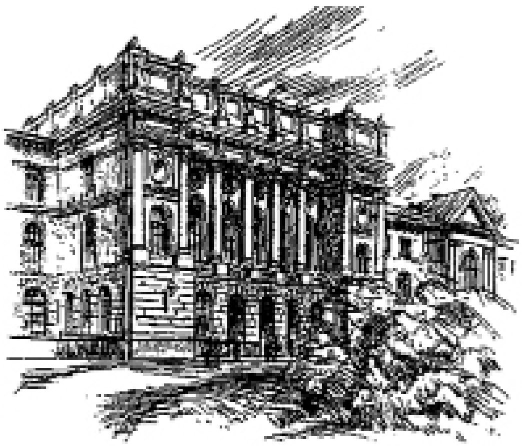
THE MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE RUSSIAN FEDERATION



ISSN 2687-0517

Computing, Telecommunications and Control

Vol. 16, No. 1
2023

Peter the Great St. Petersburg
Polytechnic University
2023

COMPUTING, TELECOMMUNICATIONS AND CONTROL

The journal is included in the List of Leading PeerReviewed Scientific Journals and other editions to publish major findings of PhD theses for the research degrees of Doctor of Sciences and Candidate of Sciences.

Open access journal is to publish articles of a high scientific level covering advanced experience, research results, theoretical and practical problems of informatics, electronics, telecommunications, and control.

The journal is indexed by Ulrich's Periodicals Directory, Google Scholar, EBSCO, ProQuest, Index Copernicus, VINITI RAS Abstract Journal (Referativnyi Zhurnal), VINITI RAS Scientific and Technical Literature Collection, Russian Science Citation Index (RSCI) database Scientific Electronic Library and Math-Net.ru databases.

The journal is registered with the Federal Service for Supervision in the Sphere of Telecom, Information Technologies and Mass Communications (ROSKOMNADZOR). Certificate ЭЛ No. ФС77-77378 issued 25.12.2019.

Editorial office

Dr. Sc., Professor A.S. Korotkov – Editor-in-Chief;

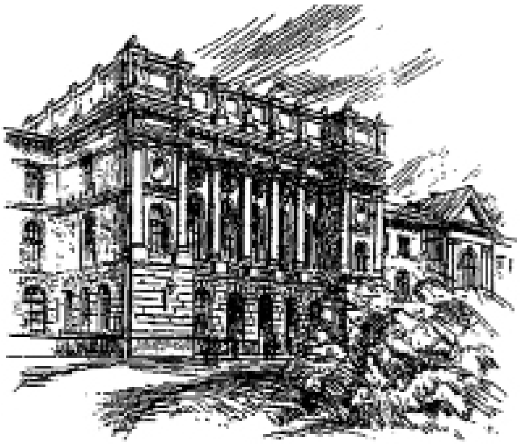
E.A. Kalinina – literary editor, proofreader; G.A. Pyshkina – editorial manager; A.A. Kononova – computer layout; D.Yu. Alekseeva – English translation.

Address: 195251 Polytekhnicheskaya Str. 29, St. Petersburg, Russia.

+7 (812) 552-6216, e-mail: infocom@spbstu.ru

Release date: 31.05.2023

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ



ISSN 2687-0517

Информатика, телекоммуникации и управление

Том 16, № 1
2023

Санкт-Петербургский политехнический
университет Петра Великого
2023

ИНФОРМАТИКА, ТЕЛЕКОММУНИКАЦИИ И УПРАВЛЕНИЕ

РЕДАКЦИОННЫЙ СОВЕТ ЖУРНАЛА

РЕДАКЦИОННАЯ КОЛЛЕГИЯ ЖУРНАЛА

Журнал с 2002 года входит в Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени доктора и кандидата наук.

Сетевое издание открытого доступа публикует статьи высокого научного уровня, освещающие передовой опыт, результаты НИР, теоретические и практические проблемы информатики, электроники, телекоммуникаций, управления.

Сведения о публикациях представлены в Реферативном журнале ВИНИТИ РАН, в международной справочной системе «Ulrich`s Periodical Directory», в Российской государственной библиотеке. В базах данных: Российский индекс научного цитирования (РИНЦ), Google Scholar, EBSCO, Math-Net.Ru, ProQuest, Index Copernicus.

Журнал зарегистрирован Федеральной службой по надзору в сфере информационных технологий и массовых коммуникаций (Роскомнадзор). Свидетельство о регистрации ЭЛ № ФС77-77378 от 25.12.2019.

Учредитель и издатель: Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Российская Федерация.

Редакция журнала

д-р техн. наук, профессор А.С. Коротков – главный редактор;

Е.А. Калинина – литературный редактор, корректор; Г.А. Пышкина – ответственный секретарь, выпускающий редактор;

А.А. Кононова – компьютерная вёрстка; Д.Ю. Алексеева – перевод на английский язык.

Адрес редакции: Россия, 195251, Санкт-Петербург, ул. Политехническая, д. 29.

Тел. редакции +7(812) 552-62-16, e-mail: infocom@spbstu.ru

Дата выхода: 31.05.2023

Contents

Содержание

Telecommunication Systems and Computer Networks
Телекоммуникационные системы и компьютерные сети

Research article
DOI: https://doi.org/10.18721/JCSTCS.16101
UDC 654.1

THROUGHPUT EVALUATION OF THE MILLIMETER-WAVE 5G COMMUNICATION SYSTEMS

A.V. Pudeev[1], *O.V. Bolkhovskaya*[2] ✉ ,
I.A. Bolotin[3], *A.A. Maltsev*[4]

[1] LG Electronics, Moscow, Russian Federation;
[2,3,4] Lobachevsky State University of Nizhni Novgorod,
Nizhni Novgorod, Russian Federation
✉ obol@rf.unn.ru

Abstract. The new, millimeter-wave (mmWave) Wi-Fi standard IEEE 802.11ay considers various indoor and outdoor scenarios, including multiple users access and backhauling with a range up to several hundred meters. Moreover, the 5G wireless communication systems are expected to adopt a heterogeneous network (HetNet) architecture, where small mmWave cells overlap a conventional macro cells network. The new applications require large antenna arrays and multi-stream transmission (MU-MIMO) with new beamforming algorithms, aimed not only at the single link quality maximization, but also at the optimization of the throughput in the whole deployment. In this paper, we evaluate the throughput of the mmWave communication systems for the main scenarios of their deployment. A comparative analysis of the different large antenna array techniques is carried out in application to MU-MIMO transmission at the small cells base station (BS) or Wi-Fi access points. The joint beamforming and scheduling algorithms utilizing the introduced antenna array architectures at the BS were developed. Finally, performance evaluation and comprehensive comparative analysis of the considered antenna array techniques are done using system level simulations for three deployment scenarios ("open space", "alleyway" and "hotel lobby") defined in the adopted millimeter-wave Wi-Fi standard IEEE 802.11ay. The proposed large antenna array architectures and the developed joint beamforming and scheduling algorithms for MU-MIMO transmission may find practical applications in millimeter wave Wi-Fi and 5G NR wireless communication systems.

Keywords: 5G, IEEE 802.11ay, mmWave, communication systems, hybrid beamforming, antenna arrays

Citation: Pudeev A.V., Bolkhovskaya O.V., Bolotin I.A., Maltsev A.A. Throughput evaluation of the millimeter-wave 5G communication systems. Computing, Telecommunications and Control, 2023, Vol. 16, No. 1, Pp. 7–20. DOI: 10.18721/JCSTCS.16101

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.16101
УДК 654.1

ОЦЕНКА ПРОПУСКНОЙ СПОСОБНОСТИ СИСТЕМ СВЯЗИ МИЛЛИМЕТРОВОГО ДИАПАЗОНА ДЛИН ВОЛН ПЯТОГО ПОКОЛЕНИЯ

А.В. Пудеев[1], *О.В. Болховская*[2] ✉, *И.А. Болотин*[3], *А.А. Мальцев*[4]

[1] Филиал корпорации LG Electronics Inc. (Республика Корея), Москва, Российская Федерация;

[2,3,4] Нижегородский государственный университет им. Н.И. Лобачевского, Нижний Новгород, Российская Федерация

✉ obol@rf.unn.ru

Аннотация. Новый стандарт Wi-Fi миллиметрового диапазона IEEE 802.11ay предполагается применять для высокоскоростной передачи данных как внутри, так и вне помещений, включая связь пользователей с точками доступа (малыми базовыми станциями) и радиорелейную связь между базовыми станциями с дальностью действия до нескольких сотен метров. Кроме того, ожидается, что системы сотовой связи 5G будут использовать гетерогенную архитектуру, в которой малые соты (радиусом порядка 50 м) обслуживают пользователей в миллиметровом диапазоне длин волн и перекрываются с обычными большими макросотами, работающими в диапазоне ниже 6 ГГц. Новые приложения систем связи миллиметрового диапазона длин волн требуют использования антенных решеток с большим числом элементов и многопотоковой передачи данных с алгоритмами формирования диаграмм направленности, нацеленными не только на максимизацию качества отдельной линии связи, но и на оптимизацию пропускной способности в целом во всей области развертывания. В статье дана оценка пропускных способностей систем связи миллиметрового диапазона для основных сценариев их развертывания, предусмотренных стандартом IEEE 802.11ay («открытое пространство», «узкая улица», «лобби отеля»). Проведен сравнительный анализ эффективности использования на базовых станциях (точках доступа) различных техник выполнения больших антенных решеток для реализации одновременной многопотоковой передачи данных многим пользователям. Разработаны совместные алгоритмы формирования диаграмм направленности и планирования передачи данных многим пользователям. Для оценки эффективности различных архитектур антенн использовано моделирование на системном уровне. Предположено, что базовая станция малой соты использует технологию радиосвязи с параметрами физического уровня, аналогичными стандарту IEEE 802.11ay. Предлагаемые архитектуры антенных решеток и разработанные алгоритмы могут найти практическое применение в системах связи миллиметрового диапазона Wi-Fi и 5G NR.

Ключевые слова: 5G, IEEE 802.11ay, миллиметровый диапазон, системы связи, гибридное формирование луча, антенные решетки

Для цитирования: Pudeev A.V., Bolkhovskaya O.V., Bolotin I.A., Maltsev A.A. Throughput evaluation of the millimeter-wave 5G communication systems // Computing, Telecommunications and Control. 2023. Т. 16, № 1. С. 7–20. DOI: 10.18721/JCSTCS.16101

Introduction

The rapid progress of millimeter wave technologies over the past fifteen years has enabled the development and mass production of low-cost Radio Frequency Integrated Circuits (RFIC) together with digital signal processing (baseband) chips [1]. Currently, these technologies are already used for internet access in Wi-Fi systems (IEEE 802.11ad, IEEE 802.11ay standards in the 57–71 GHz band) and they are planned to be widely used in the 5th generation of mobile communication systems (5G). The first phase of 5G New Radio (NR) will address bands below 40 GHz and the following phases will exploit bands up to 100 GHz [2].

The first millimeter-wave Wi-Fi standard IEEE 802.11ad, operating in the 57–64 GHz band, specifies the point-to-point links in the indoor scenarios [3]. In such conditions, the single small-size (about 8–16 elements) phased antenna arrays (PAA) with relatively low gain (about 12–15 dBi) [4] may be used for reliable data transmission up to several tens of meters. In this case, the beamforming algorithms may be limited to the simple sector sweep process, with exhaustive search of the optimal weight vector from the pre-defined codebook.

The new, millimeter-wave Wi-Fi standard IEEE 802.11ay adopted in 2021 year [5] considers various indoor and outdoor scenarios [6, 7], including backhauling and access with ranges up to several hundred meters. Moreover, the 5G wireless communication systems are expected to adopt a heterogeneous network (HetNet) architecture, wherein mmWave small cells are overlaid onto a conventional macro cells network [8].

Other possible deployment scenarios and use cases for mmWave systems are described in [9]. In most considered cases, the antenna gains should be significantly higher than for a usual indoor scenario, and antenna beams should be significantly narrower, which requires new approaches to the beam forming and beam selection [10–12] and new antenna technologies [13, 14]. In addition to the simple point-to-point links, IEEE 802.11ay standard will include point-to-multipoint links. The new applications require new beamforming algorithms, aimed not only to the single link quality maximization, but rather to the optimization of the throughput across the entire deployment. Maximum throughput of the new IEEE 802.11ay standard is considered to be about 100 Gbps [15]. Such challenging performance demands can be met with the help of MU-MIMO mode supporting multiple independent data streams as well as higher channel bandwidth up to 8.64 GHz [15, 16]. The multi-stream transmission is a right way of exploiting the multi-element antenna arrays for system performance improvement.

Due to the great importance of the task of the mmWave 5G communication systems development, a number of papers have been devoted to the investigation of possible approaches to large scale antenna arrays design and their performance characteristics evaluation. A large list of references in this area can be found in the overview papers [11, 12]. Along with deep theoretical investigations [17–19], several practical algorithms have been proposed for multiuser beamforming in mmWave massive MIMO (Multiple-Input-Multiple-Output) systems [20–25]. However, in the most of these papers beamforming algorithms were designed separately without taking into account the important scheduling procedure. Also the channel models used in these works for system performance evaluation did not adequately describe real multipath environment (scenarios) where the system should be deployed.

In this paper we focus on analysis of different large antenna array technologies, which may be used at mmWave small cell base stations (BSs) or Wi-Fi access points (APs). This will allow us to evaluate the mmWave communication system performance characteristics in different deployment scenarios and provide recommendations for practical system architectures.

It is assumed that BSs are equipped with antenna arrays with a large number of antenna elements. Three BS antenna array configurations were chosen as basic for consideration: the Multi-Stream Phased Antenna Array (MS PAA) architecture, which can be fully implemented in the RF part with codebook based analog beamforming only, the Fully Adaptive Array (FAA) configuration, fully implemented in the BB and considered as the most flexible solution with maximal available number of spatial degrees of freedom, and the Modular Antenna Array (MAA) configuration, which realizes a hybrid beamforming processing with coarse RF beamforming and fine BB beamforming [26–29]. To evaluate the performance of each antenna array technology, we will use the system level simulations approach, assuming that small cell BSs exploit the PHY layer technique with PHY layer parameters similar to IEEE 802.11ad and IEEE 802.11ay standards for orthogonal frequency division multiplexing (OFDM) modulations. The mmWave system performance investigation for these three BS antenna array configurations is done for the open area, street canyon and hotel lobby scenarios defined in the IEEE 802.11ay standard [5–7]. Therefore, this investigation will allow us to demonstrate advantages and drawbacks of the considered antenna array technologies in typical mmWave small cells deployment environments and elaborate the recommendations for their practical usage.

Base station architectures for multi-stream transmission

There are different ways for implementing multi-stream transmission/reception based on antenna array architecture.

The simplest solution lies in the division of the whole antenna array with N elements into a number K of smaller subarrays (one subarray per stream), and using these subarrays for independent transmission of K data streams (see Fig. 1*a*). The number of phase shifters in this scheme is equal to the number of the whole antenna array elements N. In reality this approach is equivalent to uniting K independent jointly controlled transmitters at one BS system. This does not introduce any additional complexity into the BB and RF parts in comparison with single stream transmission if no special inter-stream interference cancellation schemes are applied in baseband. However, such solution has one significant drawback: the apertures of the subarrays are K times smaller than the whole antenna array aperture. Therefore, the subarray antenna gains and directivities are reduced accordingly. Because of that drawback, we will not further study the multi-stream transmission with aperture division in this paper.

To avoid the aperture degradation, three multi-stream transmission antenna array schemes will be considered.

The first scheme is Multi-Stream Phased Antenna Array (MS PAA) BS architecture (see Fig. 1*b*). In this scheme, the whole phased antenna array aperture is used for transmission of each data stream, at the cost of increasing the number of phase shifters and RF end circuits' complexity. The number of phase shifters in this scheme is equal to the number of the antenna array elements N multiplied by the number of streams K. It can be seen that this BS antenna array architecture may be fully implemented in the RF part with analog beamforming only.

The second scheme is the Fully Adaptive Array (FAA) architecture with the number of RF chains equal to the number of antenna array elements N. Obviously this is the most flexible solution, because it may provide digital beamforming in the BB with maximal available number of spatial degrees of freedom N. The block diagram of the FAA for single carrier transmission is shown in Fig. 1*c*. It is realized by KxN complex multiplications (weighting) of the modulated and coded data stream signals. In this sense the FAA solution is optimal for MU-MIMO processing of K independent streams, and allows application of all existing MIMO processing techniques, like TX Maximum Ratio Transmission (MRT), RX Maximum Ratio Combining (MRC), TX and RX zero-forcing (ZF), and RX Minimal Mean Square Error (MMSE) reception [30]. However, for practical applications, especially for highly-directional steerable antenna arrays with a very large number of elements, having the same number of RF chains may be prohibitively complex and cost expensive.

To reduce the complexity of the BS antenna system and to retain both BB processing advantages and very high antenna gains, the Hybrid Beamforming technique [31] was proposed on the base of the Modular Antenna Arrays (MAA scheme), which consists of M subarray modules and RF chains (see Fig. 1*d*). The hybrid beamforming processing incudes coarse beamforming performed in the RF part by the controlled phase shifters (built-in in the subarray modules, each module comprises L antenna elements) and fine beamforming performed in the BB by applying a set of the complex weights to each RF branch of the modulated and coded data stream signals. Such MAA scheme allows benefiting from the BB processing having a relatively small number M of the RF chains – one per each subarray module. At the same time, this scheme has less adaptation possibilities in comparison with the FAA (only M digital degrees of free-

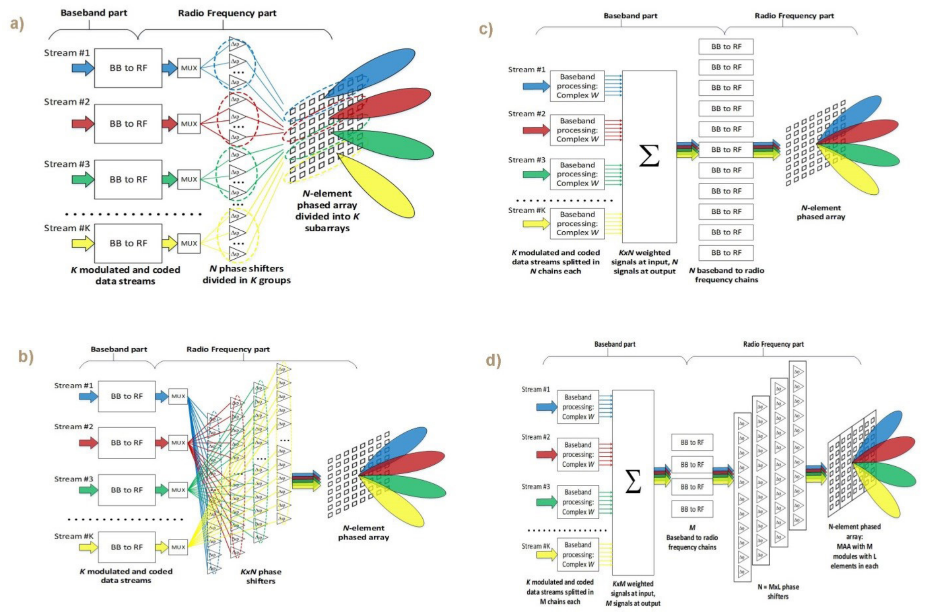


Fig. 1. *a* – Aperture division multi-stream transmission;
b – Multi-stream phased antenna array (MS PAA) scheme; *c* – Fully adaptive array (FAA) scheme;
d – Hybrid beamforming technique for MAA scheme

dom versus N degrees in the FAA, $M = N/L$). Therefore, the MAA may also be referred to as the partially adaptive array.

MU-MIMO – joint beamforming and scheduling algorithms at the BS

In practical implementation of multi-user transmission in MU-MIMO mode, the beamforming task is inseparable from the scheduling procedure. Indeed, for some environments groups of simultaneously scheduled users determine the beam directions and required inter-stream interference cancellation processing. And vice versa, the beamforming determines potential user data rates and, finally, the throughput metrics that the scheduler should take into account for transmission assignments. Thus, a joint beamforming and scheduling algorithm for simultaneous transmission to several users should be applied at the BS [26–28].

In this paper, we use a joint BF and scheduling algorithm [28] that dynamically solves the problem of maximizing the total throughput at the BS with an antenna array for DL MU-MIMO transmission. This algorithm will assign the available space-time-frequency communication system resources to several users based on some pre-defined criterion (metrics) calculated for each user. The widely used criterion is based on the proportional fair (PF) metric. For each DL packet transmission act, the PF scheduling algorithm assigns available space-time-frequency resources for transmission of K independent modulated and coded data streams to K different users (see Fig. 1). At the same time, preference in service is given to a group of K users demonstrating for the current state of the channel the maximum value of the PF metric, which is the sum of K available individual user throughputs, each normalized by the amount of data already received by the particular user. The optimal number of users and individual users involved in the given group may be determined "step-by-step" as describe below.

The most suitable for practical implementation of the PF scheduling is the so called greedy scheduling algorithm. It is an iterative algorithm which makes the "best choice" at each step for including a new user in the group of users for simultaneous transmission. At each iteration step of the MU-MIMO greedy scheduling one user is picked up considering the maximization of the total group PF metric. The user selection

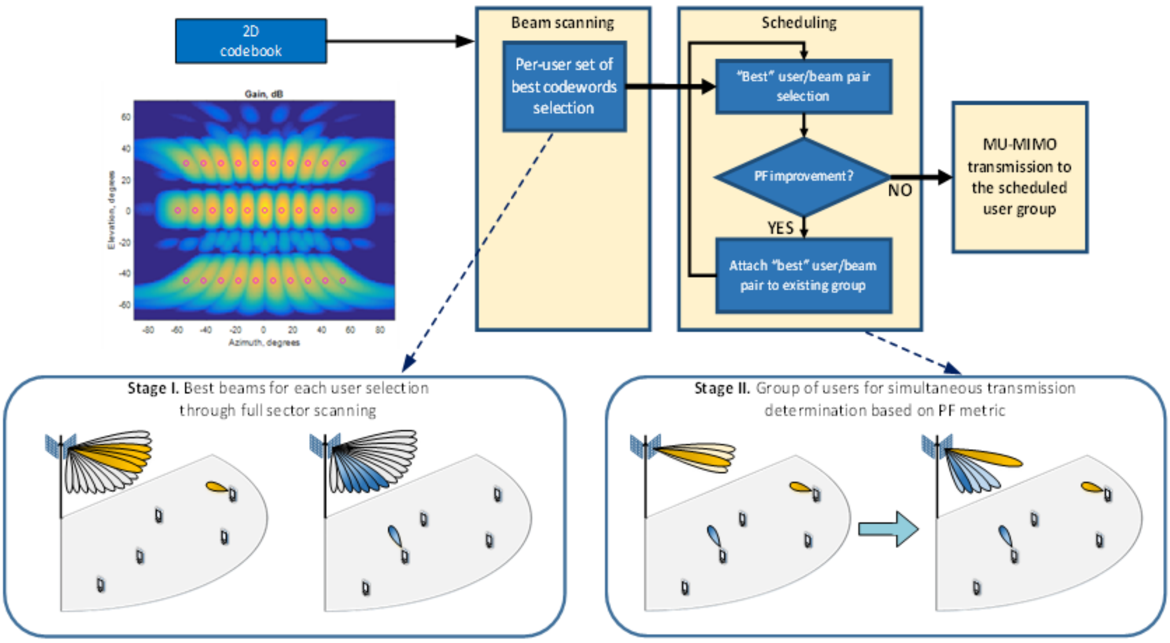


Fig. 2. Codebook-based joint beamforming and scheduling algorithm for MS PAA scheme

on each iteration should take into account available throughputs of all already chosen users, recalculating their SINRs and therefore their modulation and coding schemes (MCSs) and throughputs, considering possible new inter-stream interference and transmit power splitting, and afterwards recalculating their PF metrics. The iterative process stops when addition of any further new user reduces the total group PF metric [28].

However, it should be noted that making the "best choice" at each step of the PF greedy scheduling algorithm does not necessarily produce a global optimal solution to the overall maximization problem. However, in many instances it is rather close to the optimal one, which may be found by an exhaustive searching algorithm that has exponential complexity.

Codebook-based analog MU-MIMO beamforming for multi-stream phased antenna array. In this assumption, the BS is equipped with a multi-stream phased antenna array (MS PAA) with a number of possible beams (or partial antenna patterns) defined in the codebook. For example, a 2D DFT codebook can be used for RF beamforming both in elevation and azimuthal plane. The codebook-based MU-MIMO beamforming flowchart with the illustration of per-user codebook construction and PF greedy scheduling procedure is shown in Fig. 2.

At the first stage of the beam scanning procedure, each user will store and feedback to BS several indexes of the best beams with maximum received signal power, for creating a set of "perspective" codewords for each user. Then, at the second stage of the PF greedy scheduling procedure, both the best user and its beam most suitable for MU pairing are chosen at each step. Once the new user with the certain beam (codeword) was selected for including in the group of users for simultaneous transmission at any step of greedy scheduling, it will not change that beam at the subsequent steps.

Baseband digital MU-MIMO beamforming techniques. In a rich scattering environment, the full advantages of the massive MIMO system can be exploited using such beamforming strategies as maximum ratio transmission (MRT) or zero-forcing (ZF).

In case of the user equipment (UE) with one antenna element, the MRT algorithm uses the measured channel matrices to calculate the first eigenvector for establishing beamforming (a partial antenna pattern) for each user separately. Then the PF greedy scheduler tries to find a suitable MU group of users without

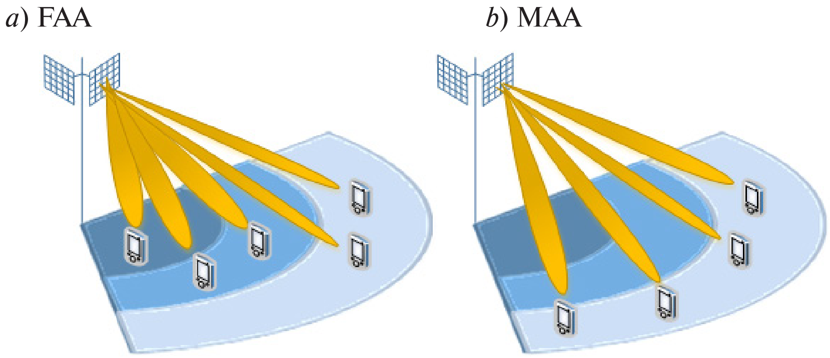


Fig. 3. MU-MIMO for FAA case (*a*) and MAA case (*b*)

any additional TX beamforming adjustment. Therefore, using the MRT algorithm for TX antenna array beamforming does not allow combining the users located a shot distance away from each other in one MU group for simultaneous data transmission. For example, in case of the one-path line-of-sight (LOS) channel, these users should be at an angular distance from each other at least 3-4 times greater than the half power beam width (HPBW) of the BS antenna array pattern to successfully avoid inter-stream interference.

The purpose of ZF technique is to suppress the mutual MU inter-stream interference arising due to simultaneous multiuser data transmissions, by modifying the initial MRT beamforming vectors using the inverse covariance interference matrix with regularization by the minimal mean square error (MMSE) algorithm [28]. Physically, this beamforming modification will lead to establishing near to zero values of one UE partial antenna pattern in the directions of main beams of all other UEs included in the same MU-MIMO group for simultaneous data transmission. It is obvious that using a more complex ZF beamforming algorithm allows simultaneously transmitting data to users located more densely in space. For example, in case of the LOS channel, these users may be located at an angular distance of about one HPBW.

There are huge beamforming capabilities for the full adaptive antenna array (FAA) because the number of degrees of freedom is equal to the number of antenna elements. Therefore, the FAA may effectively form equivalent beams in any direction and provide better selection of scheduled users from the whole cell (see Fig. 3*a*).

For the MAA with vertical module placement, the array may have limited ability for vertical beamforming to several users and almost full adaptation ability in the horizontal plane. So, in this case, the average direction in the elevation plane is used for all users in the group and per-user directions in the horizontal plane are used for MRT and ZF operations. Thus, in the multi-user mode in the mmWave small cell with such type of MAA, the users for simultaneously scheduled group should be selected from the same "ring" around the BS, with the same elevation angle and distance (see Fig. 3*b*).

One of the significant advantages of digital MU-MIMO beamforming is that the baseband processing of the OFDM (OFDMA) signals allows using frequency selective beamforming for the a more accurate communication system adaptation to the frequency selective millimeter-wave channel. In the following section, a comparison between wideband and per-subcarrier beamforming will also be provided.

mm-WAVE communication system performance analysis

The investigation of mm-Wave 5G communication systems performance was made by direct modeling of physical (PHY) and medium access control (MAC) layers in accordance with the IEEE 802.11ay specification. These also included the multipath channel modeling and usage of the channel state information for near to optimal BS antenna array beamforming and data multi-stream scheduling for MU transmission.

The efficiency of the considered MU-MIMO antenna array technologies and beamforming algorithms was studied on the system level for three scenarios with the growing complexity of environment and channel models exactly defined in the IEEE 802.11ay evaluation methodology documents [32]:

Open area – minimal number of passes (rays). Only one-two quasi-deterministic (Q-D) rays give significant impact (direct LOS and ground reflected rays).

Street canyon – the number of significant rays increases up to four (additional reflections from building walls). Azimuthal diversity of these rays gives significant impact to the channel.

Hotel lobby – the environment causes a big number of significant rays (all rays up to second order wall and ceiling reflections).

The consideration of these scenarios enables a full assessment the capabilities of different MU-MIMO modes. For these scenarios, the Q-D channel models developed in [7] were used. More detailed simulation assumptions are provided in Table 1. The same type of 8x16 elements antenna arrays (8 vertical and 16 horizontal elements size) are used at the BS/AP for all antenna techniques. The MAA consisted of 16 subarrays, each containing 8×1 elements. The basic IEEE 802.11ay PHY layer OFDM mode parameters for channel bandwidth 2.16 GHz were used for simulations. The OFDM is based on a 512-point FFT with 336 active data subcarriers, and 16 fixed pilot tones. The subcarriers at DC and on either side of DC are nulled to avoid any issues with carrier feed-through and the cyclic prefix is fixed at 25 % of the OFDM symbol period.

Table 1

Scenarios for Q-D channel models

Parameters		Assumption
Channel model/Pathloss		Q-D channel model
Carrier / BW		60 GHz / 2 GHz
BS antenna array	Array types	MS PAA / FAA / MAA
	Height	6 m (5.5 m for Hotel lobby scenario)
	Configuration / TX power	8x16 elements / 19 dBm
	Element gain	5 dBi
User antenna	Height	1.5 m
	Gain	Omni, 0 dBi
MU-MIMO Transmission scheme		SVD-based MRT/ZF, 2D-128 DFT codebook-based MRT, 8 spatial streams maximum for any antenna technique
Beamforming granularities		Wideband / Per subcarrier
Channel estimation		Perfect (the overhead is considered)
Scheduling type		Greedy PF MU scheduling

The deployment assumptions for considered scenarios are illustrated in Fig. 4–6. The open area scenario is very close to the "pure" free space LOS environment. The outdoor street canyon scenario represents a more complex millimeter-wave propagation environment. The users are mostly grouped on the relatively narrow sidewalks. So, from the physical point of view, the realization of MU-MIMO transmissions in the street canyon scenario are much more difficult due to the denser UEs deployment and mutual interference between spatial streams produced by reflected rays.

The system level simulation main results for all considered scenarios are summarized in Table 2.

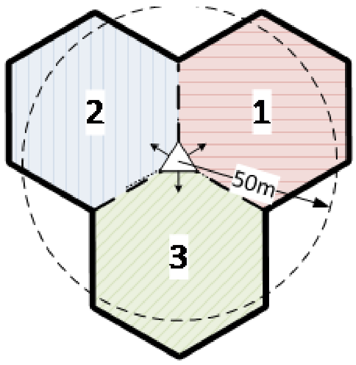


Fig. 4. Open area scenario deployment assumptions

Isolated small cell with three BSs (antenna height 6 m) each serving its own 120° sector and operating in its own channel (frequency reuse-3).

Users are placed randomly (with uniform distribution) within the cell considering 50 users per cell sector.

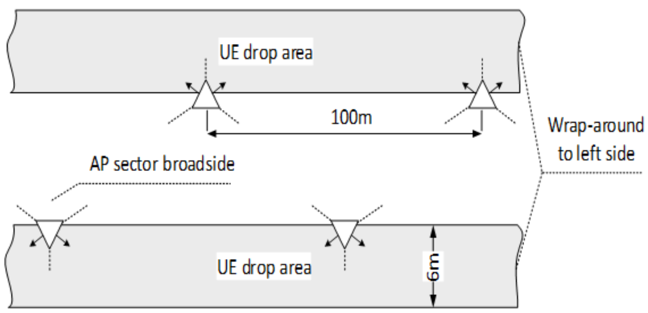


Fig. 5. Street canyon scenario deployment assumptions

Two BSs are mounted at each lamppost serving two sectors (frequency reuse-2) along the sidewalk.

Users are placed randomly (uniform distribution) within the sidewalks considering 45 users per each BS sector.

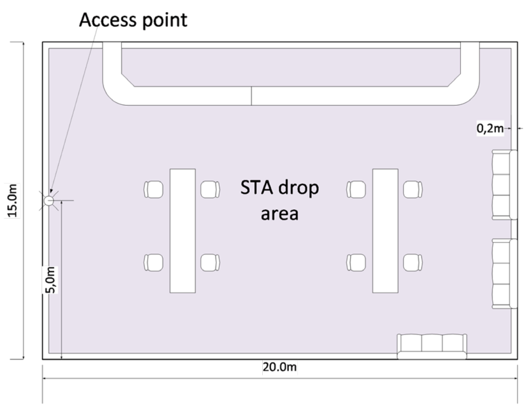


Fig. 6. Hotel lobby scenario deployment assumptions

One BS, mounted at the center of the shortest side, has a single sector, directed to the center of the lobby.

40 users are placed randomly (with uniform distribution) within the lobby area.

Table 2

Simulation results for considered scenarios

BS antenna array model	Beamforming scheme	Beamforming frequency granularity	Open area scenario				Street canyon scenario				Hotel lobby scenario			
			BS throughput, Gbp	Avg user throughput, Mbps	Cell-edge throughput, Mbps	Avg size of MU group	BS throughput, Gbps	Avg user throughput, Mbps	Cell-edge throughput, Mbps	Avg size of MU group	BS throughput, Gbps	Avg user throughput, Mbps	Cell-edge throughput, Mbps	Avg size of MU group
MS PAA	CBB	Wideband	7.2	144	70	3.76	7.8	173	75	2.26	11.5	575	367	3.48
FAA	MRT	Wideband	10.2	205	107	4.78	10.0	222	81	2.52	17.0	848	484	4.46
		Per subcarrier	10.2	205	115	4.64	11.2	249	112	2.66	16.5	826	494	4.54
	ZF	Wideband	11.4	229	110	5.23	12.4	275	81	3.16	29.4	1471	645	7.34
		Per subcarrier	**13.9**	**278**	**131**	**6.03**	**22.3**	**495**	**180**	**6.11**	**41.2**	**2060**	**1201**	**9.66**
MAA	MRT	Wideband	8.8	175	95	4.16	6.0	134	68	1.46	10.0	503	377	2.54
		Per subcarrier	8.7	173	98	4.13	6.5	145	102	1.44	9.8	492	378	2.54
	ZF	Wideband	9.3	186	100	4.44	6.3	139	68	1.56	11.2	559	400	2.88
		Per subcarrier	9.9	198	107	4.71	7.7	171	109	2.08	11.9	597	424	2.98

The table includes main characteristics (metrics) specified in the official mmWave communication system evaluation methodology [32], such as: BS total throughput, average UE throughputs, average size of MU group (average number of UEs grouped for simultaneous transmission) and cell-edge UE throughput (only 5 % of the served UEs may have the throughput below this threshold).

The numbers in the Table 2 are illustrated by histograms shown in Fig. 7, 8, for more convenient representation of the simulation results. For simplicity, only wideband scheduling/beamforming results are presented in the illustrative histograms.

It can be seen from presented simulation results that the ideal FAA antenna predictably demonstrates better performance metrics in comparison with the MAA and MS PAA for all beamforming schemes for all scenarios.

The comparison of the practical schemes such as MS PAA and MAA looks more interesting. In the LOS-dominant scenario of the Open area, the MAA demonstrates 20–30 % better system performance (depending on the applied beamforming technique) for all metrics: BS and UE throughputs, as well as cell-edge throughput and MU aggregation metrics. However, for more complex multipath scenarios, such as

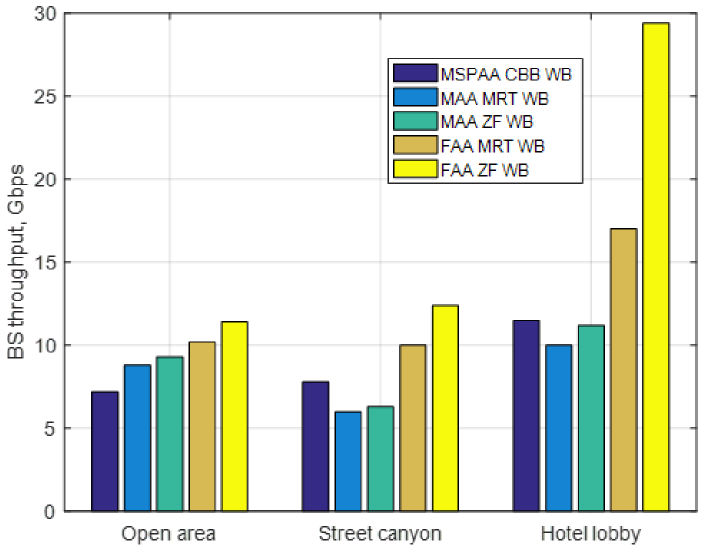


Fig. 7. BS aggregate throughput comparison for different scenarios and techniques

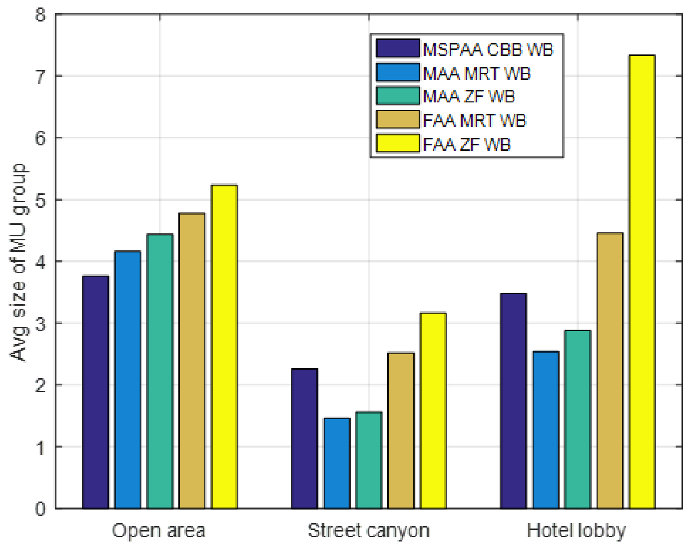


Fig. 8. Average number of UE in MU group comparison for different scenarios and techniques

Street Canyon and Hotel Lobby, the main system performance metrics of analog-based MS PAA scheme and MAA with ZF algorithm looks very similar. The reduction of the MAA system performance in these scenarios can be explained by the fact that MAA can schedule in the same MU group only the UEs that are seen from the BS/AP on nearly the same elevation angle (see Fig. 6), while MS PAA has a freedom to select different beams in azimuth and elevation for the same MU group. In the rich multipath and compact scenarios, the elevation angles may be very different, which gives the MS PAA scheme an upper hand for simultaneous data transmission to a larger number of UEs (see Table 2). So, "the freedom" of the MAA beamforming in the frequency domain is counterpoised by "the freedom" of the MS PAA beamforming in space domain.

The maximum spectral efficiency is provided by the most complex FAA with an SVD-based ZF beamforming scheme and per subcarrier frequency granularity. This scheme allows achieving very high throughputs due to very flexible resource allocations in the space-frequency domain and denser grouping of UEs for simultaneous transmission. There are about 6 users per MU group in average even in the most complex (from the MU grouping point of view) open area and street canyon scenarios, and up to 9 users in the hotel lobby scenario with maximum total BS throughput 41 Gbps.

From the general point of view it is clear that such behavior of the considered MU-MIMO schemes and beamforming algorithms may be explained by different multipath propagation channels for the open area, street canyon and hotel lobby scenarios. The richer multipath channel theoretically has greater MIMO capacity, but to realize of this opportunity for increasing the system throughput we need to implement a very complex FAA MU-MIMO scheme.

Finally, it should be noted that BS and UE throughputs represented in Table II and by histograms in Fig. 7, 8 are given for omni user antennas with 0 dBi gain and channel bandwidth of 2.16 GHz. The mounting of the small PAA with 4–8 elements at the user equipment and application of the effective asymmetric links beamforming technique [10] will make it possible to support data transmission in 4.32 GHz and 8.64 GHz channels with the same modulation and coding schemes, thereby to multiply throughputs by 2 and 4 times respectively.

Summary

In this paper we investigated large antenna array technologies for MU-MIMO transmission in mmWave small cells with different involvement of radiofrequency and baseband parts for signal processing. The investigation in different environments was done for the Multi-Stream Phased Antenna Array (MS PAA) architecture, which can be fully implemented in the RF part with codebook based analog beamforming only, the Fully Adaptive Array (FAA) configuration, fully implemented in the BB and considered as the most flexible solution with the maximum available number of spatial degrees of freedom, and the Modular Antenna Arrays (MAA) configuration, which realizes a hybrid beamforming processing with coarse RF beamforming and fine BB beamforming.

The FAA scheme predictably demonstrates better performance metrics in comparison with the MS PAA and the MAA for all beamforming schemes and frequency granularities in all considered scenarios. The maximum spectral efficiency is provided by the most complex FAA with ZF beamforming scheme and per subcarrier frequency granularity, but for the large antenna arrays with a large number of antenna elements such FAA scheme has high complexity and cost. The MS PAA with 8 streams (2D-128 DFT) and the MAA with 16 degrees of freedom demonstrate acceptable performance metrics in typical environments and therefore may be considered as good candidates for further practical implementations at the BSs or APs of the future mmWave small cells.

We should emphasize that the main results of the paper were obtained by direct system simulations based on strict compliance with the parameters of the IEEE 802.11ay standard and the usage of generally accepted signal processing, beamforming and scheduling algorithms. Therefore, these results are practically important for cellular operators, communication equipment manufacturers and vendors to make decisions on the implementation of available antenna technologies for deploying the new mmWave cellular and Wi-Fi communication systems of the 5th and further generations.

REFERENCES

1. **Liu D., Pfeiffer U., Grzyb J., Gaucher B.** *Advanced millimeter-Wave technologies: Antennas packaging and circuits*. USA NY: Wiley, 2009.

2. **Dahlman E., Parkvall S., Skold J.** *5G NR: The next generation wireless access technology*, 2nd ed., 2020.

3. **Nitsche T., Cordeiro C., Flores A.B., Knightly E.W., Perahia E., Widmer J.C.** IEEE 802.11ad: Directional 60 GHz communication for multi-gigabit-per-second Wi-Fi. *IEEE Commun. Mag.*, 2014, vol. 12, no. 4, pp. 132–141.

4. **Pan H.K., Horine B.D., Ruberto M., Ravid S.** Mm-Wave phased array antenna and system integration on Semi-Flex packaging. *IEEE AP-S Conference on Antenna and Propagation for Wireless Communications*, 2011, pp. 2059–2062.

5. IEEE Standard for Information Technology—Telecommunications and Information Exchange between Systems Local and Metropolitan Area Networks—Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 2: Enhanced Throughput for Operation in License-exempt Bands above 45 GHz, *https://standards.ieee.org/ieee/802.11ay/6142/*

6. IEEE 802.11-14/1185r0: NG60 usage scenarios. *http:// mentor.ieee.org/802.11/dcn/14/11-14-1185-00-ng60-ng60-usage-scenarios.pptx*

7. *Channel models for IEEE 802.11ay*. IEEE doc.: 802.11-15/1150r9, 2017.

8. **Okasaka S., et al.** Proof-of-concept of a millimeter-Wave integrated heterogeneous network for 5G cellular. *SENSOR*, Aug. 2016.

9. **Sakaguchi K., et al.** Where, when, and how mmWave is used in 5G and beyond. *IEICE Transactions on Electronics*, Oct. 2017, Vol. E100-C, no. 10, Pp. 790–808.

10. **Maltsev A., Pudeyev A., Bolotin I., Bolkhovskaya O.** Asymmetric links beamforming for mmWave overlay in LTE-based heterogeneous network. *2017 IEEE 86th Vehicular Technology Conference* (*VTC-Fall*), Toronto, Canada, 2018, pp. 1–5.

11. **Heath R.W., et al.** An overview of signal processing techniques for millimeter Wave MIMO systems. *IEEE J. Sel. Topics Signal Processing*, 2016, vol. 10, no. 3, pp. 436–453.

12. **Molisch A.F., et al.** Hybrid beamforming for massive MIMO: A survey. *IEEE Communications Magazine*, 2017, vol. 55, no. 9, pp. 134–141. DOI: 10.1109/MCOM.2017.1600400

13. **Maltsev A., Sadri A., Pudeyev A., Bolotin I.** Highly directional steerable antennas: High-Gain antennas supporting user mobility or beam switching for reconfigurable backhauling. *IEEE Veh. Tech. Mag.*, 2016, vol. 11, no. 1, pp. 32–39.

14. **Maltsev A.A., Lomaev A., Pudeyev A.V., Bolotin I.A., Bolkhovskaya O.V., Seleznev V.M.** Millimeter-wave toroidal lens-array antennas experimental measurements. *2018 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting*, 2018, pp. 607–608.

15. **Ghasempour Y., da Silva C.R.C.M., Cordeiro C., Knightly E.W.** IEEE 802.11ay: Next-generation 60 GHz communication for 100 Gb/s Wi-Fi. *IEEE Communications Magazine*, 2017, vol. 55, no. 12, pp. 186–192.

16. **da Silva C.R.C.M., Lomayev A., Chen C., Cordeiro C.** Analysis and simulation of the IEEE 802.11ay single-carrier PHY. *2018 IEEE International Conference on Communications* (*ICC*), 2018, pp. 1–6. DOI: 10.1109/ICC.2018.8422532

17. **El Ayach O., Rajagopal S., Abu-Surra S., Pi Z., Heath R.** Spatially sparse precoding in millimeter wave MIMO systems. *IEEE Trans. Wireless Commun.*, 2014, vol. 13, no. 3, pp. 1499–1513.

18. **Alkhateeb T.A., El Ayach O., Leus G., Heath R.W.** Channel estimation and hybrid precoding for millimeter Wave cellular systems. *IEEE Journal of Selected Topics in Signal Processing*, 2014, vol. 8, no. 5, pp. 831–846. DOI: 10.1109/JSTSP.2014.2334278

19. **Sohrabi F., Yu W.** Hybrid digital and analog beamforming design for large-scale antenna arrays. *IEEE Journal of Selected Topics in Signal Processing*, 2016, vol. 10, no. 3, pp. 501–513. DOI: 10.1109/JSTSP.2016.2520912

20. **Feng Y., Kim S.C.** Hybrid precoding based on matrix-adaptive method for multiuser large-scale antenna arrays. PLoS ONE 12(12):e0188723, 2017. *https://doi.org/10.1371/journal.pone.0188723*

21. **Shaban A.W., Damen O., Xin Y., Au E.** Statistically-aided codebook-based hybrid precoding for millimeter Wave channels. *IEEE Access*, 2020, vol. 8, pp. 101500–101513. DOI: 10.1109/ACCESS.2020.2997190

22. **He S., Wang J., Huang Y., Ottersten B., Hong W.** Codebook based hybrid precoding for millimeter Wave multiuser systems. *IEEE Transactions on Signal Processing*, 2017, vol. 65, no. 20, pp. 5289–5304. DOI: 10.1109/TSP.2017.2723353

23. **Chen J.** Efficient codebook-based beamforming algorithm for millimeter-Wave massive MIMO systems. *IEEE Transactions on Vehicular Technology*, 2017, vol. 66, no. 9, pp. 7809–7817. DOI: 10.1109/TVT.2017.2677957

24. **Liang L., Xu W., Dong X.** Low-complexity hybrid precoding in massive multiuser MIMO systems. *IEEE Wireless Commun. Lett.*, 2014, pp. 653–656.

25. **Ruan H., Xiao P., Xiao L., Kelly J.R.** Joint iterative optimization based low-complexity adaptive hybrid beamforming for massive MU-MIMO systems. *IEEE Transactions on Communications*, 2021, vol. 69, no. 3, pp. 1707–1722. DOI: 10.1109/TCOMM.2021.3053021

26. **Nam J., Adhikary A., Ahn J.-Y., Caire G**. Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling. *IEEE J. Sel. Topics Signal Process*, 2014, vol. 8, no. 5, pp. 876–890.

27. **Saqib N.U., Cheon K.-Y, Park S., Jeon S.-W.** Joint optimization of 3D hybrid beamforming and user scheduling for 2D planar antenna systems. *2021 International Conference on Information Networking* (*ICOIN*), 2021, pp. 703–707. DOI: 10.1109/ICOIN50884.2021.9334028

28. **Sergeev V., Bolkhovskaya O., Maltsev A.** A practical approach to hybrid beamforming for 5G Wi-Fi systems. *2021 Wave Electronics and Its Application in Information and Telecommunication Systems* (*WECONF*), St. Petersburg, Russia, pp. 1–9. DOI: 10.1109/WECONF51603.2021.9470612

29. **Maltsev A., Sadri A., Pudeyev A., Bolotin I., Davydov A., Morozov G., Weiler R.** Partially adaptive arrays application for MU-MIMO mode in a mmWave small cells. *IEEE PIMRC 2015*, pp. 315–319.

30. **Paulraj A., Nabar R., Gore D.** *Introduction to space-time wireless communications*. U.K. Cambridge: Cambridge Univ. Press, 2008.

31. **Maltsev A., Pudeyev A., Sadri A., Cordeiro C.** *Millimeter-Wave transceiver with coarse and fine beamforming with interference suppression and method*. US Patent 20130308717, Pub.: Nov. 2013.

32. *TGay Evaluation Methodology*. IEEE doc.: 802.11-15/0866r2, 2015.

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Пудеев Андрей Вячеславович
Andrey V. Pudeev
E-mail: andrey.pudeev@lge.com

Болховская Олеся Викторовна
Olesya V. Bolkhovskaya
E-mail: obol@rf.unn.ru

Болотин Илья Александрович
Ilya A. Bolotin
E-mail: ilya.a.bolotin@gmail.com

Мальцев Александр Александрович
Alexander A. Maltsev
E-mail: maltsev@rf.unn.ru

Submitted: 18.02.2023; Approved: 16.05.2023; Accepted: 17.05.2023.
Поступила: 18.02.2023; Одобрена: 16.05.2023; Принята: 17.05.2023.

Computing, Telecommunication and Control, 2023, Vol. 16, No. 1, Pp. 21–34.
Информатика, телекоммуникации и управление. 2023. Том 16, № 1. С. 21–34.

Circuits and Systems for Receiving, Transmitting and Signal Processing

Устройства и системы передачи, приема и обработки сигналов

Research article
DOI: https://doi.org/10.18721/JCSTCS.16102
UDC 621.37

IMPACT OF SIGNAL DISTORTION IN A POWER AMPLIFIER ON TELECOMMUNICATION SYSTEM EFFICIENCY

A.O. Pergushev[1] ✉ , *V.A. Sorotsky*[2], *S.V. Zavjalov*[3]

[1,2,3] Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation

✉ pergushev.aleksey@mail.ru

Abstract. The need to increase information transfer rate in up-to-date telecommunication systems led to a wide spread of OFDM and SEFDM signals with a relatively high peak-to-average power ratio. In case of power amplifiers operating at a fixed supply voltage, this power ratio feature can result in up to 20–30 % decrease in efficiency, which predetermined the interest in the use of more efficient solutions. These methods, including envelope tracking, outphasing, etc., are based on non-linear transformations of a radio frequency signal and/or its envelope. The resulting signal distortions can be accompanied by a noticeable increase in error rate at the receiver, which negatively affects the telecommunication system performance. The paper considers a set of factors with the most significant impact upon signal distortion in a power amplifier with tracking supply. An analytical model of a multi-cell tracking power supply was developed taking into account the errors of envelope and reference voltage digital conversion, the transistors' inertial parameters, the nonlinearities of the magnitude- and phase-frequency responses of the output low-pass filter. The impact of these factors on the envelope signal spectrum distortion and the bit error rate at receiver were considered. Proceeding from the results obtained, the authors proposed requirements to the transistors switching performance and the permissible dispersion of the parameters of envelope tracking power supply.

Keywords: modulated power supply, harmonic distortion, envelope tracking, power amplifier, efficiency, analytical model

Citation: Pergushev A.O., Sorotsky V.A., Zavjalov S.V. Impact of signal distortion in a power amplifier on telecommunication system efficiency. Computing, Telecommunications and Control, 2023, Vol. 16, No. 1, Pp. 21–34. DOI: 10.18721/JCSTCS.16102

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.16102
УДК 621.37

ВЛИЯНИЕ ИСКАЖЕНИЙ СИГНАЛОВ В УСИЛИТЕЛЕ МОЩНОСТИ НА ЭФФЕКТИВНОСТЬ СИСТЕМЫ ПЕРЕДАЧИ СООБЩЕНИЙ

А.О. Пергушев[1] ✉ , *В.А. Сороцкий*[2], *С.В. Завьялов*[3]

[1,2,3] Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

✉ pergushev.aleksey@mail.ru

Аннотация. В современных телекоммуникационных системах, исходя из необходимости повышения скорости передачи информации, широко используются многочастотные сигналы (OFDM и SEFDM), для которых характерен сравнительно высокий пик-фактор. Последнее обстоятельство в случае усилителей мощности, работающих при фиксированном напряжении питания, может привести к снижению КПД до 20–30 % и предопределило интерес к использованию более эффективных в плане энергетической эффективности технических решений. В их число входят методы следящего питания, дефазирования и другие, основанные на нелинейных преобразованиях радиочастотного сигнала и/или его огибающей. Возникающие при этом искажения сигналов могут сопровождаться заметным возрастанием количества ошибок на приемной стороне, что негативно отразится на эффективности системы передачи сообщений. В статье рассмотрены основные факторы, оказывающие наиболее существенное влияние на уровень искажений сигналов в усилителе мощности, реализованном на основе метода следящего питания. Разработана аналитическая модель многоячейкового модуляционного источника питания, учитывающая погрешности дискретизации сигнала огибающей и опорного напряжения, инерционные параметры транзисторов, нелинейности амплитудно-частотной и фазо-частотной характеристик выходного фильтра нижних частот и проведена оценка степени их влияния на искажения спектра сигнала огибающей и вероятность битовой ошибки при приеме радиосигналов. На основе полученных результатов сформулированы требования к быстродействию транзисторов и допустимому разбросу параметров модуляционного источника питания.

Ключевые слова: модуляционный источник питания, нелинейные искажения, метод следящего питания, усилитель мощности, КПД, аналитическая модель

Для цитирования: Pergushev A.O., Sorotsky V.A., Zavjalov S.V. Impact of signal distortion in a power amplifier on telecommunication system efficiency // Computing, Telecommunications and Control. 2023. Т. 16, № 1. С. 21–34. DOI: 10.18721/JCSTCS.16102

Introduction

The use of multi-frequency signals (OFDM and SEFDM) is one of the promising ways to increase the information transfer rate in telecommunication systems (TS) [1]. However, the insufficient bandwidth of the telecommunication channel, as well as the linearity disturbance of its transfer characteristic, usually lead to signal distortions, which can be accompanied by a noticeable increase in the number of errors arising at the receiver and thus reducing the TS efficiency.

The paper proposes a universal approach that allows to evaluate the effect of signal distortion, regardless of which part of the telecommunication channel (transmitter, communications line or receiver) they occur. Within the limited size of the paper, the case is considered when distortions occur only in one section of the channel – in transmitter and, accordingly, the main contribution to the signals parasitic change is made by the power amplifier (PA).

When amplifying multi-frequency signals with power amplifiers (PAs) operating in the classes AB or B, which are considered to be sufficiently linear, the relatively high peak to average power ratio (PAPR) of this type of signals leads to a decrease in efficiency as far as 20–30 %. This circumstance predetermined interest in the use of technical solutions that are more efficient in terms of power efficiency. Among the common ways to keep the efficiency of the PA at an acceptable level, it should be noted the methods of envelope tracking (ET) [2–5], envelope extraction and restoration (EER) [3, 6, 7], outphasing [3, 8–10] and Doherty [3, 11, 12]. At the same time in the first two methods pulse-width modulation (PWM) is used to amplify the envelope signal, during which, as is known, the signal is subjected to nonlinear transformations, and in the second method, in addition, the switch mode is also used to amplify the radio frequency signal, which can also be accompanied by distortions. Nonlinear distortions are also inherent in the other two known methods of amplifying signals with a high PAPR.

Taking into account the above, the purpose of the paper is to determine the parameters of the PAs which have the most significant impact on the signal distortions and caused by these distortions the bit error rate in the receiver. Solving this problem will make it possible to recommend the most rational requirements in terms of transmitter implementation.

Initial provisions

Let's consider the main criteria that will be used in assessing the influence of signal distortions on the efficiency of the TS.

As is known, any change in signals' spectrum is a very accurate indicator of undesirable distortions that have occurred in them. However, it is important not only to take into account the level of spurious spectral components, but also in what part of the spectrum they appeared. In particular, parasitic spectral components caused by the use of PWM in ET power supplies can be grouped around harmonics of the clock frequency, which is as high as ten or more times compared with the upper frequency in the signal spectrum [13].

Accordingly, these spectral components, although they can be considered as a measure for assessing the distortions that have arisen in the signal, but taking into account their influence is more interesting from the point of view of the problems that will be encountered in ensuring electromagnetic compatibility (EMC) requirements in the transmitter. So when evaluating the TS efficiency, the most interesting are those parasitic spectral components that directly occur into the frequency band of the useful signal, and, accordingly, can lead to errors in receiver.

In order to simultaneously take into account both the disposition of spurious spectral components in the frequency band of the signal as well as their level, it is advisable to use a universal scalar indicator, which can be chosen as the bit error rate. This characteristic accumulates all the negative effects caused by signal distortions. Based on the above reasoning, we will divide the solution of the problem under consideration into two stages: at first, we estimate the influence of various PA parameters on the signal distortion, then we determine the bit error rate due to these distortions.

Transmitter model

When getting on to the discussion of the transmitter model, in order to unambiguously determine it, it is necessary to specify the method that is supposed to be used to amplify signals with a high PAPR. In this regard, for definiteness, we will consider the ET method. The block diagram of the output stage of the transmitter, is shown in Fig. 1. One more clarification should be made here: since a detailed simultaneous consideration of the processes in the ET power amplifier and in the RF power amplifier can be quite voluminous and beyond the scope of one publication, we will assume that the effect of distortions introduced by the RF power amplifier can be neglected. Several reasons can be cited to justify this approach. One of the arguments is that the problem of lowering signal distortions in RF power amplifiers is not new, and currently there are effective solutions to reduce them by introducing pre-distortion into the

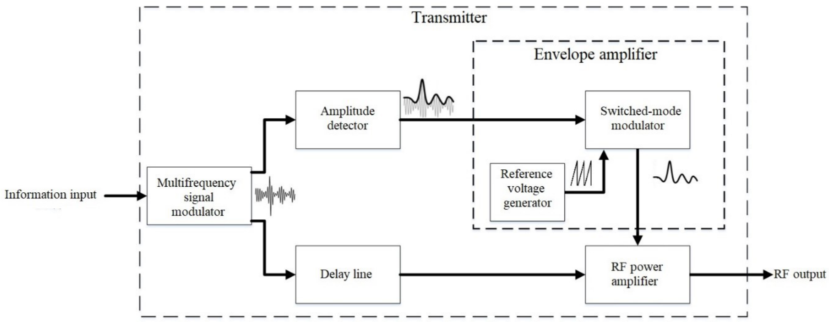


Fig. 1. The structure of transmitter output stage

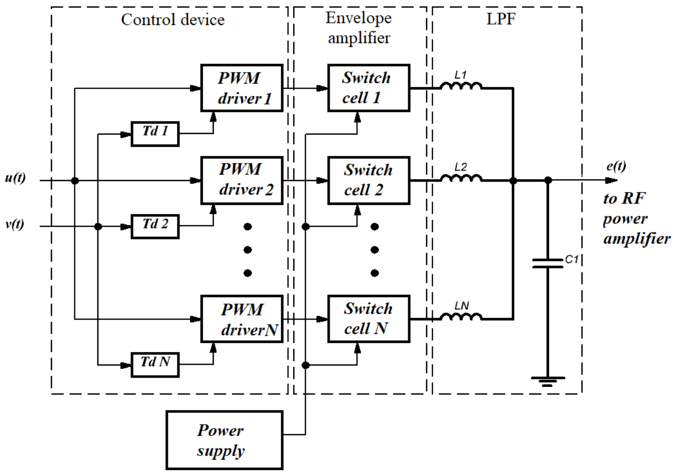


Fig. 2. Structure of a multi-cell modulated power supply

amplified signal [14–20]. In contrast, the study of the main factors that lead to envelope distortions in ET power supplies has not been adequately reflected in the known publications. Moreover, in those few publications that are devoted to this problem, the authors limit themselves to only considering distortions in the ET power amplifiers, without touching on the problem of how these distortions will affect in the TS efficiency lowering when receiving signals. And finally, taking into account the contribution that ET power sources can make to increasing the level of bit errors is also relevant due to ever-increasing requirements for expanding their bandwidth, which in itself is a non-trivial task and may be accompanied by signal distortions increase.

To expand the bandwidth of modulated power supplies (MPS), they are usually performed on the basis of a multi-cell structure, the diagram of which is shown in Fig. 2. In it, individual cells of the MPS are implemented in the form of half-bridge circuits that amplify the envelope voltage in PWM mode [4, 5].

In Fig. 2 the following designations are used: $u(t)$ – envelope voltage; $v(t)$ – reference sawtooth voltage used in PWM; $e(t)$ – output voltage. Herewith the time delay of the reference voltage for the i-th cell ($i = 1, 2, ..., N$) is equal , where T_r – the period of the reference voltage $v(t)$, N – the number of cells. The distortions caused by the ET power amplifier can be conditionally divided into three types, depending on in which part of the voltage spectrum $e(t)$ they appear.

Distortions of the first type are caused by nonlinear transformations of the voltage $u(t)$ during PWM and appear as a set of spectral components around the frequency $f \approx kF_{max}qN$ (region III in Fig. 3), where F_{max} – the upper limit of the frequency band in the voltage $u(t)$ spectrum, q is the ratio of the reference

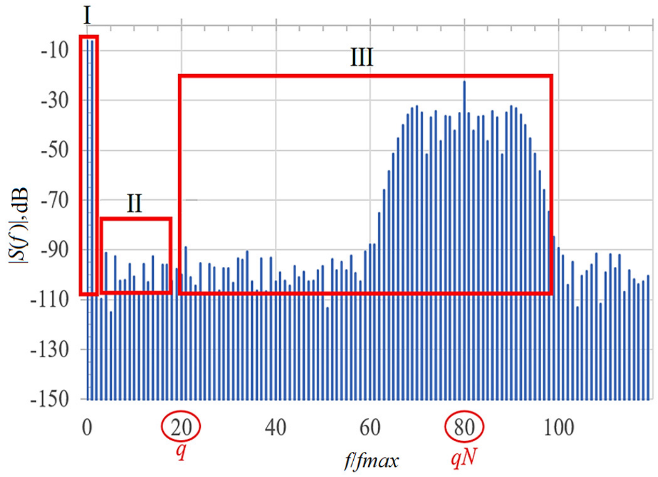


Fig. 3. Typical regions in the output voltage spectrum of a multicell MPS

voltage frequency to F_{max}, $k = 1, 2, \ldots$ – an integer. The negative impact of these spectral components can be reduced by increasing the frequency of the reference voltage and the number of cells N at a fixed bandwidth of the low-pass filter.

Distortions of the second type can be classified by the appearance of spurious spectral components in the region $F_{max} < f < F_{max}qN$ (region II in Fig. 3), which theoretically should be absent when using a multicell structure. However, studies have shown that the cause of these distortions, as a rule, is an error in setting the phases of the reference voltage in the channels of a multi-cell MPS, as well as the inductance dispersion of the coils used in the output circuits of the cells [13, 21, 22]. Spurious spectral components at the bottom of this region need to be controlled as they can get through the lowpass filter into the RF signal.

The most undesirable referring to the impact on the TS efficiency are distortions of the third type, which are accompanied by the appearance of spurious spectral components in the frequency range $f < F_{max}$ occupied by the envelope signal (region I in Fig. 3). These spectral components cannot be eliminated by filtering neither in the MPS itself nor at the output of the RF amplifier. The most noticeable contribution to the appearance of these distortions is made by:

1) envelope signal and reference voltage sampling errors;
2) inertial parameters of transistors (difference in rise and fall times in each of the switches) [21, 22];
3) turn-on and turn-off time intervals dispersion of the MPS transistors;
4) nonlinearity of the amplitude-frequency and phase-frequency characteristics of the MPS low-pass filter, and what is more the influence of both nonlinearities can increase significantly if the cutoff frequency is incorrectly chosen.

The most obvious way to take into account in the transmitter model listed above points 1–4 is simulation using CAD software. However, in the case of an MPS with an extended bandwidth, which is usually achieved by increasing N, the number of transistors in the model also increases ($2N$). As a result, the time required for one simulation iteration with a certain set of parameters increases markedly. In the case of a large number of combinations of variable parameters and, especially when studying the influence of the spread of transistor parameters (see Section 3), when the number of calculation iterations will be determined by the required fiducial probability and can reach tens for each combination of variable parameters, such simulation will become difficult to implement.

An alternative approach that makes it possible to overcome the problems noted above is the use of an analytical model [13], in which, under certain assumptions, it is possible to significantly reduce the time spent on calculations.

When creating an analytical model, we will rely on the well-known expression that describes the N-cell MPS output voltage, which was obtained under the assumption that transistors are ideal:

$$e(t,\ q,\ N)=\frac{1}{2N}\sum_{i=1}^{N}\left[1+\operatorname{sign}\left(u(t)-\mathrm{v}(i,\ t,\ q)\right)\right], \tag{1}$$

where $\mathrm{v}(t, m, q)$ – the sawtooth reference voltage; t – time; i – cell number.

In this form, the model makes it possible to adequately estimate only the spectral components that are multiples of the product qN (region III, Fig. 3).

In order to take into account the influence of errors that arise when using the digital representation of the envelope voltage and the reference voltage (see Section 1), the original model should be converted to the following form:

$$e_d\left(\tau_d,\ q,\ N,\ n_d,\ K_p\right)= \\ =\frac{1}{2N}\sum_{k=0}^{\infty}\sum_{i=1}^{N}\left\{1+\operatorname{sign}\left[\frac{\operatorname{floor}\left[u(k\tau_d)2^{n_d}\right]}{2^{n_d}}-\frac{\operatorname{floor}\left[\mathrm{v}(i,\ k\tau_d,\ q)K_p\right]}{K_p}\right]\right\}, \tag{2}$$

where τ_d – sampling time; n_d – is the capacity of the analog-to-digital converter (ADC); floor(x) – a function that rounds each element x to the nearest integer less than or equal to this element; K_p – number of quantization levels of reference voltage.

If it is necessary to consider the influence of the inertial properties of transistors (point 2), as well as the uneven gain-frequency characteristic and nonlinearity of the phase-frequency characteristic of the low-pass filter (point 4), the analytical model takes the form:

$$e_{d2}\left(\tau_d,\ q,\ N,\ n_d,\ K_p,\ \tau,\ L,\ C\right)= \\ =\frac{1}{2N}\sum_{k=0}^{\infty}\sum_{i=1}^{N}\left\{\left[1+\chi\left(\frac{\operatorname{floor}\left[u(k\tau_d)2^{n_d}\right]}{2^{n_d}},\ \frac{\operatorname{floor}\left[\mathrm{v}(i,\ k\tau_d,\ q)K_p\right]}{K_p},\ \tau\right)\right]\times\left[H(L_i,\ C)\right]\right\}, \tag{3}$$

where $\tau = \{\tau_{on}, \tau_{off}, \tau_r, \tau_f\}$ – a set of parameters' values that determine the inertial properties of transistors ((τ_{on} – turn-on delay time, τ_{off} – turn-off delay time, τ_r – rise time, τ_f – fall time); $H(L_i, C)$ – transfer function of the multi-input filter; Li – inductance at the output of the i-th cell; C – filter capacity; $L = \{L_1, L_2, ..., L_N\}$ – a set of inductances in the LPF.

Finally, if it is required to evaluate the influence of the spread of turn-on and turn-off times of transistors (point 3), then this is achieved by another modification of the model: by introducing into the set of parameters that determine the inertial properties of transistors the required distribution law of the random variable and the dispersion σ, which, as a rule, can be selected from the data provided by the manufacturer. Further, for definiteness, we will proceed from the fact that the scatter of the parameters of the MPS elements obeys the normal law.

To take into account the influence of the spread in expression (3), the set of parameter values τ that determine the transistors' inertial properties should be represented as: $\tau = \{\tau_{on} + \sigma_{on}, \tau_{off} + \sigma_{off}, \tau_r + \sigma_r, \tau_f + \sigma_f\}$, where $\sigma_{on}, \sigma_{off}, \sigma_r, \sigma_f$ – the dispersions, respectively, of the turn-on and turn-off times, rise and fall times.

Confirmation of the analytical model adequacy

To confirm the reliability of the presented analytical model of the MPS it's necessary to estimate the coincidence of the characteristics provided by model with the results of simulation. Simulation can be performed in any of the circuit simulators (*LTSpice*, *MicroCap*, *Multisim*, *Keysight ADS*, etc.), while to describe transistors, it is advisable to use their models provided by manufacturing companies.

The simulation was carried out with the following initial data: the number of cells $N = 4$, the relative duration of transistor's switching interval $t_{on}(t_{off}) \leq 0.02T$, where T is the period of the reference sawtooth voltage, $q = 15$, $n_d = 10$ bits, $K_p = 10$ bits , the second-order output LPF implemented in accordance with the Bessel approximation, its cutoff frequency was three times the frequency of the testing harmonic voltage.

Fig. 4*a* shows the normalized spectral diagram of the voltage u_{sp} obtained as a sum of voltages at the outputs of the MPS cells, and in Fig. 4*b* – normalized spectral diagram of the voltage e_{d2} at the output of the MPS filter at 10 % deviation of the first cell inductance from the nominal value. In Fig. 4, the red line corresponds to the simulation results, the blue line – to the analytical model which takes into account the inertial parameters of the transistors and the digital conversion error, and the green curves – to the results of the analytical model with digital conversion and without inertial parameters account.

From comparison of the analytical and simulation results, shown in Fig. 4, it follows that both their behavior and coincidence measure are quite convincing confirmation of the proposed analytical model authenticity. Parasitic spectral components near the frequency of the testing voltage ($f/f_{\max} = 1$) are due to the inertia of transistors switching process. In simulation (Fig. 4), their level is around 60...50 dB, and in the case of an analytical model, 70...65 dB, respectively.

The error of the analytical model in the case of taking into account the transistors inertial parameters is 15–30 %. Without considering inertial parameters the error increases up to 45–60 %. Similar results, taking into account the transfer function of the LPF, are also typical for the diagrams in Fig. 4*b*. However, in Fig. 4*b*, it can be seen also an increase in the components that are multiples of the PWM clock frequency, by an average of 15–30 dB, caused by a deviation from the nominal value of one of the inductances. At the same time, the coincidence between the results of analytical model and simulation in this part of the spectrum is rather high.

Speaking about the advantages associated with the use of an analytical model, it should be noted that the time expenditure when using it turned out to be 200–300 times less than in the case of simulation modelling. This circumstance is especially important when it becomes necessary to carry out statistical modelling, for example, when studying the influence of the spread of parameters of transistors and passive components of circuits on signal distortion. It is obvious that the introduction of parameters' dispersion into the MPS simulation model, which determine the inertial properties of active elements and their properties in the turn-on state, is a rather complicated and sometimes impossible task as some companies enable no model editing.

Research results

The proposed analytical model was used to evaluate distortions influence that occur while amplifying the envelope voltage of a multi-frequency signal in the MPS, assuming that the RF power amplifier does not introduce distortions.

The goal of modeling was to determine how bit error rate affected by the following factors:

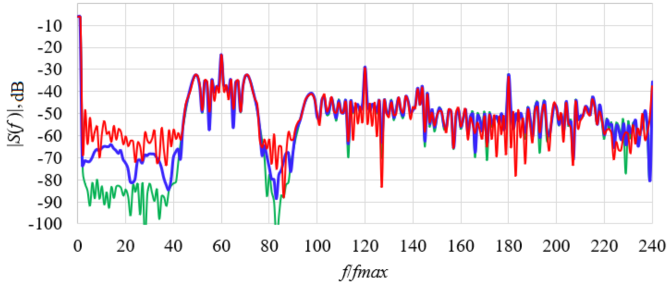
– parameters that determine the inertia of transistors (the turn-on and turn-off time, voltage rise and fall, included in the previously introduced vector τ);

– transistors' on-state voltage drop;

– statistical dispersion of the above transistors' parameters, as well as the ratings of some circuit passive elements;

– parameters of input signal and PWM reference voltage digital conversion.

a)



b)

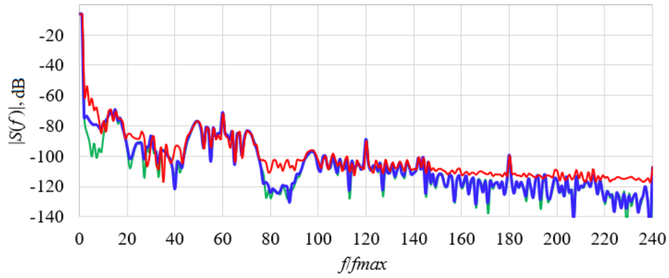


Fig. 4. Spectral diagrams of voltage u_{sp} (*a*), spectral diagrams of voltage e_{d2} (*b*)

The simulation was carried out for an OFDM signal with QAM-16. This choice was made intentionally, since it was found in [23] that QAM signals had stronger dependence of the bit error rate upon amplitude distortion than, for example, FSK signals. The results below correspond to the case of using a 4th order Bessel LPF with a bandwidth of five times the bandwidth occupied by the subcarriers of the test RF signal, which was conventionally taken as the frequency F_{max}, at $q = 15$, $n_d = 10$ bits, $K_p = 10$ bits. The filter type and its bandwidth correspond to the recommendations [2, 24].

The numerical experiment was carried out on the basis of statistical modeling, which included 1000 tests. For each test in the regions I – III, the averaged values of the corresponding spectral components were determined.

Estimation influence of transistors and MPS scheme elements parameters. Fig. 5*a* shows the dependences of the spectral components in region I (see Fig. 3) on the ratio τ/T. At the same time, in the region II, this dependence repeats the curve in Fig. 5*a*, for this reason, it is not shown. Fig. 5*b* shows a diagram of the dependence of the level of spectral components, multiples of the PWM clock frequency (region III), on the ratio τ/T. It should be noted that the study was carried out with a deviation of the parameters that determine the inertial properties of active elements within 10 % of the nominal value.

Fig. 5*a* shows that an increase in the vector τ from tenths of a percent to 10 % leads to an increase in the spectral components in the regions I and II by almost 50 dB. Moreover, a similar trend is observed with respect to components that are multiples of the PWM clock frequency. Besides, the spectral components in the III region can increase by almost 40 dB. The results obtained indicate that the use of a low-speed transistors can have a very negative impact on both the distortion of the RF signal and the level of out-of-band emission.

A study of transistors on-state voltage drop effect on spectrum distortions showed that when it changes by a factor of 50 in the range from 0.1 % to 5 % of the total voltage on the active element in the regions I and II, the spurious spectral components practically remain unchanged at the level of –105...–106 dB. For this reason, these dependencies are not shown in the paper. However, an increase in the on-state voltage drop a)

b)

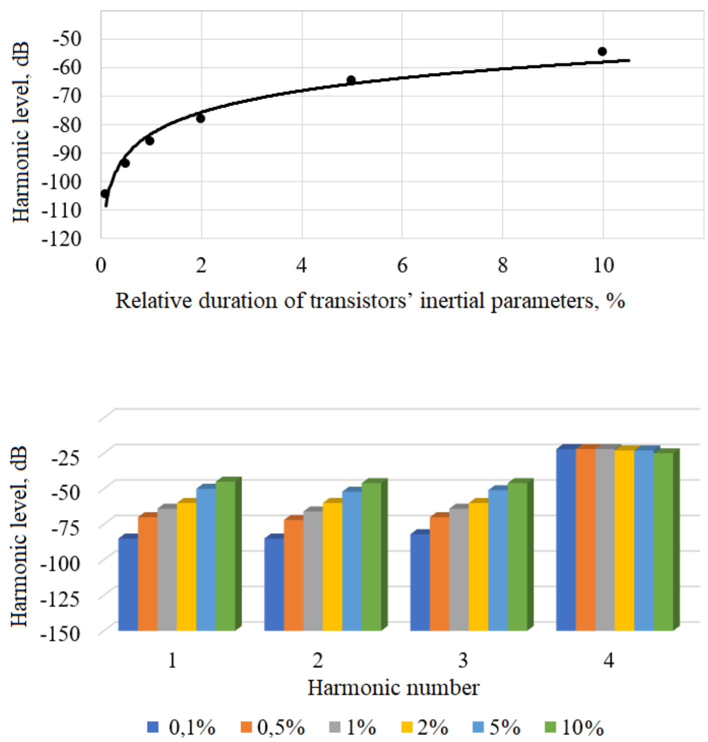


Fig. 5. The level of spectral components in the signal band (a), the level of spectral components that are multiples of the PWM clock frequency (b) for $\tau/T = 0.001$; 0.005; 0.01; 0.02; 0.05; 0.1

causes an increase in the components that are multiples of the PWM clock frequency, which is confirmed by the diagram shown in Fig. 6.

Figure 6 shows that with an increase in the normalized on-state voltage drop on transistors in the range from 0.1 % to 5 %, the first and second harmonics of the PWM clock frequency increase by almost 40 dB and can create certain problems in terms of out-of-band emissions. It should be noted that this change in the on-state voltage drop is not accompanied by an increase distortion in the signal band. This allows us to conclude that the on-state voltage drop does not affect a bit error rate when receiving a signal.

It was shown in [13] that the error in the phase setting between the output voltages of the MPS cells causes an increase in the spectral components that are multiples of the PWM clock frequency. However, it is important not only to determine the degree of influence of the error on these components, but also to find out whether it causes a distortion in the regions I and II. Figure 7 shows the results demonstrating the effect of the specified MPS parameter on the level of components that are multiples of the PWM clock frequency.

The magnitude of the phase's deviation of the cells output voltages is expressed as a percentage of the period of the PWM reference voltage. It should be noted that with a phase deviation reaching 10 %, the level of the components in the regions I and II increases by no more than 1–2 dB, remaining within the range of –90...–100 dB. Phase deviations have a more significant effect on components that are multiples of the clock frequency: for example, with an error of 10 %, the first harmonic of the clock frequency increases by almost 40 dB and can cause the problems mentioned above.

The effect of the inductances dispersion installed at the outputs of the MPS cells has a character similar to the effect of phase deviation. This is confirmed by Fig. 8, which shows the dependences of the spectral components level in the region III for a set of values of the normalized inductances dispersion from the

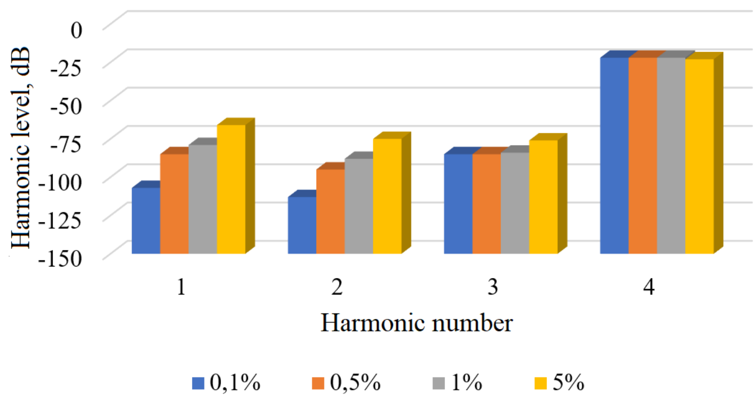


Fig. 6. Spectral components, multiples of the PWM clock frequency, depending on the magnitude of the on-state voltage drop

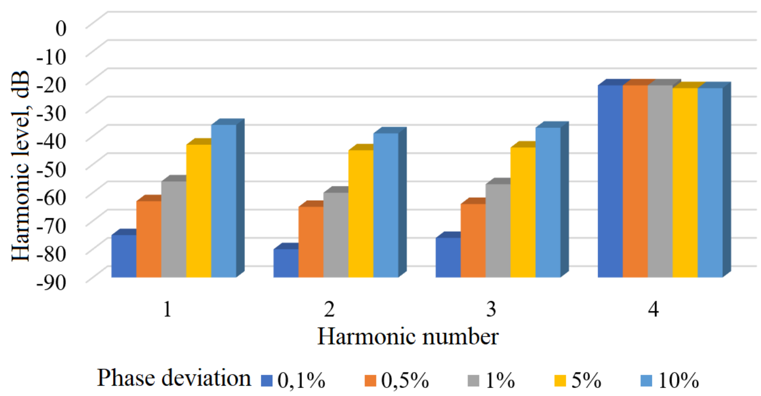


Fig. 7. The level of spectral components in the region III depending on the phases' deviation of the MPS cells' output voltages

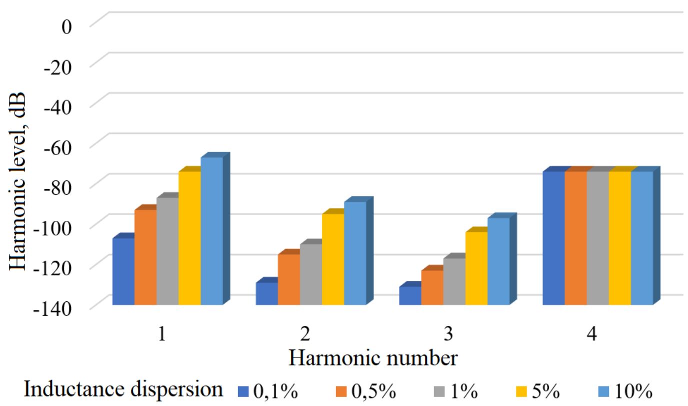


Fig. 8. The magnitudes of spectral components, multiples of the PWM clock frequency, depending on the dispersion of the inductances at the MPS cells' outputs

calculated values. In regions I and II, the spectral components level does not depend on the dispersion of inductances and is in the range of –95...–101 dB.

Fig. 8 shows that the level of spectral components that are multiples of the PWM clock frequency increases by 40 dB, which can also cause an increase in out-of-band emission.

Summarizing the obtained results, we can conclude that the most noticeable effect on the distortion of the envelope signal spectrum is caused by the insufficient speed of transistors. The transistors on-state voltage drop and the dispersion of inductances at the input of the LPF do not distort the spectrum in the band of the envelope signal, but mainly cause an increase in the multiples of the clock frequency.

Estimation bit error rate due to envelope voltage distortion. To assess how much the factors discussed above influence on the bit error rate, a numerical experiment was carried out using a communication channel model with additive Gaussian noise, while the number of tests was 10000. An OFDM signal based on 12 and 80 subcarriers with QAM-16 was used as a test signal. The numerical values of the MPS parameters used in the simulation are given in Section 5 of the paper.

Figure 9*a*, *b* shows the dependence of the bit error rate on the signal-to-noise ratio for an OFDM signal based on 12 and 80 subcarriers, respectively, calculated for different values of the parameters $\tau = \{\tau_{on} + \sigma_{on}, \tau_{off} + \sigma_{off}, \tau_r + \sigma_r, \tau_f + \sigma_f\}$, where $\sigma_{on} = 0.1\tau_{on}$, $\sigma_{off} = 0.1\tau_{off}$, $\sigma_r = 0.1\tau_r$, $\sigma_f = 0.1\tau_f$. In these figures, the zero value of the parameter along the abscissa axis corresponds to the idealized case, in which the influence of the MPS parameters is not taken into account. As the study showed, the energy loss due to the influence of other MPS parameters does not exceed 0.3 dB with a bit error rate of 10^{-3}.

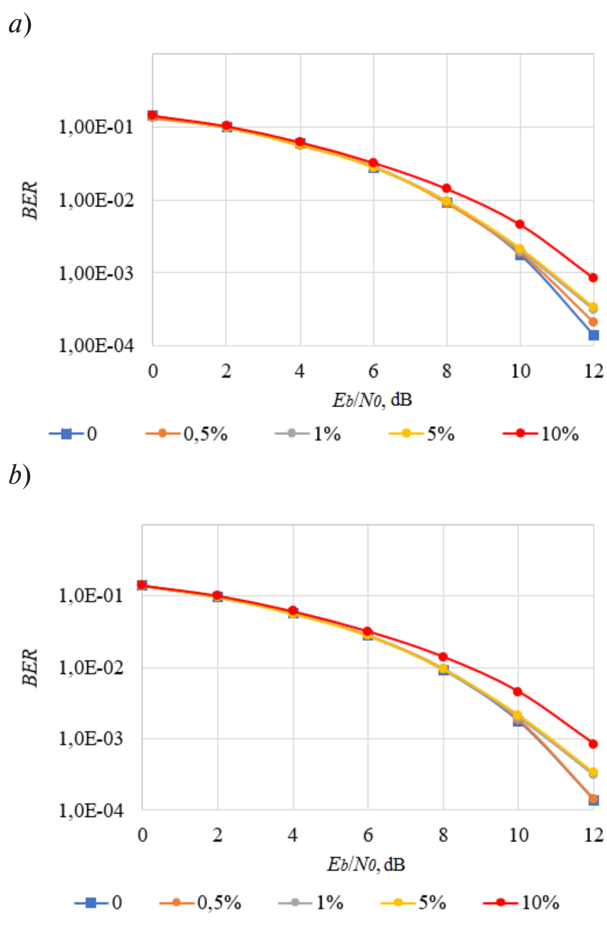


Fig. 9. BER as a function of signal-to-noise ratio for different τ/T

Analyzing the results presented in Fig. 9, the following conclusions can be made. The parameters that determine the inertial properties of active elements, as expected, affect the bit error rate. However, this is observed when the value of these parameters is in the region of 10 % of the PWM reference voltage period. In practice, this means the use of transistors with rather high switching times.

The influence of the dispersion of the MPS cells output voltages phases, the inductances at the input of the low-pass filter, the transistors' on-state voltage drop on the bit error rate is less expressed. The energy loss does not exceed tenths of a decibel.

Conclusion

Summarizing of the results presented above leads to the following conclusions.

1. The most significant influence on the distortion of the envelope signal, both in terms of increasing the level of bit error rate and in terms of spectrum degradation, is exerted by the inertia of transistors. For the QAM-16 signals considered as an example, this effect starts at $\tau/T \geq 0.005$ and, with a bit error rate of 10^{-3}, causes an energy loss of the order of 1.5 dB. Moreover, this pattern persists with an increase in the number of subcarriers in the OFDM signal from 12 to 80.

2. The transistors on-state voltage drop, the phase deviation of the output voltages of the MPS, and the dispersion of inductances at the outputs of the MPS cells do not have a pronounced effect on the bit error rate, regardless of the number of subcarriers (12 or 80).

3. It was found that the parameters listed in Section 2, under certain conditions, contributed to the appearance of spectral components that are multiples of the PWM clock frequency. Based on the allowable level of parasitic components equal to –80 dB, then the dispersion of inductance ratings should not exceed 1 %, and the phase deviation of the output voltages of the MPS cells should not exceed 0.1 %. Failure to follow these recommendations may result in an increase in the level of out-of-band emissions.

4. The proposed approach can be extended to arbitrary types of signals with a relatively high peak-to-average power ratio, which are amplified using the envelope tracking method.

REFERENCES

1. **Darwazeh I., Ghannam H., Xu T.** The first 15 years of SEFDM: A brief survey. *2018 11th International Symposium on Communication Systems, Networks & Digital Signal Processing* (*CSNDSP*). Budapest, Hungary, 2018, Pp. 1–7. DOI: 10.1109/CSNDSP.2018.8471886

2. **Wang Z.** *Envelope tracking power amplifiers for wireless communications*. Artech House, 2014. 343 p.

3. **Kazimierczuk M.K.** *RF Power amplifiers*. John Willey Sons, Chichester, 2014. 658 p.

4. **Sankman J., Song M.K., Ma D.** Switching-converter-only multiphase envelope modulator with slew rate enhancer for LTE power amplifier applications. *IEEE Transactions on Power Electronics*, 2016, Vol. 31, no. 1, Pp. 817–826. DOI: 10.1109/TPEL.2015.2406710

5. **Jin Q., Ruan X.** Switch-linear hybrid envelope-tracking power supply with multilevel structure. *IECON 2013–39th Annual Conference of the IEEE Industrial Electronics Society*, 10–13 Nov. 2013, Pp. 1325–1330. DOI: 10.1109/IECON.2013.6699324

6. **Suetsugu T., Oyama N., Kuga S., Taromaru M., Wei X.** Comparison of EPWM architectures for class E ampliers in EER system. *16th International Power Electronics and Motion Control Conference and Exposition*, Antalya, Turkey, 21–24 Sept. 2014, Pp. 331–336. DOI: 10.1109/EPEPEMC.2014.6980514

7. **Vasic M., Garcia O., Oliver J.A., Alou P., Diaz D., Cobos J.A., Gimeno A., Pardo J.M., Benavente C., Ortega F.J.** Efficient and linear power amplifier based on envelope elimination and restoration. *IEEE Transactions on Power Electronics*, 2012, Vol. 27, no. 1, Pp. 5–9. DOI: 10.1109/TPEL.2011.2162005

8. **Roblin P., Barton T.W., Chang H.C., Liang C., Sear W.** Design of a 4-Way Chireix amplifier using a nonlinear embedding device model. *2017 IEEE 18th Wireless and Microwave Technology Conference* (*WAMICON*). Cocoa Beach, FL, USA, 24–25 April 2017. DOI: 10.1109/WAMICON.2017.7930280

9. **Ghahremani A., Annema A.J., Nauta B.** Outphasing class-E power amplifiers: From theory to back-off efficiency improvement. *IEEE Journal of Solid-State Circuits*, 2018, Vol. 53, no. 5, Pp. 1374–1386. DOI: 10.1109/JSSC.2017.2787759

10. **Jang H., Wilson R., Canning T., Seebacher D., Schuberth C., Arigong B.** Self-outphasing Chireix power amplifier using device input impedance variation. *2016 IEEE MTT-S International Microwave Symposium* (*IMS*), San Francisco, CA, USA, 22–27 May 2016. DOI: 10.1109/MWSYM.2016.7540027

11. **Kim B.** *Doherty power amplifiers from fundamentals to advanced design methods*. Elseivier Inc, 2018. 174 p.

12. **Andersson C.M., Gustafsson D., Cahuana J.C., Hellberg R., Fager C.** A 1 to 3 GHz digitally controlled dual-RF input power amplifier design based on a Doherty-outphasing continuum analysis. *IEEE Transactions on Microwave Theory and Techniques*, 2013, Vol. 61, no. 10, Pp. 3743–3752. DOI: 10.1109/TMTT.2013.2280562

13. **Pergushev A.** The analytical model for calculating distortions in the envelope tracking power supply. *2019 IEEE International Conference on Electrical Engineering and Photonics* (*EExPolytech*), St. Petersburg, Russia, 17–18 Oct. 2019, Pp. 72–75. DOI: 10.1109/EExPolytech.2019.8906813

14. **Guo Y., Zhu A.** Power adaptive decomposed vector rotation based digital predistortion for RF power amplifiers in dynamic power transmission. *2017 IEEE Topical Conference on RF/Microwave Power Amplifiers for Radio and Wireless Applications* (*PAWR*), Phoenix, AZ, USA, 15–18 Jan. 2017, Pp. 8–10. DOI: 10.1109/PAWR.2017.7875559

15. **Varahram P., Jamuar S.S., Mohammady S., Hamidon M.N., Khatun S.** Power amplifiers linearization based on digital predistortion with memory effects used in CDMA applications. *2007 18th European Conference on Circuit Theory and Design*, Seville, Spain. 27–30 Aug. 2007, Pp. 488–491. DOI: 10.1109/ECCTD.2007.4529639

16. **Lee Y.S., Lee M.W., Jeong Y.H.** High-power amplifier linearization using the Doherty amplifier as a predistortion circuit. *IEEE Microwave and Wireless Components Letters*, 2008, Vol. 18, no. 9, Pp. 611–613. DOI: 10.1109/LMWC.2008.2002461

17. **Wang S., Hussein M.A., Venard O., Baudoin G.** Identification of low order cascaded digital predistortion with different-structure stages for linearization of power amplifiers. *2018 IEEE Topical Conference on RF/Microwave Power Amplifiers for Radio and Wireless Applications* (*PAWR*), Anaheim, CA, USA, 14–17 Jan. 2018, Pp. 62–65. DOI: 10.1109/PAWR.2018.8310068

18. **Wu S., Hau S.F.S., Wong Y.M.** The use of nonlinear curvefitting in digital predistortion for linearizing RF power amplifiers. *2005 5th International Conference on Information Communications & Signal Processing*, Bangkok, Thailand, 06–09 Dec. 2005, Pp. 960–963. DOI: 10.1109/ICICS.2005.1689192

19. **Yang X., Roblin P., Chaillot D., Mutha S., Strahler J., Kim J., Ismail M., Wood J., Volakis J.** Fully orthogonal multi-carrier predistortion linearization for RF power amplifiers. *2009 IEEE MTT-S International Microwave Symposium Digest*, Boston, MA, USA, 07–12 June 2009, Pp. 1077–1080. DOI: 10.1109/MWSYM.2009.5165887

20. **Han R., Jiang C., Yang G., Tan J., Liu F.** Pruned basis space search for digital predistortion of RF power amplifiers. *IEEE Transactions on Microwave Theory and Techniques*, 31 Jan. 2023, Pp. 1–12. DOI: 10.1109/TMTT.2023.3239794

21. **Pergushev A., Sorotsky V., Ulanov A.** Output voltage PWM conversion inaccuracies in envelope tracking power supply for high peak-to-average power ratio applications. *2019 IEEE International Conference on Electrical Engineering and Photonics* (*EExPolytech*), St. Petersburg, Russia, 17–18 Oct. 2019, Pp. 9–12. DOI: 10.1109/EExPolytech.2019.8906854

22. **Pergushev A., Sorotsky V., Ulanov A.** Criteria for selection envelope tracking power supply parameters for high peak-to-average power ratio applications. *2019 IEEE International Conference on Electrical Engineering and Photonics* (*EExPolytech*), St. Petersburg, Russia, 17–18 Oct. 2019, Pp. 13–16. DOI: 10.1109/EExPolytech.2019.8906793

23. **Singh A., Kaur H.** Non linearity analysis of high power amplifier in OFDM system. *International Journal of Computer Applications*, 2012, Vol. 37, no. 2, Pp. 37–41.

24. **Kahn L.R.** Single-sideband transmission by envelope elimination and restoration. *Proceedings of the IRE*, 1952, Vol. 40, no. 7, Pp. 803–806. DOI: 10.1109/JRPROC.1952.273844

INFORMATION ABOUT AUTHORS/СВЕДЕНИЯ ОБ АВТОРАХ

Пергушев Алексей Олегович
Aleksey O. Pergushev
E-mail: pergushev.aleksey@mail.ru

Сороцкий Владимир Александрович
Vladimir A. Sorotsky
E-mail: sorotsky@mail.spbstu.ru

Завьялов Сергей Викторович
Sergey V. Zavjalov
E-mail: zavyalov_sv@spbstu.ru

Submitted: 21.02.2023; Approved: 22.03.2023; Accepted: 17.05.2023.
Поступила: 21.02.2023; Одобрена: 22.03.2023; Принята: 17.05.2023.

Research article
DOI: https://doi.org/10.18721/JCSTCS.16103
UDC 62.5

REQUIREMENTS FOR COMMUNICATION AND POSITIONING SYSTEMS FOR GROUP OPERATION OF AUTONOMOUS UNMANNED UNDERWATER VEHICLES

N.N. Semenov[1], *V.G. Mikhlin*[2], *D.B. Akhmetov*[3] ✉

[1,2] St. Petersburg State Marine Technical University, St. Petersburg, Russian Federation;

[3] Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation

✉ akhmetov@spbstu.ru

Abstract. When gathering a group of AUVs in a limited area, both the safety of their work and the possibility of participation are necessary. In a group, AUV should be out of the way of the other devices in the very least, but ideally should contribute to completion of the common task. This paper discusses issues related to the parameters of communication systems and positioning of AUVs working in a group to solve a single common task. The paper presents the numerical calculation results of the dependence of the AUV efficiency on the size of the subgroup. An assessment of the delays in data transmission and restructuring of the synchronization method in the group is also given, as well as the possibility to simplify group management in the case under study.

Keywords: robotics, underwater robotics, group application, AUV, communication systems, positioning

Citation: Semenov N.N., Mikhlin V.G., Akhmetov D.B. Requirements for communication and positioning systems for group operation of autonomous unmanned underwater vehicles. Computing, Telecommunications and Control, 2023, Vol. 16, No. 1, Pp. 35–45. DOI: 10.18721/JCSTCS.16103

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.16103
УДК 62.5

ТРЕБОВАНИЯ К СИСТЕМАМ СВЯЗИ И ПОЗИЦИОНИРОВАНИЯ ДЛЯ ГРУППОВОЙ РАБОТЫ АВТОНОМНЫХ НЕОБИТАЕМЫХ ПОДВОДНЫХ АППАРАТОВ

Н.Н. Семенов[1], *В.Г. Михлин*[2], *Д.Б. Ахметов*[3] ✉

[1,2] Санкт-Петербургский морской технический университет,
Санкт-Петербург, Российская Федерация;

[2] Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

✉ akhmetov@spbstu.ru

Аннотация. При использовании группы АНПА на ограниченной территории необходимо обеспечить как безопасность их работы, так и возможность взаимодействия. АНПА в группе не должны мешать работать другим аппаратам, их цель – помогать решению общей для всей группы задачи. В статье рассмотрены вопросы необходимых параметров систем связи и позиционирования АНПА, работающих в группе для решения единой для всей группы задачи. Показаны результаты численного моделирования зависимости эффективности группы АНПА от размера подгруппы. Также дана оценка влияния задержек передачи данных и перестроения на синхронизацию в группе и показан способ упрощения управления группой в данном случае.

Ключевые слова: робототехника, подводная робототехника, групповое применение, АНПА, системы связи, позиционирование

Для цитирования: Semenov N.N., Mikhlin V.G., Akhmetov D.B. Requirements for communication and positioning systems for group operation of autonomous unmanned underwater vehicles // Computing, Telecommunications and Control. 2023. Т. 16, № 1. С. 35–45. DOI: 10.18721/JCSTCS.16103

Introduction

Autonomous unmanned underwater vehicles (AUVs), "mobilis in mobili", move in a complex mobile environment, which is rather muddy, therefore the optical methods of observation and detection work only at short distances and do not pass radio waves. This complicates communication and positioning, while acoustic communication, though allows information transmission for long distances, is significantly limited by communication speed and range. These limitations are particularly severe when using groups of AUVs – information transmission between two AUVs becomes interference for other AUVs, preventing them from exchanging information. Therefore, communication and positioning issues for AUV group operation are still unsolved, being urgent and important.

Positioning

Satellite positioning systems used for land, surface and air navigation (GPS, GLONASS, Galileo, ...) cannot work under water, since water is a conductor and electromagnetic waves quickly attenuate. Even 10 cm of the water layer do not allow to use them [1].

Creating a global system of underwater acoustic navigation proves technically difficult, although the communication ranges may be sufficient for this, since the acoustic signal propagation is affected by the layered structure of the propagation medium (changes in the speed of sound in water depending on pressure, salinity and temperature) and reflection from boundaries (bottom and surface) [2]. Therefore, the construction of such systems is not even considered at present.

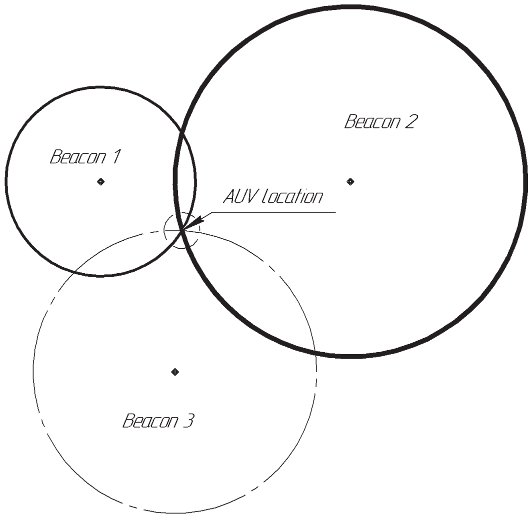


Fig. 1. Example of trilateration for three beacons

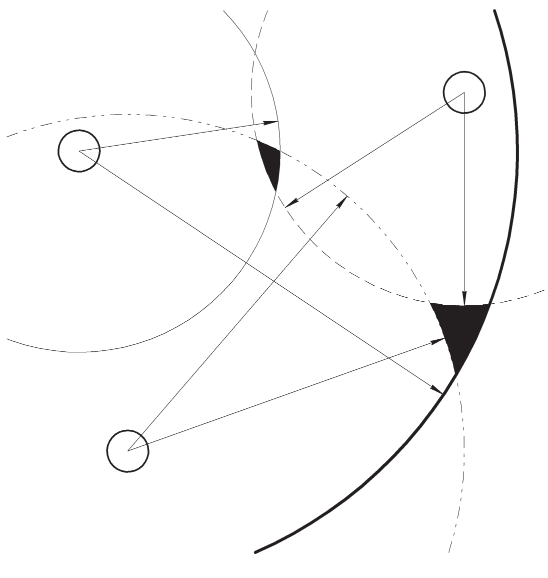


Fig. 2. Example of the beacon positioning errors structure

There are local positioning systems, which can be divided into three main directions: positioning with long base (LBL – Long Base Location), positioning with short base (SBL – Short Base Location) and positioning with ultra-short base (USBL – Ultra-Short Base Location) [3–6].

LBL is a system consisting of a set of hydroacoustic beacon responders located at the boundaries of the work area and moving AUVs. Upon request from the AUV, the beacons send a response signal with a known delay. The AUV calculates its own position by trilateration based on the response signal delays and knowledge of the exact position of the beacons, as shown in Fig. 1. The accuracy of positioning in such a system is determined by the accuracy of signal arrival delays, nonlinear signal propagation, position relative to the beacons, and the positioning inaccuracies of the beacons themselves, which may have an overhead part that receives GPS coordinates [7]. It is the most exact method for the whole area of works (accuracy of tens of centimeters when the distance from beacons is not more than 2000 m and when being inside the zone limited by beacons), but it is also the most time-consuming method for water area preparation – beacons must be placed and their coordinates adjusted beforehand. The structure of position-dependent coordinate errors is shown in Fig. 2, which shows that while the vehicle is inside the area bounded by beacons, the errors are symmetrical relative to the true position of the vehicle, and they can be reduced by repeated measurements. But when the vehicle leaves this area, the errors become asymmetrical, and the positioning accuracy drops dramatically.

The accuracy of the LBL, as seen in Fig. 2, is determined by the accuracy of distance, and as practice shows, due to the nonlinear propagation of acoustic signals is up to 5 % of the distance [8, 9]. Consequently, the further from the center of the area bounded by beacons, the greater the positioning error. Since the error is static, it cannot be removed by repeating measurements. If errors of defining distance E are expressed as vectors Δr_i, which direction coincides with the direction on the beacon, and the length is equal to the error standard deviation, the quantity of beacons N, the positioning error E can be expressed through product of these vectors:

$$E = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\overrightarrow{\Delta r_i}\,\overrightarrow{\Delta r_j}.$$

SBL is a system that has multiple hydroacoustic response beacons, located within the working area, and moving AUVs. The beacons can be positioned on the sides of the support vessel (Fig. 3), which limits the distance between the beacons. This reduces water area preparation time, but increases AUV positioning error – it depends on the distance from the beacons. The errors structure is similar to the long base location, when the device is outside the zone limited by beacons.

The ultrashort base location assumes that the receiving antennas on AUV contain several receivers, the arrival direction is determined by the phase difference, and the distance is determined by the delay. Thus, it is possible to determine the location of all AUVs in a group by one beacon, as well as to determine the mutual location of AUVs relative to other AUVs (Fig. 4).

The accuracy of determining the coordinates in the system with ultra-short base depends not only on the accuracy of distance (it depends on the distance similar to other methods and is up to 5 % of the distance, denoted as Δr, m), but also on the accuracy of angles ($\Delta\alpha$, radians), is about 1° which when using small antennas [10, 11]. Then the accuracy will be as follows:

$$E = \Delta r \Delta\alpha.$$

Inertial positioning systems are based on Newton's laws of mechanics for inertial reference systems. Such systems are divided into angle measuring systems (gyroscopes), which allow to keep a given direction for a long time, and linear acceleration sensors (accelerometers), data from which can be converted into the current AUV coordinates by double integration method. Double reintegration of acceleration in time leads to the fact that such systems quickly enough accumulate error [12], and therefore can be used only for tasks lasting units to tens of minutes.

Optical positioning systems are often used to control a large number (up to several thousands) of flying vehicles or ground robots [13]. To do this, the signal from video cameras (usually several to suppress collisions when one vehicle is behind another) is fed to a pattern recognition system, objects are selected, identified, and by analyzing their position and motion parameters, control commands are formed. But such systems cannot work in water, as water is quite muddy and refracts optical beams. There are optical communication systems, when an amplitude-modulated optical signal propagates in water for hundreds of meters and allows a high speed of information transmission. According to the propagation delays of such a signal, it is possible to measure the distance between two devices under water [14]. The advantages are high accuracy, possibility of simultaneous operation of several AUVs, the disadvantage is limited range.

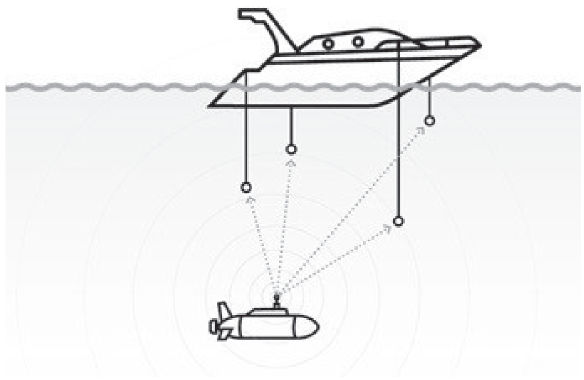


Fig. 3. Example of SBL

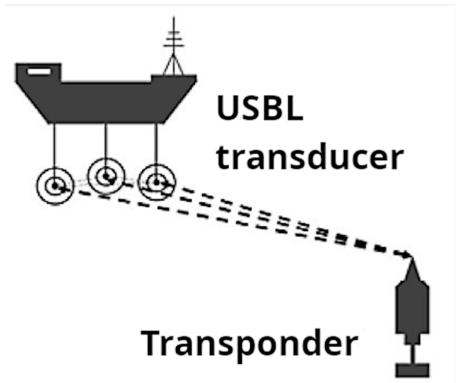


Fig. 4. Example of USBL

Thus, for a group of AUVs, the optimal solution for positioning is a system with ultra-short base location on each of the devices, supplemented with inertial systems for Kalman filtering of current coordinates. The current accuracy is no worse than 0.1° by angle, no worse than 10 cm by distance in the range of 0...4000 m in a homogeneous medium, or no more than 1.5 % of the distance when taking into account the nonlinear propagation of acoustic waves [8]. Frequency range is 25–50 kHz, the size of the antenna is 150 × 150 × 100 mm.

Collecting and transmitting information

After launching the AUV starts execution of the set program, it moves in muddy water, and thus the visual control of the state is impossible. Therefore, a communication channel is required to monitor the current state of each AUV. When the task changes, a new command should be communicated to each AUV. If it is impossible to transmit a message from one AUV to the control center (control service of the whole AUV network, located on the support vessel or on shore), it is necessary to provide retransmission of the message through other AUVs.

When working in a group the following takes place: distribution of tasks between AUVs, control of neighbors' status, joint maneuvering and avoiding obstacles. This requires a communication channel between ANPAs in the group.

Existing hydroacoustic modems have a range up to 4000–5000 m, but are severely limited by the speed of information transfer. Therefore, it is preferable to use faster modems with lower transmission speed to exchange information in a group, and a modem with maximum range, though with low transmission speed to transfer information to the control center. Thus, if each AUV in a group of AUVs is equipped with a high-speed modem with a range sufficient for information exchange within the group, and several AUVs in this group are additionally equipped with modems with a long transmission range, the whole group will be equipped with everything necessary for information exchange both between themselves and with the control center.

Hydroacoustic modems at distances up to 300 m can transmit up to 60 kbit/s and more, while at distances up to 1000 m – only up to 30 kbit/s, and the speed drops to 12–13 kbit/s at distances up to 3500 m, and to tens of bits per second at distances over 4000 m [15]. This is due to the peculiarities of signal propagation in water, the high level of natural and artificial noises as well as the restrictions on the size of the antennas on the AUV. The Shannon-Hartley theorem [15] describes the relationship of information transmission rate as a function of signal-to-interference ratio and bandwidth as follows:

$$C = B\log_2\left(1+\frac{S}{N}\right),$$

where C is the channel bandwidth, bits/s; B is the bandwidth of the communication channel, Hz; S/N is the ratio of signal power to noise.

And the signal-to-noise ratio S/N (SNR) is determined from the basic equation of hydroacoustics [16]:

$$SNR = RL - IL - PL + AD,$$

where RL is the radiation level; IL – the interference level; PL – the propagation loss; AD – the antenna directivity parameter. All values are in dB.

The radiation level is limited by the area of the antenna and is a design parameter, like the directivity parameter. Interference level is determined by the bandwidth of the receiver tuned to the signal bandwidth, making signal bandwidth expansion an unaffected factor in the S/N ratio. The propagation loss is determined by the widening of the signal wave front and the absorption of the wave by the aqueous medium:

$$RL = \beta r + 20\log(r),$$

where r is the distance, m; β is the spatial attenuation coefficient, dB/m.

Thus, the range and speed of communication are design parameters of hydroacoustic modem and cannot be better than theoretical limit.

If AUVs move in a sufficiently compact group, it is possible to use optoacoustic communication between AUVs – an acoustic modem for long-range communication and an optical modem with a range up to 200–300 m for communication of AUVs inside the group. Both suspended particles and dissolved organic matter (DOM) (fatty acids and amino acids, amino sugars, chlorine pigments, hydrocarbonates, phenols, etc.) affect absorption of light radiation. Due to differences in the DOM composition, there are differences in the estimates of the attenuation coefficient α. It was established by means of experimental studies [8] that the absorption index of dissolved organic matter changes according to the following law:

$$\alpha(\lambda) = \alpha_i e^{-\mu_i(\lambda-\lambda_0)},$$

where μ depends on concentration and on wavelength λ; α_0 – the absorption at wavelength λ_0. The blue-green light with λ_0 = 520–530 nm has the minimum absorption coefficient in seawater.

The following scattering coefficient is used for monochromatic green LED, which has a wide directionality:

$$b = \frac{d^2}{4\left(d + rtg(\alpha)\right)^2},$$

where b is the fraction of light after scattering; d – the diameter of the emitter; r – the distance; α – the given angle of illumination.

Thus, the maximum amount of light energy that will reach the receiver at a distance of 200 m will be about 1 %, but the light absorption coefficient can reach the value of 0.4 in coastal waters, then the transmission factor will be 10–89 for a distance of 200 m, which technically does not allow to transmit signals at such a distance. The maximum communication distance in muddy water will be 9.4 m with a receiver sensitivity of 10^{-6} of the transmitter power.

Increasing the transmitter power will allow us to transmit at distances of over 200 m (theoretically up to 500 m) in clear sea water, but such distances are unattainable in coastal waters because of active light scattering on suspended particles.

Such modems have high speed of information transmission, small delay, allow several messages to be transmitted simultaneously, but are limited in range.

The recommended hydroacoustic modem is a USBL modem with a broadband antenna, able to operate both in the high frequency range for high-speed exchange with other AUVs and positioning, and in the low frequency range for message transmission over long distances, combined with an optical communication system. This system allows significant expansion of high-speed data transmission between network agents.

Message relaying protocols are well considered for peer-to-peer (mesh) communication networks [9], and will not be considered in this paper.

Automatic network reconfiguration

The following abnormal situations may occur during the operation of the AUV group:

1. Failure of one or more AUVs;
2. Addition of one or more AUVs to the group;
3. Obstacles in the group's way;
4. Change of task for the group.

A group of AUVs moves in formation along a given route. We do not consider interaction protocols for route selection yet. It is necessary to control the group state and bring the changed information to all AUVs in the group taking into account peculiarities of underwater communication and positioning.

It is necessary to maintain formation in moving environment, i.e. to position each AUV relative to neighbors and relative to the bottom, to move in formation. The frequency of mutual position request depends on the speed of the vehicles and the distance between them, as well as on the accuracy of keeping the position in formation. For example, it is enough to make a request at least once every 3 s, at speed of 3 m/s, 30 m distance between units, and position accuracy of 3 m. Each packet transmits its state (serviceable / faulty / battery loaded) and the calculated coordinate.

If one of the AUVs does not respond at the next request cycle, that AUV is requested again, confirming that it is out of order, and then the group begins to realign. If one request cycle begins to rearrange only one AUV, it will take a lot of time to rearrange the whole group, which will lead to decrease of group efficiency, therefore it is necessary to inform all devices that need to be rearranged in minimum time.

When new AUVs are added to the group (or contact is restored with old ones leaving the group), the new AUVs start the group realignment procedure so that the group is lined up in a given formation in minimal time with minimal loss of resources. The optimization procedure is a mathematically solved "consensus problem" for multiagent systems.

Bypassing obstacles and changing the group's task triggers a new rearrangement with recounting of participants, after which the group moves in an organized formation.

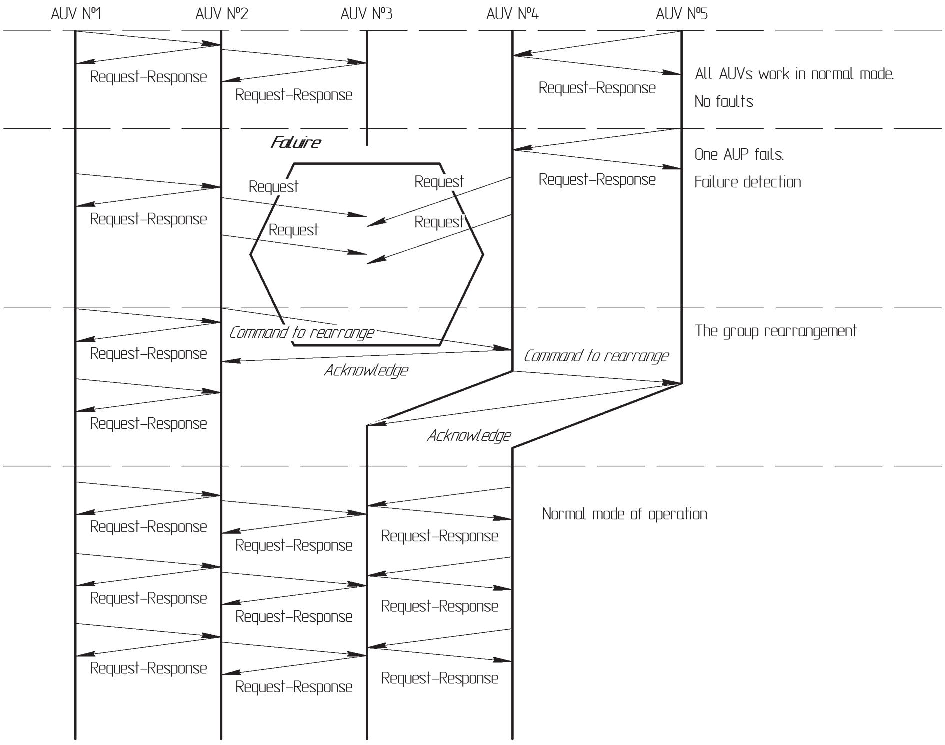


Fig. 5. Temporal diagram of exchange between AUVs when one AUV fails

The time to make a decision depends on the number of vehicles, the range and speed of communication, and also the communication protocol (how many messages to send and who to) [17].

Numerical modeling of AUV group behavior to select communication and navigation system requirements

Let us consider the necessary information exchange for movement of a group of 100 AUVs moving at a speed of 4 m/s in a line to inspect the territory of the bottom 20 × 20 km for detection of sunken objects. The distance between the AUVs in the line is 50 m. Range of communication and positioning within the group is 300 m, communication speed is 50 kbit/s. Positioning based on USBL modems. All the AUVs are the same, there are no dedicated "commanders" in the group. The example of exchange between 5 AUVs in case of failure of AUV 3 is shown in Fig. 5.

The described protocol of accident exchange and detection followed by rearrangement works well when the number of AUVs in a group does not exceed 15–20. In large groups, it is suggested to divide the groups into subgroups of up to 15–20 AUVs, monitor the state of the whole subgroup within this subgroup, and inform other subgroups only when an accident has occurred or rearrangement is required. In this case the accident of one AUV is detected within a subgroup, this subgroup receives a command to rearrange within it, and in parallel informs the other subgroups where to rearrange. Having received the command to realign, the other subgroups disseminate the information within the subgroup and further to the remaining subgroups, after which the entire subgroup realigns.

Let us perform numerical simulation of behavior of a large group of AUVs (300 units) with different number of subgroups. Let us derive group efficiency (as a ratio of probability of object detection on the bottom to time of group operation $E = P/t$), consider the probability density of failures in time constant, equal for all AUVs and chosen so that during the whole mission 5 % of AUV failures occurred as a comparison criterion. All the AUVs are in good working order at the start of the mission.

The signal processing time is not taken into account, only the time for signal propagation between AUVs is:

$$t_{ij} = \frac{r_{ij}}{c},$$

where r_{ij} is the distance between units i and j; c is the speed of sound in water.

When the AUV fails due to delays in neighbor requesting and signal propagation in the course of the mission, there are areas not surveyed by either AUV, so the resulting probability of detection P_{res} is expressed as follows:

$$P_{res} = P_d\left(1 - \frac{\sum_{i=1}^{M} S_{i,unsurv}}{S_{\Sigma}}\right),$$

where P_d is the probability of detection during site survey; S_{Σ} – the total survey area; $S_{i,unsurv}$ is the area of the section where the AUV failed and which was left unsurveyed (so as not to stop the whole group); M is the number of unsurveyed sections.

The results of the simulation are shown in Table 1 and in Fig. 6.

Table 1

Results of modeling group efficiency as a function of subgroup size

Test number	Subgroup size	Operation time, mins	Efficiency, %	Comment
1	No failures	500	1/500 = 100 %	Reference efficiency
2	0-No MAC	1827	27	Several inspections
3	1	936	53	–
4	3	895	56	–
5	10	677	74	Optimal size
6	20	706	71	–
7	50	798	62	–
8	100	863	57	–
9	300	936	53	–

As it can be seen in Table 1 and Fig. 6, the efficiency is minimal without use of MAC (group size 0). When any AUV fails, it is necessary to repeat inspection of the territory which was to be inspected by the failed AUV, which greatly increases total inspection time and decreases efficiency. When subgroup size is equal to one AUV or 300 (all AUVs in one subgroup), the efficiency is determined by the speed of message transmission between all AUVs through the chain, acknowledgement of reception, and sequential rearrangement. As the size of a subgroup increases, it is possible to divide and parallelize the procedure of requesting the AUV serviceability in each subgroup; when a failure occurs, not individual AUVs but subgroups are rearranged, which significantly increases the efficiency of the whole group. But with further increase in the number of AUVs in a subgroup, the time of the requesting AUV serviceability increases again, and efficiency decreases.

Conclusion

Based on the review of possible positioning and communication systems, the authors propose using a broadband acoustic modem with a mutual positioning system on ultra-short distances and optical commu-

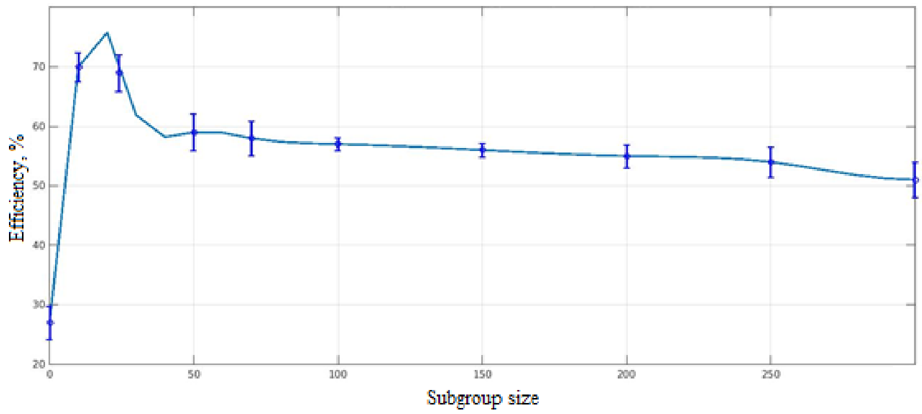


Fig. 6. Dependence of efficiency on subgroup size

nication system on short distances for each AUV. Such a solution makes it possible to transmit information both for long distances (for communication with the control center) and exchange information between a large number of AUVs in a group with small dimensions of the modem and optical system.

Since information exchange and subsequent realignments are delayed, synchronization of actions for large groups of AUVs is quite time consuming. However, controlling a large group is much easier and faster if large groups are divided into subgroups, and each subgroup is provided with state control, positioning and joint rearrangements.

The dependence of AUV group efficiency on the size of a subgroup became clear as a result of a numerical simulation: there is a number of AUVs in the group optimal for the chosen parameters and performed tasks. There are 10...20 units in the example above. Such a number of vehicles allows controlling the number and condition of all vehicles in the group, to transmit commands and to monitor their execution.

Thus, this article proposes a mathematical dependence of the effectiveness of a group of AUVs (the probability of completing a task in a given time) on the speed and range of mutual communication, as well as on the size of the subgroups the group is divided into to reduce the number of messages sent. Such an approach was not found in other materials available for analysis.

REFERENCES

1. **Alcocer A., Oliveira P., Pascoal A.** Underwater acoustic positioning systems based on buoys with GPS. *Proc. of 8th ECUA*, June 12-15, 2020, pp. 1–8.

2. **Tang Suhua, Chen Hongbin, Zhu Yi, Zhang Wenwen, Wu Kefei, Yuan Fei.** Underwater acoustic micromodem for underwater Internet of Things. *Wireless Communications and Mobile Computing*, 2022, pp. 1530–1543. DOI 10.1155/2022/9148756

3. **Batista P., Silvestre C., Oliveira P.** GES integrated LBL/USBL navigation system for underwater vehicles. *Proceedings of the IEEE Conference on Decision and Control*, 2012. DOI: 10.1109/CDC.2012.6426614

4. **Batista P., Silvestre C., Oliveira P.** A time differences of arrival-based homing strategy for autonomous underwater vehicles. *International Journal of Robust and Nonlinear Control*, 2010. DOI: 10.1002/rnc.1547

5. **Choukroun D.** Novel results on quaternion modeling and estimation from vector observations. *AIAA Guidance, Navigation, and Control Conference*, 2009. DOI: 10.2514/6.2009-6315

6. **Morgado M., Oliveira P., Silvestre C., Vasconcelos J.F.** USBL/INS tightly-coupled integration technique for underwater vehicles. *Proceedings of the American Control Conference. IEEE Xplore 2014.* DOI: 10.23919/ACC.2004.1383952

7. **Whitcomb L.L., Yoerger D.** Hanumant Singh combined Doppler/LBL based navigation of underwater vehicles. *IEEE T Autonomous Robots*, 2016.

8. **Aparicio J., Jimenez A., Alvarez F.J., Ruiz D., Urena J.** Characterization of an underwater positioning system based on GPS surface nodes and encoded acoustic signals. *IEEE T INSTRUM MEAS*, 2016. DOI: 10.1109/TIM.2016.2552699

9. **Kinsey J.C., Qingjun Yang, Howland J.C.** Nonlinear dynamic model-based state estimators for underwater navigation of remotely operated vehicles. *IEEE T CONTR SYST T*, 2014. DOI: 10.1109/TCST.2013.2293958

10. **Han Yun-Feng, Derek Li, Zheng Cui-E, Sun Da-Jun.** A precision evaluation method of USBL positioning systems based on LBL triangulation. *Acta Physica Sinica*, Chinese Ed., 2015, 64 (9). DOI: 10.7498/aps.64.094301

11. **Pennec S., Touin G., Azou S., Collin L.** On joint acoustic communication and positioning through MFSK-modulated signals and Costas arrays. *IEEE Xplore. OCEANS 2010 IEEE – Sydney*, 2020. DOI: 10.1109/OCEANSSYD.2010.5603656

12. **Korkishko Yu.N., Fedorov V.A., Prilutskiy V.E., et al.** High-precision inertial measurement unit IMU-5000. *2018 IEEE International Symposium on Inertial Sensors and Systems* (*INERTIAL*). DOI: 10.1109/ISISS.2018.8358121

13. **Kin-Hsing Hsieh.** *Visual positioning system and method for mobile user equipment*. Patent US20080153516A1. 2017.

14. **Casimer DeCusatis, Ivan Kaminow.** *The optical communications reference*. Oxford: Academic Press, 2010, 473 p.

15. **Kulik S.Yu., Rodionov A.Yu., Dubrovin F.S.** On reliability of data transmission and distance estimation using mobile underwater acoustic modems. *25th Saint Petersburg International Conference on Integrated Navigation Systems. 2018, May*. DOI: 10.23919/ICINS.2018.8405859

16. **Urik R.Dzh.** *Osnovy gidroakustiki*. Leningrad: Sudostroyeniye Publ., 1978. 448 p. (rus)

17. **Jennings N.R., Sycara K., Woolridge M.** A roadmap of agent research and development. *Autonomous Agents and Multi-Agent Systems Journal*, 2008, Vol. 1, Iss. 1, pp. 7–38.

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Семенов Николай Николаевич
Nikolay N. Semenov
E-mail: semenov@smtu.ru

Михлин Валерий Григорьевич
Valerii G. Mikhlin
E-mail: valeriy_mikhlin@mail.ru

Ахметов Денис Булатович
Denis B. Akhmetov
E-mail: akhmetov@spbstu.ru

Submitted: 09.01.2023; Approved: 11.05.2023; Accepted: 17.05.2023.
Поступила: 09.01.2023; Одобрена: 11.05.2023; Принята: 17.05.2023.

Computing, Telecommunication and Control, 2023, Vol. 16, No. 1, Pp. 46–59.
Информатика, телекоммуникации и управление. 2023. Том 16, № 1. С. 46–59.

Software of Computer, Telecommunications and Control Systems

Программное обеспечение вычислительных, телекоммуникационных и управляющих систем

Research article
DOI: https://doi.org/10.18721/JCSTCS.16104
UDC 004.43

FLOVVER: A GRAPHICAL FUNCTIONAL LANGUAGE WITH A COMPILER FOCUSED ON RECURSION OPTIMIZATION

A.A. Zavyalov[1], *S.M. Staroletov*[2] ✉

[1] Novosibirsk State University, Novosibirsk, Russian Federation;
[2] Polzunov Altai State Technical University, Barnaul, Russian Federation
✉ serg_soft@mail.ru

Abstract. Visual languages reflect many parts of textual programming languages, however, the existing visual programming solutions lack higher-order functions and recursion concepts. The article introduces the design of a visual language Flovver, which implements the concepts of graphical functional programming. We propose a programming language that supports higher-order and recursive computations. The language accepts programs in a specially designed notation with semantics which we explain in this paper using the lambda calculus. The syntactic unit of such a program is a function that can be combined in a specific way with other functions. We present a fixpoint combinator that helps to specify a recursive behavior in the graphical functional language. To obtain calculate-effective programs, we design and implement a compiler for it, which is capable to optimize recursive programs. We also discuss code generation to JavaScript using the static single assignment (SSA) form. Finally, we propose a sketch of graphical integrated environment to design programs in Flovver using pre-defined blocks, and we present the generated SSA-like code in the paper. The approach is demonstrated on well-known Factorial and Fibonacci recursive programs.

Keywords: programming language, graphical language, functional language, optimizing compiler

Citation: Zavyalov A.A., Staroletov S.M. Flovver: A graphical functional language with a compiler focused on recursion optimization. Computing, Telecommunications and Control, 2023, Vol. 16, No. 1, Pp. 46–59. DOI: 10.18721/JCSTCS.16104

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.16104
УДК 004.43

FLOVVER: ГРАФИЧЕСКИЙ ФУНКЦИОНАЛЬНЫЙ ЯЗЫК С ОРИЕНТИРОВАННЫМ НА ОПТИМИЗАЦИЮ РЕКУРСИИ КОМПИЛЯТОРОМ

А.А. Завьялов[1], *С.М. Старолетов*[2] ✉

[1] Новосибирский государственный университет, Новосибирск, Российская Федерация;

[2] Алтайский государственный технический университет им. И.И. Ползунова, Барнаул, Российская Федерация

✉ serg_soft@mail.ru

Аннотация. Визуальные языки отражают многие черты текстовых языков программирования, однако в существующих решениях визуального программирования не хватает функций высшего порядка и рекурсии. В статье описан дизайн визуального языка Flovver, реализующего концепции графического функционального программирования. Предложен язык программирования, поддерживающий рекурсивные вычисления более высокого порядка. Язык принимает программы в специально разработанной нотации с семантикой, объясняемой с использованием лямбда-исчисления. Основной синтаксической единицей такой программы является функция, которая может определенным образом комбинироваться с другими функциями. Представлен комбинатор неподвижной точки, помогающий определить рекурсивное поведение в данном графическом функциональном языке. С целью получения вычислительно-эффективных программ разработан и реализован компилятор, способный оптимизировать рекурсивные программы. Рассмотрена генерация кода в программу на JavaScript с использованием формы статического одиночного присваивания (SSA). Предложен эскиз графической интегрированной среды для разработки программ во Flovver с использованием заранее определенных блоков и представлен сгенерированный SSA-подобный код. Подход демонстрируется на известных рекурсивных программах вычисления факториала и последовательности Фибоначчи.

Ключевые слова: язык программирования, графический язык, функциональный язык, оптимизирующий компилятор

Для цитирования: Zavyalov A.A., Staroletov S.M. Flovver: A graphical functional language with a compiler focused on recursion optimization // Computing, Telecommunications and Control. 2023. Т. 16, № 1. С. 46–59. DOI: 10.18721/JCSTCS.16104

Introduction

With the involvement of more people in the process of software development, graphical or visual programming languages are beginning to regain popularity. This takes us back to 1970s when Alan Kay was developing the *Dynabook* project [1], with the aim of involving children in programming, in particular by manipulating graphic objects to construct a program, due to the fact that visual information is easier to remember. In the last decade, MIT has cultivated the *Scratch* and *App Inventor* languages, which allow users to combine blocks or graphical elements of programs that include variables, loops, conditions, and so on [2, 3] in a web interface. Accordingly, such blocks can include other blocks, enabling nested programs organization. The latter language can be even used as the initial programming language to teach children how to create mobile applications, and has been successfully applied in the development of computational thinking [4].

If we try to define a visual programming language, then it can be noted that such a language contains graphical elements as syntactic units or primitives, and allows the developer to create programs by manipulating such elements instead of specifying them in the text [5]. A. Repenning has been analyzing the experience of using existing graphical programming languages over the past twenty years [6] and noted that such languages make programming more accessible to a wide range of people without extensive programming experience. The use of graphical languages helps the developers at three levels:

• At the syntax level: instead of a boundless text, the elements of visual languages are conveniently represented in the form of icons, blocks and diagrams, which eliminates the possibility of syntactical errors in the program.

• At the level of semantics: graphical representation of language objects can visually show the purpose of program primitives and ensure control of their connections only with compatible elements, which means reducing the time of learning.

• At the application level: visual languages enable programming languages researchers to get a certain representation based on a program to explore or prove its properties.

All of the above corresponds to the modern *No-code* or *Low-code* paradigms, which implies the refusal (partial or complete) of writing textual code when building software systems. This approach also correlates with the Model-Driven Development concept, where the program construction starts with some model and the code is only a by-product.

In his 1977 Alan Turing Award lecture [7], a programming language researcher John Backus delivered a lecture "Can programming be freed from the von Neumann paradigm?" [8]. In this speech, he proposed functional languages as an alternative to traditional or imperative languages, and also presented the algebra of functional programs as a formal system of functional programming.

The use of functional languages is especially relevant in the modern era of big data since the execution process in such languages involves the calculation of functions without data dependencies; therefore, it can be parallelized without synchronization overheads, and even dynamically replaced during the calculation if necessary [9].

Creating specifically a graphical visual language is a challenge for us. There has been a long history of work in this area that has led to the design of graphical functional languages (one can mention, for example, such pioneering work as [10, 11]). However, some important questions remain regarding the construction of (i) a formalized syntax for a graphical functional language, as well as the implementation of a full-featured graphical environment, including (ii) a compiler from a graphical language to an internal representation (iii) an optimizer, and (iv) a launcher for running resulting programs and handling their interaction with graphical input-output elements. In this work, we are addressing these issues.

In conducting the present research, we focus on some key factors. The first is the design of an efficient architecture of the graphical environment, where we use the Elm [12, 13] approach, which implements an architecture for creating web-oriented functional languages to generate web applications and games. However, the design of the environment is not a subject of the present paper. The second factor is the implementation of an optimizing compiler for recursive calls. It should be noted here that functional programming is closely related to recursion, which can be used both for organizing simple loops and for solving enumeration problems of practical value.

However, in many cases, it is possible to eliminate recursive calls when generating the resulting program code [14]. In this paper, we discuss means to optimize both tail-recursive calls [15] and general recursive schemes using the memoization technique [16, 17], as applied to graphical functional language programs. Due to the native graph structure of the programs, it is easy to get an internal representation for such optimizations. The third and crucial factor is the ability to study a formal treatment of the graphical language, where the λ-calculus and fix point combinators are useful for us.

Our work is mainly inspired by classic pioneering approaches on graphical languages that were proposed in the 1980s. The thing is that at that time, the graphical interface just began to appear and a large number of researchers started to develop their own graphical languages, including functional ones. However, later interest in graphical languages faded; we attribute this to the dominant paradigms of the time, which led large programs poorly expressed in graphical languages. Nevertheless, we can state that now interest in graphical languages has begun to grow again, since by now, almost all algorithms have been written and are available as components, and the code turns simply into manipulating them. Such programs can just be well implemented in graphical languages, which is exploited by the mentioned systems like MIT App Inventor.

Therefore, in the existing work, we set the goal of creating a sketch of a visual functional language, which is intended primarily for teaching the basic concepts of functional languages and lambda calculus. It was a challenge for us to develop a fully functional graphical IDE that allows the user to create, run and view the results of programs in the browser. We designed a software so that the components (standard functions) can be extended in the future. For our purposes, it is advisable to generate an SSA (Static Single Assignment) representation of graphical programs in JavaScript: such generation makes it possible both to show the user a text representation of his graphical program in its original and optimized form, and also to interpret the graphical program directly in the browser.

We understand that the examples of programs for calculating the factorial and the Fibonacci sequence considered in the work are very speculative, since both cases are best examples not to use recursion at all. However, in this case we have two different types of recursion (tail and general), and it is possible to demonstrate compiler optimization methods on it.

Syntax and semantics of the proposed Flovver language

In this section, we propose the syntax of the developed *Flovver* language in the form of elements of a graphical diagram. As for its semantics, we denote language units as λ-calculus terms.

Representation of functions. Flovver belongs to a class of applicative languages (like, for example, LISP in its original design [18]) that is, it assumes a sequence of evaluations of a function with a given number of arguments and passes the result of such an evaluation to another function. For a discussion of the semantics of an applicative language, see [19]. Therefore, at the syntax level in the Flovver language, there is only one object: a function.

A function converts from 1 to N values of the given input types into one value of the output type (the variant of constant functions with 0 inputs is also possible). From a mathematical point of view, a function is a mapping of a domain set A to a range set B [20]:

$$f : A \to B.$$

Since a datatype in a language is a set of values that have the general structure or form [21], then by introducing $A = t_{i_1} \times t_{i_2} \ldots \times t_{i_N}$ and $B = t_o$ where $t_{i_1},\ t_{i_2}, \ldots \times t_{i_N},\ t_o \in T$ and T is the set of input and output datatypes, we define the function $f : T^n \to T$ in terms of the programming language, which has the signature $f : t_{i_1} \to t_{i_2} \to \ldots t_{i_N} \to t_o$.

In the Flovver language, elementary objects are functions or terms $\lambda x_1 \ldots x_n f\left(x_1,\ \ldots,\ x_n\right)$ that are represented by diagrams of the form shown in Fig. 1.

Here f is some function of type *input* 1 → *input N* → *output*.The left side of the block is the inputs of the function, while the right side of the block is the output of the function.

Composition of functions. On the right side of the function block, there is an arc that can be connected to the input of another function (Fig. 2). The semantics of this construction for input values $v_1 \ldots v_n$ (see an example of composition for λ-calculus in [22]) can be explained as:

$$f' = \left(\left(\left(\left(\lambda x_1 \ldots x_n . f\left(x_1, \ldots, x_n\right)\right) v_1\right) \ldots\right) v_n\right),\quad g' = \left(\left(\lambda x . g\left(x\right)\right) f'\right).$$

Here, the λ-term f is applied to all of its (given) arguments, after which the λ-term g is applied to the result.

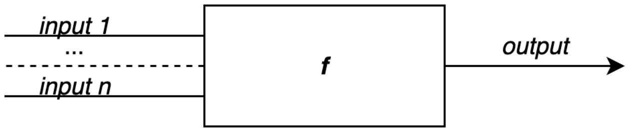


Fig. 1. Function representation in Flovver

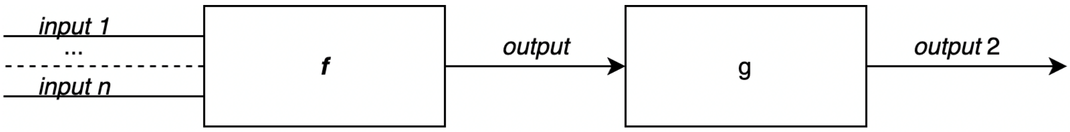


Fig. 2. Function composition in Flovver

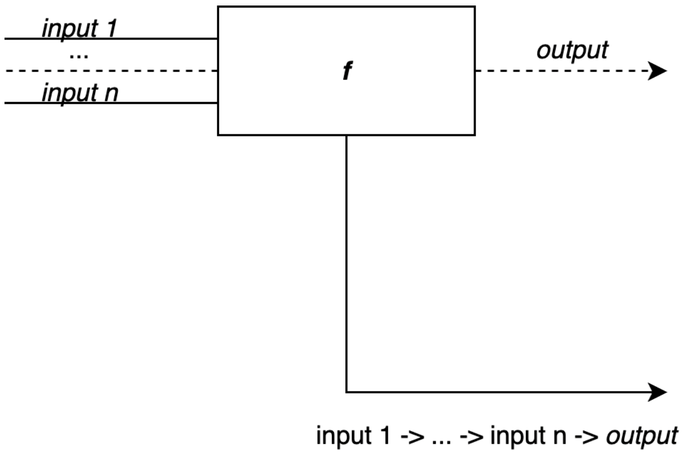


Fig. 3. Partial function application syntax

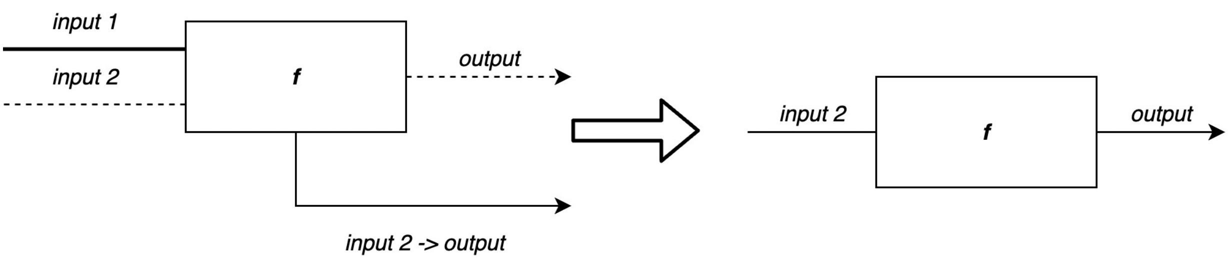


Fig. 4. An example of the partial function application

If not all arguments are passed to the input, the function is considered to be underdefined, and the output arc from *f* to *g* cannot be created in our visual editor.

Partial application of functions. A function and passed arguments can be partially applied by drawing an arc from the bottom of their block to a point of use (Fig. 3).

As a result, we get a function from a (non-strictly) smaller number of arguments, and the previously passed arguments will be fixed (Fig. 4).

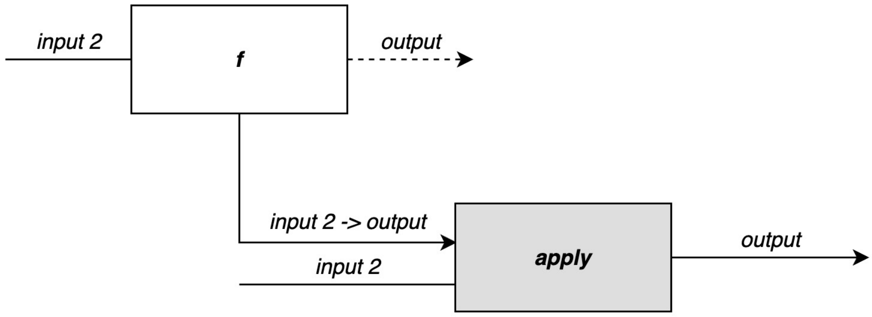


Fig. 5. An example of using the special function *apply*

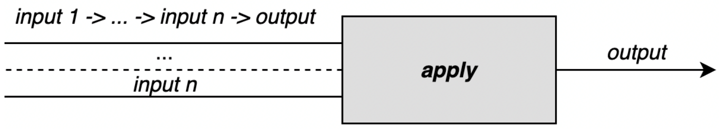


Fig. 6. Syntax of the *apply* function

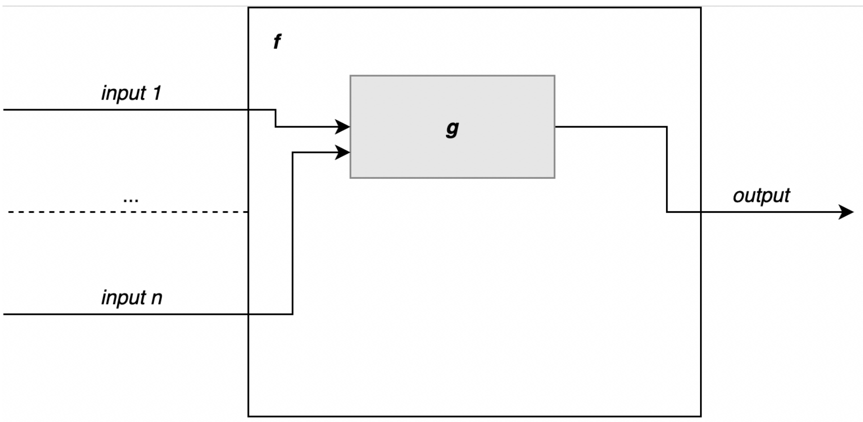


Fig. 7. Building a new function

To calculate a function passed as a value [23], a graphical language developer can use the special function *apply* (Fig. 5).

The *apply* function takes a function of N arguments as its first parameter; the next $2...N + 1$ parameters are the arguments passed to the parameter function (Fig. 6).

Compound functions. The construction of new functions from the given ones is shown in Fig. 7. Here we define a logically separate function block f with its own inputs and outputs. The left side of the block is the inputs, and the right side is the output. Inside the block, there is a function g, to which the inputs f are applied; the result of the function g is passed to the output f.

Thus, the semantics of the construction presented in Fig. 7 is defined as:

$$f := \lambda x_1 \dots x_n . g\left(x_1, \dots x_n\right).$$

In general, for an arbitrary function block fun, the semantics looks like

$$fun := \lambda x_1 \ldots x_n . T,$$

where T is a term dependent on $x_1 \ldots x_n$, and the dependence is determined as a result of graphic connections within the block.

The *Self* operator to support recursion. To support the declaration of recursive functions inside a functional block, we propose the creation of a special *self* block, which is a link to the function that is being declared. Functions with a self block can be considered applied to the fixed point combinatory [24]. Such a combinator (also known as the Y-combinator [25]) is a special higher-order function that calculates the fixed point of another function according to the rules [26]:

$$fix := \lambda f.(\lambda y.yy)(\lambda z.f(zz)),$$
$$combine\ \ self\ \ f := fix(\lambda self.\lambda arg.f).$$

The practical value of such a function lies in the ability to use recursion for anonymous functions without having to define a name for them.

In this case, the *Factorial* and *Fibonacci* functions can be expressed as following (we use LISPish parenthesized prefix notation to describe functions here):

$$fac := combine\ \ self\, x\ \left(if\,(=x0)1\left(*x\left(self\,(-x1)\right)\right)\right),$$
$$fib := combine\ \ self\, x\ \left(if\,(<x2)\,x\left(+\left(self\,(-x1)\right)\left(self\,(-x2)\right)\right)\right).$$

Schematic example for the Factorial function. In Fig. 8, we show a function block for the *Factorial* function: $N \to N$! Since the elementary syntactic unit in Flovver is a function, in this diagram, the block consists of connected function nodes (including special cases as constant functions, for this example, these are functions that return 1).

There are also *eq*?: $\mathbb{N} \to \mathbb{N} \to \mathbb{B}$ function that returns the result of comparing two arguments; *mul*: $\mathbb{N} \to \mathbb{N} \to \mathbb{N}$ returning the result of the product; minus1: $\mathbb{N} \to \mathbb{N} \to \mathbb{N}$ subtracting 1 from the argument; *if*: $\mathbb{B} \to \mathbb{N} \to \mathbb{N} \to \mathbb{N}$ verifies the first argument, and evaluates and returns the second otherwise the third; and finally, the *self* operator described earlier. This assumes that there is a set (a palette) of standard functions that a functional language developer can use. This scheme operates entirely in accordance with the rule specified in the previous section.

Building an optimizing compiler

On the internal representation. To address further issues of optimization and code generation, it is necessary to consider the internal representation (IR) of the compiler of a graphical functional language into code in a language suitable for execution (in this case, in a browser). A good discussion of IR is given in the lectures by Xavier Leroy [27].

Due to the initially chosen graphical form of programs, Flovver can use graph IR natively, following the ideas from [28], i.e.:

- there are two sources of data: nodes and their connections;
- the connections can be internal and external "by value" and "by name";
- vertexes reflect applications, function definitions, and recursive calls.

Methods for recursion optimization. After IR is determined, some optimizations can be made on it, in this case, we describe the optimizations of recursive calls. First, consider an algorithm for optimizing tail-recursive calls [29]. It occurs where a recursive call is the last operation before the call from the function. In this case, there is no need to call the function and save the execution context in the stack since the

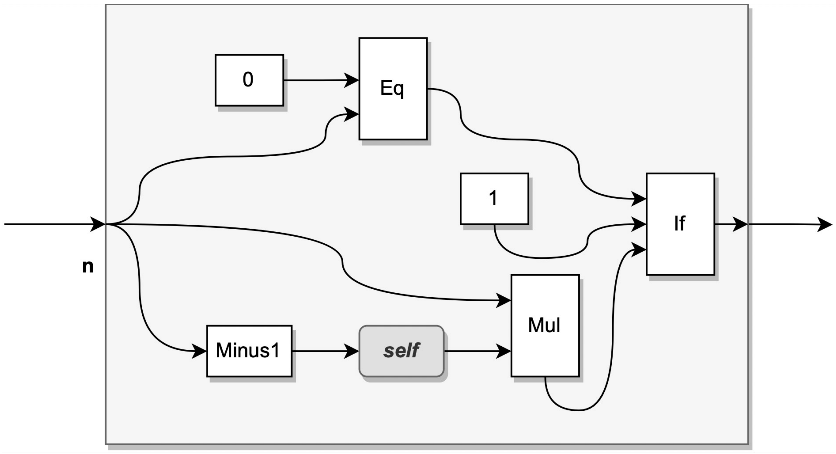


Fig. 8. A scheme for the Factorial function

parameters will not be used and the return address is already on the stack. Therefore, we can substitute passed parameters instead of function arguments, and rather than calling, go to the beginning of the function, organizing a loop.

An optimizing compiler algorithm might look like this:

1. Mark all functions that return a special tail recursion pattern with a special flag.
2. At the code generation stage, for such functions, generate the *while (...) {...}* construct for their body, in which we change the parameter and the accumulator variable, but do not generate a recursive call.

The next question arises: how to define such a tail recursion pattern? The valid tail call notation was formalized by William Clinger [30] and specified in the documents (R6RS: 5.11, 11.20) for the Scheme language [31]. The grammar of such a tail expression can be defined as:

```
<almost tail expr> ::= <rec. call> | <if> | <expr>
<if> ::= if <expr> <almost tail expr> <almost tail expr>
```

Therefore, we define *a tail recursion context* as the place where a recursive call is guaranteed to be a tail recursion call. In our work, we are interested in two contexts: (1) the end of the function and (2) the conditional expression at the end of the function.

Secondly, we consider issues of general recursion optimization. As in the previous case, for each specific recursive scheme, one can search for the corresponding context. However, we decided to optimize the general form of recursion using the memoization concept (caching previous calculations using a hash table with a key according to the passed parameters).

It is possible to memoize any calculation in Flovver since the language is purely functional [32]. In this case, we can monitor the growth of the table for potentially non-terminating functions and report this to the user before the stack overflow program crashes.

In Fig. 9, white color indicates direct calculation of values of the *Fibonacci* sequence Fib with memorization, and gray color shows getting values from memo tables. The computational complexity at the first run was reduced from a value comparable to $O(Fib(n))$ to $O(n)$, thereby approaching the complexity of the iterative algorithm for calculating the Fibonacci function. However, the memoizable version requires $O(n)$ memory for memo tables, while the iterative algorithm with two intermediate values in a loop uses just $O(1)$ memory, which leads to the conclusion that such an optimization is universal, but not completely, optimal.

Code generation. To easily emit code for a target platform, we need our intermediate representation to be transformed in a specific way. We decided to translate the Flovver programs to textual languages supporting higher-order functions and lexical closures, such as JavaScript, Scheme or Python (Haskell or Elm

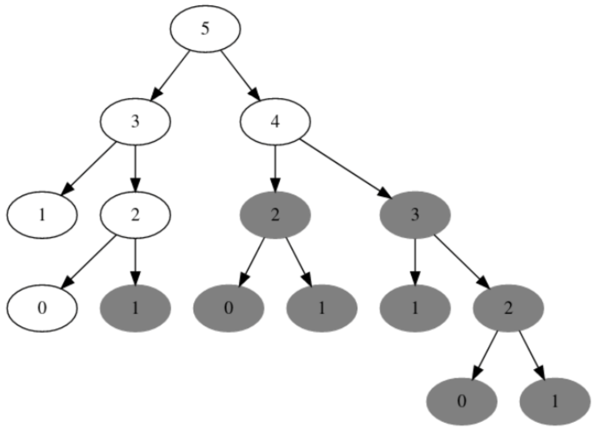


Fig. 9. Memoization for Fib (5)

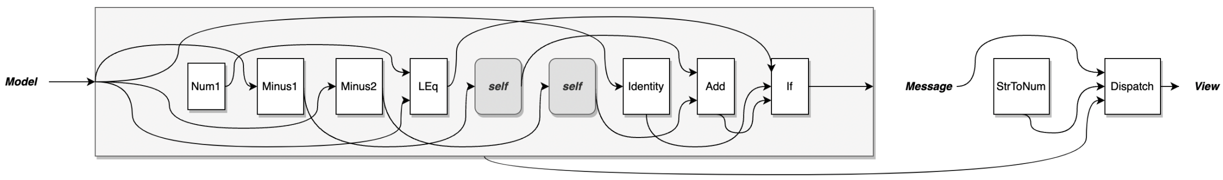


Fig. 10. A topologically sorted DAG of the Fibonacci program

would be fine too), to compile partially applied functions and function blocks easily, without reasoning about such concepts as "closure conversion" and "lambda lifting" [33].

To simplify the target code generation, it is convenient to structure the program as the chain of variable definitions in which each variable is assigned a value only once, and its identifier is used in the following part of the program. Similar ways to structure programs presented in the concepts of the Static Single Assignment (SSA) form [34, 35] used mainly in imperative programming languages, and the Administrative Normal form (A-normal form or ANF) [36] that leverages let-style of ML family languages [37].

We need to place a variable definition before its use. It can be done by reordering the vertices of the program graph since an IR graph is a dependency graph in which links represent value dependencies between objects and there are no circular dependencies in it. Thus our IR is a directed acyclic graph (DAG, see an example in Fig. 10). For a DAG, we can arrange the vertices in the following order: $source \rightarrow v_1 \rightarrow \ldots \rightarrow sink$ by performing a topological sorting [38]. Finally, the ordered graph will be fairly easy to translate into the target platform code. So, it takes three steps to convert our IR into code:

Step 1. The program graph is ordered by the topological sorting.

Step 2. Each node is mapped to the variable definition by the following rule in sort order:

1. Obtain a unique identifier for the variable.
2. Specialize the code generation for a node:

- the node is a block without input parameters → obtain a simple value (e.g. integer, string);
- the node is a fully-connected block with all inputs and output specified → generate an application of function, corresponding to the block, to its inputs;
- the node is a partial application → generate an anonymous function with unbound inputs of block used as its parameters;
- the node is a function block or a partial application → replace each use of bound input of the block with the unique name and generate code for function body.

3. Replace each use of the original node with an identifier of the corresponding variable.

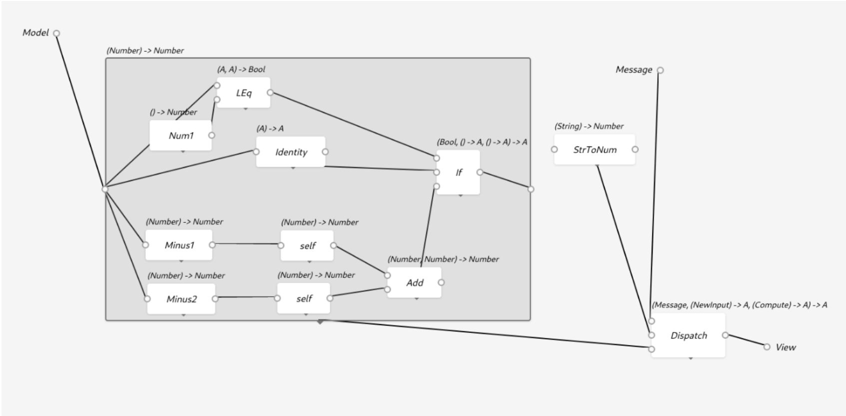


Fig. 11. F_n in Flovver environment

<u>Step 3.</u> Using the variable names, it is relatively easy to generate code for the target platform (in the case of Flovver, in JavaScript).

The *Fibonacci* program design, code generation and optimization

Figure 11 shows a program created in the Flovver interactive environment to calculate the Fibonacci function as F(0) = 0, $F(1) = 1$, $F(n) = F(n-1) + F(n-2)$ for $n > 1$.

The implementation of our approach was already discussed in Fig. 8, and the reader can see the differences for the Fibonacci function in this case. Function signatures and their purposes are mostly clear, and we can also note the *StrToNum* function, which is passed to the *Dispatch* input, receives a *Message* from the GUI environment with a value of *N* and returns the result to *View*. It all implements the Elm architecture.

With the optimizer flags disabled, the code in Listing 1 is generated for the discussed graphic program. The code starts at line 18, after which a sequence of SSA calls is defined that implements the calculation scheme.

Listing 1. Non-optimized generated code for the Fibonacci function

```
const update = (model, message) => {
  const fsa_1 = () => Num1();
  const fsa_2 = (fsa_2_arg_0) => StrToNum(fsa_2_arg_0);
  const fsa_6 = () => {
    const fsa_6_r = (fsa_6_arg_0) => {
      const fsa_7 = () => Minus2(fsa_6_arg_0);
      const fsa_8 = () => fsa_6_r(fsa_7());
      const fsa_9 = () => Minus1(fsa_6_arg_0);
      const fsa_10 = () => fsa_6_r(fsa_9());
      const fsa_11 = () => Add(fsa_10(), fsa_8());
      const fsa_12 = () => Identity(fsa_6_arg_0);
      const fsa_13 = () => LEq(fsa_6_arg_0, fsa_1());
      const fsa_14 = () => If(fsa_13(), fsa_12, fsa_11);
      return fsa_14();
    }
    return fsa_6_r(model);
  }
 const fsa_15 = () => Dispatch(message, fsa_2, fsa_6);
  return fsa_15();
}
```

With the memoization flag set, the code in Listing 2 is generated for the discussed graphics program. This code works similarly to the previous one, but additionally, a hash table is defined at line 6, which is used at lines 18 and 19.

Listing 2. Optimized generated code for the Fibonacci function

```
const update = (model, message) => {
  const fsa_1 = () => Num1();
  const fsa_2 = (fsa_2_arg_0) => StrToNum(fsa_2_arg_0);
  const fsa_6 = () => {
    const fsa_6_r = (() => {
      const fsa_6_st = {};
      const fsa_6_w = (fsa_6_arg_0) => {
        const fsa_7 = () => Minus2(fsa_6_arg_0);
        const fsa_8 = () => fsa_6_r(fsa_7());
        const fsa_9 = () => Minus1(fsa_6_arg_0);
        const fsa_10 = () => fsa_6_r(fsa_9());
        const fsa_11 = () => Add(fsa_10(), fsa_8());
        const fsa_12 = () => Identity(fsa_6_arg_0);
        const fsa_13 = () => LEq(fsa_6_arg_0, fsa_1());
        const fsa_14 = () => If(fsa_13(), fsa_12, fsa_11);
        return fsa_14();
      }
    return (fsa_6_arg_0) => fsa_6_st[[fsa_6_arg_0]] =
    fsa_6_st[[fsa_6_arg_0]] || fsa_6_w(fsa_6_arg_0);})();
    return fsa_6_r(model);
  }
  const fsa_15 = () => Dispatch(message, fsa_2, fsa_6);
  return fsa_15();
}
```

Related work

There have been years of research behind the visual programming languages since Goldstine and von Neumann proposed to represent machine-aided calculations as flow diagrams [39, 40]. The approach was firmly rooted in software modeling but was considered ineffective and unimplementable relatively to computers of those times. With the growth of computer power, this approach was abandoned in favor of a well-known textual approach to programming popularized by FORTRAN and ALGOL. However, the interest in the visual approach to program construction has begun to return back since the '70–80s with the development of declarative and applicative programming paradigms, and logical/functional programming. There is a variety of languages developed back in the '80s and '90s that present ideas similar to our work. So, one example is the Prograph language [10], in which programs were organized as a "prographs" (Prolog graphs). Prograph also supported the structuring of programs into procedures. However, Prograph provided iterations via imperative FOR, WHILE and REPEAT blocks.

There were a few visual languages based on the applicative and functional paradigms. For instance, in Viz [41] there were mechanisms to represent mathematical functions and λ-abstraction to organize a program with combinators; the discussed Backus's FP system was implemented in Pagan's graphical FP language [42].

However, Viz offers manipulation with arcs in the flowchart to organize cycles and conditionals, whereas in our work we rely on combinators. Pagan's graphical FP language, in turn, puts forward space-partitioning based syntax, which we consider impractical compared to flow diagrams. Modern ideas of the usages of graphical functional languages include their application in data science, focusing on visibility and explainability (see, for example, the Enso language [43]).

We have observed that there is a lack of syntax and semantics formalization in this area, seems it is not uncommon in mathematics to use diagrammatic reasoning. The area where it can be used is the category theory. The concept of string diagrams has attracted a lot of attention as a formal foundation for reasoning using graphical notation [44]. They allow for formal conversion between the topological point of view (boxes and wires) and algebraic (certain categorical constructions), and such diagrammatic syntax could be used to give precise control over resources. It is already presented in our work by driving wires to duplicate values of variables. Another vision on wired dataflow programming is presented in the work [45].

Conclusion

As a result of this work, we designed the visual language Flovver and developed a visual programming environment to create and run programs in this language. It includes a multi-pass visual language compiler with the ability to eliminate tail recursion, as well as to optimize general recursion through memoization. The generated code can be executed in the browser and the result of its execution is obtained in the associated controls. Therefore, the environment is self-contained but currently includes a palette of elementary blocks only for the Factorial and Fibonacci functions. In the implementation, the Scala language, Jetty server, Scalatra and Svelte frameworks were used. To provide interaction with GUI that send messages to and receive responses from a graphical program, we follow the Elm approach [13] and Model-View-Update architecture. This project is completely open and available on GitHub [46].

Preliminary information about the described approaches was published in [47], discussed at the ruSTEP seminar and defended in the form of a qualifying work at the Department of Applied Mathematics of AltSTU. Finally, the tool was demonstrated at the SEIM'22 conference.

Future research directions may include: support for reciprocal recursion; formalization of the language from the point of view of the theory of graphical and functional languages; introduction of static typing and type inference mechanism; the study of common recursive patterns by analyzing the structure of large software systems using real functional languages.

REFERENCES

1. **Kay A., Goldberg A.** Personal dynamic media. *Computer*, 1977, Vol. 10, Pp. 31–41. DOI: 10.1109/C-M.1977.217672

2. **Pokress S.C., Veiga J.J.D.** MIT app inventor: Enabling personal mobile computing. 2013. Available: *https://appinventor.mit.edu/explore/resources/personal_mobile_computing* (Accessed 17.02.2023).

3. **Patton E.W., Tissenbaum M., Harunani F.** *MIT app inventor: Objectives, design, and development, computational thinking education*, 2019, Pp. 31–49. DOI: 10.1007/978-981-13-6528-7_3

4. **Tissenbaum M., Sheldon J., Sherman M.A., et al.** The state of the field in computational thinking assessment. *13th International Conference of the Learning Sciences* (*ICLS*), 2018, Pp. 1304–1311. DOI: 10.22318/cscl2018.1304

5. **Rémi D.** The maturity of visual programming. 2015. Available: *https://web.archive.org/web/20210119062636/https://www.craft.ai/blog/the-maturity-of-visual-programming/* (Accessed 17.02.2023).

6. **Repenning A.** Moving beyond syntax: Lessons from 20 years of blocks programing in agentsheets. *J. Vis. Lang. Sentient Syst.*, 2007, Vol. 3, Pp. 68–91. DOI: 10.18293/VLSS2017-010

7. ACM, Chronological listing of Turing award winners. 2020. Available: *https://amturing.acm.org/byyear.cfm* (Accessed 17.02.2023).

8. **Backus J.** Can programming be liberated from the von Neumann style? A functional style and its algebra of programs. *Communications of the ACM*, 1978, Vol. 21, Pp. 613–641. DOI: 10.1145/359576.359579

9. **Staroletov S.M.** Functional languages for distributed systems (in Russian). 2019, 215 p. Available: *https://www.elibrary.ru/item.asp?id=41291912* (Accessed 17.02.2023). EDN VIRQBH

10. **Cox P., Mulligan I.** Compiling the graphical functional language PROGRAPH. *Proc. of the 1985 ACM SIGSMALL Symposium on Small Systems*, 1985, Pp. 34–41. DOI: 10.1145/317164.317169

11. **Cordy J.R., Graham T.N.** GVL: A graphical, functional language for the specification of output in programming languages. *Proc. 1990 International Conference on Computer Languages*, 1990, Pp. 11–12. DOI: 10.1109/ICCL.1990.63756

12. **Czaplicki E.** Elm: Concurrent FRP for functional GUIs, Senior thesis. Harvard University. 2012, 44 p. Available: *https://elm-lang.org/assets/papers/concurrent-frp.pdf* (Accessed 17.02.2023)

13. **Czaplicki E.** The Elm architecture. 2021. Available: *https://guide.elm-lang.org/architecture/* (Accessed 17.02.2023).

14. **Shilov N.V.** Etude on recursion elimination. *Modeling and Analysis of Information Systems*, 2018, Vol. 25, Pp. 549–560. DOI: 10.18255/1818-1015-549-560

15. **Debray S.K.** Optimizing almost-tail-recursive Prolog programs. *Conference on Functional Programming Languages and Computer Architecture*, 1985, Pp. 204–219. DOI: 10.1007/3-540-15975-4_38

16. **Michie D.** "Memo" functions and machine learning. *Nature*, 1968, Vol. 218, Pp. 19–22. DOI: 10.1038/218019a0

17. **Norvig, P.** Techniques for automatic memoization with applications to context-free parsing. *Computational Linguistics*, 1991, Vol. 17, Pp. 91–98.

18. **McCarthy J., Levin M.I., Abrahams P.W., Edwards D.J., Hart T.P.** LISP 1.5 programmer's manual. 1965. DOI: 10.5555/1096473

19. **Turner D.A.** The semantic elegance of applicative languages. *Proc. of the 1981 Conference on Functional Programming Languages and Computer Architecture*, 1981, Pp. 85–92. DOI: 10.1145/800223.806766

20. **Kirkinskiy A.S.** Mathematical analysis (in Russian). 2006. Available: *https://www.elibrary.ru/item.asp?id=25759717* (Accessed 17.02.2023). EDN VRSEXL

21. **Amadio R.M., Cardelli L.** Subtyping recursive types. *ACM Transactions on Programming Languages and Systems* (*TOPLAS*), 1993, Vol. 15, Pp. 575–631. DOI: 10.1145/155183.155231

22. **Fischer M.J.** Lambda calculus schemata. *ACM SIGPLAN Notices*, 1972, Vol. 7, Pp. 104–109. DOI: 10.1145/942578.807077

23. **Plotkin G.D.** Call-by-name, call-by-value and the lambda-calculus. *Theoretical Computer Science*, 1975, Vol. 1, Pp. 125–159. DOI: 10.1016/0304-3975(75)90017-1

24. **Pierce B.C., Benjamin C.** *Types and programming languages*. 2002. DOI: 10.5555/509043

25. **Park D.** *The Y-combinator in Scott's lambda-calculus models*. 1976. DOI: 10.5555/901310

26. Rosetta code, Y combinator / Scheme. 2022. Available: *https://rosettacode.org/wiki/Y_combinator#Scheme* (Accessed 17.02.2023).

27. **Leroy X.** Functional programming languages Part V: Functional intermediate representations. 2017. URL: *https://xavierleroy.org/mpri/2-4/fir.2up.pdf* (Accessed 17.02.2023).

28. **Leißa R., Köster M., Hack S.** A graph-based higher-order intermediate representation. *2015 IEEE/ACM International Symposium on Code Generation and Optimization* (*CGO*), 2015, Pp. 202–212. DOI: 10.1109/CGO.2015.7054200

29. **Abelson H., Sussman G.J.** Structure and interpretation of computer programs.1996. Available: *http://library.oapen.org/handle/20.500.12657/26092* (Accessed 17.02.2023).

30. **Clinger W.D.** Proper tail recursion and space efficiency. *Proc. of the ACM SIGPLAN 1998 Conference on Programming Language Design and Implementation*, 1998, Pp. 174–185. DOI: 10.1145/277650.277719

31. **Dybvig R.K.** *The Scheme programming language*. 2009. DOI: 10.5555/525334

32. **Sabry A.** What is a purely functional language? *Journal of Functional Programming*, 1998, Vol. 8, Pp. 1–22. DOI: 10.1017/S0956796897002943

33. **Johnsson T.** Lambda lifting: Transforming programs to recursive equations. *Conference on Functional Programming Languages and Computer Architecture*, 1985, Pp. 190–203. DOI: 10.1007/3-540-15975-4_37

34. **Kelsey R.A.** A correspondence between continuation passing style and static single assignment form. *ACM SIGPLAN Notices*, 1995, Vol. 30, Pp. 13–22. DOI: 10.1145/202530.202532

35. **Cooper K., Torczon L.** *Engineering a compiler*. 2011. DOI: 10.5555/1526330

36. **Flanagan C., Sabry A., Duba B., Felleisen M.** The essence of compiling with continuations. *Proc. of the ACM SIGPLAN 1993 Conference on Programming Language Design and Implementation*, 1993, Pp. 237–247. DOI: 10.1145/155090.155113

37. **Leroy X., Doligez D., Frisch A., Garrigue J., Rémy D., Vouillon J.** The OCaml system release 4.11: Documentation and user's manual. 2020. Available: *https://ocaml.org/releases/4.11/ocaml-4.11-refman.pdf* (Accessed 17.02.2023).

38. **Cormen T.H., Leiserson C.E., Rivest R.L., Stein C.** *Introduction to algorithms*. 2009. DOI: 10.5555/500824

39. **Morris S.J., Gotel O.** Flow diagrams: Rise and fall of the first software engineering notation. *International Conference on Theory and Application of Diagrams*, 2006, Pp. 130–144. DOI: 10.1007/11783183_17

40. **Glimm J.G., Impagliazzo J., Singer I.** *The legacy of John von Neumann*. 2006, 344 p. DOI: 10.5555/549689

41. **Holt C.M.** Viz: A visual language based on functions. *Proc. of the 1990 IEEE Workshop on Visual Languages*, 1990, Pp. 221–226. DOI: 10.1109/WVL.1990.128410

42. **Pagan F.G.** A graphical FP language. *ACM SIG-PLAN Notices*, 1987. Vol. 22, Pp. 21–39. DOI: 10.1145/24697.24699

43. New Byte Order Inc., Enso Language Syntax. 2021. Available: *https://enso.org/docs/syntax* (Accessed 17.02.2023).

44. **Bonchi F., Pavlovic D., Sobocinski P.** Functorial semantics for relational theories. 2017. Available: *https://arxiv.org/pdf/1711.08699.pdf* (Accessed 17.02.2023).

45. **Nilsson H., Courtney A., Peterson J.** Functional reactive programming continued. *Proc. of ACM SIGPLAN Workshop on Haskell*, 2002, Pp. 51–64. DOI: 10.1145/581690.581695

46. **Zavyalov A.** Flovver (WIP). 2021. Available: *http://github.com/flovver/* (Accessed 17.02.2023).

47. **Zavyalov A.A.** Designing a visual functional language with recursion description capabilities (in Russian). Science and youth. 2021, Pp. 173–176. Available: *https://www.elibrary.ru/item.asp?id=46680361* (Accessed 17.02.2023). EDN VWVNOC.

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Завьялов Антон Алексеевич
Anton A. Zavyalov
E-mail: a.zavyalov.98@yandex.ru

Старолетов Сергей Михайлович
Sergey M. Staroletov
E-mail: serg_soft@mail.ru

Submitted: 18.02.2023; Approved: 03.04.2023; Accepted: 17.05.2023.
Поступила: 18.02.2023; Одобрена: 03.04.2023; Принята: 17.05.2023.

Information, Control and Measurement Systems
Информационные, управляющие и измерительные системы

Research article
DOI: https://doi.org/10.18721/JCSTCS.16105
UDC 62.5

EFFECTIVENESS EVALUATION OF MULTI-AGENT CONTROL SYSTEMS FOR AUTONOMOUS UNDERWATER VEHICLES FOR UNDERWATER OPERATION

N.N. Semenov[1] ✉ , *V.G. Mikhlin*[2], *D.B. Akhmetov*[3]

[1,2] St. Petersburg State Marine Technical University, St. Petersburg, Russian Federation;

[3] Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation

✉ semenov@smtu.ru

Abstract. The need to use groups of homogeneous and heterogenous robots in a confined space leads to the need for robots to interact with each other to prevent accidents and interfere with the work of other robots. And limited in speed and range communication channels do not allow remote control of each robot separately, that leads to the need to create multi-agent control systems or the ability of a group of robots to solve emerging problems without human intervention. This article discusses the effectiveness of such a group depending on the technical constraints of each robot and the number of robots in the group. The paper shows that an increase in the number of AUVs in a group leads to a significant increase in efficiency, but when a certain number is reached, the efficiency drops, because large groups of AUVs spend much more time changing lanes, and the increase in efficiency with an increase in the number of AUVs disappears.

Keywords: robotics, underwater robotics, group application, AUV, multi-agent control, efficiency

Citation: Semenov N.N., Mikhlin V.G., Akhmetov D.B. Effectiveness evaluation of multi-agent control systems for autonomous underwater vehicles for underwater operation. Computing, Telecommunications and Control, 2023, Vol. 16, No. 1, Pp. 60–68. DOI: 10.18721/JCSTCS.16105

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.16105
УДК 62.5

ОЦЕНКА ЭФФЕКТИВНОСТИ ПРИМЕНЕНИЯ СИСТЕМ МУЛЬТИАГЕНТНОГО УПРАВЛЕНИЯ АВТОНОМНЫМИ НЕОБИТАЕМЫМИ ПОДВОДНЫМИ АППАРАТАМИ ДЛЯ РАБОТЫ ПОД ВОДОЙ

Н.Н. Семенов[1] ✉, *В.Г. Михлин*[2], *Д.Б. Ахметов*[3]

[1,2] Санкт-Петербургский морской технический университет, Санкт-Петербург, Российская Федерация;

[3] Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Российская Федерация

✉ semenov@smtu.ru

Аннотация. Необходимость применения групп как однородных, так и разнородных роботов в ограниченном пространстве приводит к необходимости взаимодействия роботов между собой с целью предотвращения аварий и помех другим роботам. Ограниченность каналов связи по скорости и дальности не позволяет удаленно контролировать каждого робота, что приводит к необходимости создания систем мультиагентного управления, то есть способности группы роботов решать возникающие проблемы без участия человека. В статье рассмотрен вопрос эффективности такой группы в зависимости от технических ограничений каждого робота и числа роботов в группе. Показано, что увеличение числа автономных необитаемых подводных аппаратов (АНПА) в группе приводит к существенному увеличению эффективности, но при достижении определенного числа эффективность падает, поскольку большие группы АНПА тратят на перестроение значительно больше времени, и прирост эффективности при увеличении числа АНПА пропадает.

Ключевые слова: робототехника, подводная робототехника, групповое применение, АНПА, мультиагентное управление, эффективность

Для цитирования: Semenov N.N., Mikhlin V.G., Akhmetov D.B. Effectiveness evaluation of multi-agent control systems for autonomous underwater vehicles for underwater operation // Computing, Telecommunications and Control. 2023. Т. 16, № 1. С. 60–68. DOI: 10.18721/JCSTCS.16105

Introduction

The issues of group application of underwater vehicles for various tasks have recently been paid close attention to [1–6]. This includes search for submerged objects, including search for minerals, technical inspection of underwater facilities and their maintenance.

The underwater vehicles in the group are supposed to have a channel of information exchange with each other and a positioning system [4]. But the peculiarities of such channels underwater are the limited range and speed of information transfer.

Implemented autonomous underwater vehicles (AUV) groups are known, for example, for the submerged objects search [7]. It is proposed to use leading vehicles (nominally Bluefin-12) there, they directly explore the bottom in search of submerged objects, then an intermediate link is introduced – communication and navigation facilities (each equipped with an inertial navigation system and communication system). These vehicles (nominally Bluefin-21) provide the small (lead) vehicles with the information and communication channel necessary for their navigation. Finally, there are vehicles that re-find the marked objects and further investigate them if necessary.

Three levels are introduced when building a control system for a group of vehicles: strategic, tactical and operational ones [4]. A group of vehicles has a common task (mission), to solve which all the resources of the group are used. As a rule, it is possible to accomplish the mission by different algorithms. The optimality criterion (if there are several criteria, then their convolution) is used to select a specific algorithm from the set.

The upper, strategic management level determines the division of the entire mission for a AUV group into subtasks for each AUV separately, determines ways to control the mission accomplishment, makes distribution of subtasks (goals) among specific vehicles of the group. The strategic management level is not the property of a single vehicle, but is the property of the group. Moreover, both division of the mission into subtasks and distribution of subtasks between vehicles can be changed promptly if the general task changes or, for example, one vehicle of the group fails and functions redistribution among the remaining vehicles of the group is required. The tactical level receives information which determines how a particular vehicle performs a subtask (mission) from this level. This subtask contains a goal which is independent of the other vehicles, and allows to control its execution. For example, the trajectory of the vehicle, its place in the formation, the area surveyed, and the amount of information collected.

The tactical level is part of each vehicle's control system, it breaks down the current goal into a set of actions or trajectories needed to achieve the goal, and monitors their execution. This level exchanges information with neighboring vehicles in the group to clarify, for example, their position in the group and adjusts the trajectory transmitted to the operational level. The task of the tactical level is to make a decision to bypass arising obstacles and return to their place in the formation.

At the operational level, control actions are generated on the available control resources of the vehicle to maintain the specified trajectory and collect the required information. Such resources include thrusters (marching, vertical and horizontal thrusters), rudders, stabilizers (passive or active), roll levelling mechanisms, buoyancy change system and shifted center of gravity.

The decentralized control systems for a group of vehicles, such as a multi-agent control system, are undoubtedly of great interest [2, 7].

In this case, one mission is assigned to the entire group of vehicleson the strategic level, for example, search for submerged objects in a limited area, which is pre-marked with underwater beacons, and the size of the survey strip (area) by one vehicle. The width of the survey strip is assigned based on the capabilities of the acoustic complex of the given vehicle to detect a submerged object with a given probability P_d. The same considerations determine the depth and speed of the vehicle. The mission will be understood as: geographical coordinates of the territory to be surveyed, characteristics of the flooded object (criteria for deciding that it is a flooded object) and the required probability of its detection. Each vehicle evaluates its efficiency in surveying the areas closest to it and transmits the information to neighboring vehicles, which in turn report their efficiency in surveying these areas, after which the areas are surveyed by the vehicle that can survey them in the minimum time and having spent the minimum amount of resources. Resolution of conflicts between vehicles is considered in works on the "consensus problem" [7, 13–15].

Since the main task of the AUV group is to accomplish the mission (for example, detect objects on the bottom) in the minimum time with the maximum probability, let us set the AUV group "efficiency" criterion as the ratio of the detection probability to the time of the entire group operation

$$Eff_{gr}(n) = \frac{P_{gr}(n)}{t_{gr}(n)},$$

where n is the group size, P is a given probability of object detection, t – group operating time (maximum operating time of the AUV in the group).

If the economic component, i.e. the necessity to take into consideration an increase of cost of the group and its service with increase of its number, is taken into account, then there appears one more criterion – *economic efficiency*, i.e. efficiency of use of each AUV in the group:

$$Eff_{gr}(n) = \frac{P_{gr}(n)}{t_{gr}(n) \cdot n}.$$

The following approach is used to estimate the probability of finding an object in a given area [9]: the object (objects) sought for is on the bottom of the area surveyed by AUV with width L and length d (Fig. 1). The area is rectangular. The AUV passing along the strip with its locator captures the whole surveyed strip (by width). All values are evenly distributed.

The scheme has the following notations: d is the width of the strip surveyed by one AUV; L is the length of the strip surveyed by one AUV; dL is the distance covered by AUV during the time dt; $S_s = d \cdot L$ is the area of the site surveyed by AUV.

The area surveyed by the AUV during the time T_s (search time) is shaded in Fig. 1. Then the probability P of finding an object in the strip according to the geometric definition (probability) can be defined as follows:

$$P = \frac{d}{S_s} \int_0^{T_s} V_{\text{AUV}}(t)dt = \frac{\int_0^{T_s} V_{\text{AUV}}(t)dt}{L},$$

where $V_{\text{AUV}(t)}$ is the speed of AUV during the site survey.

The operating time of the whole group t_{gr} is determined by the longest operating time of each AUV in the group. Working time of each AUV t_i^{AUV}:

$$t_i^{\text{AUV}} = t_{w\_i}^{\text{AUV}} + t_{re\_i}^{\text{AUV}}, \tag{1}$$

where t_i^{AUV} is the survey time of the area allocated for this vehicle for the selected trajectory; $t_{w\_i}^{\text{AUV}}$ – the time of the search for flooded objects (work); $t_{re\_i}^{\text{AUV}}$ – the time of rearrangement to enter the working trajectory.

Calculations of rearrangement time and exit to a given trajectory are considered in the moving objects control theory and are well reviewed in [10, 11].

Probability of contact with an object caught in the range of observation P_c means (GAS AUV):

$$P_c = 1 - e^{\frac{kD_d\Omega}{V_{\text{AUV}}T_0}},$$

where $k = \frac{\pi}{2 \cdot 360} = 0.00436$ is the conversion factor when Ω is given in degrees; Ω – the sector of the survey by the observation vehicle (AUV); T_0 – the time of the sector survey by AUV SONAR; D_d the mathematical expectation of the range of the technical means of detection (AUV SONAR) – the average expected detection range.

Probability of detecting an object in a given time interval P_d:

$$P_d = 1 - e^{-\frac{uT_s}{S_s}},$$

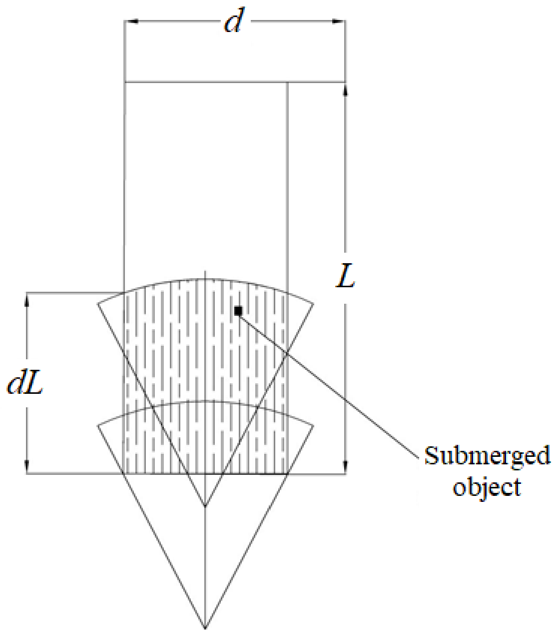


Fig. 1. Scheme for calculating the location of an object in a given area

where u is the search performance, taking into account the probability of obtaining contact with it; T_s – the search time (the time of AUV being in the search site with area S_s).

The search performance is defined as $u = W_{sb}V_{\text{AUV}}P_c$, where W_{sb} is the effective width of the AUV SONAR survey band. In case of one AUV W_{sb} is equal to $W_{sb} = 2D_d$.

For a group AUV consisting of N_{AUV}:

$$W_{sb} = \left(N_{\text{AUV}} - 1\right) d_{\text{AUV}} + 2D_d,$$

where d_{AUV} is the distance between the AUVs when surveying the area.

The mathematical expectation of object detection time T_d shows how long from the start of the search an object can be expected to be detected on average:

$$T_d = \frac{S_s}{u}.$$

The mathematical expectation of the number objects MO_d, detected during the search time T_s is as follows:

$$MO_d = N_{a1}P_{d1} + N_{a2}P_{d2} + \ldots + N_{an}P_{dn} = \sum_{i=1}^{n} N_{ai}P_{di},$$

where N_{ai} is the number of objects in the i-th search area; P_{di} is the probability of detecting objects in the i-th search area.

It is planned to search objects by m types of AUV with different search productivity in the site of S_s area.

N_{t_1} is the number of the first type AUVs with search capacity u_1;

N_{t_2} is the number of the second type AUVs with search capacity u_2;

...
N_{t_n} is the number of the n-th type AUVs with search performance u_n.

$$\frac{S_1}{S_s} = \frac{N_{t\_1}u_1}{N_{t\_1}u_1 + N_{t\_2}u_2 + \ldots + N_{t\_n}u_n},$$

$$\frac{S_2}{S_s} = \frac{N_{t\_2}u_2}{N_{t\_1}u_1 + N_{t\_2}u_2 + \ldots + N_{t\_n}u_n}, \ \ldots,$$

$$\frac{S_n}{S_s} = \frac{N_{t\_n}u_n}{N_{t\_1}u_1 + N_{t\_2}u_2 + \ldots + N_{t\_n}u_n}.$$

Note: $S_s = S_1 + S_2 + \ldots + S_n$.

It is assumed that the detection of each of n flooded objects are independent events A_i, where $i \in 1, 2, ..., n$. Then the probability of detecting of at least one flooded object is as follows:

$$P_d\left(A_1 + A_2 + \ldots + A_n\right) =$$
$$= 1 - \left(1 - P_d\left(A_1\right)\right)\left(1 - P_d\left(A_2\right)\right)\ldots\left(1 - P_d\left(A_n\right)\right).$$

Modeling the behavior of a group of AUVs with efficiency calculations

Having received the general mathematical model of the work efficiency of a group of AUVs for detecting underwater objects, let us carry out numerical simulation of the received model taking into account technical opportunities and restrictions at search in real conditions of the Baltic sea.

Suppose the search area is 20 × 20 km. The search width with one AUV based on requirements of detection probability ($P = 0.95$) is 50 m. The autonomous operation time of one AUV is 4 hours, the speed is 5 m/s.

Survey with one AUV would take 10 km·(10 km / 50 m)/5 m/s = 111 hours or 4.6 days. The survey is practically impossible when it is necessary to return the AUV every 4 hours and recharge [12]. But suppose the time of survey (taking into account descent/rise and accumulators recharging 28 times) is not less than 6 days, and take 0.95 probability of detection by serviceable device, the efficiency would be equal to $Eff =$ $= 0.95/111 = 0.00855$.

Increasing the number of AUVs in group up to N means decreasing time of territory survey but in this case if the group moves by lines, the more group the more time will be required for rearrangement. Trajectories of AUV movement can be different, but they all become more complicated with increasing the number of AUVs in a limited territory. The probability of failure of each robot is constant and conditionally accepted as 0.1. Therefore, the efficiency of group operation when moving in a line, depending on the number of AUVs in the group, is as follows:

Figure 2 shows that in the left part of the graph increasing the number of AUVs in the group leads to a significant increase in efficiency, but the efficiency drops when reaching a certain number. This is due to the fact that large groups of AUVs spend considerably more time for rearrangements, and the increase in efficiency with increasing number of AUVs disappears, thus, the efficiency drops.

If the economic contribution in efficiency calculation, i.e. efficiency of each AUV in the group, is taken into account, the following characteristic are obtained (see Fig. 3).

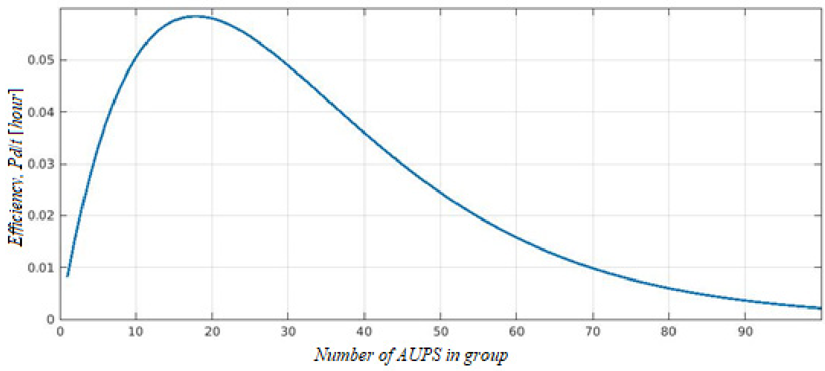


Fig. 2. Efficiency of the AUV group depending on the number of agents

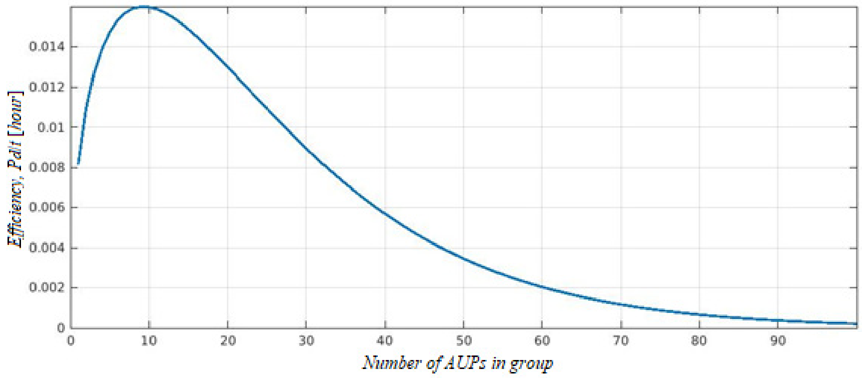


Fig. 3. Efficiency of the AUV group depending on the number of agents

Figure 3 shows that there is the maximum efficiency of application of each AUV application with the smaller number of AUVs in the group than at calculating efficiency of all the group. It is due to the fact that although addition of new AUVs increases the efficiency of the whole group, the efficiency increase turns out to be small and it does not cover economic expenses for purchasing and servicing additional AUVs.

Conclusion

It is reasonable to choose either such a group of AUVs that will survey the territory in the minimum time with a given probability, or will do this with a minimum number of AUVs, maximizing the economic effect. At the same time, the efficiency of using each AUV to solve a common problem is maximum.

In the available literature, there is no mathematical criterion for optimizing the composition of the group, the composition of the group was assumed to be predetermined, therefore, this article proposes a new method and a new optimization criterion that allows, with minimal financial costs (the minimum number of devices used), to perform the search task with a given probability in a time close to the minimum, that is, to ensure the maximum efficiency of the group.

REFERENCES

1. **Moore B.J., Passino K.M.** Decentralized redistribution for cooperative patrol. *Int. J. Robust Nonlinear Control*, 2008, Vol. 18, Pp. 165–195.

2. **Michel A.N., Kaining W.K., Bo H.** *Qualitative theory of dynamical systems*. N.Y.: Marcel Dekker, Inc., 2001, 706 p.

3. **Kiselev L.V., Inzartsev A.V., Bychkov I.V., Maksimkin N.N., Khmelnov A.Ye., Kenzin M.Yu.** Situatsionnoye upravleniye gruppirovkoy avtonomnykh podvodnykh robotov na osnove geneticheskikh algoritmov. *Podvodnyye Issledovaniya i Robototekhnika* [*Underwater Investigations and Robotics*], 2009, No. 2/8, Pp. 34–43. (rus)

4. **Fax J.A. Murray R.M.** Information flow and cooperative control of vehicle formations. *IEEE Transactions on Automatic Control*, 2004, Vol. 49, Iss. 9, Pp. 1465–1476.

5. **Casalino G., Aicardi M., Bicchi A., Balestrino A.** Closed loop steering and patch following for underactuated marine vehicles: A simple Lyapunov function based approach. *IEEE Robotics and Automation Magazine*, 2005, No. 2 (1), Pp. 27–35.

6. **Mayevskiy A.M., Gaykovich B.A.** Razrabotka gibridnykh avtonomnykh neobitayemykh apparatov dlya issledovaniya mestorozhdeniy uglevodorodov [Designing hybrid autonomous unmanned vehicles for exploration of hydrocarbon fields]. *Vesti Gazovoy Nauki: Collected Scientific Technical Papers*, 2019, No. 2 (39), Pp. 29–40. (rus)

7. **Willcox S., Goldberg D., Vaganay J., Curcio J.** *Multi-vehicle cooperative navigation and autonomy with the bluefin cadre system*. Massachusetts Institute of Technology, No. 05, 2023.

8. **Curcio J., Leonard J., Vaganay J., Patrikalakis A., Bahr A., et al.** Grund experiments in moving baseline navigation using autonomous surface crafts. *Proceedings of OCEANS 2005 MTS/IEEE*, Washington DC, Sept. 17-23, 2005. DOI: 10.1109/OCEANS.2005.1639839

9. **Vasilyev S.N.** Metod reduktsii i kachestvennyy analiz dinamicheskikh sistem, I-II. *Izv. RAN. Teoriya i Sistemy Upravleniya*, 2006, No. 1, Pp. 21–29; No. 2, Pp. 5–17. (rus)

10. **Kozhemyakin I.V., Blinkov A.P., Rozhdestvenskiy K.V., Ryzhov V.A., Melentyev V.D.** Perspektivnyye platformy morskoy robototekhnicheskoy sistemy i nekotoryye varianty ikh primeneniya. *Izvestiya SFedU. Engineering Sciences*, 2016, No. 1 (174), Pp. 59–77. (rus)

11. **Boris G., Kulchenko A., Maevskiy A., Beresnev M.** The structure of automatic control systems for underwater gliders. *Proceedings of the 4th ICCMA, Mechatronics and Automation*. Dec. 2016, Pp. 88–91. DOI: 10.1145/3029610.3029640

12. **Ageyev M.D., Kiselev L.V., Matviyenko Yu.V.** *Avtonomnyye podvodnyye roboty: Sistemy i tekhnologii*. Moscow: Nauka Publ., 2005, 398 p. (rus)

13. **Ambartsumyan A.A., Potekhin A.I.** Gruppovoye upravleniye v diskretnosobytiynykh sistemakh. *Problemy Upravleniya* [*Control Sciences*], 2012, No. 5, Pp. 46–53. (rus)

14. **Inzartsev A.V., Pavin A.M., Bagnitskiy A.V.** Planirovaniye i osushchestvleniye deystviy obsledovatelskogo podvodnogo robota na baze povedencheskikh metodov. *Podvodnyye Issledovaniya i Robototekhnika* [*Underwater Investigations and Robotics*], 2013, No. 1 (15), Pp. 4–16. (rus)

15. **Kapustyan S.G.** Algoritmy kollektivnogo uluchsheniya plana pri reshenii zadachi raspredeleniya tseley v gruppe robotov. *Iskusstvennyy Intellect* [*Artificial Intelligence*], 2006, No. 3, Pp. 679–690. (rus)

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Семенов Николай Николаевич
Nikolay N. Semenov
E-mail: semenov@smtu.ru

Михлин Валерий Григорьевич
Valerii G. Mikhlin
E-mail: valeriy_mikhlin@mail.ru

Ахметов Денис Булатович
Denis B. Akhmetov
E-mail: akhmetov@spbstu.ru

Submitted: 03.01.2023; Approved: 04.05.2023; Accepted: 17.05.2023.
Поступила: 03.01.2023; Одобрена: 04.05.2023; Принята: 17.05.2023.

Computing, Telecommunication and Control, 2023, Vol. 16, No. 1, Pp. 69–##.
Информатика, телекоммуникации и управление. 2023. Том 16, № 1. С. 69–##.

Intellectual Systems and Technologies
Интеллектуальные системы и технологии

Research article
DOI: https://doi.org/10.18721/JCSTCS.16106
UDC 004.85

IMPLEMENTATION OF MACHINE LEARNING ALGORITHMS FOR PARKINSONIAN GAIT DATA

O. Unal[1] ✉ , *V.V. Potekhin*[2]

[1,2] Peter the Great St. Petersburg Polytechnic University,
St. Petersburg, Russian Federation

✉ ogulunal@gmail.com

Abstract. In this study, we used the Physionet gait database and extracted gait features such as step/stride regularities and symmetries to build a classifier for Parkinson's disease (PD) subjects and healthy controls. We also improved the number of features using the mean and standard deviation of step times during their usual, self-selected pace for approximately 2 minutes on level ground. Extracted features were used in three different machine learning algorithms.

PD is a neurodegenerative disorder caused by the neurodegeneration of regions of the basal ganglia. Gait abnormality is one of the main symptoms of PD. Motor symptoms in Parkinson's disease cause a lack of control over movements and difficulty initiating muscle movements such as shuffling steps, quicker strides, or moving slower than expected for the corresponding age. The proposed approach can be used for the diagnosis of PD that can be automated or performed remotely.

Keywords: machine learning, supervised learning, Parkinson's disease, gait, feature analysis

Citation: Unal O., Potekhin V.V. Implementation of machine learning algorithms for Parkinsonian gait data. Computing, Telecommunications and Control, 2023, Vol. 16, No. 1, Pp. 69–78. DOI: 10.18721/JCSTCS.16106

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.16106
УДК 004.85

РЕАЛИЗАЦИЯ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ДАННЫХ О ПОХОДКЕ ПРИ БОЛЕЗНИ ПАРКИНСОНА

О. Юнал[1] ✉, *В.В. Потехин*[2]

[1,2] Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Российская Федерация

✉ ogulunal@gmail.com

Аннотация. Изучена база данных походки Physionet и рассмотрены такие характеристики походки, как регулярность шага и симметрии, для построения классификатора для пациентов с болезнью Паркинсона (БП) и здоровых людей контрольной группы. Увеличено количество функций, используя среднее значение и стандартное отклонение времени выполнения шагов в их обычном, самостоятельно выбранном темпе, в течение примерно двух минут на ровном месте. Извлеченные функции использованы в трех различных алгоритмах машинного обучения.

БП – нейродегенеративное заболевание, вызванное нейродегенерацией областей базальных ганглиев. Нарушение походки является одним из основных симптомов БП. Двигательные симптомы при болезни Паркинсона вызывают отсутствие контроля над движениями и трудности с инициированием мышечных движений, таких как шаркающие шаги, более быстрые шаги или движение медленнее, чем ожидалось для соответствующего возраста. Предложенный подход рекомендуется для диагностики БП, которая может быть автоматизирована или выполнена удаленно.

Ключевые слова: машинное обучение, контролируемое обучение, болезнь Паркинсона, походка, анализ признаков

Для цитирования: Unal O., Potekhin V.V. Implementation of machine learning algorithms for Parkinsonian gait data // Computing, Telecommunications and Control. 2023. Т. 16, № 1. С. 69–78. DOI: 10.18721/JCSTCS.16106

Introduction

Almost all neurodegenerative diseases of the brain begin insidiously and progress over the span of years. Parkinson's disease (PD) is a chronic and progressive neurological disorder that results in rigidity, tremor, postural instability, and slowness. The main symptoms appear gradually and worsen over time. As the disease progresses, people may have difficulty walking, mental and behavioral changes, depression, memory problems, sleep problems, and fatigue may also occur. PD is caused by selective cell death of dopamine-producing neurons in the brain. There are a variety of theories about what causes the neurons to die. Parkinson's disease has no cure, per se, and current treatments consist of externally supplying the dopamine that the dying neurons stop producing naturally [18].

The main problems of Parkinsonian gait are step length reduction with speed reduction impaired coordination, a decrease in the length of the step with an increase in the frequency of steps, the inability to produce effective steps at the beginning of walking, or complete cessation of steps during gait, and problems with dual tasking while gait. In this study, we extracted 12 different features using the Physionet database. Extracted features are correlated to Parkinsonian gait abnormalities therefore, we applied machine learning algorithms to make classification using a new dataset. Machine learning is a powerful technique for effectively analyzing data like gait signals. In this study, machine learning is used to analyze the gait data of the PD patients to classify them into "healthy" and "not healthy" classes based on extracted gait features.

Literature review

The human brain & neurotransmitters. Approximately one and a half kilogram organ that organizes every function to manage the body itself. The brain is also responsible to analyse the information as well as controlling emotions, intelligence, memory and movement. The main parts of the brain are Cerebrum, Cerebellum, and Brainstem [1].

Neurons are communicating through the body and communicate with one another to transmit signals. Although, neurons are not simply connected physically. Each neuron's end is a small gap that is a synapse and to communicate with the next cell, the signal has to be able to cross the space [5]. This process is known as neurotransmission. There are excitatory and inhibitory neurotransmitters. Excitatory will increase the likelihood that the neuron will fire an action potential while inhibitory will decrease it. Neurotransmitters and their functions are explained as [2]:

- Adrenaline (excitatory): also called epinephrine, it increases blood flow and heart rate.
- Norepinephrine (NE): also known as noradrenaline (excitatory) increases blood flow and attention.
- Dopamine (both excitatory and inhibitory): is mostly responsible for feelings of reward and pleasure. Low dopamine level is related to a specific disorders such as Parkinson’s disease [3].
- Serotonin (inhibitory): feelings of happiness and well-being, stable mood, sleep cycle, and digestive system.
- Gamma-aminobutyric acid (GABA) (inhibitory): inhibits neuron firing in the Central Nervous System, high levels improve sleep quality and provide calming effect.
- Acetylcholine (excitatory): learning, memory, muscle contraction, awakening.
- Glutamate (excitatory): learning, memory, and creating new nerve pathways.
- Endorphins: natural pain killer, excitement, and exercise.

Neurodegeneration. Different brain diseases attack different brain regions such as: Cerebrovascular diseases (primarily arteries), Infectious diseases (various substrates), Demyelinating diseases (primarily myelin) and Neurodegenerative diseases (primarily neurons). Neurodegeneration is progressive loss of structure/function of neurons and the death of neurons. Each neurodegenerative disease of the brain is caused by the progressive accumulation of a specific protein inclusion (proteinopathy) [6]. Over time this accumulation becomes toxic to the brain leading to irreversible degeneration (death) of neurons and atrophy. There are many neurodegenerative diseases, some of them are; Alzheimer’s, Huntington’s and Parkinson’s disease (see Fig. 1).

Parkinson’s disease. Parkinson's disease, also called primary parkinsonism, paralysis agitans, or idiopathic parkinsonism is a degenerative neurological disorder that is characterized by the onset of tremor, muscle rigidity, slowness in movement (bradykinesia), and stooped posture [7]. The disease was first described in 1817 by British physician James Parkinson in his Essay on the Shaking Palsy [7]. In PD, neurons in the part of the brain called the basal ganglia to begin to die off and produce less dopamine causing dopamine levels to fall. As dopamine starts to decrease, signs and symptoms of Parkinson's disease begin to appear [4] (see Fig. 2).

There are various kinds of research aiming at early detection of PD. In [19], focuses on vocal detection of Parkinson's disease while in [20] their research focuses on finger alterations during keyboard usage. In [21], gait data were used however their research was based on stride time variability and swing time variability. In our study, we use mean, and variance to detect the variabilities and we applied the coefficient of variation that shows the degree of variability relative to the mean value of the data.

Our study provides another possible usage of machine learning classifiers for PD patient diagnosis that is cost-effective, and provides more information for PD.

Gait in Parkinson’s disease. Gait is one of the neurological examinations for neurodegenerative diseases as well as for Parkinson’s disease. Motor symptoms in PD come from a lack of control over movements and difficulty initiating muscle movements [8]. Some features of Parkinsonian gait are:

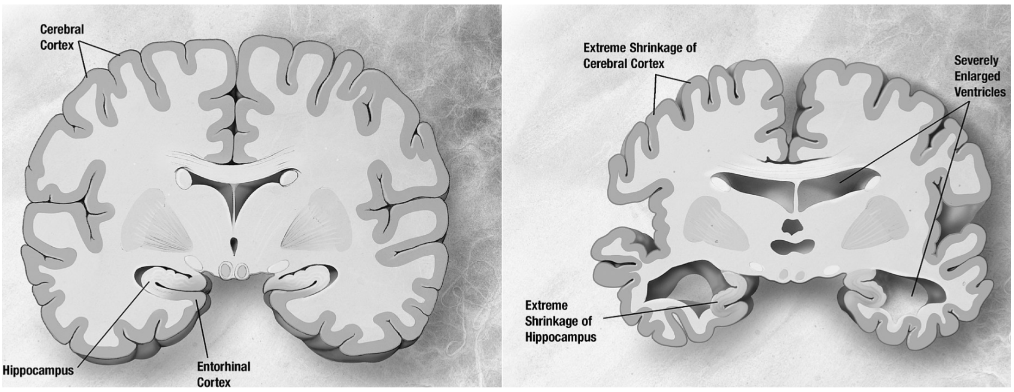


Fig. 1. Example diagram of normal and neurodegenerated brain

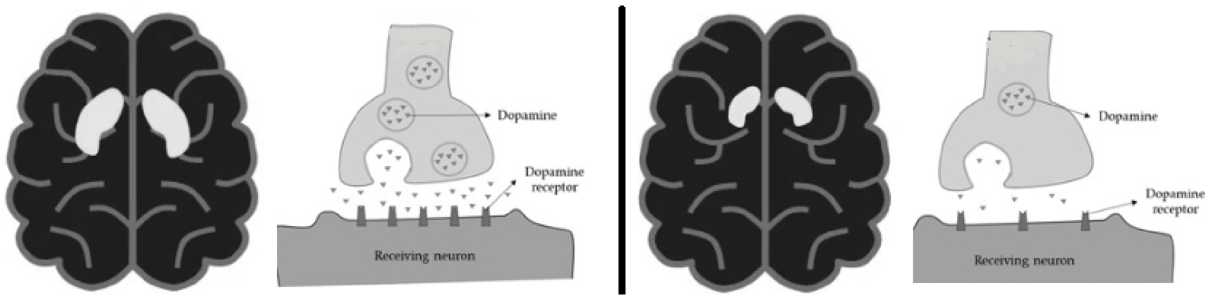


Fig. 2. SPECT scan and synaptic terminal of Healthy controls (left) and Parkinson patients on the right

- Taking small and shuffling steps.
- Move more slowly than expected for the corresponding age.
- Strides becoming quicker.
- Freezing of gait.

In this research, Gait data was used to analyse Parkinsonian gait characteristics, Afterward, Features related to gait data were extracted and a new dataset was created to test with different machine learning algorithms for classification. Extracted features were explained in detail in the Preprocessing section below.

Methods & materials

Physionet gait dataset. The database contains measures of gait from 93 patients with idiopathic PD (mean of age: 66.3 years; 63 % men), and 73 healthy controls (mean age: 66.3 years and 55 % men). The database includes the vertical ground reaction force (VGRF) records of subjects as they walked at their usual pace for 2 minutes on the level ground [9]. Underneath each foot were 8 sensors (Ultraflex Computer Dyno Graphy) that measure the force (in Newtons) as a function of time [9]. The output of each of these 16 sensors has been digitized and sampling was 100 per second [9]. In the database, two signals indicate the sum of the 8 sensor outputs for each foot. Among 8 sensors of each foot, sensors close to the toe were chosen for further preprocessing (1 sensor for the left and 1 for the right foot).

Column 1: Time (in sec).

Columns 2-9: Vertical ground reaction force on each of 8 sensors located under the left foot.

Columns 10-17: VGRF on each of the 8 sensors located under the right foot.

Columns 18,19: Total force under the left and right foot respectively. Ga, Ju, or Si in the database indicate the study from which the data originated:

Ga: Galit Yogev et al (dual-tasking in PD; Eur. J. Neuro, 2005);

Ju: Hausdorff et al. (RAS in PD; Eur. J. Neuro, 2007);

Si: Silvi Frenkel-Toledo et al. (Treadmill walking in PD; Mov. Disorders, 2005).

Feature extraction & preprocessing. A step is the movement made from one foot to the other while stride is a long step. The coefficient of variation (CV) is the ratio of the standard deviation to the mean value and shows the degree of variability relative to the mean value of the data. These features were generated using Physionet gait data in order to use in ML models. The higher the CV indicates greater variance. mean, standart deviation and CV formulas are as follows:

$$\mu = \frac{\sum_{i=1}^{n} xi}{n};$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (xi - \mu)^2}{n}};$$

$$Cv = \frac{\sigma}{\mu}.$$

Like many other motor symptoms that can occur in patients with PD, gait changes and freezing are caused by a loss of control in areas of the brain responsible for producing intentional movements. There are mainly 2 different gait disturbances that are episodic and continuous [10]. The episodic gait disturbances are freezing of gait and hesitation, the continuous changes indicates alterations in the walking pattern that appear, at least at first glance, to be more or less consistent from one step to the next, i.e., they persist and are apparent all the time [10]. Step, stride regularity and step symmetry features were generated to analyse and compare differences between PD patients and healthy controls.

Coordination of steps has also been shown to be dysfunctional in those with PD during gait therefore step times in PD for subjects in the database were used [11]. In the Fig. 3 represents peak detection using gait data. Peak times and peak values were also used to derive other features (step/stride regularity and step symmetry).

Gait data can be analyzed by unbiased autocorrelation procedures to give cadence, step length and measures of gait regularity and symmetry [12]. Step timing information can be calculated using the peak

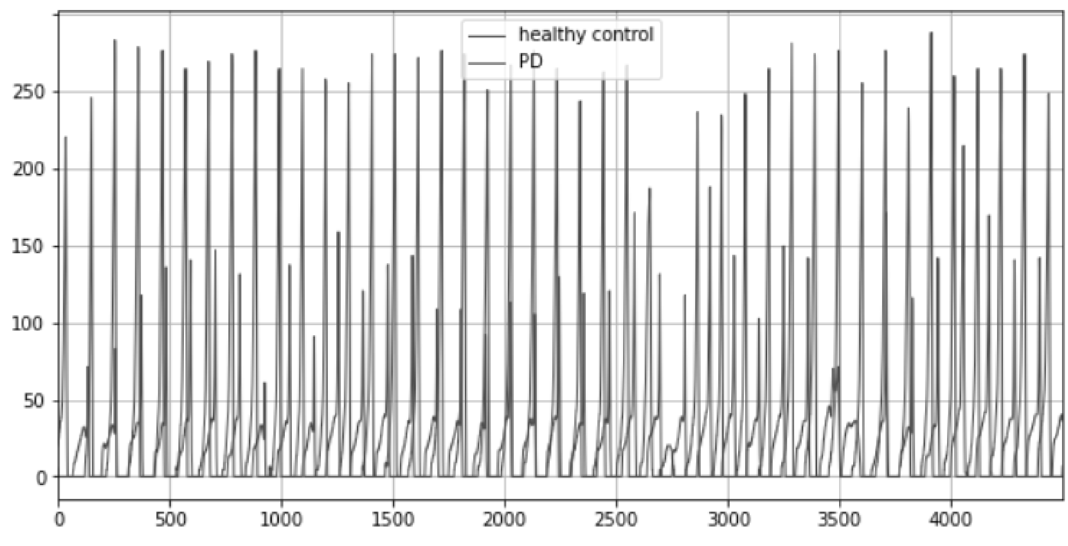


Fig. 3. Comparison of heathy and PD signals

times in raw gait data. Therefore, three features were generated that are: mean time between steps, standard deviation of step time, and the coefficient of variation (see Fig. 4).

As a result of feature extraction, there were 12 new columns representing step regularity, stride regularity, step symmetry, step time mean, step time standard deviation, and step time coefficient of variation for left and right leg electrodes. Ga, Ju, and Si studies were tested individually. This means the new dataset for Ga (dual tasking) has 113, Ju (Rhythmic Auditory Stimulation) has 129 and Si has 64 (Treadmill walking) rows that are extracted from each trial in the Physionet dataset. Thus, the dataset has 214 PD and 92 healthy controls. Below generated features are shown using the seaborn library in python programming language (see Fig. 5).

Machine learning algorithms

Random forest (RF). RF is an ensemble learning method, namely a random forest model consisting of a large number of small decision trees (estimators) that creates their predictions. The RF combines the result to obtain accuracy [13]. A random forest builds trees in parallel (bagging in ensemble learning) while boosting builts trees sequentially. Hyperparameters chosen for RF algorithm: max_depth (longest path between the root and the leaf node) = 3, max_features (RF is allowed to try in an individual tree) = 1, min_samples_leaf (minimum number of samples required to split an internal node) = 3.

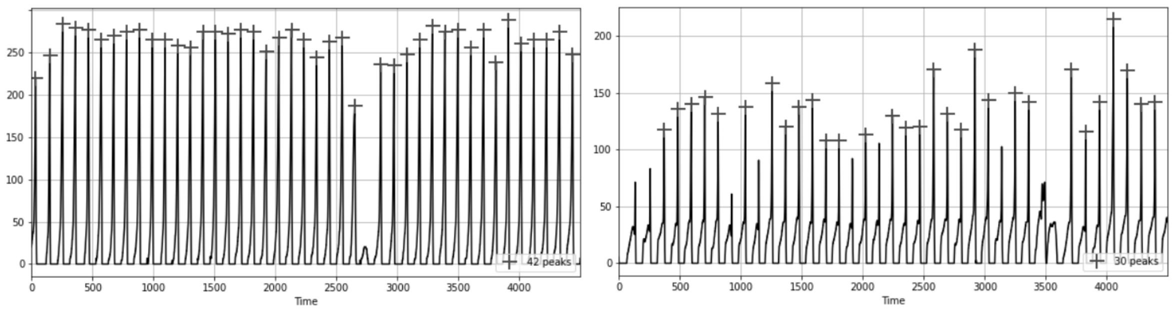


Fig. 4. Detected peaks of healthy controls and PD patients during 45 seconds interval

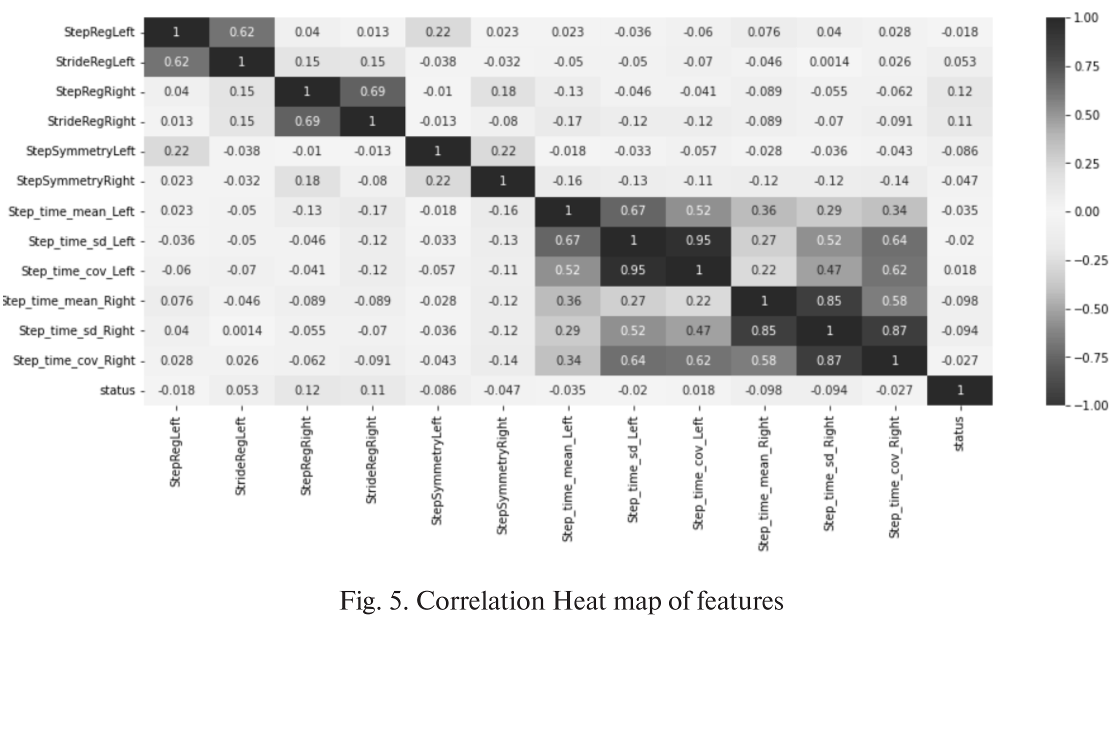


Fig. 5. Correlation Heat map of features

The model gave best performance using RF that construct many individual decision trees at training. Predictions from all trees are pooled to make the final prediction, to calculate Gini importance:

$$ni_j = w_j c_j - w_{left(j)} c_{left(j)} - w_{right(j)} c_{right(j)}.$$

Afterwards, importance for each feature on a decision tree is then calculated:

$$fi_j = \frac{\sum j \ Node \ j \ splits \ on \ feature \ i \ Nij}{\sum k \in all \ nodes \ ni_k}.$$

The final feature importance, is the average over all the trees:

$$RFfi_j = \frac{\sum j \in all \ trees^{normfi_{ij}}}{T}.$$

KNN. K-Nearest Neighbour is a supervised learning algorithm. KNN measures the similarity between the new case/data and training cases therefore the algorithm can classify the new case into the categories. KNN can be used for regression tasks or classification tasks. The default number of neighbors (five) was chosen.

AdaBoost. AdaBoost is an ensemble learning method, namely, an AdaBoost classifier is a meta-estimator that begins by fitting a classifier on the original dataset and then fits additional copies of the classifier on the same dataset but where the weights of incorrectly classified instances are adjusted such that subsequent classifiers focus more on difficult cases [14].

Results

Selected algorithms (RF, KNN, AdaBoost) were applied to three different data studies (Ga, Ju and Si). Each of them gave a different result. After different combinations, among 12 extracted features, only six of them were used in algorithms. Those are: StrideRegLeft, StrideRegRight, StepSymmetryLeft, StepSymmetryRight, Step_time_mean_Left Step_time_sd_Left gave highest accuracies for each data study. Random forest gave the highest accuracy for each task of binary classification. Note that this result indicates 100 % accuracy and Healthy vs PD which are presented in the matrix confusion, along with evaluation formulas below using test data (see Fig. 6).

Accuracy (TP + TN/(TP + TN + FP + FN)).

Precision = TP/(TP+FP).

Recall = TP/(TP+FN).

F1 Score = 2*(Recall * Precision) / (Recall + Precision).

Ga is dual-tasking for healthy controls and PD patients. Ju's study contains PD records using Rhythmic Auditory Stimulation (RAS) which is a treatment technique for patients. RAS is designed to improve gait by providing the patient with auditory cues throughout the gait. Treadmill walking can improve gait stability in patients who have Parkinson’s disease [15]. Combining Ga, Ju, and Si studies provided the worst result with high false positives for each algorithm. Different ML models were created and results were compared for each different Physionet database study (Ga, Ju, Si).

Patients having PD are supposed to have reduced step length people with Parkinsonian gait usually take small and shuffling steps. It may be difficult for them to lift their legs [16]. However, Parkinsonian gait is not the only symptom. There are other symptoms such as tremor, which usually begins in the hand and is more likely to occur when the limb is relaxed. Other symptoms are slowness of movement, Bradykinesia, and rigidity. It is also difficult to understand the common characteristics of gait because it might be differ-

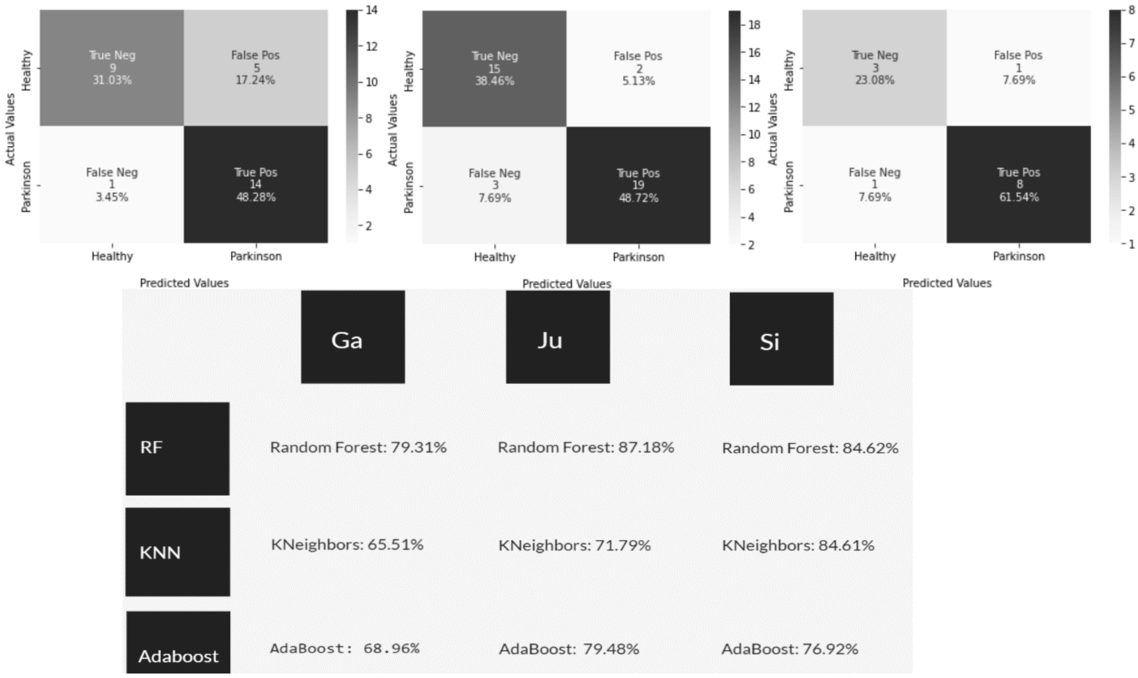


Fig. 6. Test accuracy results of algorithms and confusion matrix of RF using test data of Ga, Ju and Si respectively

ent for each person. As a result, using gait data alone might not be enough to classify or analyze PD using machine learning.

Because of having small datasets, we did not use neural networks in this study. Neural networks require a bigger size of data to achieve better results. The random forest algorithm provides a higher level of accuracy in predicting outcomes than the decision tree algorithm. RF performs well when there are imbalanced datasets. RF has methods for balancing errors in class population unbalanced data sets. Random forest minimizes the overall error rate when there is a data set. Thus, the larger class will get a low error rate while the smaller class will have a higher error rate. Hyperparameters of RF were tuned (explained in the previous section) KNN is mainly based on feature similarity. It is sensitive to outliers and noise. The performance of the KNN algorithm gets worse as the number of features increases. Another algorithm, AdaBoost, is another ensemble learning algorithm that is also sensitive to outliers and noise [17]. As a result, among chosen algorithms, RF gave the highest accuracies.

Conclusions

In this study, we presented a method to classify Parkinsonian gait data using extracted features from the Physionet gait database. Gait disorder is one of the main symptoms of Parkinson’s disease. Features based on gait characteristics were extracted and we used supervised learning algorithms to implement three machine learning models based on three datasets. The proposed models gave the highest accuracies using Random forest compared to KNN and AdaBoost algorithm. The results of our approach can be further improved by implementing it on a larger dataset. The classification results based on gait data can be improved by combining patients’ EEG and EMG data. As future work, EEG data of PD patients and healthy controls will be used for preprocessing and ML algorithms. Parkinson's disease is a neurodegenerative disease thus, it can be helpful to observe the effects of dopamine alterations in the brain using EEG signals. Another future work will be using EMG data to analyze muscle signals during the tremor stages. The gait data analysis and implemented ML models for the classification of Parkinson's disease can provide support for patients to improve their quality of life.

Acknowledgments

Firstly, I would like to thank my father Dr. Turgay Ünal for his help in Parkinson studies also would like to thank Associate professor Vyacheslav Potekhin for all his help during my studies. Finally, I thank my mom Dr. Seher Gülçin Ünal for her assistance in giving all the support in my life.

REFERENCES

1. **Marieb H., Hoehn K.** *Brain anatomy adapted from human anatomy & physiology*. 9th ed., Pearson, 2009, Vol. 9.

2. **Marianne K., Daniella S., Ariel R., David N., Jackson C., Ricardo G.** Dopamine: Functions, signaling, and association with neurological diseases. *Cellular and Molecular Neurobiology*, 2019, Pp. 1–30.

3. **Meder D., Herz D.M., Rowe J.B., Lehéricy S., Siebner H.R.** The role of dopamine in the brain – lessons learned from Parkinson's disease. *NeuroImage*, 2019, Pp. 79–93.

4. **Sergei G.** Dopamine function and the efficiency of human movement. *Journal of Cognitive Neuroscience*, 2016, Pp. 1–10.

5. **Potekhin V.V., Ünal O.** Analysis of emotions using EEG data and machine learning. *Proceedings of the 32nd DAAAM International Symposium*, Publ. by DAAAM International, Vienna, Austria, 2021, Pp. 158–167.

6. **Li D., Liu C.** Conformational strains of pathogenic amyloid proteins in neurodegenerative diseases. *Nat. Rev. Neuroscience*, 2022.

7. Parkinson disease. *Encyclopedia Britannica*, 21 Jan. 2021, *https://www.britannica.com/science/Parkinson-disease* (Accessed 2 June 2022).

8. **Lazzaro D.B.** Gait analysis in Parkinson's disease. *Overview of the Most Accurate Markers for Diagnosis and Symptoms Monitoring, Sensors*. Basel, Switzerland, 2020, Vol. 20.

9. **Goldberger A., Amaral L., Glass L., Hausdorff J., Ivanov P., Mark R., Stanley H.E.** *PhysioBank, PhysioToolkit and PhysioNet: Components of a new research resource for complex physiologic signals*, 2000, Pp. 215–220.

10. **Hausdorff J.M.** Gait dynamics in Parkinson's disease: Common and distinct behavior among stride length, gait variability, and fractal-like scaling. *Chaos*. Woodbury, N.Y., 2009. DOI: 10.1063/1.3147408

11. **Plotnik M., Giladi N., Hausdorff J.M.** Bilateral coordination of walking and freezing of gait in Parkinson's disease. *Eur. J. Neurosci*, 2008. DOI: 10.1111/j.1460-9568.2008.06167.x

12. **Nilssen R., Helbostad L.J.** Estimation of gait cycle characteristics by trunk accelerometry. *Journal of Biomechanics*, 2004, Vol. 37.

13. **Ali J., Khan R., Ahmad N., Maqsood I.** Random forests and decision trees. *International Journal of Computer Science Issues*, 2012.

14. **Freund Y., Schapire R.** *A decision-theoretic generalization of on-line learning and an application to boosting*, 1995.

15. **Earhart G.M., Williams J.** Treadmill training for individuals with Parkinson disease. *Physical Therapy*, 2012, Pp. 893–897. DOI: 10.2522/ptj.20110471

16. **Noh B., Youm C., Lee M., Cheon S.M.** *Gait characteristics in individuals with Parkinson's disease during 1-minute treadmill walking*, 2020. DOI: 10.7717/peerj.9463

17. **Chengsheng T., Liu H., Bing X.** AdaBoost typical algorithm and its application research. *MATEC Web of Conferences*, 2017.

18. **López-Grueso M.J., Padilla C.A., Bárcena J.A.** Deficiency of Parkinson's related protein DJ-1 alters Cdk5 signalling and induces neuronal death by aberrant cell cycle re-entry. *Cell Mol. Neurobiology*, 2022. DOI: 10.1007/s10571-022-01206-7

19. **Parisi L., RaviChandran N., Manaog M.L.** Feature-driven machine learning to improve early diagnosis of Parkinson's disease. *Expert Systems with Applications*, 2018, Vol. 110, Pp. 182–190. DOI: 10.1016/j.eswa.2018.06.003

20. **Giancardo L., Sánchez-Ferro A., Arroyo-Gallego T., Butterworth I., Mendoza C.S., et al.** *Computer keyboard interaction as an indicator of early Parkinson's disease*, 2016. DOI: 10.1038/s41598-018-32121-x

21. **Herman T., Gruendlinger L., Peretz C., Frenkel-Toledo S., Giladi N., Hausdorff J.M.** *Effect of gait speed on gait rhythmicity in Parkinson's disease: Variability of stride time and swing time respond differently*, 2005. DOI: 10.1186/1743-0003-2-23

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Юнал Огюл
Ogul Unal
E-mail: ogulunal@gmail.com

Потехин Вячеслав Витальевич
Vyacheslav V. Potekhin
E-mail: Slava.Potekhin@spbstu.ru
ORCID: https://orcid.org/0000-0001-9850-9558

Submitted: 24.11.2022; Approved: 10.05.2023; Accepted: 17.05.2023.
Поступила: 24.11.2022; Одобрена: 10.05.2023; Принята: 17.05.2023.