL and consideration of the highly dynamic nature intrinsic of the environments where SUTIL operates.

4. Analysis and Discussion

Figure 3 presents the approaches described in this paper. At the top of the flow chart is the NS problem, in the second row different approaches and their categories, namely MADM, game theory, fuzzy logic and utility functions are placed. Each category has its own specific methods
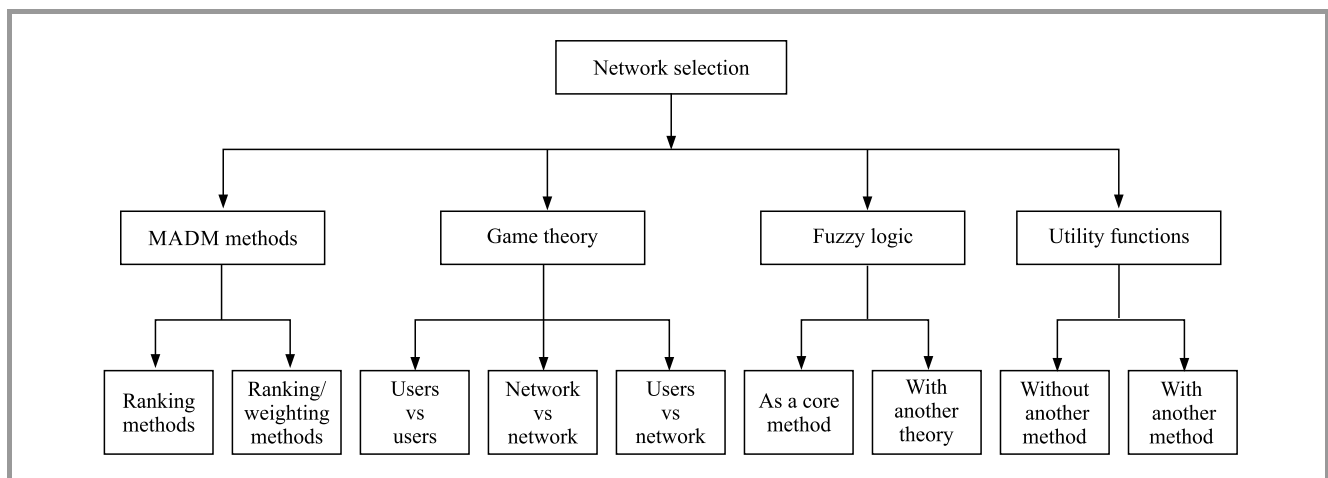

Fig. 3. Summary of NS methods used.

Table 5
Summary of the discussed methods

Type	Advantages	Disadvantages
Ranking methods	Easy to understand and use Relatively good results Low complexity	Rank reversal phenomenon Lack of a weighting process Different behavior with various applications
Weighting/rating methods	Weighting process	Rank reversal phenomenon High complexity
Users vs. users Users vs. networks Networks vs. networks	No rank reversal phenomenon	Selfish behavior of users lead to the congestion
As a core method With another approach	Overcome the MADM's drawbacks	Few works in the literature Time consuming due to coupling two approaches
Without other methods With another approach	Solving simple problems Powerful and used in many works Overcome the MADM's drawbacks	Few works in the literature

(if such exist), which are regrouped in the bottom row and are summarized in Table 5:

- Ranking methods. TOPSIS, SAW, WPM and others – methods which require another method to obtain the weight vector. Such methods suffer from a lot of problems, such as the rank reversal phenomenon, and the influence of a bad value of a criterion on the good value of another criterion.

- Ranking/weighting methods. These methods suffer from high complexity and rank reversal, which can occur here as well, because it is a problem affecting the entire MADM approach.

- Game theory. Game theory is a good tool to model the network selection problem. The main concern is the computation time, due to the relatively high complexity degree. This issue may be avoided in the case of a game between networks.

- Fuzzy logic. It has been adapted to solve the network selection problem but, generally, it is not used alone, it needs to be accompanied by MADM methods or genetic algorithms.

- Utility functions. The utility function is used with other methods, such as fuzzy logic or MADM, and it can also be used in a unilateral way.

References

[1] M. Lahby, S. Baghla, and A. Sekkaki, "Survey and comparison of MADM methods for network selection access in heterogeneous networks", in *Proc. 7th Int. Conf. on New Technol., Mobil. and Secur. NTMS 2015*, Paris, France, 2015, pp. 1–6 (doi: 10.1109/NTMS.2015.7266522).

[2] S. Scherzer and T. Scherzer, "Method and system for selecting a wireless network for offloading", US Patent 9,148,843, Sept. 29, 2015.

[3] S. K. Kovvali, C. Boyle, R. Ghai, J. S. Amand, and G. Alden, "Content and ran aware network selection in multiple wireless access and small-cell overlay wireless access networks", US Patent 9,001,682, Apr. 7, 2015.

[4] J. Wu, B. Cheng, C. Yuen, Y. Shang, and J. Chen, "Distortion-aware concurrent multipath transfer for mobile video streaming in heterogeneous wireless networks", *IEEE Trans. on Mob. Comput.*, vol. 14, no. 4, pp. 688–701, 2015 (doi: 10.1109/TMC.2014.2334592).

[5] E. H. Watanabe, D. S. Menasché, E. De Souza e Silva, and R. M. Leao, "Modeling resource sharing dynamics of VoIP users over a WLAN using a game-theoretic approach", in *Proc. 27th Conf. on Comp. Commun. INFOCOM 2008*, Phoenix, AZ, USA, 2008 (doi: 10.1109/INFOCOM.2008.144).

[6] M. Alkhawlani and A. Ayesh, "Access network selection based on fuzzy logic and genetic algorithms", *Adv. in Artif. Intell.*, vol. 8, no. 1, p. 1–12, 2008 (doi: 10.1155/2008/793058).

[7] F. Bendaoud, F. Didi, and M. Abdennebi, "A modified-SAW for network selection in heterogeneous wireless networks", *ECTI Trans. on Elec. Engin., Electron., and Commun.*, vol. 15, no. 2, pp. 8–17, 2017.

[8] Y. K. Salih, O. H. See, R. W. Ibrahim, S. Yussof, and A. Iqbal, "A user-centric game selection model based on user preferences for the selection of the best heterogeneous wireless network", *Annals of Telecommunications*, vol. 70, no. 5-6, pp. 239–248, 2015 (doi: 10.1007/s12243-014-0443-6).

[9] A. Sgora, D. D. Vergados, and P. Chatzimisios, "An access network selection algorithm for heterogeneous wireless environments", in *Proc. of the IEEE Symp. on Comput. and Commun. ISCC 2010*, Riccione, Italy, 2010, pp. 890–892 (doi: 10.1109/ISCC.2010.5546556).

[10] L. Abdullah and C. R. Adawiyah, "Simple additive weighting methods of multi criteria decision making and applications: A decade review", *Int. J. of Inform. Process. and Manag.*, vol. 5, no. 1, p. 39–49, 2014.

[11] K. Savitha and C. Chandrasekar, "Trusted network selection using SAW and TOPSIS algorithms for heterogeneous wireless networks", *Int. J. of Comp. Appl.*, vol. 26, no. 8, pp. 22–29, 2011 (doi: 10.5120/3125-4300).

[12] Q.-T. Nguyen-Vuong, Y. Ghamri-Doudane, and N. Agoulmine, "On utility models for access network selection in wireless heterogeneous networks", in *Proc. IEEE Netw. Operations and Manag. Symp. NOMS 2008*, Salvador, Bahia, Brazil, 2008, pp. 144–151 (doi: 10.1109/NOMS.2008.4575128).

[13] D. L. Olson, "Comparison of weights in TOPSIS models", *Mathem. and Comp. Modelling*, vol. 40, no. 7–8, pp. 721–727, 2004 (doi: 10.1016/j.mcm.2004.10.003).

[14] B. Bakmaz, Z. Bojkovic, and M. Bakmaz, "Network selection algorithm for heterogeneous wireless environment", in *Proc. of IEEE 18th Int. Symp. on Pers., Indoor and Mob. Radio Commun. PIMRC 2007*, Athens, Greece, 2007, pp. 1–4 (doi: 10.1109/PIMRC.2007.4394315).

[15] M. Lahby, L. Cherkaoui, and A. Adib, "An intelligent network selection strategy based on MADM methods in heterogeneous networks", *Int. J. of Wirel. & Mob. Netw. (IJWMN)*, vol. 4, no. 1, pp. 83–96, 2012 (doi: 10.5121/ijwmn.2012.4106).

[16] K. Savitha and C. Chandrasekar, "Vertical handover decision schemes using SAW and WPM for network selection in heterogeneous wireless networks", *arXiv preprint arXiv:1109.4490*, 2011.

[17] P. TalebiFard and V. C. Leung, "Context-aware mobility management in heterogeneous network environments", *J. of Wirel. Mob. Netw., Ubiquitous Comput., and Dependable Appl. (JoWUA)*, vol. 2, no. 2, pp. 19–32, 2011 (doi: 10.22667/JOWUA.2011.06.31.019).

[18] D.-M. Shen, H. Tian, and L. Sun, "The QoE-oriented heterogeneous network selection based on fuzzy AHP methodology", in *Proc. of the 4th Int. Conf. on Mob. Ubiquit. Comput., Systems, Serv. and Technol. UBICOMM 2010*, Florence, Italy, 2010, pp. 275–280.

[19] C.-T. Lin and P.-F. Hsu, "Adopting an analytic hierarchy process to select internet advertising networks", *Marketing Intell. & Plann.*, vol. 21, no. 3, pp. 183–191, 2003 (doi: 10.1108/02634500310474993).

[20] Y.-M. Wang and Y. Luo, "On rank reversal in decision analysis", *Mathem. and Comp. Modell.*, vol. 49, no. 5–6, pp. 1221–1229, 2009 (doi: 10.1016/j.mcm.2008.06.019).

[21] Á. Huszák and S. N. Imre, "Eliminating rank reversal phenomenon in GRA-based network selection method", in *Proc. IEEE Int. Conf. on Commun. ICC 2010*, Cape Town, South Africa, 2010, pp. 1–6 (doi: 10.1109/ICC.2010.5502475).

[22] Y. B. Shin, S. Lee, S. G. Chun, and D. Chung, "A critical review of popular multi-criteria decision making methodologies", *Issues in Inform. Syst.*, vol. 14, no. 1, pp. 358–365, 2013.

[23] F. Bendaoud, M. Abdennebi, and F. Didi, "Network selection using game theory", in *Proc. 3rd Int. Conf. on Control, Engin. & Inform. Technol. CEIT 2015*, Tlemcen, Algeria, 2015, pp. 1–6 (doi: 10.1109/CEIT.2015.7233014).

[24] M. J. Osborne and A. Rubinstein, *A Course in Game Theory*. MIT Press, 1994 (ISBN: 9780262650403).

[25] D. E. Charilas and A. D. Panagopoulos, "A survey on game theory applications in wireless networks", *Comp. Netw.*, vol. 54, no. 18, pp. 3421–3430, 2010 (doi: 10.1016/j.comnet.2010.06.020).

[26] K. Mittal, E. M. Belding, and S. Suri, "A game-theoretic analysis of wireless access point selection by mobile users", *J. of Comp. Commun.*, vol. 31, no. 10, pp. 2049–2062, 2008 (doi: 10.1016/j.comcom.2008.01.025).

[27] M. Cesana, N. Gatti, and I. Malanchini, "Game theoretic analysis of wireless access network selection: models, inefficiency bounds, and algorithms", in *Proc. of the 3rd Int. Conf. on Perform. Eval. Methodol. and Tools*, Athens, Greece, 2008, Article no. 6 (doi: 10.4108/ICST.VALUETOOLS2008.4341).

[28] J. Antoniou and A. Pitsillides, "4G converged environment: Modeling network selection as a game", in *Proc. 16th IST Mob. and Wirel. Commun. Summit*, Budapest, Hungary, 2007, pp. 1–5 (doi: 10.1109/ISTMWC.2007.4299242).

[29] M. S. Z. Khan, S. Alam, and M. R. H. Khan, "A network selection mechanism for fourth generation communication networks", *J. of Adv. in Inform. Technol.*, vol. 1, no. 4, pp. 189–196, 2010 (doi: 10.4304/jait.1.4.189-196).

[30] D. E. Charilas, O. I. Markaki, and P. T. Vlacheas, "Admission control as a non-cooperative multi-stage game between wireless networks", in *Proc. 16th Int. Conf. on Sys., Sig. and Image Process. IWSSIP 2009*, Chalkida, Greece, 2009, pp. 1–5 (doi: 10.1109/IWSSIP.2009.5367804).

[31] R. Trestian, O. Ormond, and G.-M. Muntean, "Reputation-based network selection mechanism using game theory", *Phys. Commun.*, vol. 4, no. 3, pp. 156–171, 2011 (doi: 10.1016/j.phycom.2011.06.004).

[32] M. A. Khan, U. Toseef, S. Marx, and C. Goerg, "Game-theory based user centric network selection with media independent handover services and flow management", in *Proc. 8th Ann. Commun. Netw. and Serv. Res. Conf. CNSR 2010*, pp. 248–255, Montreal, QC, Canada, 2010 (doi: 10.1109/CNSR.2010.40).

[33] M. Chatterjee, H. Lin, and S. K. Das, "Non-cooperative games for service differentiation in CDMA systems", *Mob. Netw. and Appl.*, vol. 10, no. 6, pp. 939–946, 2005 (doi: 10.1007/s11036-005-4450-8).

[34] L. A. Zadeh, "Fuzzy sets and information granularity", in *Advances in Fuzzy Set Theory and Applications*, M. M. Gupta, R. K. Ragade, and R. R. Yager, Eds. Amsterdam: North-Holland, 1979, pp. 3–18 (ISBN: 9780444853721).

[35] J. Godjevac, *Idées nettes sur la logique floue*. Presses Polytechniques et Universitaires Romandes (PPUR), 1999 (ISBN: 9782880743789) [in French].

[36] S. Kher, A. K. Somani, and R. Gupta, "Network selection using fuzzy logic", in *Proc. 2nd Int. Conf. on Broadband Netw. BroadNets 2005*, Boston, MA, USA, 2005, pp. 876–885 (doi: 10.1109/ICBN.2005.1589698).

[37] S.-B. Cho and J. H. Kim, "Multiple network fusion using fuzzy logic", *IEEE Trans. on Neural Netw.*, vol. 6, no. 2, pp. 497–501, 1995 (doi: 10.1109/72.363487).

[38] C. Kahraman, U. Cebeci, and Z. Ulukan, "Multi-criteria supplier selection using fuzzy AHP", *Logist. Inform. Manag.*, vol. 16, no. 6, pp. 382–394, 2003 (doi: 10.1108/09576050310503367).

[39] Y.-M. Wang and T. M. Elhag, "Fuzzy TOPSIS method based on alpha level sets with an application to bridge risk assessment", *Expert Syst. with Appl.*, vol. 31, no. 2, pp. 309–319, 2006 (doi: 10.1016/j.eswa.2005.09.040).

[40] P. C. Fishburn, *Utility Theory for Decision Making*. Wiley, 1970 [Online]. Available: http://www.dtic.mil/dtic/tr/fulltext/u2/708563.pdf

[41] F. Bari and V. C. M. Leung, "Use of non-monotonic utility in multi-attribute network selection", in *Wireless Technology*, S. Powell and J. P. Shim, Eds. *LNEE*, vol. 44, pp. 21–39. Springer, 2009.

[42] M. A. Senouci, S. Hoceini, and A. Mellouk, "Utility function-based TOPSIS for network interface selection in heterogeneous wireless networks", in *Proc. IEEE International Conference on Communications ICC 2016*, Kuala Lumpur, Malaysia, 2016, pp. 1–6 (doi: 10.1109/ICC.2016.7511563).

[43] O. Ormond, J. Murphy, and G.-M. Muntean, "Utility-based intelligent network selection in beyond 3G systems", in *Proc. IEEE Int. Conf. on Commun. ICC 2006*, Istanbul, Turkey, 2006, vol. 4, pp. 1831–1836 (doi: 10.1109/ICC.2006.254986).

[44] L. Pirmez, J. C. Carvalho Jr, F. C. Delicato, F. Protti, L. F. Carmo, P. F. Pires, and M. Pirmez, "Sutil–network selection based on utility function and integer linear programming", *Comp. Netw.*, vol. 54, no. 13, pp. 2117–2136, 2010 (doi: 10.1016/j.comnet.2010.03.007).

Fayssal Bendaoud received his M.Sc. degree in Networks and Distributed Systems from University of Tlemcen, Algeria. In 2017, he received his Ph.D. in networks and services from University of Tlemcen. Now his is an Assistant Professor at Ecole Superieure en Informatique ESI-SBA and a member of LabRI-SBA Laboratory.

E-mail: f.bendaoud@esi-sba.dz
LabRI-SBA Lab.
Ecole supèrieure en Informatique
Sidi Bel Abbes, Algeria

Marwen Abdennebi is a researcher at the L2TI Laboratory and Associate Professor at the Department of Physics, Paris 13 University. His research interests include: radio resource management, cellular, mobile networks, 2G, 3G and B3G air interfaces, broadband radio networks, MAC layer, as well as validation, analytical models and simulation tools.
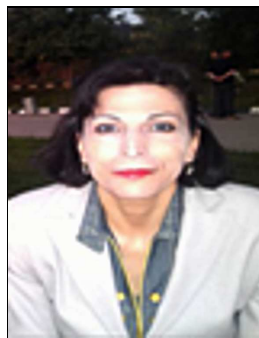
E-mail: marwen.abdennebi@univ-paris13.fr
L2TI Lab
Paris 13 University
Paris, France

Fedoua Didi obtained her Ph.D. in Computer Science from University Abou Bekr Belkaid, Algeria in 2009. She is currently the Director Chief of Laboratoire de Recherche en Informatique de Tlemcen (LRIT). She is a Professor at University Abou Bekr Belkaid.

E-mail: fedouadidi@yahoo.fr
LRIT Lab
University of Tlemcen
Tlemcen, Algeria

Fateh Bouguerra and Lamir Saidi

LAAAS Laboratory, Department of Electronics, University of Batna 2, Batna, Algeria

Abstract—High order modulation (HOM) presents a key challenge in increasing spectrum efficiency in 4G and upcoming 5G communication systems. In this paper, two non-linear adaptive equalizer techniques based on multilayer perceptron (MLP) and radial basis function (RBF) are designed and applied on HOM to optimize its performance despite its high sensitivity to noise and channel distortions. The artificial neural network's (ANN) adaptive equalizer architectures and learning methods are simplified to avoid more complexity and to ensure greater speed in symbol decision making. They will be compared with the following popular adaptive filters: least mean square (LMS) and recursive least squares (RLS), in terms of bit error rate (BER) and minimum square error (MSE) with 16, 64, 128, 256, 512 and 1024 quadrature amplitude modulation (QAM). By that, this work will show the advantage that the MLP equalizer has, in most cases, over RBF and traditional linear equalizers.

Keywords—*adaptive filter, channel equalization, M-QAM, MLP, RBF, symbol decision making.*

1. Introduction

Many key technologies for enhancing spectrum efficiency are planned for 5G, such as new non-orthogonal access scheme, generalized frequency division access, filter bank multicarrier and universal filtered multicarrier [1], massive multiple-input multiple-output and high order modulation (HOM), e.g. 256 QAM [2]. 256 QAM (8 bits/symbol) increases the maximum peak rate by 33% compared to 64 QAM (6 bits/symbol). Several studies are conducted that are concerned with employing HOM. In [3], HOM improves the system's capacity in small cells. In [4], transmission performance of HOM-based 256 and 1024 QAM over OFDM is investigated with the impact on error vector magnitude (EVM). A practical FPGA implementation of HOM is already done in [5], using the conventional constant modulus algorithm and LMS equalizers. Even though linear equalizers are widely used, they have poor performance regarding HOM complexity and sensitivity to distortions.

HOM offers better performance than lower order modulation, but still has some external and internal impairments, such as inter-cell interference (ICI) and sensitivity to multipath propagation effects (external), or thermal noise and quality degradation caused by practical manufacturing constraints influencing the EVM (internal). Employing HOM comes with the expense of a higher signal-to-noise (SNR) requirement.

ANNs, being non-linear equalizers, are gaining importance in channel equalization for their flexible architecture, optimization and the learning process [6]. In this paper, multilayer perception (MLP) and radial basis function (RBF) are considered for the simplicity of their architecture and for their ease of use, and especially for their different learning techniques, namely back-propagation, extended Kalman filter, genetic algorithm, and particle swarm optimization [7]. In the case of MLP, we will focus on the back-propagation algorithm. RBF is based on the Euclidian norm. The equalization problem is treated as a classification process that tells to what sets the received symbol belongs, and makes a decision. We will show the effectiveness of the developed MLP and RBF compared to conventional receivers with HOM, by diminishing MSE and BER criteria.

In literature, many types of channels were presented. Different models and distributions were mathematically developed under two categories: channel with line of sight (LOS), and channels with non-line of sight (NLOS). The modeling of NLOS channels is a calculation of physical processes (reflection, diffraction, scattering) affecting the signal during transmission. Nakagami-*m* represents a generalized distribution that can be applied as a LOS or NLOS channel. Rice and Rayleigh distributions are special cases of Nakagami-*m* that can be considered as LOS and NLOS channels, respectively [8]. They are mostly used for modeling those channels. In this work, the Rayleigh fading channel is used due to its good and fair approximations of multipath fading in real life. Model generation will be depicted in Section 4.

The equalizers were designed and trained by M-QAM symbols (M is modulation order) over an Orthogonal Frequency Division Multiplexing (OFDM) wave carrier in a noisy Rayleigh flat-fading channel. Simulations are carried out using an Intel Core i5 (2.2 GHz) 8 GB-RAM computer, in the Matlab 2016 environment.

The paper is structured as follows. Section 2 presents the basics of MLP and RBF equalizer designs. Section 3 de-

scribes the model of the system used for the transmission with HOM over the Rayleigh fading channel. Section 4 illustrates simulation results and presents the discussion. The paper is concluded by Section 5.

2. MLP and RBF Equalizers Design

ANN equalizers are working with either real or complex values. A real ANN has a simpler architecture and its training is fast and not difficult compared to a complex ANN. The simplicity and speed we are striving to achieve are reached by dividing complex modulated symbols into real and imaginary parts, and treating them separately, one after another. Then, the symbols are reshaped at the output in a complex form. The strategy used in the learning process is that total symbols are split randomly between learning (70%), validation (15%) and test (15%), where each step is evaluated with MSE. The architectures are designed to be simple, in order to avoid computational complexity and to minimize the decision time factor. In this case, three layers are fixed: the input layer, the hidden layer and the output layer. The input layer has one input for both real and imaginary values. The hidden layer contains only one layer of neurons (10 for MLP and 20 to 40 for RBF, depending on M-QAM). The output layer has one output only.

2.1. MLP

MLP learning is assured by the back-propagation algorithm (simple, fast, and depends on iterative laws). The cost function (MSE) is minimized by adaptation of weight coefficients, through:

$$j(w) = \frac{1}{2} \sum_{p=1}^{T} \sum_{i=1}^{n} |y_d(t) - y(t)|^2 , \qquad (1)$$

where $y_d(t)$ is the desired output, $y(t)$ is the output, n is the number of iterations, and T is the dimension of the training assembly. Weight adaptation (adjustments and updating) is based on the iterative formula:

$$w_{ij}^k(n+1) = w_{ij}^k - \mu \frac{\partial j(w)}{\partial w_{ij}^k(n)} , \qquad (2)$$

where μ represents the learning step and k is the number of the hidden layer. The activation function is chosen to be hyperbolic-tangent. An initial value of μ is fixed respecting $\mu < 0.1$. Its optimal value is determined gradually through the evolution of the MSE error of MLP.

2.2. RBF

For RBF, the learning process is divided into two stages: unsupervised nonlinear, where the parameters update techniques are quite fast, and supervised linear problem solving at the output, with the advantage of dodging the local minima issue, often experienced when employing MLP. The training of RBF can be partitioned into three phases.

Firstly, one has to find centers C_i of the Gaussian radial functions (activation function). Secondly, widths σ_i have to be fixed. Finally, the network's weights λ_j between the radial function and the output layers (Fig. 1) are determined. The simplified function that rules the RBF network, without the independent term reported in formula (21) in [9], is:

$$F(x) = \sum_{j=1}^{M} \lambda_j e^{-\frac{\|x - c_j\|^2}{2\sigma_j^2}} + \sum_{i=1}^{d} a_i x_i , \qquad (3)$$

where $\|(.)\|$ represents the Euclidean norm, x is the input value, M is number of centers C_j, d is the length of input variables, a_i are the coefficients of the linear terms, and $e^{(.)}$ is the exponential function that represents the nonlinear term.
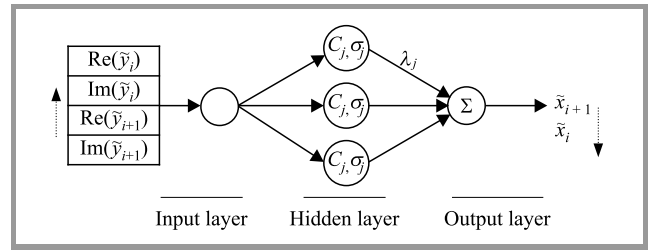
Fig. 1. Simplified RBF structure for symbol decision making with one input, one output and a hidden layer with an adequate number of neurons.

Centers C_j are adjusted and fixed utilizing competitive learning (unsupervised) with the Euclidian distance calculations expressed by:

$$d_2(x^i, C^j) = (x^i - C^j)^T (x^i - C^j)$$
$$= \sqrt{\sum_{k=1}^{D} (x_k^i - C_k^j)^2} , \quad (4)$$

$$C^k(t+1) = C^k(t) + \alpha(t)(x^i - C^k) , \qquad (5)$$

where $\alpha(t)$ is the time adaptation factor, with $0 < \alpha(t) < 1$, and σ_j is calculated with:

$$\sigma = \frac{d_{\max}}{\sqrt{2M}} , \qquad (6)$$

where $d_{\max}$ is the maximum distance between any couple of centroids.

When C_j and σ_j are fixed, Eq. (3) becomes linear, and λ_j is estimated by the pseudoinverse.

Table 1 summarizes the architecture design and all parameters of MLP and RBF equalizers for all modulation schemes.

For better performance and fair comparison, filter length of LMS has $10w_n$ coefficients, and its step μ is fixed for each

M-QAM separately (Table 2). Similarly, RLS filter length has $10w_n$ coefficients, and $\lambda_{RLS} = 0.9$ for all M-QAM.

Table 1
Summary of ANN equalizers parameters

Parameters	MLP	RBF
Criterion function	MSE	MSE
Hidden layers	1	1
Hidden layer neurons	10	20...40
Input layer neurons	1	1
Output layer neurons	1	1
Parameters	Weights w Step μ Iteration n	Centroids C_i Widths σ_i Weights λ_i
Algorithm	Back-propagation	Euclidian norm
Activation function	Hyperbolic-tangent	Gaussian
Learning	Supervised	Supervised + unsupervised

Table 2
LMS parameters

M-QAM	Step μ	w_n number
16	0.0039	10
64	0.0031	10
128	0.0021	10
256	$4.54 \cdot 10^{-4}$	10
512	$3 \cdot 10^{-4}$	10
1024	$9.99 \cdot 10^{-5}$	10

3. System Model

In the transmitter part, 300,000 sequences of a random signal $x(n)$ are generated with zero-mean and unit variance, and are supposed to be independent and identically distributed [10]. $x(n)$ is modulated with HOM and shaped with OFDM carriers. In M-QAM, we will have $300000/K$ symbols, where $K = \log_2(M)$ is the number of bits per symbol. $M = 16, 64, 128, 256, 512, 1024$. After modulation, the received signal can be expressed by:

$$\tilde{y}(n) = x(n) \cdot H + \eta(n) \,, \qquad (7)$$

where $x(n)$ is the signal emitted through H (H is the Rayleigh flat-fading channel transfer function), and $\eta(n)$ is a complex additive white Gaussian noise generated with the energy per bit-to-noise power spectral density ratio $E_b/N_0 = 1 : 28$ dB, depending on M.
The Rayleigh flat-fading channel H is considered as a modeled finite impulse-response filter that can be obtained from a random, uncorrelated, complex Gaussian processes.
In the receiver part, $\tilde{y}(n)$ in Eq. (7) is reconstructed by removing the cyclic prefix, applying the fast Fourier transform and converting parallel stream data to a serial form. After OFDM demodulation, comes the delicate part of equalizing with the use of LMS, RLS, MLP and RBF at different times, and calculating MSE of each of the above for different M-QAM and different SNRs. Finally, the symbols are demodulated to bits to calculate BER. Figure 2 represents the model for symbol decision making for all equalizers referred to in this paper.

4. Simulations and Results

All HOM schemes and their MSE values are depicted in Figs. 3 and 4, where MSE values are expressed in terms of multiple SNRs. LMS and RLS have excessive MSE values in the order of 10^2 to 10^4. This shows that linear equalizers have poor performance, particularly on low SNR values.
MLP and RBF show great resistance to channel effects, where MSE is in the order of 10^{-4} to 5 with 16, 64 and 128 QAM, and in the order of 5 to 10^2 for 256, 512 and 1024 QAM. In 16 and 512 QAM, MSE of MLP and RBF are almost the same. In 64 and 128 QAM, MSE of RBF is better, and in 256 and 1024 QAM, MSE of MLP is better.
In terms of BER, we first lay out the theoretical BER of every M-QAM. Here, the theoretical BER is considered as a reference to measure the effectiveness of equalizers coupled with signal degrading factors, such as SNR, channel
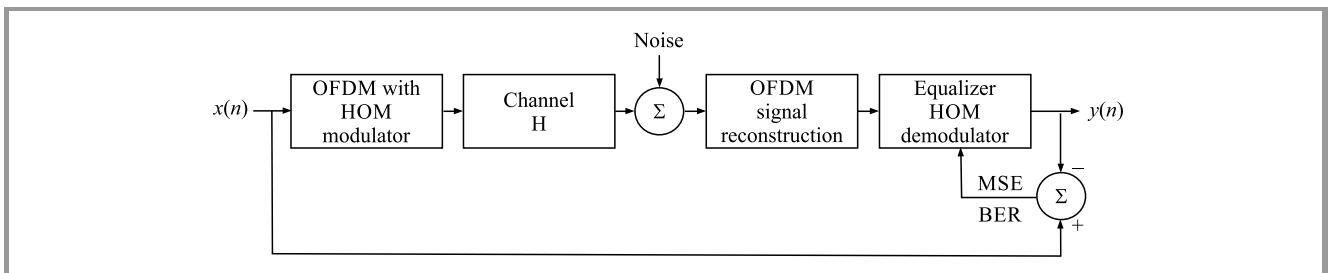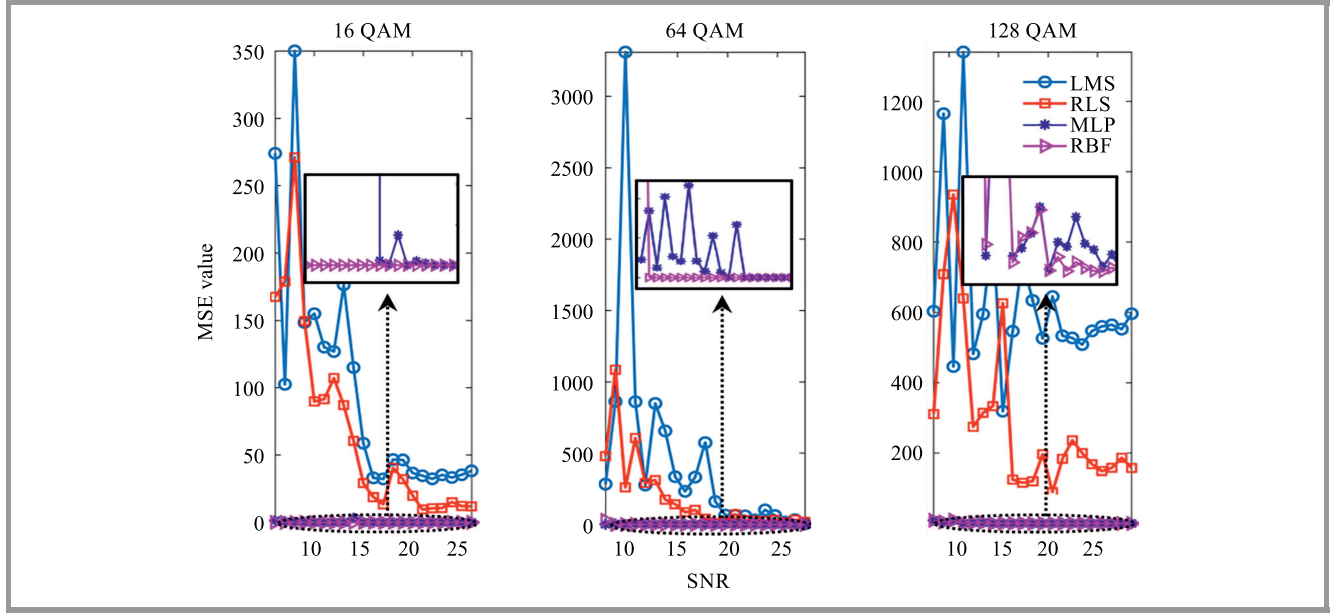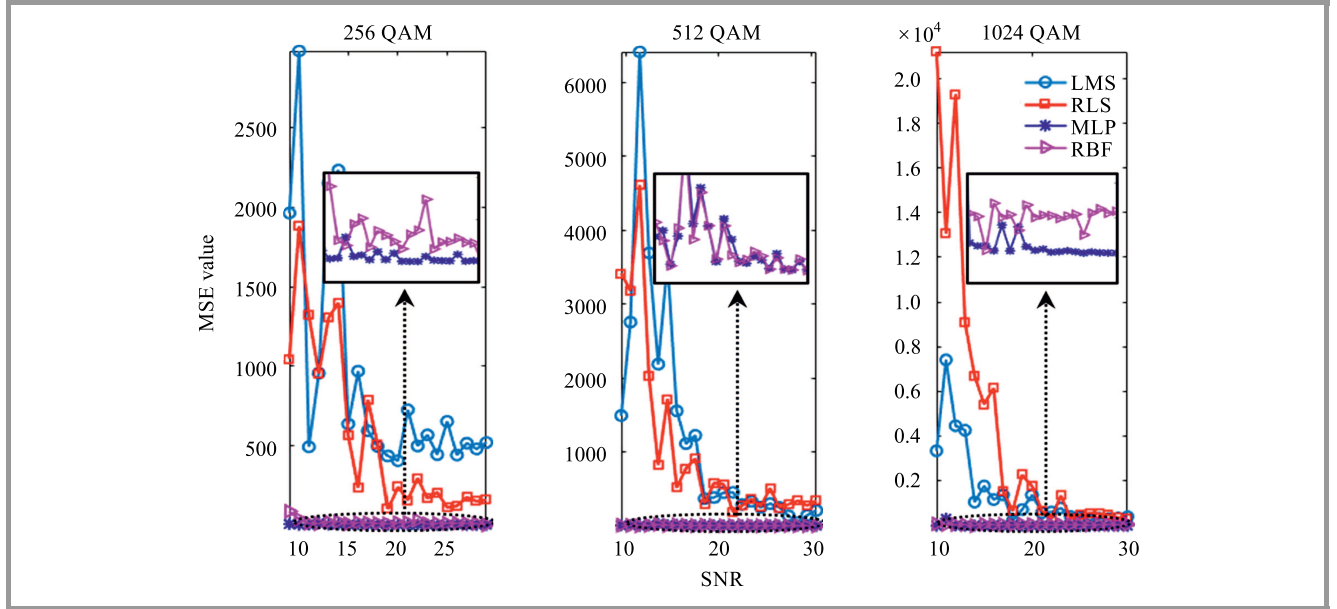

Fig. 2. Channel equalization model.

Fig. 3. Effect of SNR on MSE of equalizers, with 16, 64, and 128 QAM. (For color pictures visit www.nit.eu/publications/journal-jtit)

Fig. 4. Effect of SNR on MSE of equalizers, with 256, 512, and 1024 QAM.

type, and sensitivity of the modulation scheme used. The theoretical BER is calculated via equation [11]:

$$
P_b = \frac{2}{\sqrt{M}\log_2\sqrt{M}} \sum_{K=1}^{\log_2\sqrt{M}}
$$

$$
\sum_{i=0}^{(1-2^{-K})\sqrt{M}-1} \left\{ (-1)^{\left\lfloor \frac{i2^{K-1}}{\sqrt{M}} \right\rfloor} \left(2^{K-1} - \left\lfloor \frac{i2^{K-1}}{\sqrt{M}} + \frac{1}{2} \right\rfloor \right) Q(\phi) \right\} \quad (8)
$$

$$
\phi = \left((2i+1)\sqrt{\frac{6\log_2 M}{2(M-1)}\frac{E_b}{N_0}} \right) , \quad (9)
$$

where P_b is the BER and $Q(x)$ is the Q function given by:

$$
Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{\frac{-t^2}{2}} \, dt
$$

$$
= \frac{1}{\sqrt{\pi}} \int_x^\infty e^{-z^2} \, dz = \frac{\text{erfc}(x)}{2} , \quad (10)
$$

where $\text{erfc}(x)$ is the complementary error function, which is accessible in Matlab software.

With 16 QAM (Fig. 5), RLS and LMS are starting to diverge from the theoretical curve at 11 dB, at the same E_b/N_0 value a slight improvement is noticed in the case of MLP and RBF. At $E_b/N_0 = 12$ dB, an improvement

of 1 dB is observed for both. The two have identical performance in this modulation scheme.
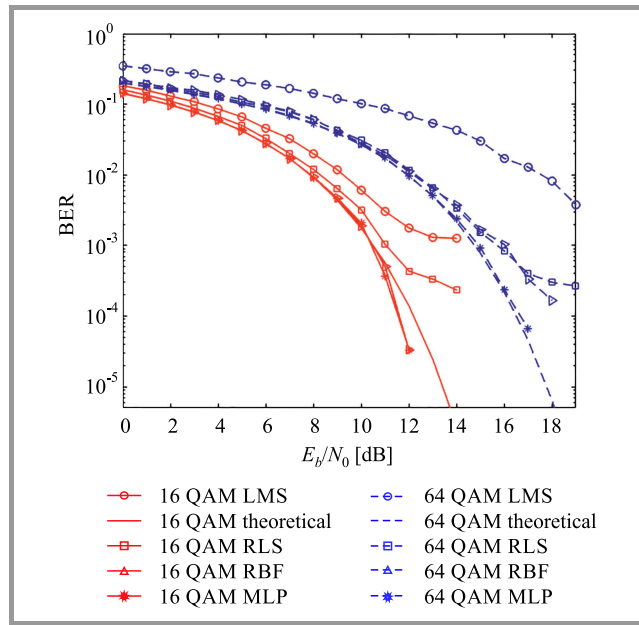
Fig. 5. Equalizer BER performance curves with 16 and 64 QAM.

With 64 QAM (Fig. 5), LMS performance is not good enough for the theoretical curve, RLS and RBF are better but start to diverge at 15 dB. MLP is performing as the theoretical curve, and is the most suited for 64 QAM.

With 128 QAM (Fig. 6), LMS and RLS start to diverge from 13 dB and 18 dB, respectively. MLP is performing better than linear equalizers, gaining almost 1 dB, but RBF is the most suited for this scheme of modulation by improving by almost 2 dB compared to the theoretical curve. With 256 QAM (Fig. 6), MLP and RBF are performing
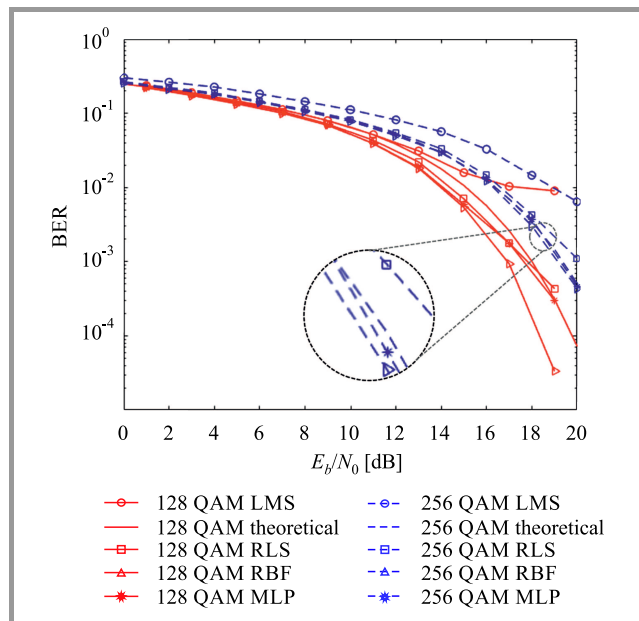
Fig. 6. Equalizer BER performance curves with 128 and 256 QAM.

nearly the same as the theoretical curve. A slight improvement is noticed on RBF performance compared to MLP. RLS is better than LMS but starts to diverge at 18 dB.

With 512 and 1024 QAM, spectral efficiency is really high (9 and 10 bits per symbol, respectively). As mentioned before, they are sensitive to excessive noise and need more SNR to be investigated compared to lower case modulations. Their performance is shown in Fig. 7. At 512 QAM, the two linear equalizers are diverging, RBF is trailing the theoretical curve but diverges at 23 dB. MLP is performing really good to gain 1.5 dB, and is better suited to 512 QAM than others.
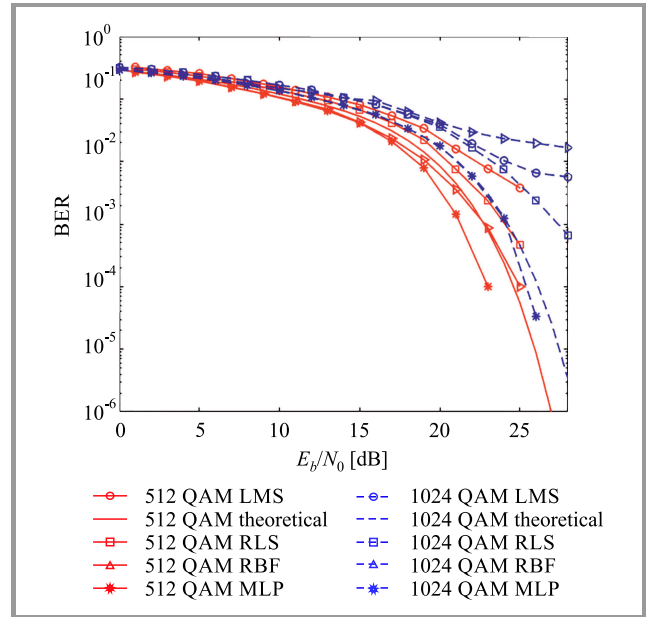
Fig. 7. Equalizer BER performance curves with 512 and 1024 QAM.

With 1024 QAM, MLP is more efficient than RLS, LMS and RBF (Fig. 7) which are not trailing the theoretical curve. RBF performance is poorer than that of LMS and RLS. We tried to train it differently and expand the architecture of RBF by adding more neurons (up to 40 neurons). Unfortunately, there was no improvement. MLP is most suited for this high spectral efficiency.

The developed neural equalizers show great resistance and excellent robustness in terms of adaptation. Symbols are appropriately equalized from the starting point with a better BER, in most cases, than theoretical simulation. When the channel is in a diminutive SNR state, MLP and RBF equalizers do not suffer from any performance degradation, and they are more appropriate for this type of channel coupled with HOM. However, RLS and LMS equalizers offer better performance for low and moderate levels of noise with HOM.

Table 3 represents the profiling of equalizer algorithms during the processing of data (all sequences and all stages of training). Initialization and MSE calculations are not considered. The average of 10 measurements is reported in the table for each algorithm. 256 QAM is chosen be-

cause – in terms of BER – the equalizers are performing almost similarly. LMS has the shortest duration because of its simplicity. RLS is more complex and it requires more time, which means that it performs similarly to MLP. The latter, at a much better speed, can have a good result for both MSE and BER. RBF takes a long time, since its architecture design is extended (more neurons) to meet better BER results.

Table 3
Profiling the equalizer speed performance
with 256 QAM at $E_b/N_0 = 20$ dB

Equalizers	LMS	RLS	MLP	RBF
Time [s]	0.239644	0.451853	0.542335	0.972604
w_n/neurons	10	10	10	20

5. Conclusion

In this paper, MLP and RBF equalizers are developed in the simplest form for HOM equalization, serving as a key for achieving more spectral efficiency. It is shown through simulation results that MSE values of the two ANN equalizers are negligible in comparison with LMS and RLS. Also the improved BER, in most M-QAM scenarios, offers great performance in HOM processing, despite its complexity (compared to linear equalizers). These models are more reliable and efficient in terms of canceling noise (MSE criterion) and error rate performance (BER criterion).

MLP is better suited than RBF with 64, 512 and 1024 QAM, and is performing almost identically to RBF with 16 and 256 QAM. RBF is performing well only with 128 QAM. Through the planning of small cells (cancelation of ICI), HOM associated with MLP equalizer are an accessible option for enhancing spectral efficiency and for increasing data rate peaks. Further investigations are to be performed to augment the simplification of learning strategies, architectures and complexity levels.

References

[1] Y. Tao, L. Liu, S. Liu, and Z. Zhang, "A survey: Several technologies of non-orthogonal transmission for 5G", *China Commun.*, vol. 12, no. 10, pp. 1–15, 2015 (doi: 10.1109/CC.2015.7315054).

[2] Y. Kishiyama, A. Benjebbour, T. Nakamura, and H. Ishii, "Future steps of LTE-A: evolution toward integration of local area and wide area systems", *IEEE Wirel. Commun.*, vol. 20, no. 1, pp. 12–18, 2013 (doi: 10.1109/MWC.2013.6472194).

[3] A. Darwish and A. S. Ibrahim, "Capacity improvement via indoor small cells", in *Proc. Int. Wirel. Commun. and Mob. Comput. Conf. IWCMC 2014*, Nicosia, Cyprus, 2014, pp. 69–73 (doi: 10.1109/IWCMC.2014.6906334).

[4] M. Iwamoto, S. Matsuoka, H. Iwasaki, and H. Otsuka, "Transmission performance of OFDM with 1024-QAM in the presence of EVM degradation", in *Proc. 2014 IEEE Asia Pacific Conf. on Wirel. and Mob.*, Bali, Indonesia, 2014, pp. 12–16 (doi: 10.1109/APWiMob.2014.6920262).

[5] S. Ma and Y. Chen, "FPGA implementation of high-throughput complex adaptive equalizer for QAM receiver", in *Proc. 8th Int. Conf. on Wirel. Commun., Netw. and Mob. Comput.*, Shanghai, China, 2012, pp. 1–4 (doi: 10.1109/WiCOM.2012.6478527).

[6] K. Burse, R. N. Yadav, and S. C. Shrivastava, "Channel equalization using neural networks: a review", *IEEE Trans. on Syst., Man, and Cybernet., Part C (Appl. and Rev.)*, vol. 40, no. 3, pp. 352–357, 2010 (doi: 10.1109/TSMCC.2009.2038279).

[7] A. Rubaai and P. Young, "Hardware/software implementation of fuzzy-neural-network self-learning control methods for brushless DC motor drives", *IEEE Trans. on Industry Appl.*, vol. 52, no. 1, pp. 414–424, 2016 (doi: 10.1109/TIA.2015.2468191).

[8] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005 (ISBN: 9780511841224, doi: 10.1017/CBO9780511841224).

[9] F. Bouguerra, I. Benacer, and L. Saidi, "MLP and RBF symbol tracking with 16 QAM modulation over multipath distorted channel", in *Proc. Int. Conf. on Adv. Syst. and Elec. Technol. IC_ASET 2017*, Hammamet, Tunisia, 2017, pp. 182–187 (doi: 10.1109/ASET.2017.7983688).

[10] H. Cai, Q. Zhang, Q. Li, and J. Qin, "Proactive monitoring via jamming for rate maximization over MIMO Rayleigh fading channels", *IEEE Commun. Lett.*, vol. 21, no. 9, pp. 2021–2024, 2017 (doi: 10.1109/LCOMM.2017.2715337).

[11] K. Cho and D. Yoon, "On the general BER expression of one- and two-dimensional amplitude modulations", *IEEE Trans. on Commun.*, vol. 50, no. 7, pp. 1074–1080, 2002 (doi: 10.1109/TCOMM.2002.800818).

Fateh Bouguerra received his B.Eng. in Electronics and Communication Engineering in 2006 and M.Sc. (Eng.) in Microwave and Telecommunication in 2011 from University of Batna 2. Now he is pursuing his Ph.D. degree as a researcher in the field of Mobile Communications. His interests include digital signal processing, wireless communication networks, artificial intelligence, and biomedical engineering.
E-mail: Bouguerra_fateh@yahoo.fr
LAAAS Laboratory
Department of Electronics
University of Batna 2
Batna, Algeria

Lamir Saidi received his Ph.D. from University of Savoie, France, in 1996. Currently, he is a full professor at the Electronics Engineering Department, University of Batna 2, Algeria. He is at the head of the Advanced Automatic and Systems Analysis Laboratory. His interests include digital signal processing and communication systems.

E-mail: l.saidi@univ-batna2.dz
LAAAS Laboratory
Department of Electronics
University of Batna 2
Batna, Algeria

Implementation of Selected Spectrum Sensing Systems for Cognitive Radio Networks using FPGA Platform

Hikmat N. Abdullah and Hadeel S. Abed

Al Nahrain University, College of Information Engineering, Baghdad, Iraq

Abstract—The energy efficient spectrum sensing method is very important in cognitive radio (CR), since high power drain may limit its implementation in mobile applications. The spectrum sensing feature consumes more energy than other functional blocks, as it depends on continuous detection of the presence or absence of the primary user (PU). In this paper, we proposed two methods to reduce energy consumption of the spectrum sensing feature. The first is of a single stage variety with a reduced number of sensed samples. The other uses two stages. The first stage performs coarse sensing for many subchannels, and the best subchannel is forwarded for fine sensing in the second stage. The performance of the proposed methods is evaluated in AWGN channel and compared with the existing approach. The proposed methods are simulated using Matlab and ModelSim and are then hardware implemented using the Altera Cyclone II FPGA board. Simulation results show that the proposed methods offer an improvement in energy consumption with an acceptable reduction in the probability of detection. At E_b/N_0 of 0 dB, the energy consumption is reduced by 50% and 72% in the first and second proposed method, respectively, compared to the traditional method (100% sensing).

Keywords—*cognitive radio, energy consumption, energy detection, FPGA.*

1. Introduction

Cognitive radio (CR) is defined as a wireless technology that changes its operating parameters (e.g., power, carrier frequency, bandwidth, modulation and coding) based on the environment conditions. CRs allow unlicensed (secondary) users (SUs) to access spectrum bands at the time they are not used by licensed (primary) users (PUs). The channels that remain unused, on a temporary basis, are called spectrum holes [1]. Cognitive radio networks suffer from many problems, i.e. the energy efficiency problem which has received a lot of research attention during recent years [2], [3].
CR networks consist of a number of units sensing PU activity by using a specific sensing technique. All CR units consume energy, but the sensing stage draws more energy than others, since the spectrum sensing techniques depend on detecting the activities of the PU. Many techniques are used to sense the spectrum (energy detector, cyclostationary, matched filter and others).

A lot of methods have been investigated to improve energy consumption. Sequential sensing is a method to decrease the average number of sensors required to make a decision [4]–[7]. The authors of [8] proposed a fixed size censoring scheme, where each sensor senses fixed size samples of the spectrum and employs a censoring policy to send the results of local decisions to the fusion center (FC). In this method, only energy consumption in the transmission stage is reduced and there is no improvement in the sensing stage, since it senses a fixed number of samples. The authors of [9] implemented an algorithm-based energy detection technique in the Xilinx Virtex2pro FPGA board, but the algorithm did not aim to reduce energy consumption. In [10], a method for improving energy consumption is proposed based on two stages: coarse and fine, but was implemented only in the course of a simulation. In this paper, we have proposed two energy detector-based techniques to reduce energy consumption of the sensing stage. In the first method, energy savings are achieved in a single stage, by reducing the number of sensed samples, while the other method consists of two stages: coarse and fine, with an algorithm that is modified compared to that from [10]. The main contribution of this paper is that the proposed methods are simulated using the ZigBee wireless standard as a case study for cognitive user (CU), and are then implemented using the Altera FPGA Cyclone II board.

2. Energy Detector

Energy detection is one of the most popular spectrum sensing techniques, because of its simplicity and because it does not need any knowledge about the PU signals. In CR networks, CU senses the spectrum and detects the absence of a PU signal. After finding a hole, CU begins to transmit data to CU receiver. The samples received at CU are [11]:

$$X(n) = h\theta S(n) + N_o(n) \; , \qquad (1)$$

where h is channel gain, $S(n)$ is the samples transmitted by PU, $N_o(n)$ is the additive white Gaussian noise samples collected from the channel, θ is the activity indicator, which can take one out of two values:

$$\theta = \begin{cases} 0 & \text{for } H_0 \text{ hypothesis} \\ 1 & \text{for } H_1 \text{ hypothesis} \end{cases} . \qquad (2)$$

The presence and absence of PU is referred to as hypothesis H_1 and H_0, respectively. The probabilities of a false alarm and of detection are calculated by comparing the detector decision metric with a pre-defined threshold λ. The decision metric E_j is defined as the accumulated energy of the tested samples during the monitoring window t:

$$E_j = \frac{1}{N} \sum_{n=1}^{N} |x(n)|^2 . \qquad (3)$$

In the above equation, N is the total number of sensed samples, $N = tF_s$, where F_s is the sampling frequency. The probability of a false alarm and the probability of detection are:

$$P_f = P_r(E_j > \lambda | H_0) , \qquad (4)$$

$$P_d = P_r(E_j > \lambda | H_1) . \qquad (5)$$

3. Design of Spectrum Sensing

3.1. The Single Stage Energy Efficient Spectrum Sensing

In this case the energy consumption improvement is achieved by decreasing the number of sensed samples to reduce the computational resources relied upon by the energy detector technique before making a decision. It also leads to a decrease in the sensing time, and it is worthy if the sensing reliability of the CR network is maintained within satisfactory limits. The reliability of sensing means that the maximum probability of detection and the minimum probability of a false alarm can be obtained. Figure 1 explains the way of choosing the samples to be sensed. Sensing of 50% of all samples is performed by sensing only even (or odd) numbered samples. To compute the amount of energy consumed by the cognitive user C_j, we use the procedure from [12], [13]:

$$C_j = N_s C_{ssj} + (1 - \rho_i) C_{tj} , \qquad (6)$$

where C_{ssj} and C_{tj} are the energies consumed by the j-th cognitive radio in sensing per sample and transmission per bit, respectively, ρ_i is the average censoring rate, and N_s is the total number of sensed samples. As we compute the
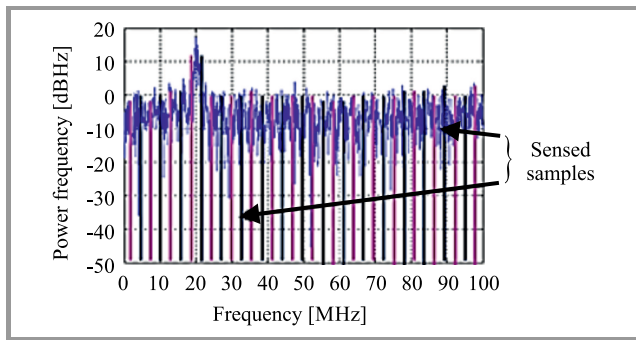
Fig. 1. The first method proposed (selecting 50% of the sensing samples). (For color pictures visit www.nit.eu/publications/journal-jtit)

amount of energy consumption in the C_{sj} sensing stage only, Eq. (6) reduces to [12], [13]:

$$C_{sj} = N_s C_{ssj} . \qquad (7)$$

Let's assume that $P_d = 1$, i.e. $C_{ssj|P_d=1}$. When P_d is decreased, C_{ssj} is increased, since the energy detector will produce a false decision which leads to repeating the sensing sequence. Hence, C_{ssj} can be written as:

$$C_{ssj} = C_{ssj|P_d=1} + C_{ssj|P_d=1}(1 - P_d) . \qquad (8)$$

After substituting Eq. (8) in (7) we have:

$$C_{sj} = kN_s \left[C_{ssj|P_d=1} + C_{ssj|P_d=1}(1 - P_d) \right] . \qquad (9)$$

One can note that when $P_d = 1$ Eq. (9) becomes $C_{sj} = N_s C_{ssj|P_d=1}$ where k is the sensing ratio, i.e. the number of selected samples sensed to the total number of samples in the band. Energy consumption per bit in the transmission stage is computed as described below. According to [14], the energy consumed in relation to the signal processing part of the transmission mode, for a data rate of 250 kbps and a 2.1 V/17.4 mA power supply, is approximately 150 nJ/bit. To transmit one bit over a distance d, the radio spends:

$$C_{tbj} = C_{t\text{-}elecj} + e_{ampj} d^2 , \qquad (10)$$

where $C_{t\text{-}elecj}$ is the level of energy consumption in the transmitter's electronics, and e_{ampj} is the amplification needed to satisfy a given receiver sensitivity level. Assuming a data rate of 250 kbps and a transmitter power of 20 mW, $C_{t\text{-}elecj} = 80$ nJ with a receiver sensitivity of –90 dBm at SNR of 10 dB e_{ampj} is 40.4 pJ/m². So, the transmitted energy per bit per for the j-th CU would be:

$$C_{tj} = C_{tbj} + 150 \text{ nJ/bit} . \qquad (11)$$

3.2. Two Stage Energy Efficient Spectrum Sensing

The second method consists of two stages: in the coarse stage, the CU senses a fixed spectrum length and a constant number of channels using a small number of sensed samples (e.g. 16%). The energy levels obtained from all channels during coarse sensing are collected, and then fine sensing is performed for the channel that has the maximum energy level. Fine sensing uses all samples in the selected channel to confirm the existence of PU, as shown in Fig. 2. To ensure that this signal is truly generated by PU and is not noise only, the level of energy accumulated in the channel is compared with threshold λ to produce a final decision about the presence of a PU signal. While detecting a spectrum hole, the same procedure is repeated, but fine sensing is employed to the channel with a minimum energy level. Such a method can be applied to all channels that may potentially have PUs (or holes), by finding the maximum (or minimum) energy channels and subtracting the accumulated energy contained in these channels from

all other channels. If the difference is small, the particular channel also has a PU or a hole. The energy consumption level achieved by the cognitive user C_j is [15]:

$$C_j = \left(k + \frac{1}{m}\right)Ns C_{ssj} + C_{tbj} + 150 \text{ nJ/bit} , \qquad (12)$$

where k is the sensing ratio N_s, and m is the number of channels in the spectrum.
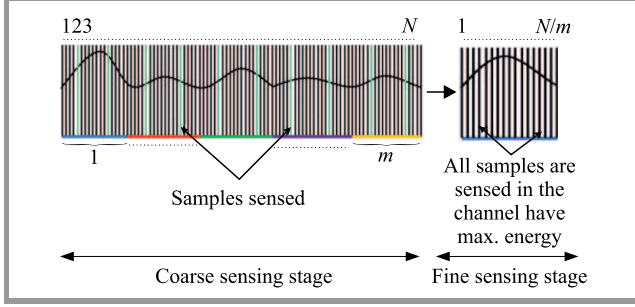
Fig. 2. Division of the spectrum using the second method proposed.

The threshold value comes from [16]:

$$\lambda = \left[Q^{-1}(Pf) + \sqrt{Ns}\right]2\sqrt{Ns}(N_s)^2 , \qquad (13)$$

where Q^{-1} is the inverse of the complimentary error function $Q(.)$, Ns is the total number of sensed samples and Pf is the probability of a false alarm.

4. Matlab Simulation Results

This section shows the results of a simulation of the energy detector's performance using the Receiver Operating Characteristic (ROC) curve, and energy consumption for a single user CR in AWGN, with the two methods proposed applied. The simulation parameters used are shown in Table 1.

Table 1
Simulation parameters

Parameter name	Value
Bit rate	2 Mbps
Carrier frequency	20 MHz
Modulation type	QPSK (PU signal)
Probability of false alarm	10^{-3}
Spectrum band	0–100 MHz
Sampling frequency	200 MHz
Bits per symbol	2
Sample per symbol	100
C_{ssj}	8 nJ [15]

Figure 3 shows the energy consumption performance curves of CU versus E_b/N_0 in AWGN, as a comparison between known and proposed methods. It can be seen that an im-
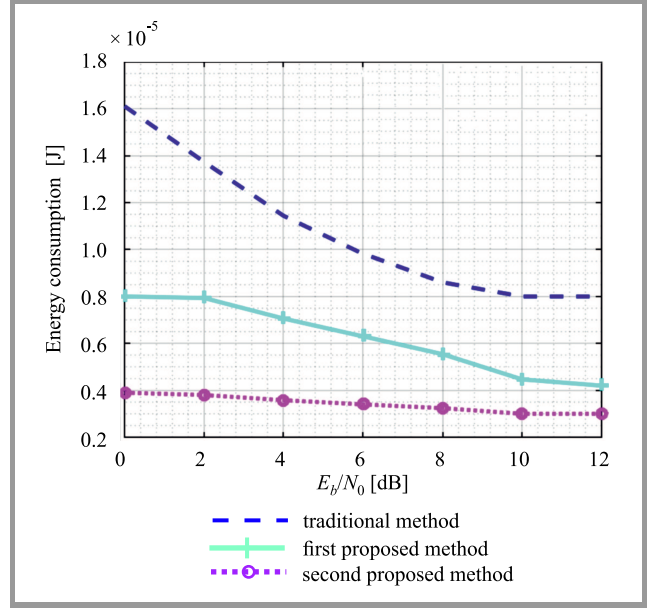
Fig. 3. Energy consumption versus E_b/N_0 in an AWGN channel.
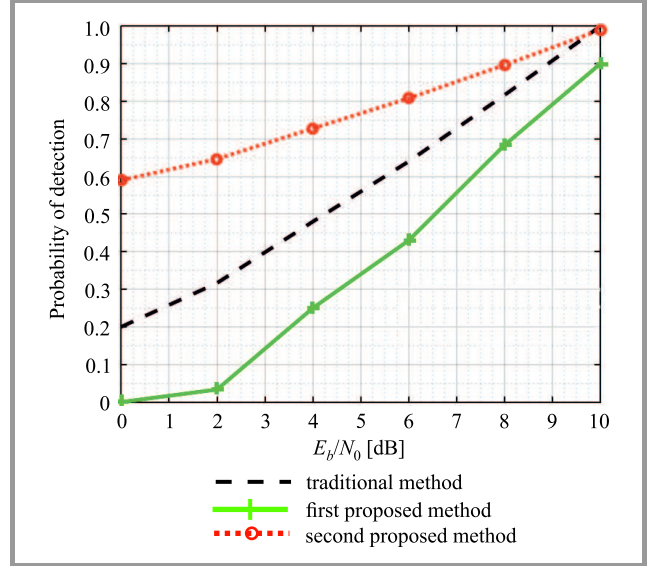
Fig. 4. P_d versus E_b/N_0 in an AWGN channel.

provement in energy consumption is obtained when E_b/N_0 is increased since at high values of E_b/N_0 a low number of sensing samples is required by CU to detect the PU signal. For example, at E_b/N_0 0 dB, energy consumption decreases by 50% and 72% in the first and second method proposed, respectively (compared to traditional methods). Hence, the second proposed method is more efficient at low E_b/N_0. Figure 4 shows the detection probability versus E_b/N_0. It can be noted that the detection probability is increased at higher E_b/N_0. Also, it shows that when comparing the first proposed method and the traditional method, P_d is increased, as the number of sensed samples is increased since the number of sensed samples in the traditional method is 100% ($k = 1$), and in the first proposed method it equals

50% ($k = 0.5$). The improvement in detection probability introduced by the second proposed method become more significant at low E_b/N_0, because of the use of two stages increases detection probability.

5. Hardware Implementation

Figure 5 shows the design procedure relied upon to implement the spectrum sensing methods using FPGA. It consists of five subsequent steps. First, the system specifications are set up, and next the VHDL description language code and the test benches of the traditional and proposed methods are written. Functionality of the design is then tested once again, using ModelSim. After successful verification, the generated VHDL is synthesized using the Quartus II tool. The synthesis produces a bit stream file which is downloaded into the FPGA board and experimental testing is performed.
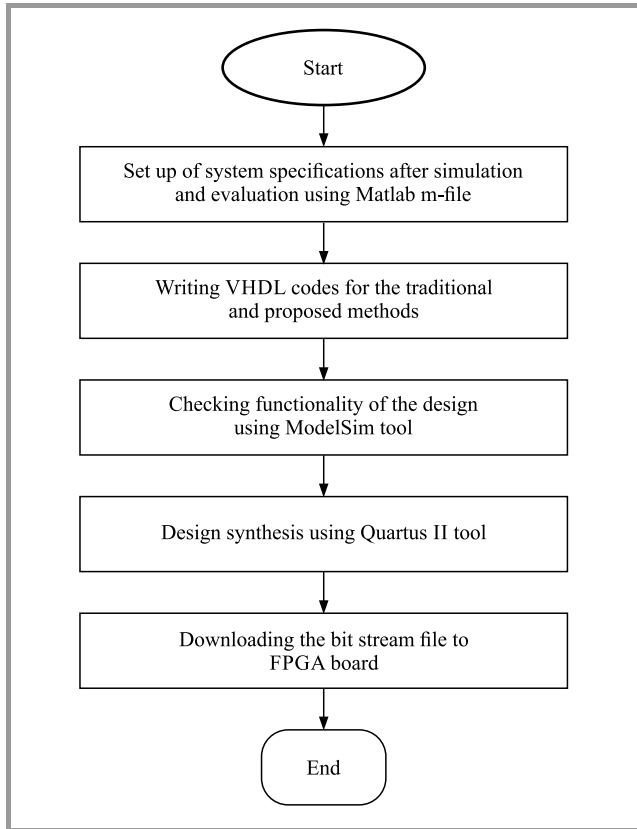
Fig. 5. Spectrum sensing method design flow using FPGA.

Figure 6 shows the design procedure relied upon to implement the traditional and proposed spectrum sensing methods using FPGA. First, the frequency samples of the sensed channels are generated. Then, the method rules are applied according to the selected switch of a special control circuit. The 8 LEDs available on the board are used to display the presence or absence of a PU in 8 channels. Figures 7 and 8 show the block diagrams that explain the design of the two proposed spectrum sensing methods. It
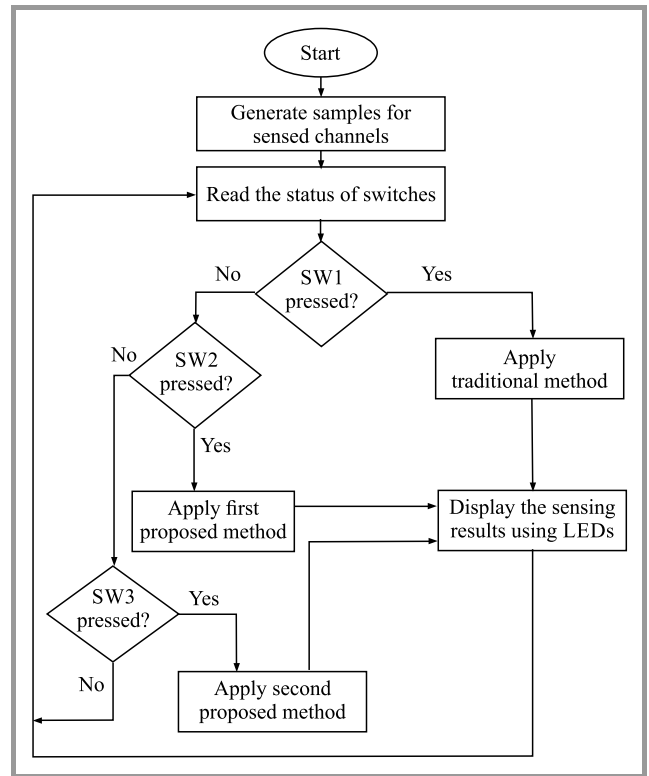
Fig. 6. Flow chart depicting implementation of spectrum sensing methods.

can be seen in Fig. 7 that 50% of samples are sensed (marked green), and then the energy accumulated in each channel is compared with a predefined threshold to make a decision about the presence or absence of PU. The tra-
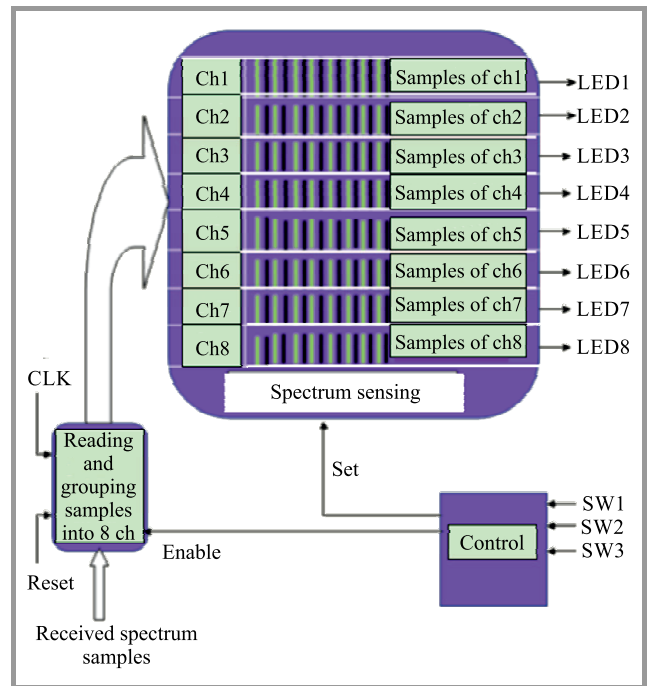
Fig. 7. Block diagram depicting implementation of the first method proposed on FPGA.
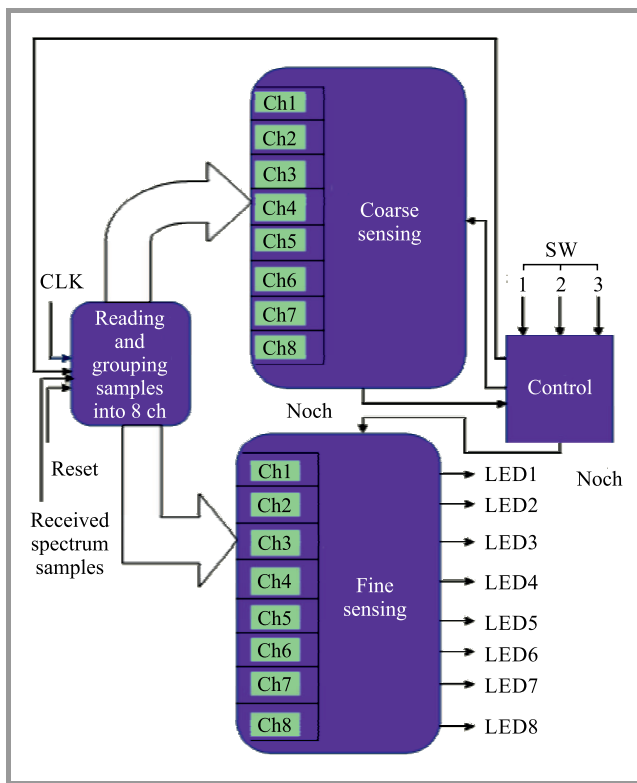
Fig. 8. Block diagram depicting implementation of the second method proposed on FPGA.

ditional method design procedure is similar, but with all samples being sensed.

It can be seen from Fig. 8 that there are two sensing blocks – one for coarse sensing and the other for fine sensing. In the coarse sensing block, the samples of each channel are

sensed with a k sensing ratio, and then the index of the "noch" channel that has the maximum accumulated energy is selected in order for the fine sensing block to perform the fine sensing procedure. The final decision about presence or absence of PU in each channel is produced through the board's LED indicators.

5.1. ModelSim and Real Hardware Results

Figures 9–11 show the ModelSim results of the traditional, the first and the second method proposed, respectively. In all results, the clock signal (CLK) drives the process of sensing samples, so that at each positive edge of the clock, one sample is sensed. Board switches SW1, SW2 and SW3 are responsible for selecting the sensing method. It can be seen that in Fig. 9 dec3, 4 and 8 equal to 1. This means that channels 3, 4 and 8 have a PU signal. In Fig. 10, the PU signal is present in channel 3 only, because of the reduction in detection performance when decreasing the number of sensed samples. Figure 11 shows the results after applying two stage sensing and finding the maximum energy channel. It can be seen that the index of the channel that has the maximum level of accumulated energy "noch" is 3, and dec3 is 1.

The experimental test results obtained with the use of the FPGA evaluation board are very close to those generated by ModelSim.

6. Conclusion

This paper has explained the way in which energy consumption in CR may be improved by applying two spectrum sensing methods. It has also evaluated the effects of
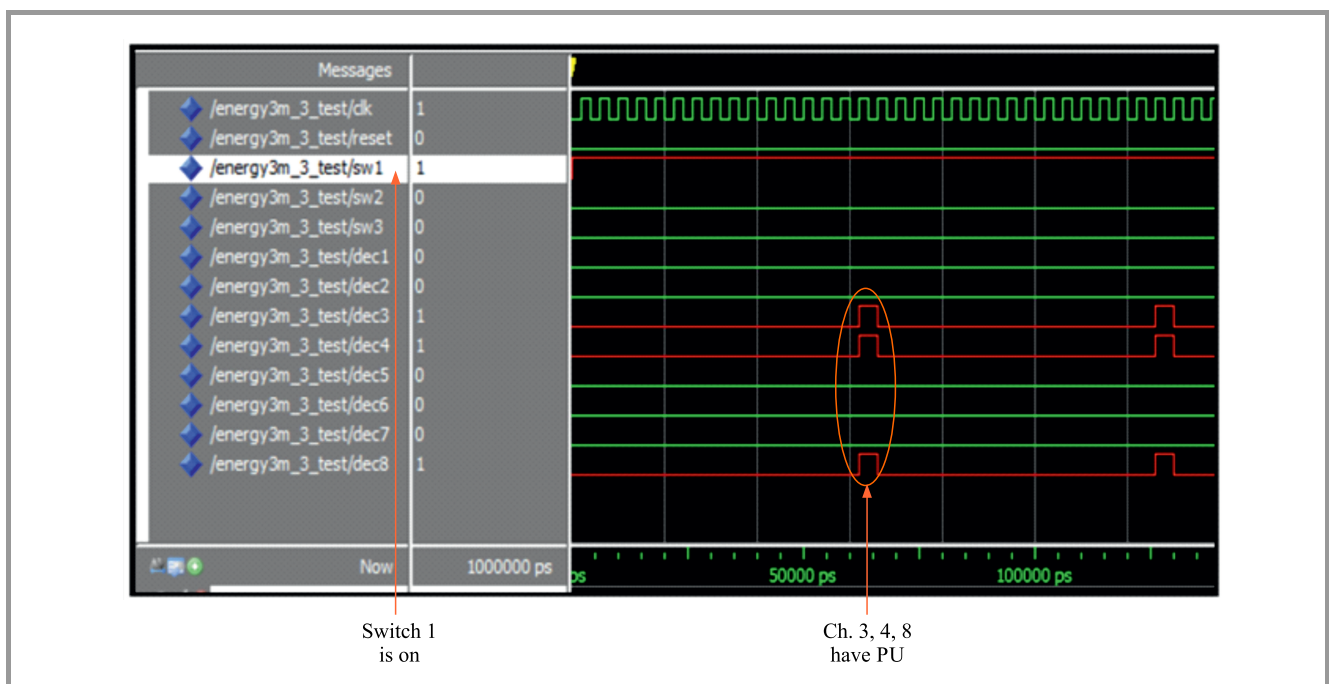
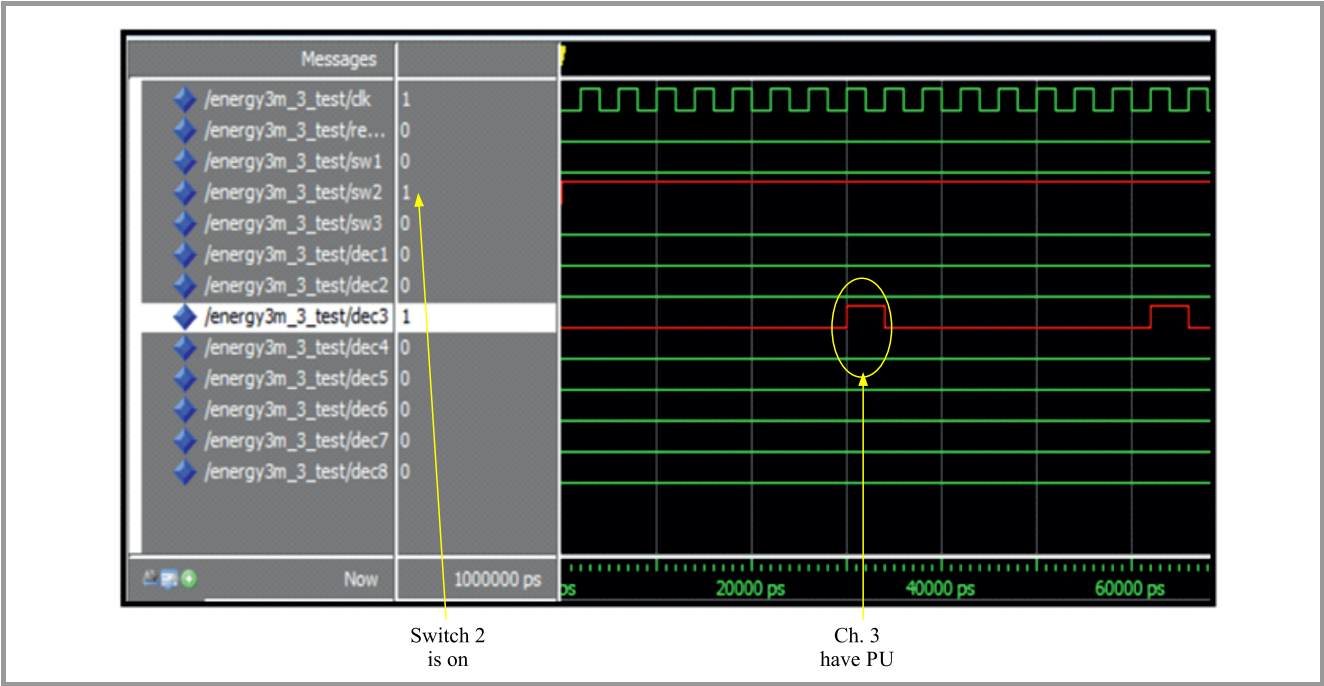

Fig. 9. ModelSim results for the traditional method.

Switch 2
is on

Ch. 3
have PU

Fig. 10. ModelSim results for the first method proposed.

Switch 3
is on

After fine sensing
Ch. 3 have PU

Result of coarse sensing, the index of channel
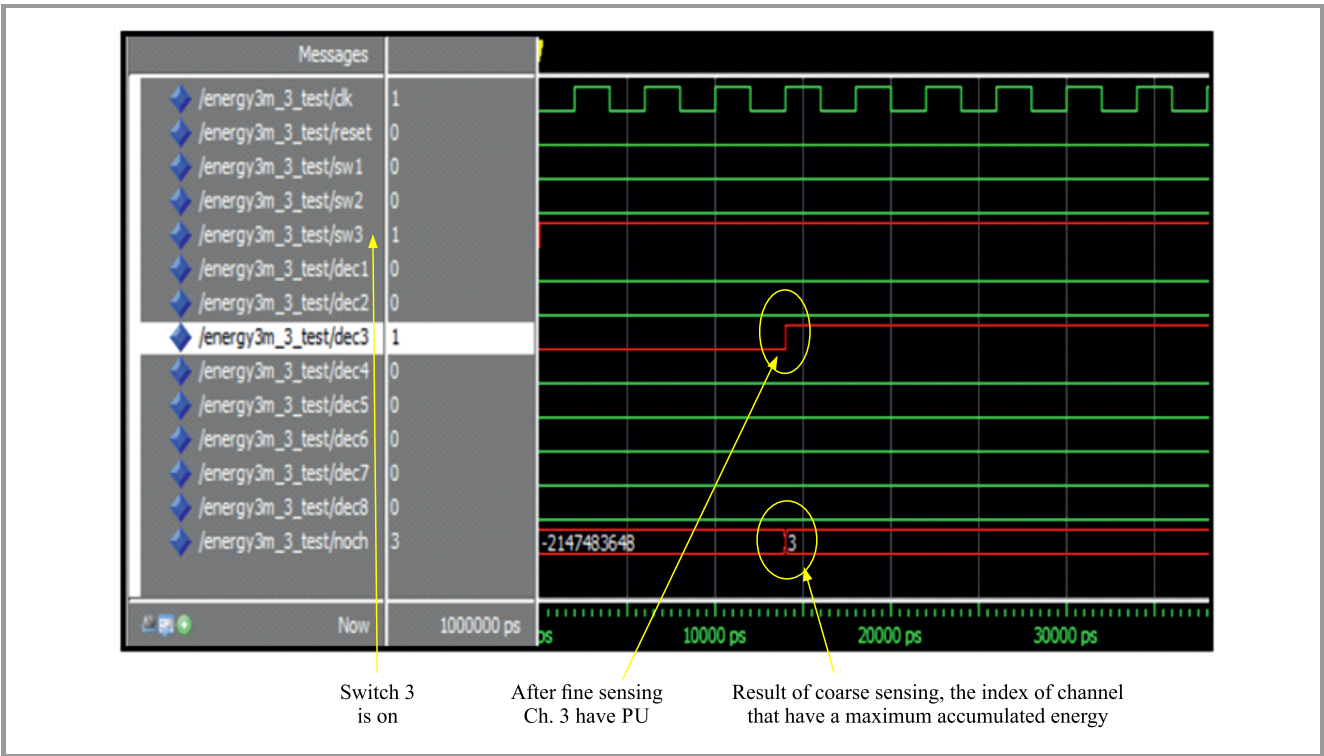that have a maximum accumulated energy

Fig. 11. ModelSim results for the second method proposed.

that improvement on the receiver's operating characteristics, by comparing them with the traditional method, using Matlab simulations. The energy consumption improvement introduced by the two methods proposed has been proved by simulation results, with no corresponding lack of de-

tection probability observed. The results of tests involving experimental implementation of the proposed methods, obtained using the Altera FPGA board, have also been confirmed by the results of simulations obtained via Matlab and ModelSim. This work can be further extended to in-

clude implementation of the fusion center for cooperative sensing, as well as to include the entire filtering and frequency transformation of the sensed channels using a single FPGA chip.

References

[1] Y. H. Chye, E. Dutkiewicz, and R. Vesilo, "Adaptive Spectrum Sensing for Cognitive Radio Systems in a Fading Environment", in *Proc. 17th Int. Symp. on Wirel. Pers. Multim. Commun. WPMC 2015*, Sydney, Australia, 2015, pp. 451–456 (doi: 10.1109/WPMC.2014.7014861).

[2] M. Matinmikko *et al.*, "Cognitive radio: An intelligent wireless communication system", Research Rep. No. VTT-R-02219-08, Technical Research Center of Finland (VTT), 2008 [Online]. Available: https://www.vtt.fi/inf/julkaisut/muut/2008/CHESS_Research_Report.pdf

[3] S. Althunibat and F. Granelli, "On results' reporting of cooperative spectrum sensing in cognitive radio networks", *J. of. Telecommun. Syst.*, vol. 62, no. 3, pp. 569–580, 2015 (doi: 10.007/s11235-015-0095-5).

[4] S. J. Kim, G. Li, and G. B. Giannakis, "Minimum-delay spectrum sensing for multi-band cognitive radios", in *Proc. IEEE Global Telecommun. Conf. GLOBECOM 2010*, Miami, FL, USA, 2010 (doi: 10.1109/GLOCOM.2010.5683914).

[5] S. J. Kim and G. B. Giannakis, "Sequential and cooperative sensing for multi-channel cognitive radios", *IEEE Trans. on Sig. Process.*, vol. 58, no. 8, pp. 4239–4253, 2010 (doi: 10.1109/TSP.2010.2049106).

[6] S. J. Kim and G. B. Giannakis, "Rate-optimal and reduced-complexity sequential sensing algorithms for cognitive OFDM radios", *EURASIP J. on Adv. in Sig. Process.*, vol. 2009, no. 2, Article ID 421540, 2009 (doi: 10.1155/2009/421540).

[7] Q. Zou, S. Zheng, and A. H. Sayed, "Cooperative sensing via sequential detection", IEEE Trans. on Sig. Process., vol. 58, no. 12, pp. 6266–6283, 2010 (doi: 10.1109/TSP.2010.2070501).

[8] S. Maleki and G. Leus, "Censored truncated sequential spectrum sensing for cognitive radio networks", *IEEE J. on Selec. Areas in Commun.*, vol. 31, no. 3, pp. 364–378, 2013 (doi: 10.1109/JSAC.2013.130304).

[9] S. Srinu and S. L. Sabat, "FPGA implementation of spectrum sensing based on energy detection for cognitive radio", in *Proc. IEEE Int. Conf. on Commun. Control and Comput. Technol. ICCCCT 2010*, Ramanathapuram, India, 2010 (doi: 10.1109/ICCCCT.2010.5670540).

[10] H. N. Abdullah and H. S. Abed, "Improvement of energy consumption in cognitive radio networks using an efficient coarse-fine sensing method", in *Proc. Int. Conf. on Change, Innov., Inform. and Disruptive Technol. ICCIIDT'16*, London, United Kingdom, 2016, pp. 218–230.

[11] M. Emara *et al.*, "Spectrum sensing optimization and performance enhancement of cognitive radio networks", *Wirel. Pers Commun.*, vol. 86, no. 2, pp. 925–941, 2015 (doi: 10.1007/s11277-015-2962-5).

[12] H. N. Abdullah and H. S. Abed, "Improvement of energy consumption in cognitive radio by reducing the number of sensed samples", in *Proc. of IEEE Al-Sadiq Int. Conf. on Multidisc. in IT and Commun. Sci. and Technol. AIC-MITC 2016*, Baghdad, Iraq, 2016, pp. 301–306 (doi: 10.1109/AIC-MITCSA.2016.7759954).

[13] H. N. Abdullah and H. S. Abed, "Energy consumption control in co-operative and non-cooperative cognitive radio using variable spectrum sensing sampling", *J. of Telecommun., Electron. and Comp. Engin. (JTEC)*, vol. 8, no. 9, pp. 7–12, 2016.

[14] S. Maleki, A. Pandharipande, and G. Leus, "Energy-efficient distributed spectrum sensing for cognitive sensor networks", *IEEE Sensor J.*, vol. 11, no. 3, pp. 565–573, 2011 (doi: 10.1109/JSEN.2010.2051327).

[15] H. N. Abdullah and H. S. Abed, "Improvement of energy consumption in spectrum sensing cognitive radio networks using an efficient two stage sensing method", *Acta Polytechnica*, vol. 47, no. 4, pp. 235–244, 2017 (doi: 10.14311/AP.2017.57.0235).

[16] S. Atapattu, "Analysis of energy detection in cognitive radio networks", Ph.D. Thesis, Department of Electrical and Computer Engineering, University of Alberta, 2013 (doi: 10.7939/R33X37).

Hikmat N. Abdullah obtained his B.Sc. in Electrical Engineering and M.Sc. in Communication Engineering from the University of Al-Mustansiryah, Iraq, in 1995 and 1998, respectively, and Ph.D. in Communication Engineering from the University of Technology, Iraq, in 2004. From 1998 to 2015 he worked as a lecturer/associate professor at the Electrical Engineering Department of Al-Mustansiryah University, Iraq. Since the beginning of 2015 he has been working as an associate professor/full professor in the college of Information Engineering, at Al-Nahrain University, Iraq. In the period of 2011–2013 he received research awards from the International Institute of Education (IIE/USA) at Bonn-Rhein-Sieg University of Applied Sciences, Germany. He is interested in wireless communication systems and cognitive radio networks.
E-mail: dr.h.abdullah@ieee.org
College of Information Engineering
Al Nahrain University
Baghdad, Iraq

Hadeel S. Abed obtained her B.Sc. from the Information and Communication Engineering Department at the Al-Khawarizmi College of Engineering, University of Baghdad, Iraq, in 2014, and M.Sc. in Network Engineering and Internet Technologies from the University of Al Nahrain, Baghdad, Iraq in 2017. Since 2016 she has been working as a part time lecturer at the Information and Communication Engineering Department of the Al-Khawarizmi College of Engineering, University of Baghdad. Her research interests include cognitive radio networks.
E-mail: hs_eng1992@yahoo.com
College of Information Engineering
Al Nahrain University
Baghdad, Iraq

Optimal Spectrum Sensor Assignment in Multi-channel Multi-user Cognitive Radio Networks

Nandini K. S.[1] and S. A. Hariprasad[2]

[1] *Jain University, Bangalore, India*
[2] *School of Engineering and Technology, Jain University, Bangalore, India*

Abstract—Accurate detection of spectrum holes is the most important and critical task in any cognitive radio (CR) communication system. When a single spectrum sensor is assigned to detect a specific primary channel, then the detection may be unreliable because of noise, random multipath fading and shadowing. Also, even when the primary channel is invisible at the CR transmitter, it may be visible at the CR receiver (the hidden primary channel problem). With a single sensor per channel, a high and consistently uniform level of sensitivity is required for reliable detection. These problems are solved by deploying multiple heterogeneous sensors at distributed locations. The proposed spectrum hole detection method uses cooperative sensing, where the challenge is to properly assign sensors to different primary channels in order to achieve the best reliability, a minimum error rate and high efficiency. Existing methods use particle swarm optimization, the ant colony system, the binary firefly algorithm, genetic algorithms and non-linear mixed integer programming. These methods are complex and require substantial pre-processing. The aim of this paper is to provide a simpler solution by using simpler binary integer programming for optimal assignment. Optimal assignment minimizes the probability of interference which is a non-linear function of decision variables. We present an approach used to linearize the objective function. Since multiple spectrum sensors are used, the optimal constrained assignment minimizes the maximum of interferences. While performing the optimization, the proposed method also takes care of the topological layout concerned with channel accessibility. The proposed algorithm is easily scalable and flexible enough to adapt to different practical scenarios.

Keywords—channel accessibility matrix, cognitive radio network, optimal assignment, probability of interference, secondary users, spectrum sensing.

1. Introduction

Cognitive Radio Networks (CRNs) have become attractive solutions that are suitable for different applications [1]–[4]. At present, with a few exceptions, almost the entire radio spectrum available is regulated and allocated exclusively to licensed users who are called primary users (PUs), i.e. TV broadcasters, mobile communication service providers. In general, the radio spectrum is not fully utilized by its PUs. The unused spectrum bands with respect to time, space and frequency, are called spectrum holes [5]. These spectrum holes are utilized by secondary users (SUs) for communication-related purposes.

In a CRN, SUs have to detect the presence or the absence of a primary radio transmission within a specified channel, in order to identify spectrum holes. Efficient sensing of spectrum holes is an important prerequisite in a CRN. Spectrum hole sensing by measuring the received energy level over a certain time interval is a well-known basic method [6]. The received signal strength depends on existing noise levels, which are random and may vary considerably. Therefore, statistical methods for efficient spectrum sensing are used here.

When cooperative sensing is used to sense multiple primary channels, optimal matching of secondary sensors to primary channels is an optimization problem encountered in the permutation space. Several articles have already been published in this regard. In [7], binary particle swarm optimization (BPSO) is used. This method provides channel assignment that maximizes the total bandwidth utilization by SUs. BPSO is basically an iterative algorithm that may take relatively more time to achieve convergence. The binary firefly (BF) algorithm is used in [8]. Here, both bandwidth utilization and fairness among SUs are fitness functions that are maximized. In BF, authors use the coding method to reduce the search space. This is also an iterative algorithm. The ant colony system is described in [9]. The genetic algorithm is used in [10]–[11]. Non-linear mixed integer programming is used in [12]. A survey of various mathematical programming methods for solving channel selection in CRNs is given in [13].

2. Basic CRN Model

Consider a multi-channel multi-user CRN, over a certain geographical area, with M licensed primary transmitters (PTs) designated as PT(1), PT(2), ..., PT(M). The non-overlapping radio frequency channels used by these M PTs are respectively designated as ch(1), ch(2), ..., ch(M).

Each transmitter uses a single primary channel. Each PT covers a certain spatial region designated as its coverage area as shown in Fig. 1. When PT(j) covers area(j) using ch(j), it means that the primary and secondary receivers within this area can receive the transmitted signal with a sufficient signal-to-noise ratio (SNR) for successful reception. In this context, when the radio signal from PT(j) covers area(j), we also say that ch(j) covers area(j) for $j = 1$ to M.
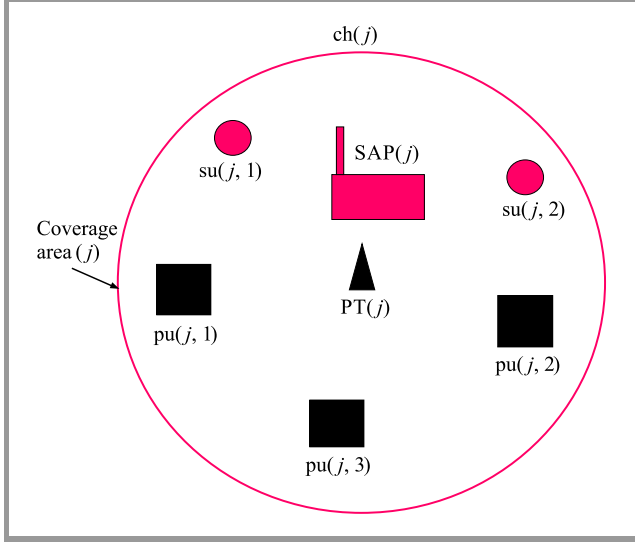
Fig. 1. Layout of a primary wireless channel ch(j) with SUs and SAP(j).

In Fig. 1, the coverage area(j) is indicated by the enclosing circle designated as ch(j). SUs within coverage area(j) are su(j, 1), su(j, 2). pu(j, 1), pu(j, 2) and pu(j, 3) are the PUs. We assume that the SUs within ch(j) are served by their own secondary access point SAP(j), for $j = 1$ to M.
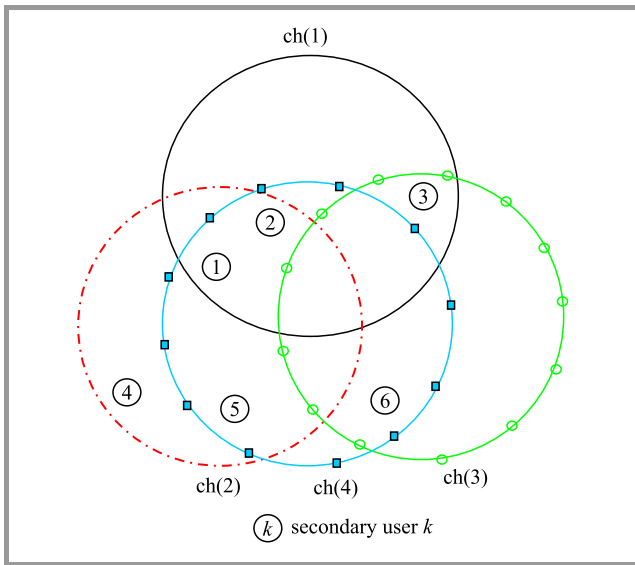
Fig. 2. Primary channels covering the SUs. (For color pictures visit www.nit.eu/publications/journal-jtit)

2.1. Multi-channel Multi-user Layout

A multi-channel multi-user layout is shown in Fig. 2. The areas covered by individual channels are overlapping, while their frequency bands are non-overlapping. The SUs (also called nodes) are globally identified by 1, 2, ..., N and is simply the integer set $\{1 : N\}$. We use su(k) to denote the SU whose ID is k. The SUs are assumed to be static. In Fig. 2, the number of primary wireless channels is $M = 4$, and the number of secondary users is $N = 6$. The secondary access points are not shown here.

The subset of SUs covered by channel ch(j) is denoted by SS(j) for $j = 1$ to M. The subsets of SUs covered by different channels (subnets) are shown in Table 1. An SU can belong to more than one channel. Thus, in Fig. 2 SUs 1 and 2 belong to ch(1), ch(2) and ch(4).

Table 1
Channels and their corresponding subsets

Channel	Boundary color	Subset of SUs
ch(1)	Black ——	SS(1) = {1, 2, 3}
ch(2)	Red —·—·—	SS(2) = {1, 2, 4, 5}
ch(3)	Green —○—	SS(3) = {3, 6}
ch(4)	Blue —■—	SS(1) = {1, 2, 5, 6}

2.2. Channel Model

A specific wireless communication channel is a radio frequency band with center frequency f_c and a bandwidth BW. Each wireless transmitter has an isotropic antenna. We assume that the primary transmission power and its antenna gain are known directly or indirectly, so that the coverage area of each transmitter is made available to the secondary network. The propagation delay is neglected. Received power is calculated using the simplified two-ray ground propagation loss model [14] as:

$$Pr(d) = Pr(1) \cdot d^{-4} , \qquad (1)$$

where $Pr(0)$ is the received power in watts at a distance of 1 m from the source, d is the distance from the source to the receiver in meters, and $Pr(d)$ is the received power in watts. We assume that the SNR is above the minimum threshold of the receivers within the coverage area. For simplicity, we assume that all devices are static and the parameters of the system do not change with time.

2.3. Secondary Users or Cognitive User Nodes

These nodes are equipped with Software Defined Radio (SDR) [15], where transceivers can dynamically adjust their radio communication parameters. Each secondary node is fitted with special spectrum sensing circuits (spectrum sensors) to detect the presence or the absence of available primary channels.

2.4. Channel Accessibility Matrix

At a specified time t, an SU may have access to several primary channels. This information is represented by the Channel Accessibility Matrix $\mathbf{CA}(t)$ [16], sized $M \times N$. The rows represent the channels, while the columns represent the SUs. $\mathbf{CA}(t)$ is a binary matrix and its element $ca(j, k, t)$ at row j and column k at time t is defined as:

$$ca(j, k, t) = \begin{cases} 1 & \text{if su}(k) \text{ is within the communication} \\ & \text{range of ch}(j) \\ 0 & \text{otherwise} \end{cases} \quad . \quad (2)$$

For the configuration shown in Fig. 2, matrix $\mathbf{CA}(t)$ is given by the values as shown in Table 2. The element $\mathbf{CA}(t)$ depends on the geographical locations of SUs and on the locations of primary base stations. For non-mobile SUs, $\mathbf{CA}(t)$ is independent of t. Then, we use the symbol $\mathbf{CA}$ instead of $\mathbf{CA}(t)$. Matrix $\mathbf{CA}$ represents the topological information.

Table 2
Channel accessibility matrix elements for the topology shown in Fig. 1

	su(1)	su(2)	su(3)	su(4)	su(5)	su(6)
ch(1)	1	1	1	0	0	0
ch(2)	1	1	0	1	1	0
ch(3)	0	0	1	0	0	1
ch(4)	1	1	0	0	1	1

2.5. Channel Sensing Model

For a given primary channel, we introduce two mutually exclusive and exhaustive hypotheses H(0) and H(1) as:

- H(0) – the channel is not occupied by the PU and it is available for SUs,

- H(1) – the channel is being engaged by the PU and is not available for SUs.

The received signal sample at the specific SU is given by [6]:

$$y(i) = \begin{cases} \eta(i) & : \text{H}(0) \\ s(i) + \eta(i) & : \text{H}(1) \end{cases} \quad . \quad (3)$$

Here, $\eta(i)$ is the noise sample, $s(i)$ is the received signal sample and $y(i)$ is the resultant, for $i = 1, 2, \ldots, n$, where n is the number of signal samples. The decision variable W at the sensing receiver after normalization of $y(i)$'s is given by [6]:

$$W = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{y(i)}{\sigma} \right)^2 \quad . \quad (4)$$

Here, σ is the standard deviation of the noise samples. Let T be the detection threshold, which is assumed to be known. If $W > T$, then the SU declares the presence of the primary channel, else its absence. Because of the dynamic nature of the radio propagation environment and the ambient noise, $W > T$ is a probabilistic event.

2.6. Probability of Detection and Probability of Interference

The conditional probability of detection under H(1) is defined as:

$$P_D = \text{prob}\big(W > T | H(1)\big) \quad . \quad (5)$$

The conditional probability of interference with the primary channel [6] is:

$$P_I = \text{prob}\big(W \le T | H(1)\big) \quad . \quad (6)$$

P_I is the probability that the spectrum sensor at that SU decides that the primary channel is free while it is actually occupied. P_I is same as the probability of misdetection [17]. Note that $P_D + P_I = 1$. The ideal case is $P_D = 1$ and $P_I = 0$. A large value of P_I means more interference with the PU. The P_I value of a specific SU with respect to a specific PTs depends on the statistical properties of W, which depend on several factors, such as distance between the secondary sensor and the transmitter, transmitter power, radio environment, etc.

2.7. Probability of Interferences in Multi-channel Multi-user Layout

Our multi-channel multi-user optimization problem has M primary channels (transmitters) and N SUs. In general, $N \ge M$. The spatial layout is such that multiple SUs can access signals from multiple PTs, as shown in Fig. 3.
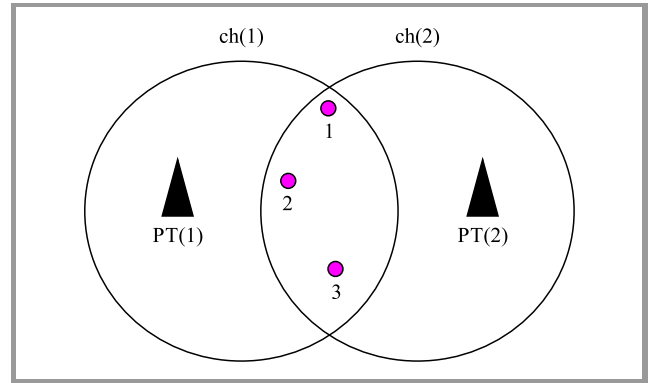

Fig. 3. Multi-channel multi-user layout ($M = 2$, $N = 3$).

Let $P_I(j, k)$ represent the probability of interference with the PT(j) by SU(k). This occurs when the SU(k) fails to detect ch(j) under H(1). Therefore, from Eqs. (5) and (6):

$$P_I(j, k) = 1 - P_D(j, k) \quad . \quad (7)$$

We assume that the $P_I(j, k)$'s values are known (either by calculations or estimation) for $j = 1$ to M and $k = 1$ to N. The collection of $P_I(j, k)$'s in a matrix format is denoted by the $\mathbf{PI}$ matrix, where j is the row and k is the column.

Example 1. As an example, $P_I(j, k)$ values for $M = 2$ and $N = 3$ are shown in a matrix form in Table 3.
In this example, all sensors can access all primary channels, and all $\mathbf{CA}$ elements are ones.

<table>
<tr><td colspan="4" align="center">Table 3
Example of a PI matrix</td></tr>
</table>

(j,k)	1	2	3
1	0.45	0.1	0.2
2	0.35	0.3	0.4

To improve channel detection capability, one or more sensors can be assigned to detect a single channel [1]. Let $SS(j)$ be a set of sensors simultaneously assigned to sense $ch(j)$, same as $PT(j)$. Here, $SS(j)$ is a subset of SUs. Theoretically, all possible simultaneous assignments will form the power set of $\{1:N\}$, that is, $\mathcal{P}(\{1:N\})$. The number of subsets in a power set is 2^N. Therefore, $SS(j)$ can be any one of the possible subsets out of 2^N, except for the empty set. These $SS(j)$ sensors send their sensed information to a fusion centre (FC). The FC determines the overall probability of detection of $ch(j)$ based on *hard decision combining*. There are several combining rules, such as AND, OR, the majority rule [1], etc. In the presented scheme, we use the OR combining rule.

We assume that sensor detection events are statistically independent. Therefore, the probability that all simultaneously assigned sensors fail to detect $ch(j)$, denoted by the *overall probability of interference*, represented by $OPI(j)$, is the joint probability due to the SUs that belong to the subset $SS(j)$. Hence, $OPI(j)$ can be represented using the multiplication rule for independent events as:

$$OPI(j) = PI\big[j, u(1)\big] \cdot PI\big[j, u(2)\big] \cdot \ldots \cdot PI\big[j, u(|SS(j)|)\big] , \tag{8}$$

where $u(1), u(2), \ldots, u(|SS(j)|)$ are the members of set $SS(j)$, and $|SS(j)|$ is the size of $SS(j)$ or the number of SUs in $SS(j)$.

Since $0 \leq PI(j,k) \leq 1$ for all j's and k's, the product term $OPI(j)$ also lies between 0 and 1. Thus, $OPI(j)$ is non-negative.

Equation (8) can be concisely expressed as:

$$OPI(j) = \prod_{u(i) \in SS(j)} PI\big[j, u(i)\big] . \tag{9}$$

3. Basic Objective and Constraints

The objective is to minimize the worst case interference at PTs due to SUs. The worst case scenario occurs when $OPI(j)$ is at its maximum, and the optimal assignment is one that minimizes the maximum of OPI. This is a minimax assignment problem [17]–[21].

From Eq. (9) one can see that the $OPI(j)$ value depends on the composition of $SS(j)$. The objective is to choose the $SS(j)$ set optimally, so that the maximum of $OPI(j)$ over j is minimized. Therefore, the objective function to be minimized can be expressed as:

$$OF_1 = \max_{j \in \{1:M\}} OPI(j) . \tag{10}$$

3.1. Assignment Matrix **X**

We introduce the assignment matrix **X** of size $M \times N$ with $N > M$. Its elements $x(j,k)$ are defined for $j = 1$ to M and $k = 1$ to N as:

$$x(j,k) = \begin{cases} 1 & \text{if } su(k) \text{ is assigned to } PT(j) \\ 0 & \text{otherwise} \end{cases} . \tag{11}$$

The objective is to determine **X** to minimize the objective function given by Eq. (10).

3.2. Constraints on **X**

There are several constraints on $x(j,k)$.

Removal of unassignable SUs: When $ca(j,k) = 0$, that assignment is impossible because $su(k)$ is outside the coverage area of $PT(j)$. To take care of this condition, we make $x(j,k) = 0$ (no assignment) when $ca(j,k) = 0$ and $x(j,k) = 0$ or 1 (can take any value) when $ca(j,k) = 1$. This can be expressed by the constraint as:

$$\big[1 - ca(j,k)\big] \cdot x(j,k) = 0 , \tag{12}$$

for $j = 1$ to M and $k = 1$ to N.

Equation (12) can be expressed in a matrix notation as:

$$(1 - \mathbf{CA}) \cdot \mathbf{X} == 0 . \tag{13}$$

Here " . *" represents the element-wise multiplication in the Matlab notation.

Row sum of X: The number of 1s in row j of **X** represents the number of SUs assigned to $PT(j)$. We have to assign at least one SU (sensor) to each j, otherwise we cannot monitor that channel. Of course, we can assign more than one to reduce the probability of interference. Therefore, the row-sum of **X** should be greater than or equal to one:

$$\sum_{k=1}^{N} x(j,k) \geq 1 \quad \text{for} \quad j = 1 \text{ to } M . \tag{14}$$

This constraint can be expressed, using the matrices, as:

$$\mathbf{X} * \mathbf{E}_{N \times 1} \geq \mathbf{F}_{M \times 1} , \tag{15}$$

where $\mathbf{E}_{N \times 1}$ and $\mathbf{F}_{M \times 1}$ are column vectors of all 1s of size $N \times 1$ and $M \times 1$. The corresponding transposes are:

$$(\mathbf{E}_{N \times 1})^T = [1, 1, \ldots, 1] ,$$

$$(\mathbf{F}_{M \times 1})^T = [1, 1, \ldots, 1] ,$$

where N number of 1's, and M number of 1's.

Column sum of X: An SU assigned to, say $PT(j)$, has to monitor $PT(j)$ regularly. It cannot monitor other channels. Therefore, not more than one $ch(j)$ or $PT(j)$ can be assigned to each SU. In addition, when $N > M$, some SUs may be left out of assignment because of some other con-

siderations. Therefore, a sensor may be assigned either to one primary channel or to none, as:

$$\sum_{j=1}^{M} x(j,k) \leq 1 \quad \text{for} \quad k = 1 \text{ to } N > M \ . \tag{16}$$

This means that any column cannot contain more than one 1, i.e. the sum of columns is either 0 or 1. The constraint represented by inequality (16) can be expressed in the matrix notation as:

$$(\mathbf{F}_{M \times 1})^T * \mathbf{X} \geq (\mathbf{E}_{N \times 1})^T \ . \tag{17}$$

Linearization of the objective function: The objective function given by Eq. (10) contains OPI(j) which is a product term given by Eq. (8). Therefore, it is non-linear and its optimization is complex and may not converge. Hence, we convert the non-linear objective function into an equivalent linear format by taking the logarithm of OPI(j). From Eq. (8):

$$\log\big[\text{OPI}(j)\big] = \log\big[\text{PI}\big(j, u(1)\big)\big] + \log\big[\text{PI}\big(j, u(2)\big)\big]$$
$$+ \log\big[\text{PI}\big(j, u(|\text{SS}(j)|)\big)\big] \ . \tag{18}$$

Since OPI(j) is non-negative, $\log\big[\text{OPI}(j)\big]$ is a monotonically increasing function of OPI(j). Therefore, maximization or minimization of OPI(j) is the same as maximization or minimization of $\log\big[\text{OPI}(j)\big]$, respectively. That is, if V is a vector of real numbers, $\log\big[\max(V)\big] = \max\big[\log(V)\big]$. Hence, the objective function given by Eq. (10) can be reformulated as:

$$\text{OF}_2 = \max_{j \in \{1:M\}} \big[\log\big(\text{OPI}(j)\big)\big] \ . \tag{19}$$

In the light of Eq. (10), Eq. (19) can be expressed as,

$$\text{OF}_2 = \log(\text{OF}_1) \ . \tag{20}$$

Obviously, minimization of OF_2 implies the minimization of OF_1.

Objective function in terms of X: Consider row j of matrix $\mathbf{X}$. The ones represent the SUs assigned to that PT(j). If $x(j,k) = 0$ for certain k, then that PI(j,k) term is absent on the right hand side (RHS) of Eq. (18). On the other hand, if $x(j,k) = 1$, then, PI(j,k) term is present on the RHS. Therefore, Eq. (18) can be expressed as:

$$\log\big[\text{OPI}(j)\big] = \sum_{k=1}^{N} x(j,k) \cdot \log\big[\text{PI}(j,k)\big] \ , \tag{21}$$

for $i = 1$ to M.

Let us introduce matrix $\mathbf{Y}$ of size $M \times N$ whose elements are $y(j,k)$ to represent the product term on the RHS of Eq. (21) as:

$$y(j,k) = x(j,k)\log\big[\text{PI}(j,k)\big] \ , \tag{22}$$

for $j = 1$ to M and $k = 1$ to N. We can see that the RHS of Eq. (22) is an element by element product of matrices $\mathbf{X}$ and $\log(\mathbf{PI})$. Therefore Eq. (22) implies:

$$Y = X. * \log(\mathbf{PI}) \ . \tag{23}$$

Substituting Eq. (22) in Eq. (21) gives:

$$\log\big(\text{OPI}(j)\big) = \sum_{k=1}^{N} y(j,k) \ , \tag{24}$$

for $j = 1$ to M.
The RHS of Eq. (24) is the sum of row j of matrix $\mathbf{Y}$. Therefore it can be expressed as:

$$\sum_{k=1}^{N} y(j,k) = Y(j,:) * \mathbf{E}_{N \times 1} \ . \tag{25}$$

In Eq. (24), $\mathbf{Y}(j,:)$ represents row j of $\mathbf{Y}$ in Matlab notation.
From Eqs. (25) and (24):

$$\log\big(\text{OPI}(j)\big) = Y(j,:) * \mathbf{E}_{N \times 1} \tag{26}$$

for $j = 1$ to M. Let *OPI* be the column vector made up of OPI(j)'s for $j = 1$ to M. Then Eq. (26) can be rewritten as:

$$\log(OPI) = Y * \mathbf{E}_{N \times 1} \ . \tag{27}$$

Now, $\max_{j \in \{1:M\}}\big[\log\big(\text{OPI}(j)\big)\big]$ is the same as the maximum of the column vector $\log(OPI)$. Therefore, Eq. (20) can be expressed as:

$$\text{OF}_2 = \max\big[\log\big(OPI\big)\big] \ . \tag{28}$$

From Eqs. (26) and (27):

$$\text{OF}_2 = \max\big(\mathbf{Y} * \mathbf{E}_{N \times 1}\big) \ . \tag{29}$$

Substituting $\mathbf{Y}$ from Eq. (23) in (29), we get OF_2 in terms of $\mathbf{X}$ as:

$$\text{OF}_2 = \max\big[\big(\mathbf{X}. * \log(\mathbf{PI})\big) * \mathbf{E}_{N \times 1}\big] \ . \tag{30}$$

Now, the optimization problem can be stated.
Minimize the objective function over the binary assignment matrix $\mathbf{X}$, where the objective function given by Eq. (30) is:

$$\text{OF}_2 = \max\big[\big(\mathbf{X}. * \log(\mathbf{PI})\big) * \mathbf{E}_{N \times 1}\big]$$

Minimize OF_2
subjected to: $(1 - \mathbf{CA}). * \mathbf{X} == 0$, as given by Eq. (13),
$\mathbf{X} * \mathbf{E}_{N \times 1} \geq \mathbf{F}_{M \times 1}$, as given by Eq. (15),
$(\mathbf{F}_{M \times 1})^T * \mathbf{X} \leq (\mathbf{E}_{N \times 1})^T$, as given by Eq. (17) .

This mini-max optimization problem can be solved using different methods mentioned in Section 1. Here, the binary integer programming is applied using the *intlinprog*() function from the optimization tool box in Matlab [22].

Once the optimal $\mathbf{X}_{opt}$ which minimizes OF_2 is obtained, the best assignment is found based on Eq. (11) as:

$$\left. \begin{array}{ll} su(k) \text{ is assigned to } PT(j), & \text{if } x_{\text{opt}}(j,k) = 1 \\ \text{No assignment} & , \text{ if } x_{\text{opt}}(j,k) = 0 \end{array} \right\} . \quad (31)$$

Once $\mathbf{X}_{\text{opt}}$ is found, the optimal value of $\mathbf{OPI}$ from Eqs. (27) and (23) can be obtained:

$$\log(\mathbf{OPI}_{\text{opt}}) = \mathbf{Y}_{\text{opt}} \cdot \mathbf{E}_{N \times 1} = \left[\mathbf{X}_{\text{opt}} . * \log(\mathbf{PI}) \right] \cdot \mathbf{E}_{N \times 1} . \quad (32)$$

Then the optimized $\mathbf{OPI}_{\text{opt}}$ is:

$$\mathbf{OPI}_{\text{opt}} = \exp\left[\left(\mathbf{X}_{\text{opt}} . * \log(\mathbf{PI}) \right) \cdot \mathbf{E}_{N \times 1} \right] . \quad (33)$$

The proposed algorithm called multi assignment is given as Algorithm 1 and is shown by examples 2 and 3.

Algorithm 1: Multi assignment

Input: CRN layout. Values of M and N with $N \leq M$, channel accessibility matrix $\mathbf{CA}$ and the probability interference matrix $\mathbf{PI}$.

Output: Optimal assignment matrix $\mathbf{X}_{\text{opt}}$ and the minimized maximum overall probability of interference $\mathbf{OPI}_{\text{opt}}$.

1. Formulate the multi-assign problem using Eqs. (30), (13), (15), and (17).
2. Solve multi-assign problem to get $\mathbf{X}_{\text{opt}}$ using binary integer non-linear programming provided by YALMIP.
3. Get the optimal assignment, using Eq. (31).
4. Get the optimal overall probability of interference, $\mathbf{OPI}_{\text{opt}}$ using Eq. (33).
5. Done

In example 2, $M = 4$ and $N = 6$. The $\mathbf{CA}$ matrix is taken as all ones. $\mathbf{PI}$ matrix of size $M \times N$ is taken (assumed to be given) as shown in Table 4.

Table 4
$\mathbf{PI}$ values of example 2

(j, k)	1	2	3	4	5	6
1	0.424	0.209	0.499	0.393	0.103	0.431
2	0.322	0.279	0.309	0.203	0.229	0.235
3	0.045	0.288	0.372	0.228	0.012	0.338
4	0.198	0.364	0.318	0.294	0.372	0.141

Here, $(\mathbf{E}_{N \times 1})^T = [1, 1, 1, 1, 1, 1]$ and $(\mathbf{F}_{M \times 1})^T = [1, 1, 1, 1]$. On solving the multi-assign problem for this $\mathbf{PI}$, using YALMIP we get the optimal assignment matrix $\mathbf{X}_{\text{opt}}$ whose elements are found to be:

$$\mathbf{X}_{\text{opt}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} .$$

The optimal assignment for $ch(j)$ is given by the locations of 1's in row j of matrix $\mathbf{X}_{\text{opt}}$. Thus, the optimal

assignment is $\left\{ \big(PT(1) \leftarrow 5 \big), \big(PT(2) \leftarrow 2, 3 \big), \big(PT(3) \leftarrow 1 \big), \big(PT(4) \leftarrow 4, 6 \big) \right\}$. The value of $\mathbf{OPI}_{opt}$ is found to be:

$$\mathbf{OPI}_{\text{opt}} = \begin{bmatrix} \mathbf{0.103} \\ 0.087 \\ 0.045 \\ 0.042 \end{bmatrix} .$$

Here the minimized max value of the objective function is 0.103. Any other combination would result in a higher value for the overall interference.

In example 3 the values are same as in example 2, except $\mathbf{CA}$, taken as shown in Table 2. Now, $\mathbf{X}_{\text{opt}}$ and the corresponding value of $\mathbf{OPI}_{\text{opt}}$ are found to be:

$$\mathbf{X}_{\text{opt}} = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{OPI}_{\text{opt}} = \begin{bmatrix} \mathbf{0.109} \\ 0.103 \\ 0.126 \\ 0.074 \end{bmatrix} .$$

Here, the minimized max value of objective function is 0.209.

4. Alternative Objectives and Constraints

4.1. Minimizing the Total Number of Assigned SUs

Consider a case where the aim is to minimize the total number of SUs assigned, subjected to the condition that $\max(\mathbf{OPI})$ should be less than a certain given upper bound, say OPIUB. This condition can be expressed as:

$$\max(\mathbf{OPI}) \leq (\text{OPIUB}) . \quad (34)$$

Taking logarithm on both sides, Eq. (33) is rewritten as:

$$\log\left[\max(\mathbf{OPI})\right] \leq \log(\text{OPIUB}) . \quad (35)$$

Since log is a monotonic function, Eq. (27) can be expressed as:

$$OF_2 = \max\left[\log(\mathbf{OPI})\right] = \log\left[\max(\mathbf{OPI})\right] . \quad (36)$$

From Eqs. (36) and (35):

$$OF_2 \leq \log(\text{OPIUB}) . \quad (37)$$

From Eqs. (30) and (37):

$$\max\left[(\mathbf{X} . * \log(\mathbf{PI})) * \mathbf{E}_{N \times 1}\right] \leq \log(\text{OPIUB}) , \quad (38)$$

where constraint (38) represents (34) in terms of $\mathbf{X}$.

Now, the optimization problem is to minimize OF_3 where:

$$OF_3 = (\mathbf{F}_{M \times 1})^T * \mathbf{X} * (\mathbf{E}_{N \times 1}) \quad (39)$$

subjected to: $(1 - \mathbf{CA}) . * \mathbf{X} == 0$, as given by Eq. (13),
$\mathbf{X} * \mathbf{E}_{N \times 1} \geq \mathbf{F}_{M \times 1}$, as given by Eq. (15),
$(\mathbf{F}_{M \times 1})^T * \mathbf{X} \leq (\mathbf{E}_{N \times 1})^T$, as given by Eq. (17),
$\max\left[\mathbf{X} . * \log(\mathbf{PI})\right] \leq \log(\text{OPIUB})$, as given by Eq. (38) .

In Eq. (39), OF_3 simply gives the total number of 1's in $\mathbf{X}$ which is the total number of SUs *assigned* to monitor the primary channels (note that the total number of *available* SUs is N). Here, the minimization of OF_3 is called the min OF_3 method. Minimization of OF_3 also means the minimization of total energy consumption by the SUs.

In example 4, the $M = 4$, $N = 11$, $\mathbf{CA} =$ all ones. OPIUB is taken as 0.02. Then, the solution for $\mathbf{X}$ is found to be:

$$\mathbf{X}_{\text{opt}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The actual optimized max($\mathbf{OPI}$) is found to be 0.0192. The minimum number of SUs assigned is found to be 8. As the OPIUB value is increased (relaxed), the minimum number of SUs required to achieve that target decreases.

In example 5, the $M = 4$, $N = 17$, $\mathbf{CA} =$ all ones. OPIUB is varied from 0.01 to 0.10 in steps of 0.01. The minimum number of SUs needed, designated as min SUs, is calculated using our method min OF_3 and the lazy method. In the lazy method, the SUs are assigned as they are available in the order su(1), su(2), and so on.

Table 5
Min SUs for different values of OPIUB

OPIUB		.01	.02	.03	.04	.05	.06	.07	.08	.09
Min SUs	Min OF_3	10	8	8	7	6	6	6	6	6
	Lazy method	16	14	13	13	11	11	9	9	7

The relationship between the min SUs needed and the OPIUB values is shown in Table 5 and the corresponding bar graph is shown in Fig. 4.
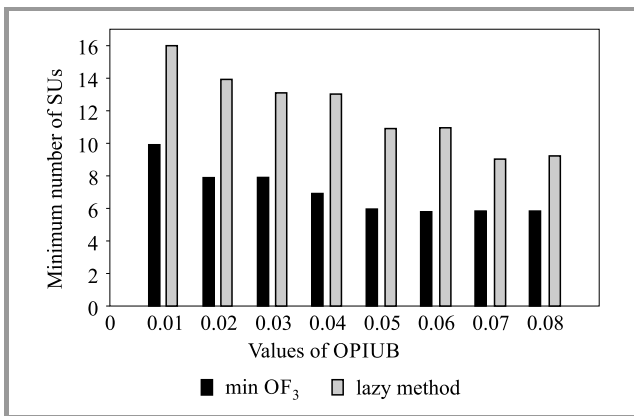
Fig. 4. Minimum number of SU's versus OPIUB.

In the case of min OF_3, the decrease in the minimum number of SUs could not fall below 6 because of other constraints (15) and (17). From Table 5, we see that min OF_3

is much better compared to the lazy method, in terms of the min SUs needed to meet the target.

4.2. Minimization of OF_2

In the next step, the minimization of OF_2 with pre-fixed number of SU's for each primary channel is analyzed. In this case, constraint (15) which fixes the number of SUs for each channel can be reframed as:

$$\mathbf{X} * \mathbf{E}_{N\times 1} = \left[g(1), g(2), \ldots, g(M) \right]^T .* \mathbf{F}_{M\times 1} . \quad (40)$$

Here, $g(j)$ is the number of SUs assigned to the primary transmitter PT(j), for $j = 1$ to M. Obviously, to satisfy Eq. (39), sum of $g(i)$'s should be less than or equal to N. Other conditions and objective functions are same as (13), (17) and (30).

5. Comparison with other Methods

Minimization OF_2, as given by Eq. (30) can be solved by the greedy method [23], [24]. We compare the greedy method method with our multi-assign method. Here, $M = 4$ and $\mathbf{CA} =$ all ones. The number of available SUs N is varied from 10 to 17. The calculated values of minimized OF_2 (minimized max($\mathbf{OPI}$)) for these N's are shown in Table 6 and the corresponding values are shown in Fig. 5.

Table 6
Variation of Min OF_2 as N varies from 10 to 17

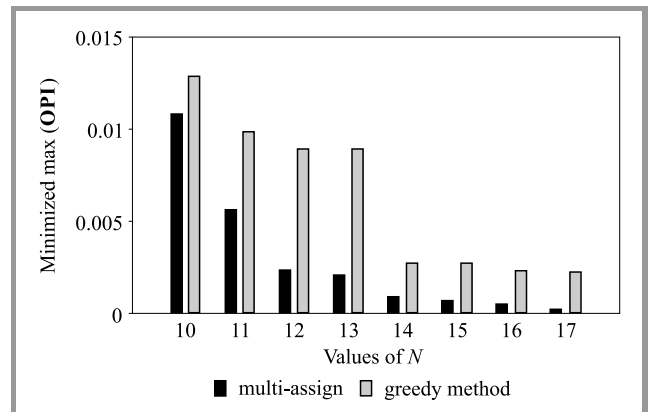| N	Min OF_2	
	Multi assign	Greedy method
10	0.01088	0.01292
11	0.00566	0.00992
12	0.00239	0.00896
13	0.00208	0.00896
14	0.00093	0.00277
15	0.00071	0.00277
16	0.00052	0.00233
17	0.00023	0.00227

Fig. 5. Minimum number of SUs versus OPIUB.

From Fig. 5 we see that our multi-assign method gives lower interference and is much better than the greedy method.

Next, the multi-assign algorithm is compared with the particle swarm optimization (PSO) method and the genetic algorithm (GA) method with reference to the time taken to solve the optimization problem. We took $M = 4$ and N is varied from 6 to 15. The result is shown in Fig. 6.
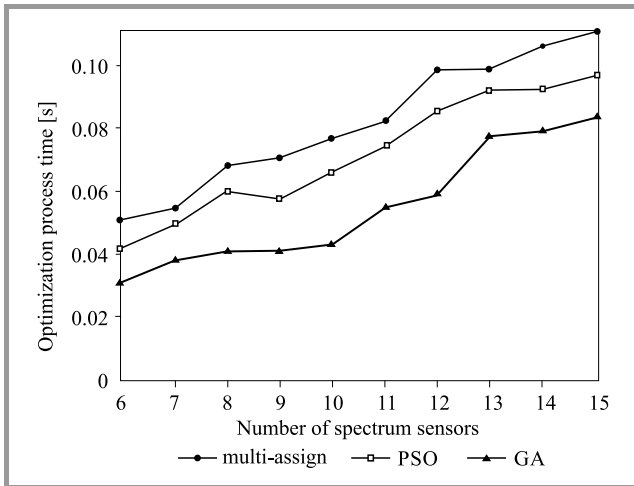

Fig. 6. Optimization process time vs. number of spectrum sensors.

From the result of Fig. 6, we find that the multi-assign algorithm is substantially faster compared to the PSO and GA algorithms. For a relatively large number of densely populated nodes, our method has not much advantage over other methods, because the complexity of our method depends on finding the best permutation of M sensors out of N.

6. Conclusion

A new method of optimal assignment, in a co-operative spectrum sensing CRN where multiple SUs and multiple primary channels are present, is described. The main contribution of this work is to convert the product term in the objective function to linear format by using the logarithm of product terms. Compared to the average of PSO and GA methods, the proposed method takes, on the average, 70% less time to calculate the optimal assignment when the number of SUs is relatively small. The minimized maximum interference value is lower in the presented method compared to the greedy algorithm or lazy methods.

References

[1] I. F. Akyildiz, B. F. Lo, and R. Balakrishnan, "Cooperative spectrum sensing in cognitive radio networks: a survey", *Phys. Commun.*, vol. 4, no. 1, pp 40–62, 2011 (doi: 101016/j.phycom.2010.12.003).

[2] I. F. Akyildiz, W. Y. Lee, M. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: a survey", *Comput. Netw.*, vol. 50, no. 13, pp. 2127–2159, 2006 (doi: 10.1016/j.comnet.2006.05.001).

[3] B. Wang and K. J. Ray Liu, "Advances in cognitive radio networks: a survey", *IEEE J. of Selec. Topics in Sig. Process.*, vol. 5, no. 1, pp. 5–23, 2011 (doi: 10.1109/JSTSP.2010.2093210).

[4] Y.-C. Liang *et al.*, "Guest editorial – cognitive radio: theory and application", *IEEE J. on Selec. Areas in Commun.*, vol. 26, no. 1, pp. 1–4, 2008 (doi: 10.1109/JSAC.2008.080101).

[5] R. Tandra, S. M. Mishra, and A. Sahai, "What is a spectrum hole and what does it take to recognize one?", *Proc. of the IEEE*, vol. 97, no. 5, pp. 824–848, 2009 (doi: 10.1109/JPROC.2009.2015710).

[6] Y. Chen, "Improved energy detector for random signals in Gaussian noise", *IEEE Trans. on Wirel. Commun.*, vol. 9, no. 2, pp. 558–563, 2010 (doi: 10.1109/TWC.2010.02.090622).

[7] H. M. Abdelsalam and A. Al-shaar, "An enhanced binary particle swarm optimization algorithm for channel assignment in cognitive radio networks", *in Proc. 5th Int. Conf. on Modell., Identif. and Control ICMIC 2013*, Cairo, Egypt, 2013, pp. 221–226.

[8] Q. Liu, W. Lu, and W. Xu, "Spectrum allocation optimization for Cognitive Radio networks using Binary Firefly Algorithm", in *Proc. of the 2014 Int. Conf. on Innov. Design and Manufact. ICIDM 2014*, Montreal, QC, Canada, 2014, pp. 257–262, 2014 (doi: 10.1109/IDAM.2014.6912704).

[9] F. Koroupi, S. Talebi, and H. Salehinejad, "Cognitive radio networks spectrum allocation: an ACS perspective", *Scientia Iranica*, vol. 19, no. 3, pp. 767–773, 2012 (doi: 10.1016/j.scient.2011.04.029).

[10] J. Elhachmi and Z. Guennoun, "Cognitive radio spectrum allocation using genetic algorithm", *EURASIP J. on Wirel. Commun. and Netw.*, pp. 133–143, 2016 (doi: 10.1186/s13638-016-0620-6).

[11] S. Chatterjee, S. Dutta, P. P. Bhattacharya, and J. S. Roy, "Optimization of spectrum sensing parameters in cognitive radio using adaptive genetic algorithm", *J. of Telecommun. and Inform. Technol.*, no. 1, pp. 21–27, 2017.

[12] R. M. Eletreby, H. M. Elsayed, and M. M. Khairy, "Optimal spectrum assignment for cognitive radio sensor networks under coverage constraint", *IET Communications*, vol. 8, no. 18, pp. 3318–3325, 2014 (doi: 10.1049/iet-com.2014.0423).

[13] A. S. Alfa, B. T. Maharaj, S. Lall, and S. Pal, "Mixed-integer programming based techniques for resource allocation in underlay cognitive radio networks: a survey", *J. of Commun. and Netw.*, vol. 18, no. 5, pp. 744–761, 2016 (doi: 10.1109/JCN.2016.000104).

[14] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Upper Saddle River, NJ: Prentice Hall PTR, 2002 (ISBN 978-0130422323).

[15] F. K. Jondral, "Software-defined radio-basic and evolution to cognitive radio", *EURASIP J. on Wirel. Commun. and Netw.*, vol. 3, pp. 275–283, 2005 (doi: 10.1155/WCN.2005.275).

[16] R. Urgaonkar and M. J. Neely, "Opportunistic scheduling with reliability guarantees in cognitive radio networks", *IEEE Trans. Mob. Comput.*, vol. 8, no. 6, pp. 766–777, 2009 (doi: 10.1109/TMC.2009.38).

[17] Y. Liang, L. Lai, and J. Halloran, "Distributed cognitive radio network management via algorithms in probabilistic graphical models", *IEEE J. on Selec. Areas in Commun.*, vol. 29, no. 2, pp. 338–348, 2011 (doi: 10.1109/JSAC.2011.110207).

[18] L. Y. Yang, M. H. Nie, Z. W. Wu, and Y. Y. Nie, "Modeling and solution for assignment problem", *Int. J. of Mathem. Models and Methods in Appl. Sci.*, vol. 2, no. 2, pp. 205–212, 2008.

[19] R. E. Burkard and E. Çela, "Linear assignment problems and extensions", in *Handbook of Combinatorial Optimization – Supplement Volume A*, and D. Z. Du and P. M. Pardalos, Eds. Boston, MA, USA: Springer, 1999, pp. 75–149 (doi: 10.1007/978-1-4757-3023-4).

[20] R. S. Garfinkel, "An improved algorithm for the bottleneck assignment problem", *Operations Res.*, vol. 19, no. 7, pp. 1747–1751, 1971 (doi: 10.1287/opre.19.7.1747).

[21] J. E. Hopcroft and R. M. Karp, "An n5/2 algorithm for maximum matching in bipartite graphs", *Soc. for Indust. and Appl. Mathem. J. of Comput.*, vol. 2, no. 4, pp. 225–231, 1973 (doi: 10.1137/0202019).

[22] MOSEK Optimization Toolbox for MATLAB [Online]. Available: http://docs.mosek.com/8.0/toolbox/intlinprog.html

[23] W. Wang and X. Liu, "List-coloring based channel allocation for open spectrum wireless networks", in *IEEE 62nd Conf. on Veh. Technol. VTC-2005-Fall*, Dallas, TX, USA, 2005 (doi: 10.1109/VETECF.2005.1558001).

[24] N. Shylashree, "Non-crossing rectilinear shortest minimum bend paths in the presence of rectilinear obstacles", *J. of Telecommun. and Inform. Technol.*, no. 3, 2018 (doi: 10.26636/jtit.2018.120417).

Nandini K. S. is a Ph.D. research scholar at Jain University, Bangalore, India. She obtained her B.Eng. and M.Eng. degrees in 2007 and 2010, respectively. She is working as an Assistant Professor at the Dept. of Electronics and Communication Engineering, RNSIT, Bangalore. She has 10 years of teaching experience and has published 3 research papers. Her area of interest is in communication and networking.
E-mail: nandinispatel@gmail.com
Jain University
Bangalore, India

S. A. Hariprasad obtained his B.Eng. in 1991 and M.Eng. in 2000. He completed his Ph.D. in 2011 and also obtained two additional D.Sc. degrees in 2013 and 2014, for post-doctoral research work on communication and embedded systems. He has published 60 papers in international and national journals/conferences and has authored a textbook on "Advanced Microprocessors". He has 22 years of teaching experience along with 9 years of research experience. He is guiding 8 Ph.D. students. Currently, he is working as a Director of the School of Engineering Technology, Jain University.
E-mail: harivat2002@yahoo.co.in
School of Engineering and Technology
Jain University
Bangalore, India

A CPW-fed Sigma-shaped
MIMO Antenna for Ka Band and
5G Communication Applications

B. T. P. Madhav, G. Jyothsna Devi, P. Lakshman, and T. Anilkumar

Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Andhra Pradesh, India

Abstract—This article presents a MIMO compact antenna measuring 45×45×1.6 mm, on the FR4 substrate, proposed for Ka band and 5G communication applications. The proposed design is suitable to overcome the issues connected with massive MIMO. It has four-sigma-shaped radiating elements and a c-shaped ground plane with coplanar waveguide feeding. Its compact dimensions suit it for most existing communications systems. The aerial operates in the 21–30 GHz range, which covers Ka and 5G communication bands. The proposed antenna exhibits the average efficiency of more than 76% within its operating band and gives a minimum signal to noise plus interference ratio. The presented antenna covers several services, such as Ka band satellite downlink applications and future 5G communication applications.

Keywords—5G, coplanar waveguide, Ka band, MIMO, satellite downlink.

1. Introduction

The unsatisfied and unresolved requirements of 4G technology gave birth to the new 5G standard. In future mobile communications, 5G is going to offer solutions to common problems, i.e. poor coverage, crowded channels, poor quality, dropped connections, etc. To achieve flexibility in communication and to fulfil the users' expectations, there is a need to move forward with 5G even further. A 5G network not only provides high speed, but also offers the capacity to carry more data with data rates of over 1 Gbps [1], [2]. Such a big capacity is going to make significant difference in communications and allows to expand the range of services offered (Fig. 1) [3], [4]. Low latency with massive connectivity and large system capacity are needed for the variety of novel applications of 5G systems [5]. 5G has the latency time on the millisecond level, and super-dense collections will improve spectral, energy as well as cost efficiency. Novel multiple access strategies, ultra-dense networking, all spectrum access, MIMO technology, full/flexible duplex and many other solutions are still under development [6]. 5G must support multiple novel applications with different requirements. The requirements include higher peaks, reduced latency, high user data

rates, increased number of devices, enhanced indoor coverage and so on. This technology will provide coverage for high-edge users [7].

The increasing demand for high speed data, together with rising congestion within the microwave spectrum are pushing for higher frequencies, including Ka, Q and V bands. The operating frequency of the Ka band is nearly 3 times higher than that of the X band [8]. This is because enhancement in satellite communications (SAT) is mainly concerned with high versatility, compactness, performance, low cost as well as low weight [9].

A fixed satellite service for wireless multimedia and internet access has generated renewed interest in the Ka band, using downlink frequencies of 17.72–21.2 GHz and uplink frequencies of 27.52–30 GHz in order to properly accommodate various future communications systems [10]. Satellite communication for the military (MKa band) offers an even wider bandwidth for high speed Internet access (30 GHz for TX and 20 GHz for RX) [11]. Satellite communication plays a major role in offering access to the Internet in rural areas. To remove high path losses from LEO and GEO frequency satellites, high-gain antennas, i.e. dish antennas of the horn type are needed [12].

2. Related Work

To achieve a high data rate and to improve signal quality and reliability, MIMO antennas are considered [13], [14]. A MIMO antenna having 2 elements for a 5G network is mentioned in [15], [16]. The Ka band single polarization MIMO satellite-earth link channel model is shown in papers [17] and [18]. A novel single layer substrate integrated waveguide corrugated technique for two high gain and low mutual coupling Ka band MIMO antennas are demonstrated in [19].

3. Massive MIMO Challenges

Massive MIMO has several advantages over conventional MIMO antennas. The throughput and radiated efficiency of
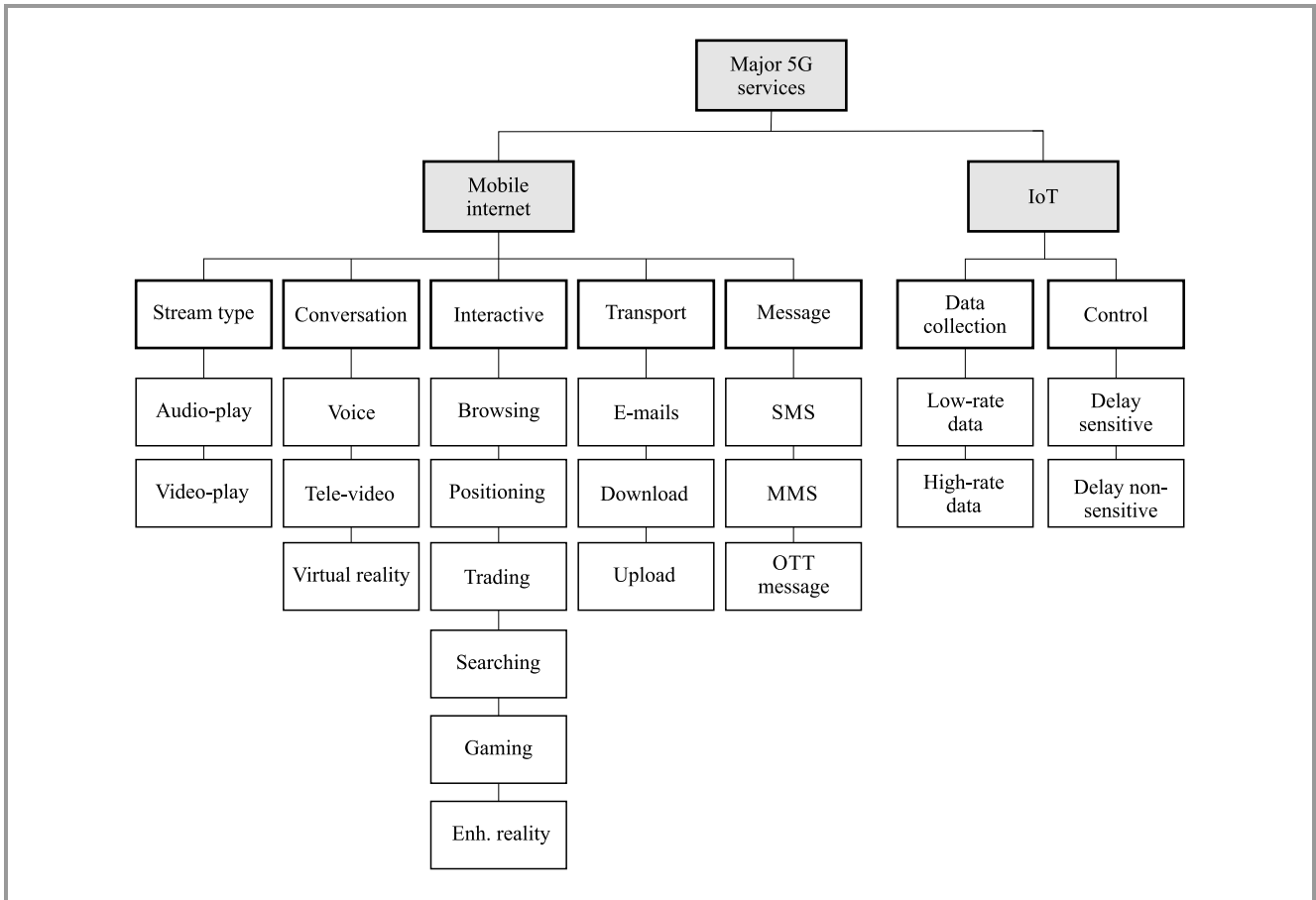
Fig. 1. Major 5G technology services.

the antenna can be improved by placing additional antennas to focus energy in each smaller region. It is an enabler to future broadband networks, or IoT applications combined with clouds.

In conventional MIMO systems like LTE, the base station sends out pilot waveforms. Based on that, channel response will be estimated and transmitted back. This is not possible in massive MIMO, at least when operating in a high mobility environment, for two following reasons:

1. The optimal downlink pilot should be mutually orthogonal between the antennas.

2. The number of channel responses each terminal must estimate should be proportional to the number of base station antennas.

Some major issues that should be addressed by massive MIMO, from the research point of view, are:

- fast and distributed coherent signal processing related downlink issues,

- the challenge of low cost hardware,

- hardware impairments,

- internal power consumption,

- channel characteristics,

- cost of reciprocity calibration,

- pilot contamination,

- non-CSI at TX,

- new deployment sceneries.

To overcome some of those issues, involving a like large coverage area, for instance, a planar MIMO antenna is designed with a 9 GHz bandwidth for Ka band and 5G applications (21–30 GHz). In comparison with massive MIMO, the proposed model is not superior with respect to throughput and efficiency, but its compact dimensions and suitable gain levels render it most suitable for the desired applications. The designed antenna comprises a sigma-shaped radiator and a c-shaped coupled ground plane with CPW feeding. Simulated and measured performance shows that the MIMO antenna plays a vital role in 5G and Ka band applications.

4. Antenna Design

Such parameters like thickness and dielectric constant of the substrate affect impedance bandwidth of the antenna.
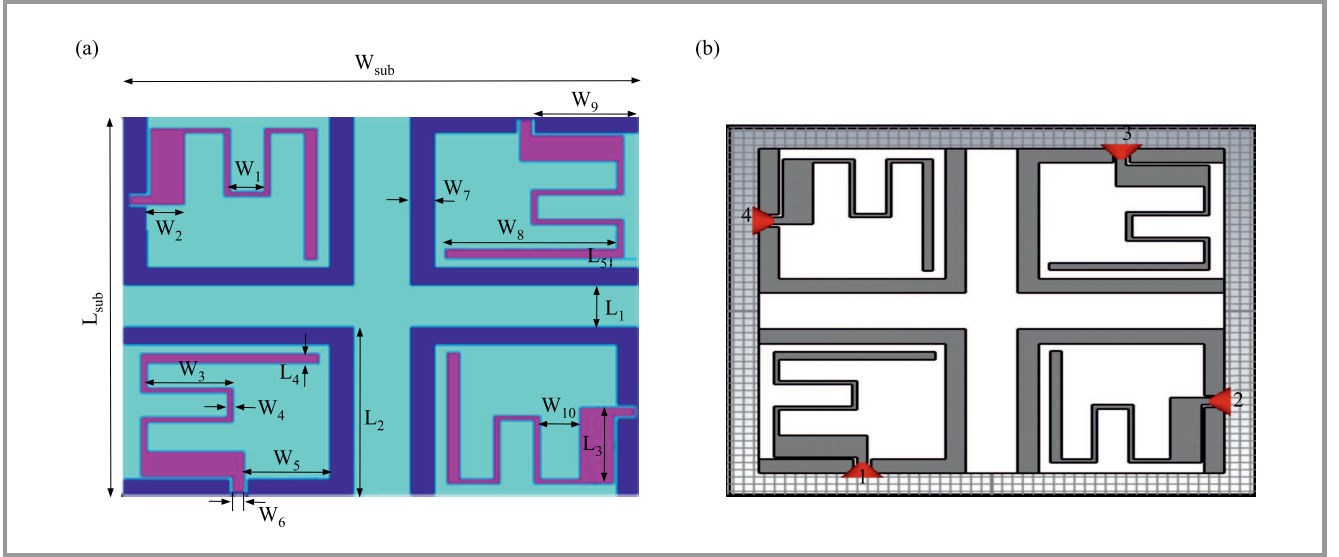
Fig. 2. MIMO antenna geometry: (a) HFSSS model, and (b) CST model.
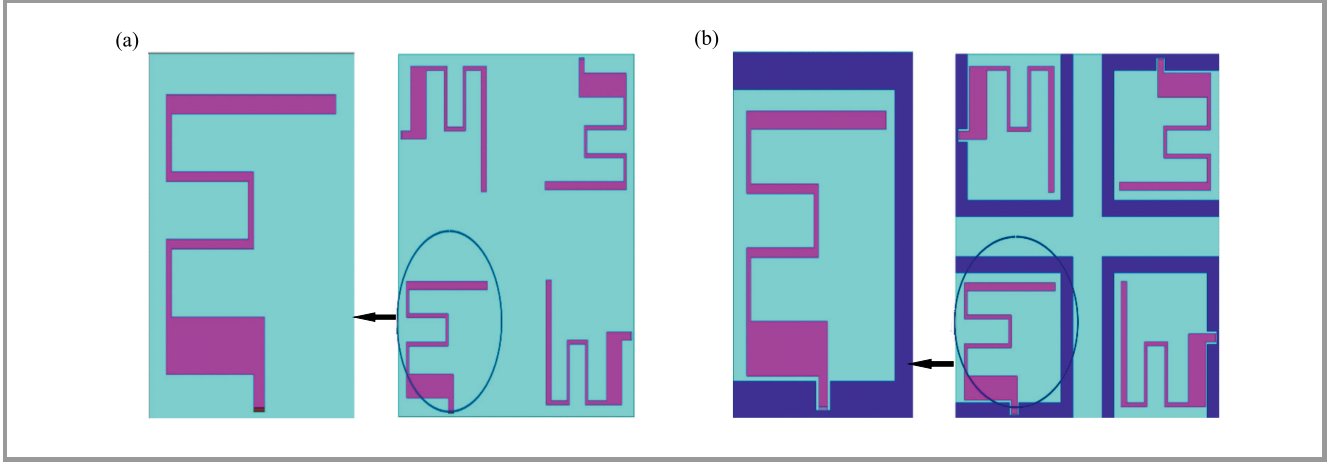
Fig. 3. MIMO antenna with: (a) patch part, and (b) patch with ground.

By increasing the height of the substrate or decreasing its permittivity, the bandwidth can be widened. The resonant frequency and impedance bandwidth will also be affected by the permittivity. Antennas manufactured on thin substrates generate more copper losses, while thicker a permittivity substrate leads to performance degradation. The configuration of the designed MIMO antenna is shown in Fig. 2. The size of the antenna proposed is $L_{sub} \times W_{sub} = 45 \times 45$ mm. A typical FR4 laminate with $\varepsilon_r = 4.3$ and $\tan \delta = 0.02$ is used. The design is composed of two parts: a sigma-shaped monopole radiator with coplanar waveguide feeding and a c-shaped ground plane. CPW feeding is used to achieve the 50 Ω impedance. When compared to microstrip line, CPW feeding has many advantages, such as easy fabrication and reduced radiation losses.

The antenna consists of four planar monopole elements with a CPW feeder through ports 1–4 respectively. The four emitters are placed perpendicularly to provide good isolation between four input ports. All sigma-shaped radiators have the same dimensions with length L_2. Each of the radiators is fed by a 50 Ω CPW line with its width equaling W6 (Fig. 3). Table 1 shows the dimensions of the design.

The design procedure starts by finding the dimension of the antenna feeder to get $Z_0 = 50$ Ω using the following equations [20]:

$$Z_0 = \frac{30\pi K^1(k)}{\sqrt{\varepsilon_e} K(k)}, \tag{1}$$

$$k = \frac{s}{s + 2w_c}, \tag{2}$$

$$\varepsilon_e = \frac{\varepsilon_r + 1}{2}(A + B), \tag{3}$$

$$A = \tanh\left(1.785 \log \frac{h}{\omega_c} + 1.75\right), \tag{4}$$

$$B = k\frac{\omega_c}{h}\left[0.04 - 0.7k + 0.01(1 - 0.1\varepsilon_r)(0.25 + k)\right], \tag{5}$$

$$K^1(k) = k\sqrt{1 - k^2}. \tag{6}$$

In Eqs. (1)–(6) s is the central conductor, w_c is the slot width of CPW, h is the substrate thickness, ε_r is the dielectric constant of the substrate laminate.

Table 1
Dimensions of the MIMO antenna design

Parameter	Value [mm]	Parameter	Value [mm]
L_{sub}	45	w3	8
L1	5	w4	0.5
L2	20	w5	9.2
L3	9	w6	1.3
L4	1	w7	2
L5	2.3	w8	15.5
W_{sub}	45	w9	9.2
w1	3	w10	3.5
w2	3	—	—

The designed MIMO antenna was analyzed and simulated using FEM-based HFSS software and CST Microwave Studio software. The comparative analysis performed with the use of both tools, as well as its results are presented in the next section.

5. Results and Discussion

This section presents HFSS. The antenna parameter S_{11} has a significant importance, which defines the impedance matching characteristics and the bandwidth. The MIMO antenna covers bandwidth of around 9 GHz (21–30 GHz)
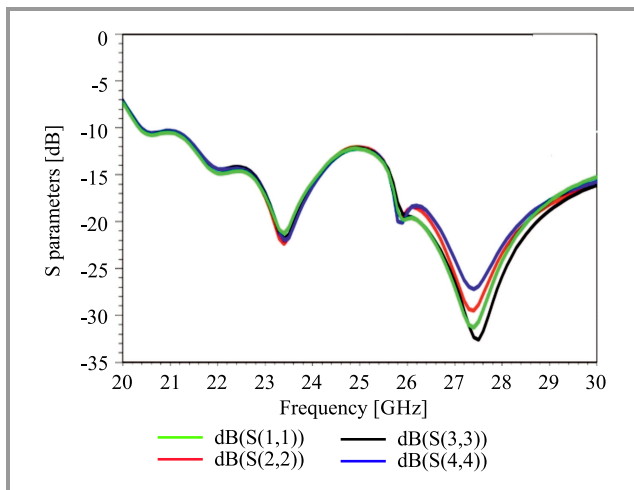
Fig. 4. Simulated reflection coefficient S_{11} of MIMO antenna. (For color pictures visit www.nit.eu/publications/journal-jtit)

with a return loss below -10 dB. The bandwidth of the antenna is satisfactory, as it covers the Ka band and 5G applications.
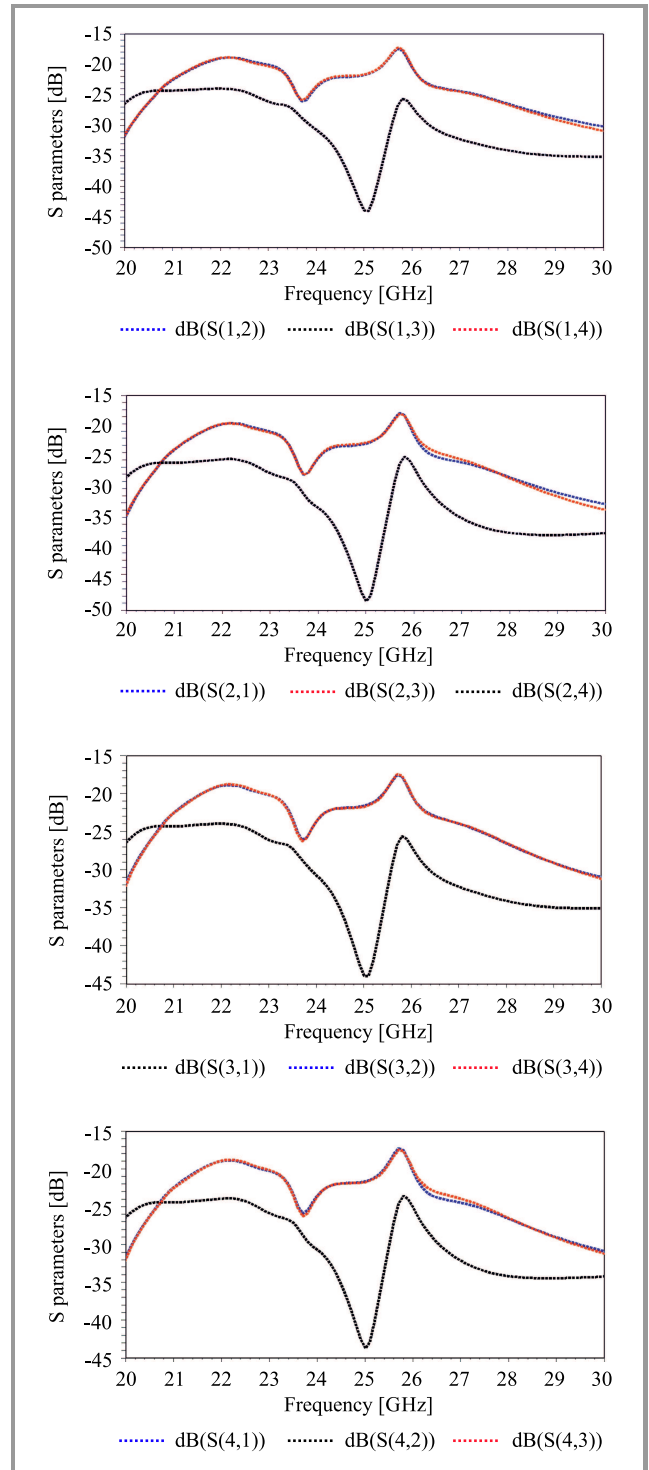
Fig. 5. Transmission coefficient result of MIMO.

The design criteria meet the satisfactory antenna return loss and bandwidth for the desired applications. The MIMO antenna has (1×4) basic elements, with the dimensions of 45×45 mm. Isolation is provided between each pair of basic elements to prevent coupling caused by transitions. The reflection and mutual coupling coefficients are is depicted in Fig. 4. The transmission coefficient values are below -15 dB, giving good significance of low mutual coupling between the elements (Fig. 5).
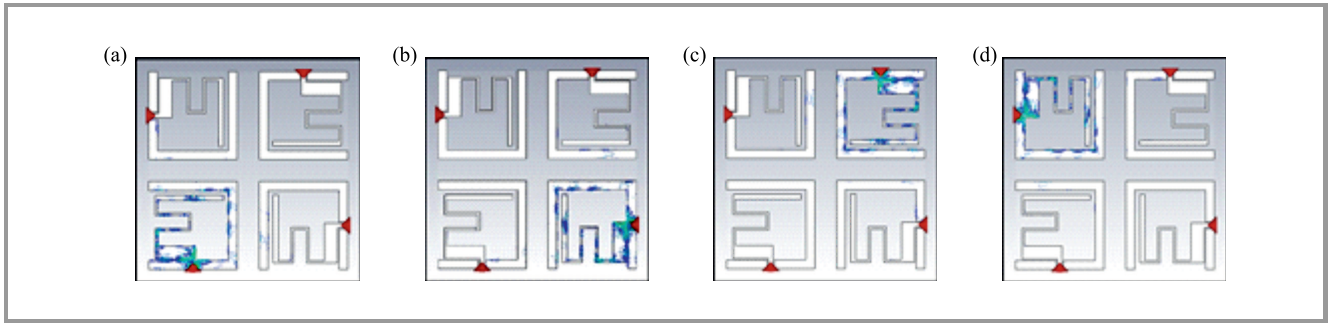
Fig. 6. Surface current distribution of a MIMO antenna at 23.5 GHz: (a) port 1 excited and other ports terminated, (b) port 2 excited, (c) port 3 excited, and (d) port 4 excited.
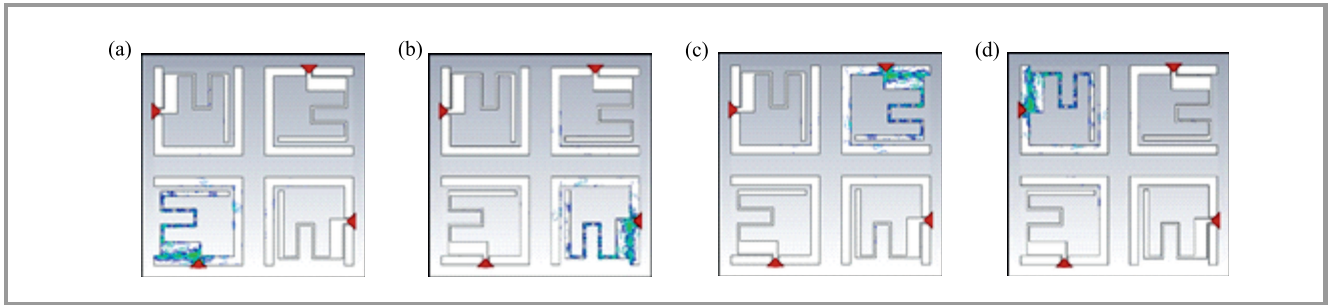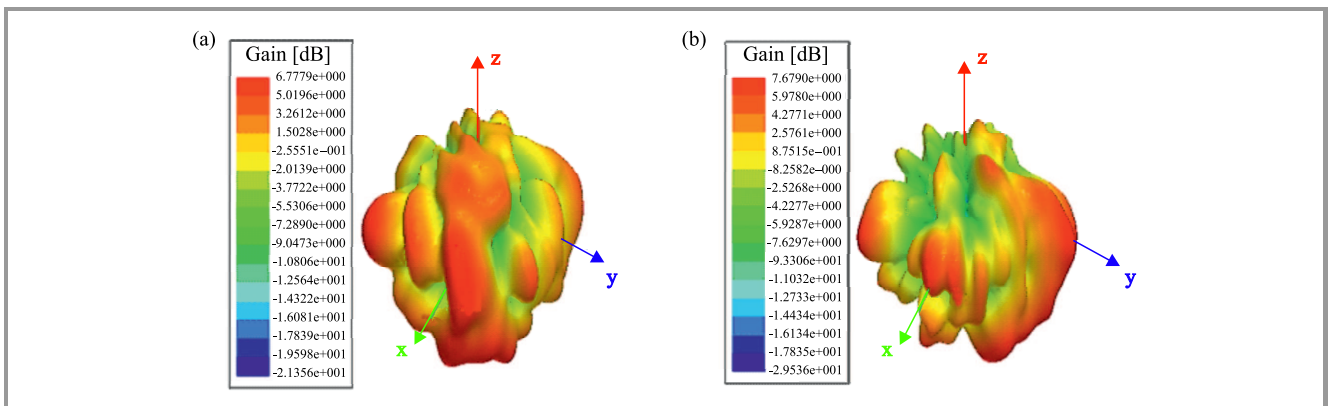

Fig. 7. Surface current distribution at 28 GHz: (a) port 1 excited and remaining ports terminated, (b) port 2 excited, (c) port 3 excited, and (d) port 4 excited.
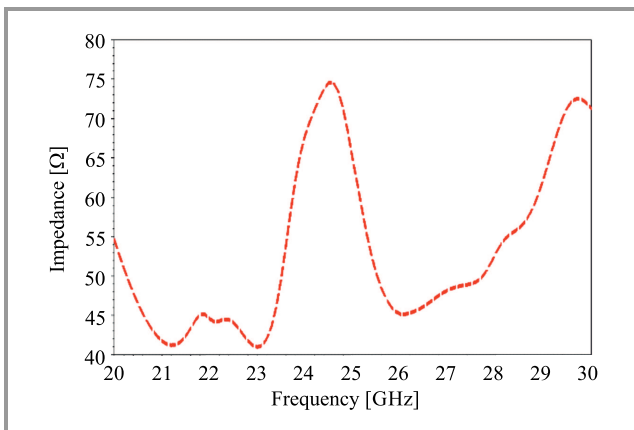
**Fig. 8.** Impedance response of the proposed MIMO antenna.

The reflection coefficients show that the elements of the proposed MIMO antenna have the average impedance bandwidth of 40%, with the reflection coefficient of less than -10 dB in 21–30 GHz. The transmission coefficients from ports 1–4 are shown in Fig. 5. Mutual coupling values are less than -15 dB over the entire band.

The current distributions at resonant frequencies of 23.5 and 28 GHz are depicted in Figs. 7 and 8. The current density on the radiating structure shows a low level of mutual coupling. When port 1 is excited and other ports are terminated, the current density is more focused on the first antenna element and very low coupling can be observed in Fig. 6a. Similarly, for other port excitations the current density on the excited port related to the radiating element and low coupling on other elements may be observed.

Fig. 9. 3D gain pattern with port 1 excited and remaining ports terminated at: (a) 23.5 GHz, and (b) 28 GHz.
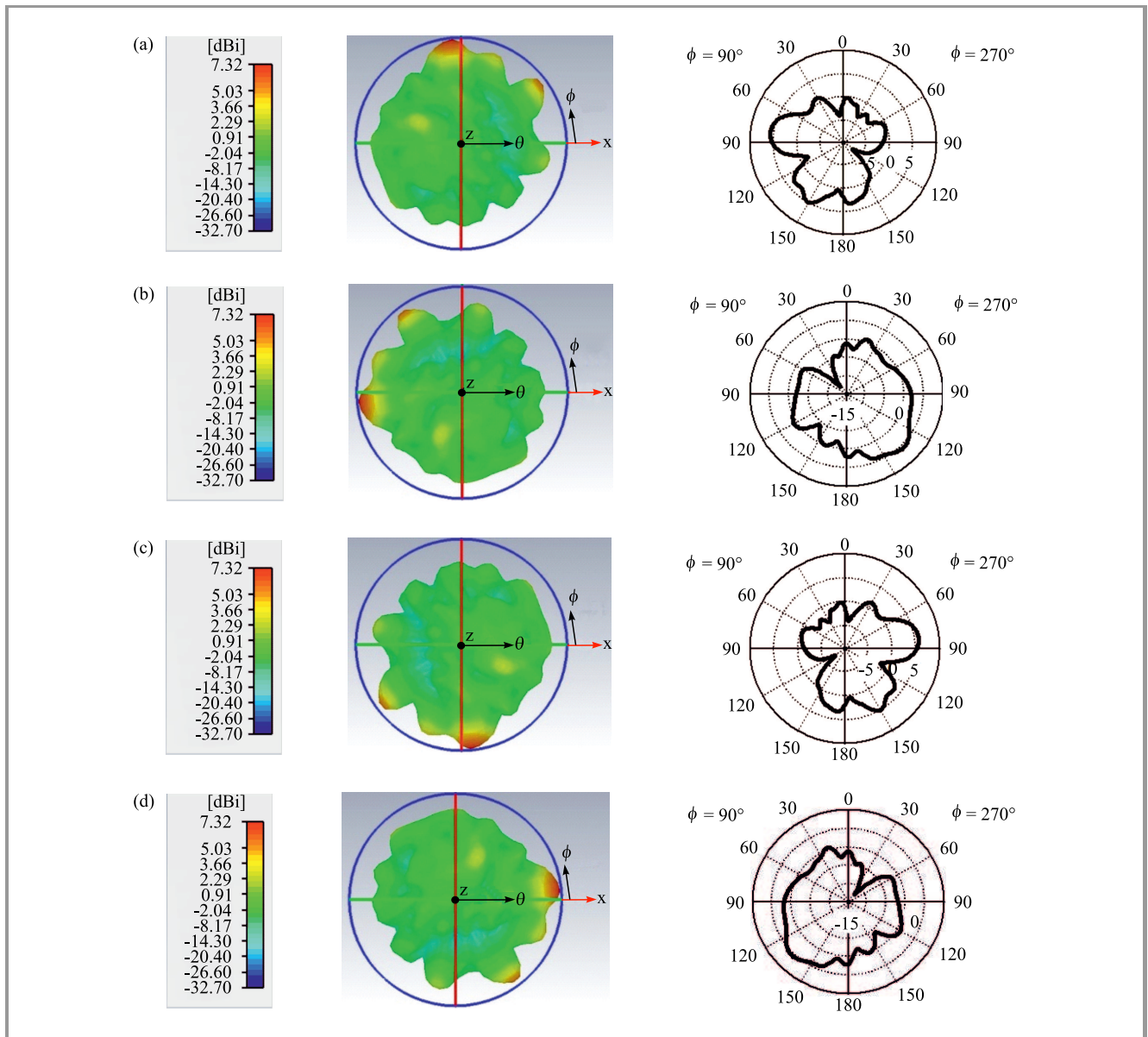
Fig. 10. Radiation pattern in 3D and polar representation at 23.5 GHz: (a) port 1 excited and remaining ports terminated, (b) port 2 excited, (c) port 3 excited, and (d) port 4 excited.

The aerial impedance at resonant frequencies is depicted in Fig. 8. The MIMO antenna shows a perfect impedance of 50 Ω at 23.5 and 28 GHz.

The MIMO antenna provides gain of over 7.6 dB within the operating range and gives minimum losses with the average efficiency of 78% over the entire band. At 23.5 and 28 GHz, the average efficiency is almost 80%. The 3D gain pattern obtained with the use of the HFSS tool is shown in Fig. 9.

To show polarization diversity performance, radiation patterns for differential ports 1–4 at 23.5 and 28 GHz are presented in Figs. 10 and 11. These patterns are almost identical with 90° rotations, which confirms the polarization diversity problem.

Figure 12 shows the frequency vs. gain and efficiency plot. The obtained gain peak is 7.3 dB at 23.5 GHz and 7.5 dB

at 28 GHz, while the average efficiency of more than 76% is achieved.

5.1. Envelope Correlation Coefficient

The envelope correlation coefficient (ECC) describes diversity-related performance. An ECC of = 0.5 is generally recognized as acceptable in various conditions. The proposed aerial provides ECC < 0.03 over its operating frequency band.
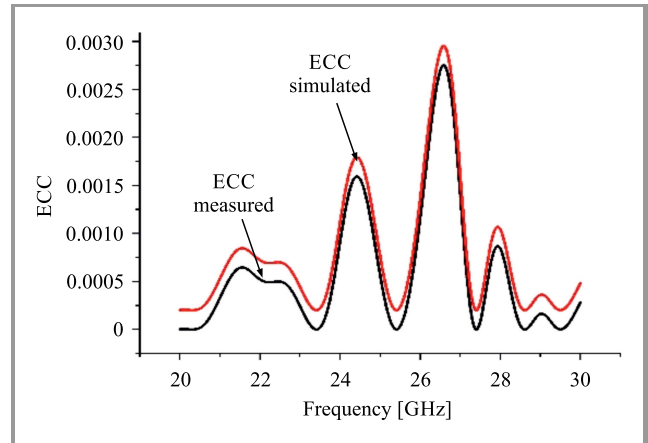
6. Parametric Analysis

The aerial dimensions were optimized using a parametric analysis. First, the antenna performance over the w2 parameter examined (Fig. 14). With w2 of 2 mm the an-
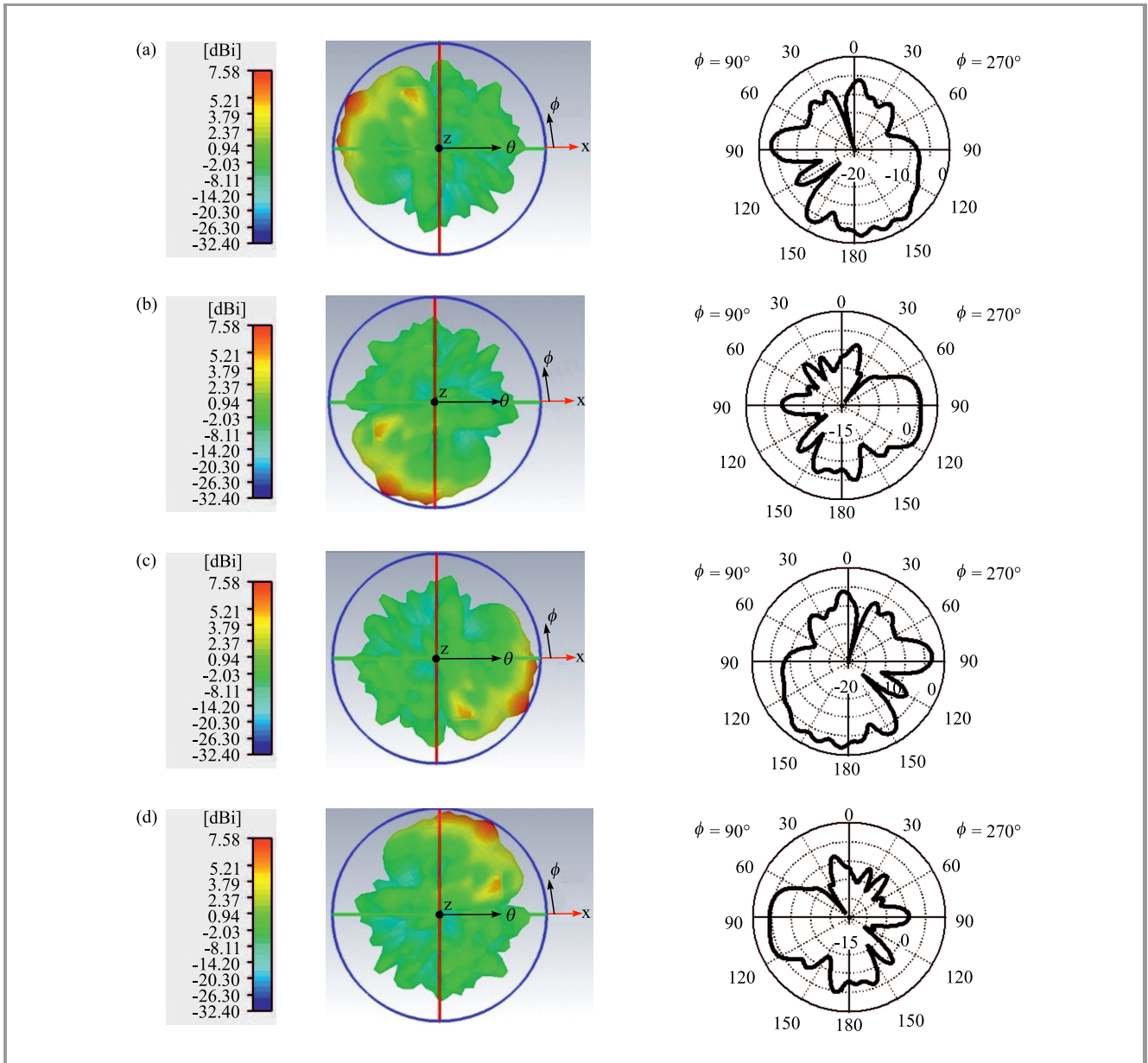
Fig. 11. Radiation pattern in 3D and polar representation at 28 GHz: (a) port 1 excited and remaining ports terminated, (b) port 2 excited, (c) port 3 excited, and (d) port 4 excited.
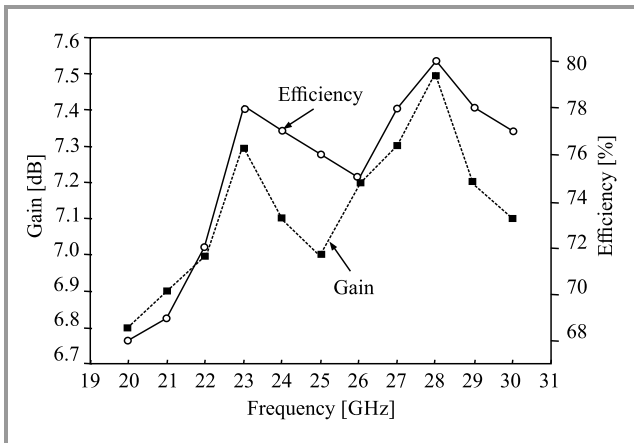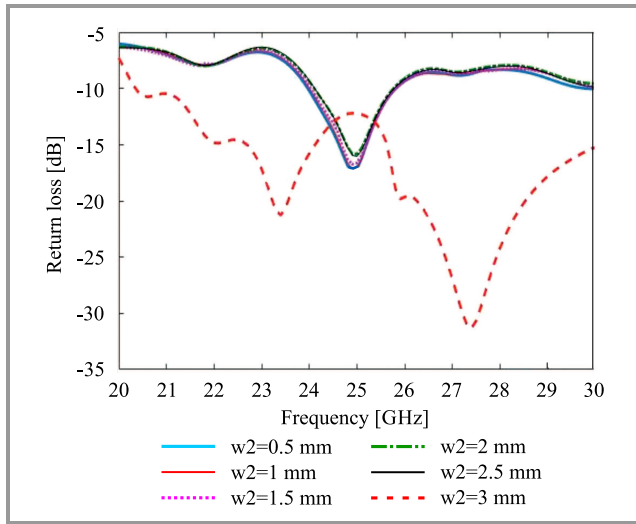

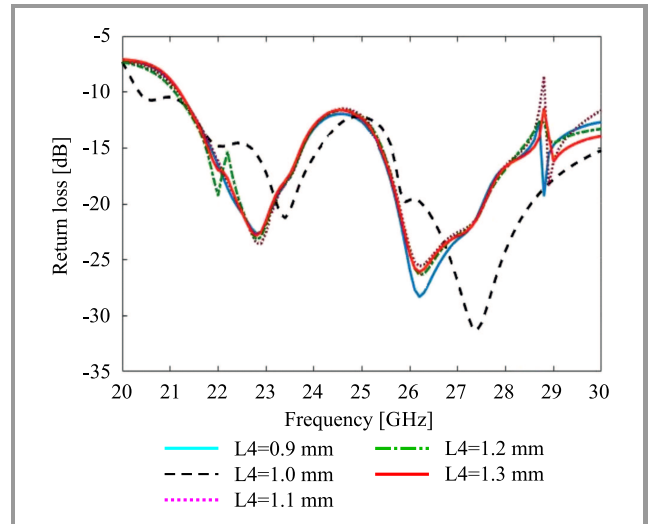
Fig. 12. Radiation efficiency and gain.

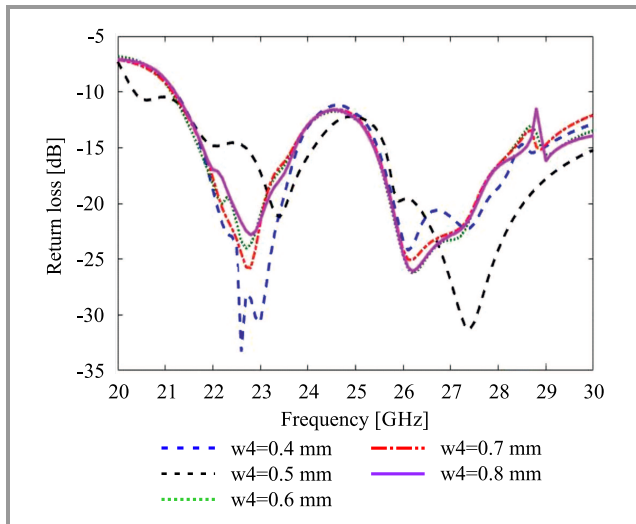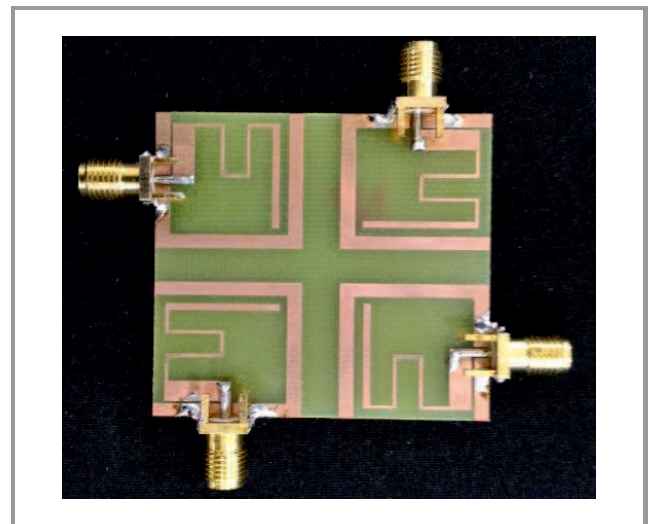**Fig. 13.** Envelope correlation coefficient.
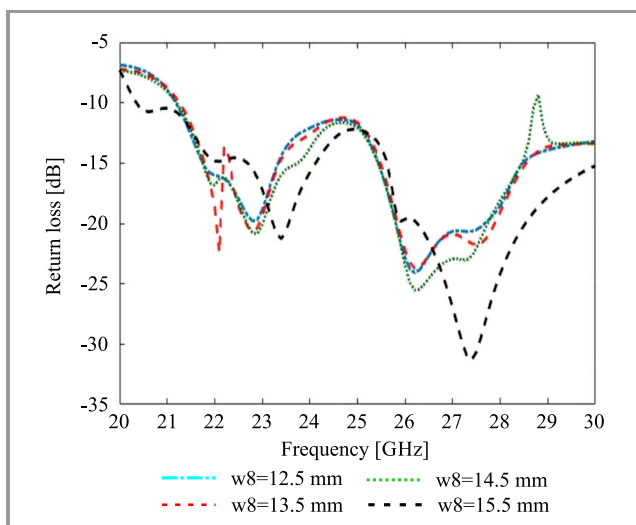
Fig. 14. Parametric analysis with change in w2.

Fig. 15. Parametric analysis with change in w4.

Fig. 16. Parametric analysis with change in w8.

Fig. 17. Parametric analysis with change in L4.

Fig. 18. MIMO antenna prototype.
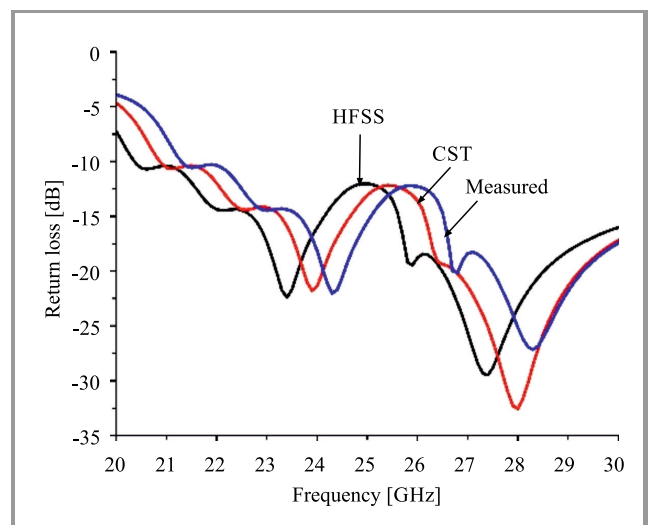
Fig. 19. Measured and simulated reflection coefficient parameter.

tenna shows superior results. Next, w4 was varied between 0.4–0.8 mm, and optimum performance was obtained at 0.5 mm (Fig. 15).

The width of radiating element's upper strip w8 was also varied from 12.5 to 15.5 mm. The optimum performance was obtained at 15.5 mm (Fig. 16). For w8 = 14.5 mm, the antenna shows a notch band from 28.5 to 29.5 GHz.

Finally, the length of the radiating element's upper strip was varied from 0.9 to 1.3 mm. For L4 = 1 mm the antenna was showing the best impedance bandwidth characteristics (Fig. 17).

The proposed MIMO antenna was prototyped on the FR4 substrate (Fig. 18), using SMA connectors soldered to all ports. The measured S_{11} reflection coefficient is compared with simulation results obtained from HFSS and CST (Fig. 19).

Table 2
Comparison of the designed MIMO
antenna with other research

Reference	Antenna dimensions [mm]	Peak gain [dB]	Average efficiency [%]	Band
[1]	$108 \times 135 \times 0.6$	7.2	70	5G
[14]	$136 \times 68 \times 1.6$	5	75	5G
[15]	$140 \times 76 \times 0.8$	3	45	5G
[16]	$130 \times 170 \times 1.6$	6.5	85	5G
[19]	$72 \times 17.2 \times 1.6$	7	68	Ka
Proposed	$45 \times 45 \times 1.6$	7.5	78	Ka and 5G

Table 2 shows a comparison between the proposed model and other antennas proposed by researchers.

7. Conclusion

The proposed antenna model provides excellent performance characteristics within the 9 GHz bandwidth, has a gain peak of 7.5 dB and achieves good polarization. Even though it is not superior to massive MIMO with respect to high-speed broadband network and advanced IoT applications, its compactness, low diversity and adaptability to advanced communication modules makes this antenna model more suitable for Ka band and 5G applications.

8. Acknowledgements

References

[1] A. Dadgarpour, M. S. Sorkherizi, and A. A. Kishk, "High efficient circularly polarized magneto-electric dipole antenna for 5G applications using dual-polarized split-ring resonator lens", *IEEE Transact. on Antenn. and Propag.*, vol. 65, no. 8, pp. 4263–4267, 2017 (doi: 10.1109/TAP.2017.2708091).

[2] Y. Yao, X. Chung, Ch. Wang, J. Yu, and X. Chen, "Wideband circularly polarized antipodal curvedly tapered slot antenna array for 5G applications", *IEEE J. on Select. Areas in Commun.*, vol. 35, no. 7, pp. 1539–1549, 2017 (doi: 10.1109/JSAC.2017.2699101).

[3] W. Hong *et al.*, "Multi-beam antenna technologies for 5G wireless communications", *IEEE Transact. on Antenn. and Propag.*, vol. 99, no. 12, pp. 6231-6249, 2017 (doi: 10.1109/TAP.2017.2712819).

[4] A. Khuran and S. R. Rosenthal, "Integrating the fuzzy front end of new product development", *MIT Sloan Manag. Rev.*, vol. 38, no. 4, pp. 103–120, 1997.

[5] Ch. Zhang, Y.-H. Huang, F. Sheikh and Zh. Wang, "Advanced baseband processing algorithms, circuits, and implementations for 5G communication", *IEEE J. on Emerging and Select. Topics in Circ. and Sys.*, vol. 7, no. 4, pp. 477–490, 2017 (doi: 10.1109/JETCAS.2017.2743107).

[6] V. Jungnickel *et al.*, "The role of small cells, coordinated multipoint, and massive MIMO in 5G", *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 44–51, 2014 (doi: 10.1109/MCOM.2014.6815892).

[7] Y. Yifei and Zh. Longming, "Application scenarios and enabling technologies of 5G", *China Commun.*, vol. 11, no. 11, pp. 69–79, 2014 (doi: 10.1109/CC.2014.7004525).

[8] R. Glogowski, J.-F. Zurcher, C. Peixeiro, and J. R. Mosig, "A low-loss planar Ka-band antenna subarray for space applications", *IEEE Transact. on Antenn. and Propag.*, vol. 61, no. 9, pp. 4549–4557, 2013 (doi: 10.1109/TAP.2013.2264452).

[9] Luo, Qi, Steven Gao, Chong Zhang, Dawei Zhou, Tobias Chaloun, Wolfgang Menzel, Volker Ziegler, and Mohammed Sobhy, "Design and analysis of a reflectarray using slot antenna elements for Ka-band SatCom", *IEEE Transact. on Antenn. and Propag.*, vol. 3, no. 4, pp. 1365–1374, 2015 (doi: 10.1109/TAP.2015.2401393).

[10] Y. Ch. Lee and Ch. S. Park, "A compact broadband PHEMT MMIC power amplifier for K through Ka-band applications", *AEU-Int. J. of Electr. and Commun.*, vol. 57, no. 4, pp. 291–294, 2003 (doi: 10.1078/1434-8411-54100174).

[11] M. R. Chaharmir and J. Shaker, "Design of a multilayer X-/Ka-band frequency-selective surface-backed reflectarray for satellite applications", *IEEE Transact. on Antenn. and Propag.*, vol. 63, no. 4, pp. 1255–1262, 2015 (doi: 10.1109/TAP.2015.2389838).

[12] B. Rohrdantz *et. al*, "An electronically scannable reflector antenna using a planar active array feed at Ka-band", *IEEE Transact. on Microwave Theory and Techniq.*, vol. 65, no. 5, pp. 1650–1661, 2017 (doi: 10.1109/TMTT.2017.2663402).

[13] D. S. Ramkiran *et al.*, "Coplanar wave guide fed dual band notched MIMO antenna", *Int. J. of Electr. and Computer Engineer.*, vol. 6, no. 4, pp. 1732–1741, 2016 (doi: 10.11591/ijece.v6i4.1057).

[14] M.-Y. Li *et al.*, "Eight-port orthogonally dual-polarized antenna array for 5G smartphone applications", *IEEE Transact. on Antenn. and Propag.*, vol. 64, no. 9, pp. 3820–3830, 2016 (doi: 10.1109/TAP.2016.2583501).

[15] M. S. Sharawi, M. Ikram, and A. Shamim, "A two concentric slot loop based connected array MIMO antenna system for 4G/5G terminals", *IEEE Transact. on Antenn. and Propag.*, vol. 65, no. 12 pp. 6679–6686, 2017 (doi: 10.1109/TAP.2017.2671028).

[16] Y.-L. Ban, Ch. Li, Ch.-Y.-D. Sim, G. Wu, and K.-L. Wong, "4G/5G multiple antennas for future multi-mode smartphone applications", *IEEE Access*, vol. 4, pp. 2981–2988, 2016 (doi: 10.1109/ACCESS.2016.2582786).

[17] J. A. Perez García, S. Kosmopoulos, and G. Goussetis, "A compact 12-way slotted waveguide power combiner for Ka-band applications", *IEEE Microwave and Wirel. Compon. Let.*, vol. 27, no. 2, pp. 135–137, 2017 (doi: 10.1109/LMWC.2016.2646903).

[18] M. V. Komandla, Gh. Mishra, and S. K. Sharma, "Investigations on dual slant polarized cavity-backed massive MIMO antenna panel with beamforming", *IEEE Transact. on Antenn. and Propag.*, vol. 65, no. 12, pp. 6794-6799, 2017 (doi: 10.1109/TAP.2017.2748239).

[19] M. Lin, P. Liu, and Zh. Guo, "Gain-enhanced Ka-band MIMO antennas based on the SIW corrugated technique", *IEEE Antenn. and Wirel. Propag. Let.*, vol. 16, pp. 3084–3087, 2017 (doi: 10.1109/LAWP.2017.2761903).

[20] J. Ch. Rao *et al.*, "Compact UWB MIMO slot antenna with defected ground structure", *ARPN J. of Engineer. and Applied Sciences*, vol. 11, no. 17, pp. 10487–10495, 2016.

B. T. P. Madhav received his B.Sc., M.Sc., MBA, M.Tech. degrees from Nagarjuna University, Andhra Pradesh, India in 2001, 2003, 2007, and 2009 respectively. He obtained his Ph.D. in the field of antennas from Konenu Lakshaiah University. Currently he is working as a Professor and Communication Systems research Head at the Electronics and Communication Engineering Department. His research interests include antennas, liquid crystal applications and wireless communications.
E-mail: btpmadhav@kluniversity.in
Department of Electronics & Communication Engineering
Koneru Lakshmaiah Education Foundation
Andhra Pradesh, India

G. Jyothsna Devi is pursuing her master's degree at the Department of Electronics & Communication Engineering of Konenu Lakshmaiah University. Her area of interest includes MIMO antennas for wireless applications, massive MIMO, compact and low profile antennas.

E-mail: jyothsnagudiseva@gmail.com
Department of Electronics & Communication Engineering
Koneru Lakshmaiah Education Foundation
Andhra Pradesh, India

Papula Lakshman received his M.Sc. degree in electronics and communication from Andhra University, Visakhapatnam, India, and his M.Tech. degree in Microwave Engineering from GITAM University, Visakhapatnam, India. He received his Ph.D. degree from IIT Bhubaneswar, Odisha, India. His research interests include antennas, array antennas, optimization methods, EMI/EMC and soft computing.
E-mail: lakshmanpappula@kluniversity.in
Department of Electronics & Communication Engineering
Koneru Lakshmaiah Education Foundation
Andhra Pradesh, India

T. Anilkumar received his B.Tech. degree in Electronics and Communication Engineering from Acharya Nagarjuna University in 2014 and M.Tech. degree in Communication Systems from Andhra University in 2016. He is currently pursuing his Ph.D. at Antennas and Liquid Crystals Research Center of ECE Department, Konenu Lakshaiah University. His current research interests include reconfigurable, tunable microstrip antennas, conformal antennas and liquid crystal polymer-based antennas for vehicular communication.
E-mail: t.anilkumar@kluniversity.in
Department of Electronics & Communication Engineering
Koneru Lakshmaiah Education Foundation
Andhra Pradesh, India

Information for Authors

Journal of Telecommunications and Information Technology **(JTIT)** is published quarterly. It comprises original contributions, dealing with a wide range of topics related to telecommunications and information technology. **All papers are subject to peer review**. Topics presented in the JTIT report primary and/or experimental research results, which advance the base of scientific and technological knowledge about telecommunications and information technology.

JTIT is dedicated to publishing research results which advance the level of current research or add to the understanding of problems related to modulation and signal design, wireless communications, optical communications and photonic systems, voice communications devices, image and signal processing, transmission systems, network architecture, coding and communication theory, as well as information technology.

Suitable research-related papers should hold the potential to advance the technological base of telecommunications and information technology. Tutorial and review papers are published only by invitation.

Manuscript. TEX and LATEX are preferable, standard Microsoft Word format (.doc) is acceptable. The authors JTIT LATEX style file is available:
http://www.nit.eu/for-authors

Papers published should contain up to 10 printed pages in LATEX authors style (Word processor one printed page corresponds approximately to 6000 characters).

The manuscript should include an abstract about 150200 words long and the relevant keywords. The abstract should contain statement of the problem, assumptions and methodology, results and conclusion or discussion on the importance of the results. Abstracts must not include mathematical expressions or bibliographic references.

Keywords should not repeat the title of the manuscript. About four keywords or phrases in alphabetical order should be used, separated by commas.

The original files accompanied with pdf file should be submitted by e-mail: redakcja@itl.waw.pl

Figures, tables and photographs. Original figures should be submitted. Drawings in Corel Draw and PostScript formats are preferred. Figure captions should be placed below the figures and can not be included as a part of the figure. Each figure should be submitted as a separated graphic file, in .cdr, .eps, .ps, .png or .tif format. Tables and figures should be numbered consecutively with Arabic numerals.

Each photograph with minimum 300 dpi resolution should be delivered in electronic formats (TIFF, JPG or PNG) as a separated file.

References. All references should be marked in the text by Arabic numerals in square brackets and listed at the end of the paper in order of their appearance in the text, including exclusively publications cited inside. Samples of correct formats for various types of references are presented below:

[1] Y. Namihira, Relationship between nonlinear effective area and mode field diameter for dispersion shifted fibres, *Electron. Lett.*, vol. 30, no. 3, pp. 262–264, 1994.

[2] C. Kittel, *Introduction to Solid State Physics*. New York: Wiley, 1986.

[3] S. Demri and E. Orłowska, Informational representability: Abstract models versus concrete models, in *Fuzzy Sets, Logics and Knowledge-Based Reasoning*, D. Dubois and H. Prade, Eds. Dordrecht: Kluwer, 1999, pp. 301–314

Biographies and photographs of authors. A brief professional authors biography of up to 200 words and a photo of each author should be included with the manuscript.

Galley proofs. Authors should return proofs as a list of corrections as soon as possible. In other cases, the article will be proof-read against manuscript by the editor and printed without the author's corrections. Remarks to the errata should be provided within one week after receiving the offprint.

Copyright. Manuscript submitted to JTIT should not be published or simultaneously submitted for publication elsewhere. By submitting a manuscript, the author(s) agree to automatically transfer the copyright for their article to the publisher, if and when the article is accepted for publication. The copyright comprises the exclusive rights to reproduce and distribute the article, including reprints and all translation rights. No part of the present JTIT should not be reproduced in any form nor transmitted or translated into a machine language without prior written consent of the publisher.
For copyright form see: http://www.nit.eu/for-authors

A copy of the JTIT is provided to each author of paper published.