ISSN 2782-2427

CONTROL SCIENCES

2/2022

[1]Russian Academy of Sciences.

CONTROL SCIENCES
Scientific Technical Journal

6 issues per year
ISSN 2782-2427
Open access

Published since 2021

Original Russian Edition
Problemy Upravleniya
Published since 2003

FOUNDER AND PUBLISHER
V.A. Trapeznikov
Institute of Control Sciences
of Russian Academy of Sciences

Editorial address
65 Profsoyuznaya st., office 410,
Moscow 117997, Russia

☎/🖹 +7(495) 198-17-20, ext. 1410

✉ pu@ipu.ru

URL: http://controlsciences.org

Published: May 5, 2022

Registration certificate of
Эл № ФС 77-80482
of 17 February 2021
issued by the Federal Service
for Supervision of Communications,
Information Technology, and Mass
Media

CONTROL SCIENCES

2.2022

CONTENTS

DOI: http://doi.org/10.25728/cs.2022.2.1

PROBLEMS OF MANAGING THE FIRE SAFETY SYSTEM OF A FACILITY. PART II: MONITORING METHODS

D.V. Shikhalev

The State Fire Academy, the Ministry of the Russian Federation for Civil Defense, Emergencies and Elimination of Consequences of Natural Disasters (EMERCOM), Moscow, Russia

✉ evacsystem@gmail.com

Abstract. This paper overviews fire safety monitoring methods for a facility and state assessment methods for socio-economic systems used in fire safety. As discovered, none of the existing fire safety monitoring systems has a decision support procedure for adjusting a parameter (or several parameters) deviating from a given range. The majority of fire safety monitoring systems only assess the state of fire protection systems and transmit information on their triggering to the operational services. Thus, fire safety monitoring is simplified to assessing the state of fire automation systems, which cannot objectively reflect the fire safety state of the facility. As established, the integrated rating procedure is a most developed tool for assessing the state of a complex socio-economic system. This procedure is widespread in the theory of active systems. Its application to fire safety assessment is described. The existing contradictions in the management of fire safety systems are revealed, and some ways to resolve them are presented.

Keywords: fire safety, management, assessment of the facility's state, fire safety system, monitoring.

INTRODUCTION

Presently, the concept of a fire safety system as a controlled object is absent: this procedure is not described, there are no criteria for assessing the efficiency of fire safety systems, and the heads of organizations do not understand what they need to manage. For details, see part I of the survey [1]. In addition, despite many approaches to fire safety assessment, they are difficult to implement for the facility's head: he or she needs deep knowledge of the subject matter and the availability of appropriate qualifications and tools (computer programs). Thus, the head cannot assess the safety of his or her facility without qualified specialists.

Monitoring is a method to assess the current state of a facility, including fire safety. It is defined as a system of continuous assessment, control, and management of the facility's condition depending on its environment [2]. Monitoring systems are widespread in various spheres of human activities [3–7], including the complex safety of buildings [8, 9].

Concluding the survey, part II considers methods to monitor the facility's fire safety and assess the state of socio-economic systems used in fire safety.

1. FIRE SAFETY MONITORING METHODS FOR FACILITIES

Fire monitoring has been developed quite recently. The first elaborated solutions are associated with the appearance of *Strelec-Monitoring*, a hardware-software complex (HSC) for emergency monitoring and warning. In the book [10], the concept of constructing a radio-channel fire safety monitoring system for facilities was proposed. The use of radio channels was justified, and their advantages over traditional telephone lines were shown. As noted, the main cause of severe consequences (mass deaths) is the inability of the existing fire detection systems to transmit signals about the fire directly to the fire departments (call delays). The requirements outlined therein concern reliability, noise immunity, and other technical parameters of monitoring systems. However, there is no clarity about the parameters to be monitored. Several fire monitoring problems in the Russian Federation were touched on, but they relate only to legal aspects of monitoring and not its type, tasks, or other technical characteristics. Thus, the proposed monitoring system is actually a system for transmitting information about the element(s) of the automatic fire alarm system triggered to the Federal Emergency Service and sending fire and rescue units based on this signal.

Another newly developed solution is *Prometheus*, the national unified analytical fire safety control system for buildings [11]. According to the official website, the system remotely controls fire protection systems in buildings, monitors their maintenance, and manages databases of all participants in the fire protection market. As supposed, *Prometheus* will form a fire safety rating of an organization. Based on the current system functionality, the rating will be formed by assessing the performance and maintenance of fire protection systems. However, it seems rash to judge the facility's fire safety by this indicator: fire safety and its state are much deeper and broader concepts. Scientifically grounded methods are implemented here; see part I of the survey [1]. This system may be considered a higher level of HSC *Strelec-Monitoring*. However, it merely controls and transmits information about the state of fire protection systems and cannot be called a monitoring system: there are no methods to control the parameters in real-time. Generally speaking, fire safety monitoring in this system is treated as the operational control of fire protection systems. Such an approach is rather superficial.

Nowadays, fire monitoring is the most elaborated area. The dissertation [12] described remote fire monitoring models with current fire state assessment in a building. When a fire occurs in a building, the task is to obtain actual data on its growth and dynamics to ensure the safety of fire and rescue units. The author developed a decision support method based on fire monitoring. It represents a multilevel procedure for analyzing and ranking decisions by preference. No doubt, the procedure is useful at the fire occurrence stage. But such a problem statement goes beyond the scope of this paper.

Another example is the logical and probabilistic approach to the state monitoring of a potentially hazardous facility, particularly scenario modeling of accidents, developed in [13]. The monitoring of potentially hazardous facilities was defined as the continuous collection of information and the observation and control of a facility, including risk analysis, measurement of technological parameters at facilities, emissions of harmful substances, and environmental conditions in the adjacent territories. Monitoring is based on the interaction of two blocks: informational (data acquisition, processing, and presentation) and expert (scenario processing, modeling, and (or) forecasting, including assessment of the results). As a monitoring indicator, the author used a set of parameters determining the facility's safety and describing the system state at a given instant. Despite the relevance of this work, its drawbacks are a general set of monitoring parameters and no information about their sources.

In addition, methods are developed to control the state of the fire environment within an automatic environmental monitoring system [14]. An approach was proposed to integrate fire alarm and extra-early detection sensors to control the gas contamination level of the premises. The ambient temperature in the control area, the level of carbon monoxide, and smokiness were monitored. Test results showed a 25% increase in the accuracy of fire detection and a 37% increase in the total reliability of the system. Despite the positive test results, the research is more about solving an engineering problem than a scientific one. As expected, increasing control means will improve the efficiency of fire detection; monitoring parameters are supposed known to fire safety experts.

Note another monitoring system used after fire occurrence and development [15]. It is intended to assess the safety of fire and rescue units while working at steel truss structures in fire. Monitoring is carried out by the structure failure index (the probability that the entire structure will collapse when one of its elements collapses). The idea is to install temperature control sensors only on the elements critical for the structure's stability. The significance of different elements for the structure's stability is calculated in advance. When a fire occurs, the sensors transmit the environmental parameters, and the structure failure index is evaluated. As demonstrated by computer simulations, the slab collapse can be predicted 180 s before its occurrence. This time is enough for the fire and rescue units to leave the hazardous area. The approach considered in [15] is valuable for fire monitoring but only at the fire occurrence stage (like the one proposed in [12]).

Modern technologies are actively introduced in monitoring. For example, we mention a conceptual fire-fighting monitoring system based on the Internet of Things (IoT) technology [16]. Using various sensors, the system monitors current information about the pressure in the fire-fighting system, the temperature and humidity of the environment, the voltage of electrical equipment, the position of control valves, and the triggering signal. An early fire detection model was proposed based on a neural network. However, the principles of its operation were not described. In addition, the five-point rating scale was introduced to assess the fire safety of a building, but the grading method was not specified. Based on the monitoring parameters, we suppose that the fire safety level is assessed by the current state of the automatic fire-fighting system. In general, the approach presented by the authors is interesting and needs development. Like in the previously considered publications, monitoring is carried out for the technical elements of fire protection systems.

There are several regulations and documents in the state standards on fire and emergency monitoring (e.g., the state monitoring system of critical and potentially hazardous facilities and dangerous cargos [17]). The object of monitoring is the safety state of such facilities and cargos. This concept does not distinguish any parameters to be monitored, defining control areas only. Its peculiarity is automated decision support to minimize the consequences of an emergency. In general, this concept describes in detail the functions, composition, and operation of the monitoring system. Despite the obvious relevance, this area has not been properly developed, remaining merely a concept.

Note the state standard [8] on the foundations and design principles of a structured system to monitor and control the engineering systems of buildings and installations structures (MSIS). Its functions, composition, information exchange procedure, and other aspects are defined. This system is designed for potentially hazardous and critical facilities. MSIS should forecast and prevent emergencies by monitoring and determining parameter deviations from the nominal values. The monitoring parameters include the following [8]:

– fire occurrence;
– failures in the heat and water supply systems;
– failures in the power supply system;
– gas supply failures;
– failures in elevator equipment;
– unauthorized entry into the premises;
– increased radiation, maximum permissible concentrations of chemically hazardous substances, etc.;
– flooding of premises, drainage systems, and technological pits;
– deviations of technological processes from standard modes;
– changes in the structural elements of the building;
– the state of fire and emergency protection systems;
– the state of engineering protection;
– the condition of areas with a high probability of dangerous natural processes (landslides, avalanches, etc.).

Structurally, MSIS consists of three subsystems: data collection and messaging, communication and crisis management, and monitoring of engineering structures. MSIS operates in the following way. The system continuously monitors factors affecting safety; in case of any deviation from the norm, it transmits information to the dispatcher for warning and decision-making. This system has no decision support modules that would suggest a set of alternative decisions. The MSIS function is to warn about changes in the facility's state.

The document [9] defines and regulates the construction of a system to monitor automatic fire protection systems and signaling on the centralized 01-112 service panel. This system should collect data on fires (accidents) and natural disasters; monitor the reliability and operability of fire protection systems. The system makes fire (fire, alarm, etc.) and service (malfunction, loop disconnection, etc.) notifications. Like MSIS, it provides no decision support during monitoring. The system's functions are limited to transmitting messages about the state of the fire protection systems to the authorized organization.

Thus, the monitoring systems considered above transmit information about the state of fire protection systems to the authorized organization (fire and rescue department). In addition, several parameters are simply monitored, and their values do not affect the control of the facility's state.

This situation can be explained by two expert opinions dating back to the late 1980s; see [18]. According to Acad. Yu.A. Izrael, monitoring is a system of observations to identify state changes under human activities, which includes observation, assessment, and forecasting of the environment state and does not include quality management for the environment and human activities [19]. His opponent, Acad. I.P. Gerasimov defined monitoring as a system of control, assessment, and management of the environment, which must be purposeful, interconnected, and efficient [20]. Moreover, the efficiency of management-free monitoring may cause some problems (the redundancy and insufficiency of information, no demand for information, etc.). As for fire safety, we accept Gerasimov's viewpoint in this paper. The primary problem in fire safety management (like in the management of any organizational system) is to monitor the system's current state, i.e., understand the starting point of management.

According to this analysis, fire safety monitoring methods for facilities are at the initial stage of development. As a rule, the state of fire protection systems is monitored, and data on their triggering are transmitted to the operational services. This is due to the following factors: no indicator characterizing the facility's fire safety state and, consequently, no technical solutions to obtain the values of monitored parameters. Moreover, none of the systems considered above monitors organizational measures (the observance of fire safety codes). Meanwhile, see part 1 of the survey [1], it is one of the most frequent reasons for fire occurrence. At the current development stage, fire safety monitoring systems for facilities need more parameters than those related to the operability of fire protection systems. The latter systems are designed for situations following fire occurrence. In addition, none of the monitoring systems in fire safety has a decision support procedure to adjust the values of parameter(s) deviating from the nominal range.

Based on the above considerations, we concretize the term "monitoring" for the facility's fire safety. Let us introduce the following definition. Fire safety monitoring for a facility is a regular purposeful activity that includes (a) assessing the facility's fire safety state based on a set of characteristic factors (in particular, organizational and technical measures), (b) control of this state by determining any deviations of the parameter values from the nominal range, and (c) decision-making in the case of such deviations. Thus, a topical problem is to identify a set of factors determining the state of fire safety and establish the relationship between them and the degree of their influence on the entire facility's state.

2. STATE ASSESSMENT METHODS FOR SOCIO-ECONOMIC SYSTEMS: APPLICATION TO FIRE SAFETY

Let us consider the existing state assessment methods for socio-economic systems. We begin with several terms.

Fire safety is the state of a facility characterized by the ability to prevent fire occurrence and development and the impact of fire hazards on people and property.[1]

Fire safety codes are special social and (or) technical conditions established to ensure fire safety by federal laws and other normative legal acts of the Russian Federation as well as regulatory documents on fire safety.[2]

Fire safety measures are actions to ensure fire safety, including the implementation of fire safety codes.[2]

Fire prevention is a set of precautionary measures to eliminate the possibility of fires and limit their consequences.[2]

Thus, no term currently characterizes the fire safety state. We define it in the following way: the facility's fire safety state is a set of organizational, social, and technical factors that determine the facility.

The state of most systems in the first approximation can be described by a set of factors (parameters) affecting their operation. The state assessment problem reduces to determining such factors (parameters), establishing their functional relationship, and obtaining qualitative or quantitative metrics of the factors. Assessment means both the process and result of measurement [21]. In this study, we comprehend assessment as the result of measuring the current state of a system. Consider several research works devoted to state assessment.

In the previous section, mechanisms for assessing the safety of potentially hazardous facilities have been described. They involve the theory of active systems. Let us consider this technology in detail.

The integrated rating procedure of complex socio-economic systems [22] is based on a hierarchical representation of the goal tree. The main idea is to disaggregate the tree vertices using the dichotomy method [23], convoluting vertex pairs up in the hierarchy. This procedure considers both quantitative (measurable or calculated) indicators and qualitative ones determined through expertise.

The procedure includes the following steps.

Step 1. Choosing n directions to assess the facility's state.

Step 2. Dividing all directions into subgroup 1 (objective estimates that can be measured, calculated, etc.) and subgroup 2 (expert estimates).

Step 3. Forming a unified rating scale for all n directions.

Step 4. Determining local estimates for the directions in subgroup 2.

Step 5. Determining the object's characteristic indicators for each direction in subgroup 2.

Step 6. Developing a scale to recalculate the indicators (Step 5) into local ratings.

[1] Federal Law of the Russian Federation of July 22, 2008, No. 123-FZ "Technical Regulations on Fire Safety Codes."

[2] Federal Law of the Russian Federation of December 21, 1994, No. 123-FZ "On Fire Safety."

Step 7. Determining the significance of the indicators (Step 6).

Step 8. Measuring or calculating the indicators of subgroup 2 (Steps 5 and 7).

Step 9. Recalculating the indicators in subgroup 2 into local ratings.

Step 10. Determining the facility's local estimates for the directions of subgroup 1.

Step 11. Determining a pair of directions to convolute the local estimates into a generalized one.

Step 12. Forming convolution matrices for the pairwise comparison of the local and generalized estimates.

Step 13. Forming the facility's integrated rating.

This procedure yields the integrated rating of a complex socio-economic system (organization, project, etc.). It was used for fire safety problems in several publications.

For example, the author [24] adopted the theory of active systems and the integrated rating procedure to develop models and mechanisms of fire safety management in a region. Fire safety in a region was characterized by three criteria: injuries and deaths in fires and the amount of material damage. The criteria were assessed using the four-point rating scale. In addition, the level of fire safety in the Voronezh Region was also assessed: as discovered, the level of fire safety decreased by 3% from 1996 to 2001. An appropriate program was developed to improve it.

This approach was further developed in [25]. The author proposed models and algorithms for fire safety management based on regional development programs. The three fire safety indicators of the previous study were supplemented by the number of fires. However, it was excluded at the next step due to exceeding the admissible relation with other indicators. Obviously, the number of fires is the main indicator affecting the others (the number of deaths, injuries, and the amount of material damage).

The approaches [24, 25] were an attempt to assess the level of fire safety, but the results seem debatable: the number of deaths and injuries and the amount of material damage are unacceptable and rather rough criteria. They can be used to assess fire safety measures in the first approximation but not the fire safety level.

The considerations presented above lead to the following conclusions. Currently, the integrated rating procedure, proposed by Prof. V.N. Burkov [22], is a most developed tool to assess the state of complex socio-economic systems. This procedure is efficient enough and can be adapted to assess the facility's fire safety system. At the same time, when the number of state indicators increases or their structural relations change, the procedure should be repeated, including expert assessment. This procedure was used to assess the fire safety level. However, the corresponding results are not applicable to assess the facility's fire safety state. Thus, fire safety state assessment methods for facilities require development.

CONCLUSIONS

Concluding part II of the survey, we note the following:

- According to the analysis, fire safety monitoring methods for facilities are at the initial stage of development. As a rule, the state of fire protection systems is monitored, and data on their triggering are transmitted to the operational services. This is due to the following factors: no indicator characterizing the facility's fire safety state and, consequently, no technical solutions to obtain the values of monitored parameters. Moreover, none of the systems considered above monitors organizational measures (the observance of fire safety codes), one of the most frequent reasons of fire occurrence. At the current development stage, fire safety monitoring systems for facilities need more parameters than those related to the operability of fire protection systems. The latter systems are designed for situations following fire occurrence. In addition, none of the monitoring systems in fire safety has a decision support procedure to adjust the values of parameter(s) deviating from the nominal range.
- Currently, the integrated rating procedure, proposed by Prof. V.N. Burkov [22], is a most developed tool to assess the state of complex socio-economic systems. This procedure is efficient enough and can be adapted to assess the facility's fire safety system. At the same time, when the number of state indicators increases or their structural relations change, the procedure should be repeated, including expert assessment. This procedure was used to assess the fire safety level for a region only, and the set of indicators was quite debatable. However, the corresponding results are not applicable to assess the facility's fire safety state. Thus, fire safety state assessment methods for facilities require development.

Let us summarize the problems of managing the fire safety system of a facility. This research area faces several serious challenges and contradictions:

- The head of a facility (organization) is charged with ensuring fire safety (managing the fire safety system), a criminal liability is stipulated for violating fire safety rules. However, no methods and algorithms are provided to manage such systems.
- The community of engineers and researchers has developed many methods to assess the facility's fire safety. Due to their complexity, the head of a facility (organization) cannot apply them without appropriate professional training to assess the current state of the fire safety system and make managerial decisions.
- Usually, studies of fire safety systems are focused on the strategic levels (region, state). At the basic level of a facility, there are still no technologies for managing fire safety systems (including problem statements and solutions) like, e.g., in the book [21].

We believe that the starting point for resolving the challenges and contradictions is the following set of problems:

– Conceptualizing the states of the facility's fire safety system at different operation stages of the system;

– Developing a methodology for assessing the state of this system at different phases of the facility's lifecycle;

– Developing methods, models, and algorithms for managing the state of the facility's fire safety system at different phases of the facility's lifecycle;

– Developing decision support methods for the head of the facility (organization) to manage the fire safety system;

– Developing an information-analytical system to support fire safety system management.

These problems should be solved in real-time.

Acknowledgments. *The author is grateful to the reviewers for careful reading of the manuscript and helpful remarks.*

REFERENCES

1. Shikhalev, D.V., Problems of Managing the Fire Safety System of a Facility. Part I: Assessment Methods, *Control Sciences*, 2022, no. 1, pp. 2–14.
2. *Grazhdanskaya zashchita: monitoring* (Civil Defense: Monitoring), Puchkov, V.A., Ed., Moscow: the All-Russian Research Institute for Civil Defense and Emergencies, 2015. (In Russian.)
3. Chepikov, N.A., A Model, Method, and Algorithm to Process Social-Hygienic Monitoring Data of a Road Transport Network for Health Risk Management, *Cand. Sci. (Eng.) Dissertation*, Kursk, 2013. (In Russian.)
4. Parashchuk, E.A., An Automated Air Monitoring System as Information Support for Making Managerial Decisions, *Cand. Sci. (Eng.) Dissertation*, Belgorod, 2014. (In Russian.)
5. Noritsa, V.A., Information Support for an Infrastructure Monitoring System of a Research and Education Institution, *Cand. Sci. (Eng.) Dissertation*, Moscow, 2013. (In Russian.)
6. Belov, A.V., Raster Models and an Algorithm to Visualize Spatial Data in Monitoring Systems for Natural and Anthropogenic Disasters, *Cand. Sci. (Eng.) Dissertation*, Kursk, 2013. (In Russian.)
7. Mikov, D.A., Information Risk Management in Remote Monitoring Systems for the Facility's State, *Cand. Sci. (Eng.) Dissertation*, Moscow, 2018. (In Russian.)
8. *GOST* (State Standard) *22.1.12-2005: Safety in Emergencies. A Structured Monitoring and Control System for the Engineering Systems of Buildings and Installations. General Requirements*, Moscow: the All-Russian Research Institute for Civil Defense and Emergencies, 2005. (In Russian.)
9. *GOST* (State Standard) *56935-2016. Production Services. Services on Constructing a System to Monitor Automatic Fire Protection Systems and Signaling on the Centralized 01-112 Service Panel*, Moscow: the Public Fire Safety Ministry, 2016. (In Russian.)
10. *Istoriya sozdaniya i sovershenstvovaniya besprovodnykh sistem monitoringa* (History of Creation and Improvement of Wireless Monitoring Systems), Zikov, V.I., Ed., Moscow: State Fire Academy of EMERCOM of Russia, 2015. (In Russian.)
11. *Prometheus. The National Unified Analytical Fire Safety Control System for Buildings*. URL: https://fortespro.ru. (Accessed January 14, 2021.)
12. Tarakanov, D.V., Multicriteria Models and Methods to Support the Management of Fire Departments Based on Fire Dynamics Monitoring in a Building, *Dr. Sci. (Eng.) Dissertation*, Moscow, 2018. (In Russian.)
13. Mosyagin, A., Monitoring of Potentially Hazardous Facilities Based on Logical and Probabilistic Modeling, *Cand. Sci. (Eng.) Dissertation*, Moscow, 2009. (In Russian.)
14. Glushko, V.S., An Integrated Environmental Monitoring System for Fire Prevention on Oil and Gas Facilities, *Cand. Sci. (Eng.) Dissertation*, Moscow, 2014. (In Russian.)
15. Jiang, S., Zhu, S., Guo, X., Chen, C., and Li, Z. Safety Monitoring System of Steel Truss Structures in Fire, *Journal of Constructional Steel Research*, 2020, no. 172, pp. 1–11.
16. Li, Y., Yi, J., Zhu, X., Wang, Z., and Xu, F., Developing a Fire Monitoring and Control System Based on IoT, *Proceedings of 2nd International Conference on Artificial Intelligence and Industrial Engineering*, 2016, pp. 174–178.
17. Decree of the Government of the Russian Federation of August 8, 2005, No. 1314-r "On Approving the Concept of the Federal Monitoring System for Critical Facilities and (or) Potentially Hazardous Infrastructure Facilities of the Russian Federation and Dangerous Cargoes," Moscow, 2005. (In Russian.)
18. Rychikhina, E.N., *Monitoring kak obshchaya funktsiya upravleniya* (Monitoring as a General Management Function), Uhta: Ural State Technical University, 2008. (In Russian.)
19. Izrael', Yu.A., The Concept of Biosphere Monitoring, in *Monitoring sostoyaniya okruzhayushchei prirodnoi sredy* (Monitoring of the Environment), Leningrad: Nauka, 1977. (In Russian.).
20. Gerasimov, I.P., *Struktura nauchnogo issledovaniya: filosofskii analiz poznavatel'noi deyatel'nosti* (Research Structure: Philosophical Analysis of Cognition), Moscow: Mysl', 1985. (In Russian.)
21. Novikov, D.A., *Control Methodology*, New York: Nova Science, 2013.

22. Burkov, V.N., Gratsianskii, E.V., Dzyubko, S.I., and Shchepkin, A.V., *Modeli i mekhanizmy upravleniya bezopasnost'yu* (Models and Mechanisms of Safety Management), Moscow: Sinteg, 2001. (In Russian.)
23. Anokhin, A.M., Glotov, V.A., Pavel'ev, V.V., and Cherkashin, A.M., *Tselenapravlenyi vybor: modeli, otnoshenia, algoritmy* (Purposeful Choice: Models, Relations, Algorithms), Moscow: Trapeznikov Institute of Control Sciences RAS, 1996. (In Russian.)
24. Polovinkina, A.I., Models and Mechanisms of Optimal Fire Safety Management in a Region, *Cand. Sci. (Eng.) Dissertation*, Moscow, 2003. (In Russian.)
25. Kuzovlev, A.V., Fire Safety Management Models and Algorithms Based on Regional Development Programs, *Cand. Sci. (Eng.) Dissertation*, Moscow, 2014. (In Russian.)

This paper was recommended for publication by V.V. Kul'ba, a member of the Editorial Board.

Received July 19, 2021, and revised November 1, 2021.
Accepted November 22, 2021.

Author information

Shikhalev, Denis Vladimirovich. Cand. Sci. (Eng.), the State Fire Academy, the Ministry of the Russian Federation for Civil Defense, Emergencies and Elimination of Consequences of Natural Disasters (EMERCOM), Moscow, Russia
✉ evacsystem@gmail.com

Cite this paper

Shikhalev, D.V. Problems of Managing the Fire Safety System of a Facility. Part II: Monitoring Methods. *Control Sciences* **2,** 2–8 (2022). http://doi.org/10.25728/cs.2022.2.1

Translated into English by *Alexander Yu. Mazurov*,
Cand. Sci. (Phys.–Math.),
Trapeznikov Institute of Control Sciences,
Russian Academy of Sciences, Moscow, Russia
✉ alexander.mazurov08@gmail.com

DOI: http://doi.org/10.25728/cs.2022.2.2

A MATHEMATICAL MODEL OF MANAGING A REGULATED MONOPOLY DISTRICT HEATING MARKET[1]

V.A. Stennikov[1], O.V. Khamisov[2], and A.V. Penkovskii[3]

Melentiev Energy Systems Institute, Siberian Branch, Russian Academy of Sciences, Irkutsk, Russia

[1]✉ SVA@isem.irk.ru, [2]✉ Khamisov@isem.irk.ru, [3]✉ penkoffsky@isem.irk.ru

Abstract. This paper formulates an approach to managing the district heating of consumers in a hierarchical two-level system. The upper level is a regulator (e.g., a regional tariff service) that adjusts the heat energy tariff for consumers. The lower level is a district heating system that technologically and organizationally combines heat energy production and transport within a unified heat supply organization. The interaction of all participants in the process of heat supply to consumers is described. Optimization criteria are proposed for the upper and lower levels. A bi-level mathematical model of the district heating system is developed using the theory of hydraulic circuits and bi-level programming. This model operates in the conditions of a regulated monopoly heat energy market. The developed approach is applied to the real district heating system of Angarsk.

Keywords: district heating systems, heat energy market, mathematical modeling, hierarchical management, optimization.

INTRODUCTION

District heating plays an important role in the heat energy markets of many countries. Today, there are about 80 thousand district heating systems (DHSs) in the world: 50 thousand DHSs are located in Russia [2], 6 thousand large DHSs are operating in Europe [3], and the other 24 thousand DHSs are located in China, the USA, Canada, and former USSR countries (Ukraine, Kazakhstan, Belarus, etc.).

According to Global Market Insights[2], the global district heating market was worth more than $150 billion in 2019, producing about 3300 million Gcal of heat, with nearly 1300 million Gcal (40%) in Russia.

In world practice, two organizational models are used to manage district heating markets: the competitive model and the natural monopoly model.

Like in other spheres, competition is very important in the district heating market: it contributes to the efficiency of heat production and its quality. As a consequence, the price for heat energy reduces, favorably affecting the development of the entire industry. The competitive model is technologically based on several independent heat sources connected to consumers via heating networks. Heating networks have to be organizationally separated from heat generation and combined into a single heating network company with an independent sphere of activity. Such an organizational model is commonly referred to as the Unified Purchaser [4, 5]. The competitive model in the district heating market successfully operates in some European countries, such as Germany [3], Finland [6], and Sweden [7].

The natural monopoly heating market model is most widespread for heat supply to consumers. This model includes tariff regulation for consumers and operates in many EU countries, e.g., the Netherlands [8],

[1] This work was performed at Melentiev Energy Systems Institute (SB RAS) within state orders FWEU-2021-0002 (registration no. AAA-A21-121012090012-1) and FWEU-2021-0006 (registration no. AAA-A21-121012090034-3) of the fundamental research of SB RAS.

[2] URL: https://www.gminsights.com/industry-analysis/district-heating-market (Accessed September 26, 2021.)

Poland [9], Lithuania [10], Latvia [11, 12], Norway [13], and Estonia [14] as well as in Russia, China, etc. In a particular country, heat energy tariffs are usually controlled by an executive authority on the state regulation of prices (tariffs) or a local authority (if vested with the corresponding powers). For details, see Table 1.

The organizational heating management model in the form of a natural monopoly with tariff regulation for consumers can be represented as a hierarchical vertically integrated system with two levels (Fig. 1).

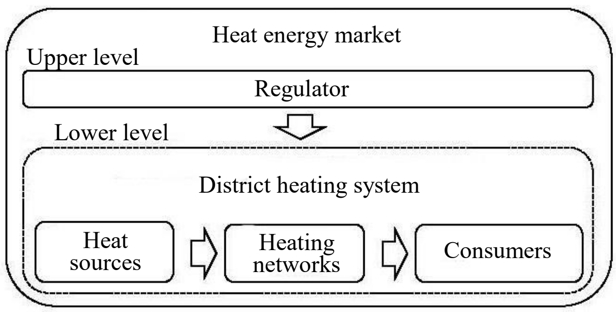

Fig. 1. Two-level heating management model.

The upper level here is the regulator responsible for adjusting the heat supply tariff for consumers. The lower level is a district heating system that technologically and organizationally combines heat energy production and transport within a unified heat supply organization (UHSO). The two-level management scheme for a monopoly district heating market separates the subsystems corresponding to particular market agents to model their behavior when implementing the objectives.

Heat energy market participants interact with each other as follows. Based on the forecasted heat demand from consumers, the UHSO produces and sells heat energy to them under several requirements. First, heat sources shall collectively produce a total heat volume to cover the demand from consumers and simultaneously obtain the maximum profit. Second, the available capacities of heat sources and the physical and technical limitations of heating networks shall be considered. In turn, protecting the rights of consumers, the regulator sets a heat energy tariff for motivating heat sources to satisfy the consumer demand (on the one hand) and maximize their profits while maintaining the optimal modes in heating networks.

Such a system is mathematically described using two-level modeling [16]. We pass to a one-level optimization problem by replacing the convex optimization problem of the second (lower) level with first-order optimality conditions. Note that these studies develop the existing approaches and methods and mostly rest on the paper [17], including the main provisions in mathematical modeling of district heating system objects (particularly heat sources, heating networks, and consumers). Nevertheless, the problem statement below fundamentally differs from that of [17] with the equal interaction of heat supply participants: it is hierarchical and, therefore, has completely different mathematical properties.

Table 1

Heat energy tariff regulators in different countries

Country	The Netherlands	Poland	Lithuania	Latvia
Regulator	Authority for Consumer and Market[3]	Energy Regulatory Office[4]	National Control Commission for Prices and Energy[5]	Sabiedrisko pakalpojumu regulēšanas Komisijas[6]
Country	Norway	Estonia	Russia	China
Regulator	Norwegian Water Resources and Energy Directorate[7]	Republic of Estonian Competition Authority[8]	Federal Anti-Monopoly Service, regional authorities[9]	Municipal authorities [15]

[3] URL: https://www.acm.nl/en/about-acm/our-organization/the-netherlands-authority-for-consumers-and-markets (Accessed October 1, 2021.)
[4] URL: http://www.ure.gov.pl/en/about-us/presidents-duties/22,Presidents-duties.html (Accessed October 1, 2021.)
[5] URL: http://www.regula.lt/en/Pages/default.aspx (Accessed October 1, 2021.)
[6] URL: https://www.sprk.gov.lv/content/siltumenergija (Accessed October 1, 2021.)
[7] URL: https://www.nve.no/energy-market-and-regulation/?ref=mainmenu (Accessed October 1, 2021.)
[8] URL: https://www.riigiteataja.ee/en/eli/ee/Riigikogu/act/520062017016/consolide (Accessed October 1, 2021.)
[9] Federal Law of the Russian Federation of July 27, 2010, no. 190-FZ “On Heat Supply.”

1. MATHEMATICAL MODELING OF A REGULATED DISTRICT HEATING MARKET

1.1. Modeling of district heating systems

The DHS topology is described by an incidence matrix A, where the number of rows coincides with the number of nodes (i = 1,..., m), and the number of columns coincides with the number of branches (j = 1,..., n). The elements a_{ji} of the matrix A are given by

$$a_{ji} = \begin{cases} 0 \text{ if arc } i \text{ is not connected to node } j, \\ 1 \text{ if arc } i \text{ has outgoing flow from node } j, \\ -1 \text{ if arc } i \text{ has incoming flow in node } j, \end{cases}$$

$$i \in I, j \in J,$$

where I and J denote the sets of heating network edges and heating network nodes, respectively.

A DHS is modeled on a time interval with an initial instant $\tau_0 = 1$ (corresponding to the total heat load calculated) and a terminal instant τ_{fin} (e.g., the calendar number of hours in a year, 8760). The set $T = \{1, 2, ..., \tau_{fin}\}$ consists of all instants.

1.2. Modeling of heat sources

The behavior of heat sources in market conditions is modeled under the following requirements: at each instant $\tau \in T$, they shall collectively produce a total heat volume to cover a given consumer demand and simultaneously obtain the maximum profit under their capacity constraints.

Let the variable $Q_{\tau j}{}^G$ describe the heat volume produced by heat source $j \in J_G$ at an instant $\tau \in T$, where J_G denotes the set of all heat sources. Also, we introduce the following notations: $w_{\tau j}{}^{HE}$ is the unit price of heat energy received by generator $j \in J_G$ at an instant τ; $w_j^P = \gamma_j / \overline{Q}_j^G$ is the fixed charge (rate) per unit of installed capacity, where γ_j is the semi-fixed costs of heat source j, and $\overline{Q}_j^G$ is the installed (maximum) capacity of heat source j. Similarly to [17], the profit of heat source j at an instant τ considering its heat capacity constraints and additional proceeds for the provided heat power is determined by

$$F_{\tau j}^G(Q_{\tau j}^G) = w_j^P \overline{Q}_j^G + w_{\tau j}^{HE} Q_{\tau j}^G - Z_{\tau j}^G(Q_{\tau j}^G) \to \max, \quad (1)$$

$$\underline{Q}_j^G \le Q_{\tau j}^G \le \overline{Q}_j^G. \quad (2)$$

In this optimization problem, $\underline{Q}_j^G$ is the minimum allowable used capacity of the heat source, and $Z_{\tau j}^G(Q_{\tau j}^G)$ is the costs of heat source j to produce the heat volume $Q_{\tau j}^G$. These costs are represented as a quadratic polynomial [17].

The price $w_{\tau j}{}^{HE}$ in problem (1), (2) is treated as an external parameter. The price w_j^P is a given constant. The profit function $F_{\tau j}{}^G(Q_{\tau j}{}^G)$ is strongly concave. Hence, problem (1), (2) has a unique solution $Q_{\tau j}{}^{G,*}(w_{\tau j}{}^{HE})$:

$$Q_{\tau j}^{G,*}(w_{\tau j}^{HE}) = \begin{cases} \underline{Q}_j^G, & w_{\tau j}^{HE} < \underline{w}_{\tau j}^{HE}, \\ \dfrac{w_{\tau j}^{HE} - \beta_j}{2\alpha_j}, & \underline{w}_{\tau j}^{HE} \le w_{\tau j}^{HE} \le \overline{w}_{\tau j}^{\mathrm{HE}}, \\ \overline{Q}_j^G, & w_{\tau j}^{HE} > \overline{w}_{\tau j}^{HE}, \end{cases} \quad (3)$$

where $\underline{w}_{\tau j}^{HE} = 2\alpha_j \underline{Q}_j^G + \beta_j$ and $\overline{w}_{\tau j}^{HE} = 2\alpha_j \overline{Q}_j^G + \beta_j$ are the heat prices of source j corresponding to the minimum and installed (maximum) capacities, and α_j and β_j are approximation coefficients in the cost function $Z_{\tau j}^G(Q_{\tau j}^G)$ [17].

1.3. Modeling of heat consumers

The set of all heat consumers J_D can be written as the union $J_D = J_{DH} \cup J_{DIG} \cup J_{DIN}$, where: J_{DH} is the consumers of housing and communal services (HCS); J_{DIG} is the industrial consumers connected to heating networks not representing source nodes (collectors); J_{DIN} is industrial consumers located on the collectors of heat sources. Clearly,

$$J_{DIN} \subset J_G,\ J_G \cap J_{DIG} = \varnothing,\ J_{DIN} \cap J_{DIG} = \varnothing. \quad (4)$$

Let $Q_{\tau j}{}^D$ be the total demand of consumers in node $j \in J_D$ at an instant $\tau \in T$. For brevity, we introduce the following notations: $Q_{\tau j}{}^{DH}$ is the demand of HCS consumers, i.e., $Q_{\tau j}{}^D$, $j \in J_{DH}$, where J_{DH} is the set of HCS consumers; $Q_{\tau j}{}^{DIG}$ is the demand of industrial consumers connected to heating networks, i.e., $Q_{\tau j}{}^D$, $j \in J_{DIG}$; $Q_{\tau j}{}^{DIN}$ is the demand of industrial consumers located on source collectors, i.e., $Q_{\tau j}{}^D$, $j \in J_{DIN}$. Due to the properties (4), we have:

$$Q_{\tau j}^D = \begin{cases} Q_{\tau j}^{DH} + Q_{\tau j}^{DIG}, & j \in J_{DH} \cap J_{DIG}, \\ Q_{\tau j}^{DH} + Q_{\tau j}^{DIN}, & j \in J_{DH} \cap J_{DIN}, \\ Q_{\tau j}^{DH}, & j \in J_{DH} \setminus (J_{DIG} \cap J_{DIN}), \\ Q_{\tau j}^{DIG}, & j \in J_{DIG} \setminus J_{DH}, \\ Q_{\tau j}^{DIN}, & j \in J_{DIN} \setminus J_{DH}. \end{cases}$$

For any instant τ, the total heat demand from HCS consumers, $Q_{\tau j}{}^{DH}$, is the demand for heating $Q_{\tau j}{}^{DHH}$ (this value varies during the heating period depending on the ambient temperature) and the demand for hot water supply $Q_{\tau j}^{DHW}$ (a fixed value during the year):

$$Q_{\tau j}^{DH} = Q_{\tau j}^{DHH} + Q_{\tau j}^{DHW}, \quad j \in J_{DH}.$$

The demand for heating needs from HCS consumers is determined using the Rossander equation [17].

The market principles of Russia's energy industry, particularly in district heating systems, led to the appearance of a new variable (the price of heat energy). From the economic point of view, demand for any good or service is characterized by a demand curve (the demand-price dependence).

Most district heating markets have inelastic demand: during the heating season, demand in the heat energy markets is a fixed value, being therefore rigidly bound to a particular system. Heat prices are not adjusted in real time depending on the heat source load but are calculated and approved for the medium or long term. At the same time, a market of modern and efficient equipment of small and medium capacity (usually boiler houses) has already formed in heat supply. With its expansion, an economically feasible approach is often to build heat sources (on the sites of industrial objects or individual heat sources in private residential buildings) of small capacity with acceptable one-time costs and minimum payback period. As a result, the demand for heat energy may become elastic.

Consumer behavior in the industrial sector is modeled using the inverse demand function. As a rule, this function is constructed based on real calculations for separate industrial consumers by approximating retrospective data considering the forecasted heat consumption volumes and prices. For industrial consumers connected to heating networks, the inverse demand function $w_{\tau j}^{DIG} = \Phi^{-1}(Q_{\tau j}^{DIG})$, where $w_{\tau j}^{DIG}$ is the consumer's purchase price for the heat volume $Q_{\tau j}^{DIG}$ (in RUB/Gcal), has a linear form [17].

For industrial consumers located on heat source collectors, the inverse demand function takes the form

$$w_{\tau j}^{DIN} = \xi_j^{DIN} - \vartheta_j^{DIN} Q_{\tau j}^{DIN}, \ j \in J_{DIN},$$

with the following notations: ξ_j^{DIN} and ϑ_j^{DIN} are the constants obtained by approximating the actual heat volume purchased by industrial enterprise $j \in J_{DIN}$ depending on its price; $w_{\tau j}^{DIN}$ is the purchase price determined only by the cost of heat production (in RUB/Gcal); $Q_{\tau j}^{DIN}$ is the heat volume purchased by industrial enterprise $j \in J_{DIN}$ located on heat source collectors (in Gcal/h).

Fluctuations in heat demand depending on the time of day and weather conditions are among the main problems of the heat energy market. Therefore, it is proposed to study the interaction between producers and consumers during each hour on a given period. Such a discrete modeling approach is practically important due to considering both the daily and seasonal heat demand factors. These factors significantly affect the volumes of heat demand and production for each heat source (consequently, its profit).

1.4. Modeling of heating networks

Unlike the competitive heat energy market model (optimization-based statement) [17], the mathematical model describing heating networks with the UHSO is a system of linear and nonlinear equations. The material balance (the first Kirchhoff law), acting as a constraint in the cost optimization of heating networks [17], is supplemented by the second Kirchhoff law (5) and closing relations (6):

$$y_\tau - A^{\mathrm{T}} p_\tau = 0, \tag{5}$$

$$y_{\tau i} - s_i x_{\tau i} \left| x_{\tau i} \right| = -H_{\tau i}, \ i \in I. \tag{6}$$

Formulas (5) and (6) have the following notations: $y_\tau = (y_{\tau 1}, ..., y_{\tau n})^{\mathrm{T}}$, where $y_{\tau i}$ is the pressure drop on network edge i at an instant τ, in mAq; A^{T} is the transposed incidence matrix; $p_\tau = (p_{\tau 1}, ..., p_{\tau m})^{\mathrm{T}}$, where $p_{\tau j}$ is the pressure in node j at an instant τ, in mAq; s_i is the hydraulic resistance coefficient of all branches on network edge i, in mh^2/t^2; $x_\tau = (x_{\tau 1}, ..., x_{\tau n})^{\mathrm{T}}$, where $x_{\tau i}$ is the flow rate of the heat-transfer agent on network edge i at an instant τ, in t/h; finally, $H_{\tau i}$ is the effective head on network edge i at an instant τ, in mAq.

The vectors x_τ, y_τ, and p_τ are the variables determining the optimal flow distribution in the heating network. If the material balance

$$\sum_{j \in J_G} Q_{\tau j}^{G} = \sum_{j \in J_{DH}} Q_{\tau j}^{DH} + \sum_{j \in J_{DIG}} Q_{\tau j}^{DIG} + \sum_{j \in J_{DIN}} Q_{\tau j}^{DIN}$$

holds at each instant $\tau \in$ T, the system of equations has a solution (x_τ^*, y_τ^*, p_τ^*) [18]. This solution will always be unique in the variables (x_τ^*, y_τ^*) and nonunique in the variables p_τ^*. The uniqueness in p_τ^* can be ensured by fixing the pressure in one node [18]. The optimal flow distribution in heating networks is obtained by solving the system of linear and nonlinear equations. The corresponding methodology is well developed [18] and can be applied without considerable difficulties.

However, finding the optimal flow distribution in heating networks becomes complicated (in comparison with the traditional technical and economic calculation) in market conditions: the fixed loads $Q_{\tau j}^{DH}$ of

HCS consumers are the only available data in the material balance relations. The heat volumes $Q_{\tau j}{}^{G}$ produced by heat sources and the loads $Q_{\tau j}{}^{DIG}$ and $Q_{\tau j}{}^{DIN}$ of industrial consumers are unknown. Therefore, the problem becomes underspecified.

Assume that the heat production prices $w_\tau{}^{H}$ = $\{w_{\tau j}{}^{HE}: j\in J_G\}$ are given. In this case, formula (3) yields the production volume for each source,

$$Q_{\tau j}{}^{G} = Q_{\tau j}{}^{G,*}(w_{\tau j}{}^{HE}), j \in J_G,$$

and the total heat supply $\sum_{j\in J_G} Q_{\tau j}^{G} = \sum_{j\in J_G} Q_{\tau j}^{G,*}(w_{\tau j}^{HE})$.
Hence, the heat volume $Q_\tau{}^{TSI}(w_\tau{}^{HE})$ offered to the industrial sector is calculated as

$$Q_\tau^{TSI}(w_\tau^{HE}) = \sum_{j\in J_G} Q_{\tau j}^{G,*}(w_{\tau j}^{HE}) - \sum_{j\in J_{DH}} Q_{\tau j}^{DH} \ . \tag{7}$$

Finally, we have to distribute the total supply $Q_\tau{}^{TSI}(w_\tau{}^{HE})$ among the industrial heat consumers, i.e., determine the nodal values $Q_{\tau j}{}^{DIN}$ and $Q_{\tau j}{}^{DIG}$ satisfying

$$\sum_{j\in J_{DIG}} Q_{\tau j}^{DIG} + \sum_{j\in J_{DIN}} Q_{\tau j}^{DIN} = Q_\tau^{TSI}(w_\tau^{HE}) \,.$$

To find them, we apply an approach based on redundant circuits of district heating systems [18]. A redundant circuit is constructed from the design one by introducing a dummy node with number $(m + 1)$ and dummy arcs outgoing from nodes $j \in J_{DIN} \cap J_{DIG}$ (the locations of industrial consumers) and incoming to the dummy node $(m + 1)$. The dummy arcs are assigned numbers from $(n + 1)$ to $(n + r)$, where r denotes the number of industrial consumers (the elements of the set $J_{DIN} \cup J_{DIG}$). The dummy node is assigned the total demand of industrial consumers $Q_\tau{}^{TSI}(w_{\tau j}{}^{HE})$ (7), and the demand of industrial consumers in nodes $j \in J_{DIN} \cup J_{DIG}$ is set to zero. The heat flow from each node $j \in J_{DIN} \cup J_{DIG}$ to node $(m + 1)$ is denoted by $x_{\tau\eta(j)}$, where $\eta(j) \in \{n + 1,..., n + r\}$ is the number of the dummy arc outgoing from node j. The resistances and heads of the dummy arcs are also assumed to be zero: $s_{\eta(j)} = 0$ and $H_{\tau\eta(j)} = 0, j \in J_{DIN} \cup J_{DIG}$. Due to the closure relations (6), the pressure drops are also zero: $y_{\tau\eta(j)} = 0$, $j \in J_{DIN} \cup J_{DIG}$. Next, formula (5) yields $(A)^{\mathrm{T}}{}_{\tau\eta(j)}\, p_\tau = 0$, where $(A)^{\mathrm{T}}{}_{\tau\eta(j)}$ is the $\eta(j)$th row of the transposed incidence matrix. The relations $(A)^{\mathrm{T}}{}_{\tau\eta(j)}\, p_\tau = 0, j \in J_{DIN} \cup J_{DIG}$, are equivalent to $p_{\tau j} = p_{\tau(m+1)}$, $j \in J_{DIN} \cup J_{DIG}$ (the pressure in the industrial sector nodes equals the pressure in the dummy node). As mentioned above, it suffices to fix the pressure in one DHS nodes to ensure the unique solution in the variable p_τ . Without loss of generality, we traditionally [18] let $p_{\tau(m+1)} = 0$. Then the analogs of equations (5) and (6) corresponding to dummy branches disappear. The final system of equations of the redundant circuit with the balance (material) relations (the first Kirchhoff law (8)–(16)) takes the form

$$A_j x_\tau + x_{\tau\eta(j)} = Q_{\tau j}^{G,*}(w_{\tau j}^{HE}) - Q_{\tau j}^{DH}\,,\ j \in J_{DH} \bigcap J_{DIN}\,, \tag{8}$$

$$A_j x_\tau = Q_{\tau j}^{G,*}(w_{\tau j}^{HE}) - Q_{\tau j}^{DH}\,, j \in J_G \bigcap J_{DH} \setminus J_{DIN}\,, \tag{9}$$

$$A_j x_\tau + x_{\tau\eta(j)} = Q_{\tau j}^{G,*}(w_{\tau j}^{HE}), j \in J_{DIN} \setminus J_{DH}\,, \tag{10}$$

$$A_j x_\tau = Q_{\tau j}^{G,*}(w_{\tau j}^{HE}), j \in J_G \setminus J_{DH} \bigcup J_{DIN}\,, \tag{11}$$

$$A_j x_\tau = -Q_{\tau j}^{DH}\,, j \in J_{DH} \setminus (J_G \bigcup J_{DIG}), \tag{12}$$

$$A_j x_\tau + x_{\tau\eta(j)} = 0, j \in J_{DIG} \setminus J_{DH}\,, \tag{13}$$

$$A_j x_\tau + x_{\tau\eta(j)} = -Q_{\tau j}^{DH}\,, j \in J_{DH} \setminus J_{DIG}\,, \tag{14}$$

$$A_j x_\tau = 0, j \in J_0, \tag{15}$$

$$-\sum_{j\in J_{DIN}\bigcup J_{DH}} x_{\tau\eta(j)} = -Q_\tau^{TSI}(w_{\tau j}^{HE}), \tag{16}$$

$$y_\tau - \mathrm{A}^{\mathrm{T}} p_\tau = 0, \tag{17}$$

$$y_{\tau i} - s_i x_{\tau i}\left|x_{\tau i}\right| = -H_{\tau i},\ i \in I. \tag{18}$$

Equation (16) corresponds to the dummy node. Due to the balance relation (7), this system always has a solution ($x_\tau^*(w_\tau^{HE})$, $y_\tau^*(w_\tau^{HE})$, $p_\tau^*(w_\tau^{HE})$). The heat volumes of industrial consumers are given by

$$\begin{aligned} Q_{\tau j}^{DIG}(w_{\tau j}^{HE}) &= \overset{*}{x}_{\tau\eta(j)}(w_\tau^{HE}),\ j \in J_{DIG}, \\ Q_{\tau j}^{DIN}(w_{\tau j}^{HE}) &= \overset{*}{x}_{\tau\eta(j)}(w_\tau^{HE}),\ j \in J_{DIN}. \end{aligned} \tag{19}$$

We illustrate the method of redundant circuits on an example of the district heating system in Fig. 2*a*. In Figs. 2*a* and 2*b*, the numbers indicate the nodes, and the numbers in circles indicate the branches. The other notations are as follows: $Q_1{}^{G}$ and $Q_5{}^{G}$ are the loads of heat sources; $Q_1{}^{DIN}$ is the load of the consumer located on the heat source collector; $Q_3{}^{DIG}$ is the load of the industrial consumer connected to the heating networks; $Q_2{}^{DH}$ and $Q_4{}^{DH}$ are the loads of HCS consumers; $Q_6{}^{TSI}$ is the total load of industrial consumers. The district heating system (Fig. 2*a*) consists of m = 5 nodes and n = 8 branches: J = {1, 2, 3, 4, 5} and I = {1, 2, 3, 4, 5, 6, 7, 8}. The correspondence between the branches and nodes is presented in Table 2.

For the circuit in Fig. 2*a*, the incidence matrix is

$$A = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & -1 & -1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

We have the following sets of nodes: the heat sources, J_G = {1, 5}; the HCS consumers, J_{DH} = {2, 4}; the industrial consumers without heat sources, J_{DIG} = {3}; the industrial consumers with heat sources, J_{DIN} = {1}; simple branching nodes, $J_0 = \varnothing$. For the

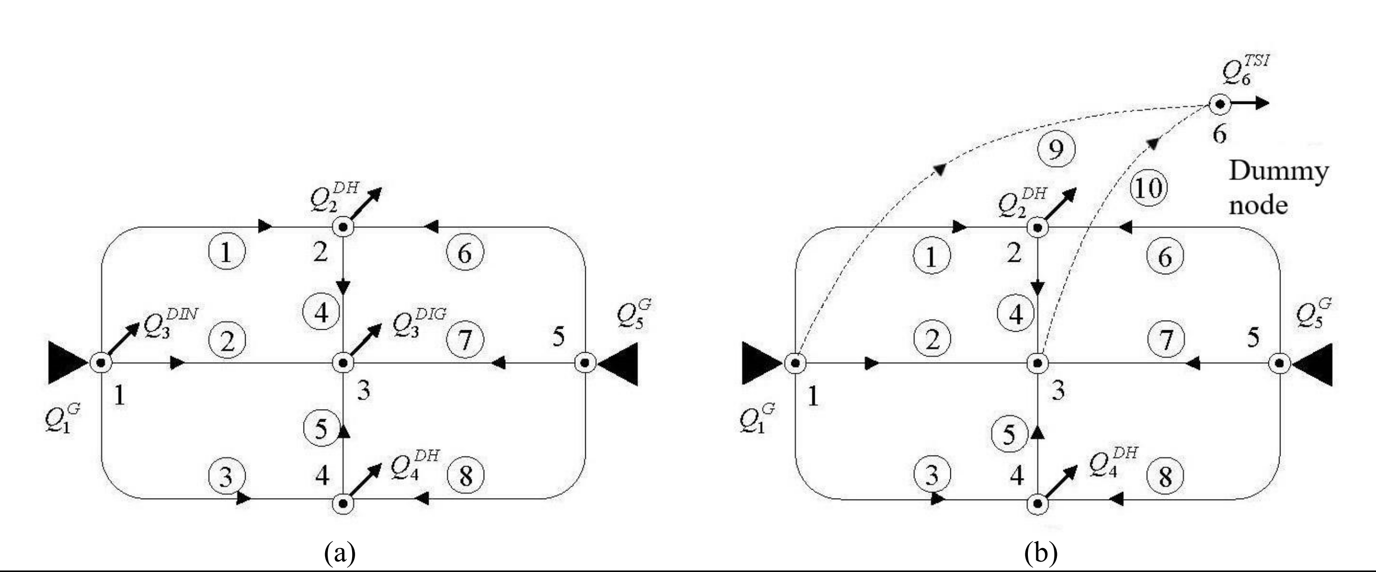

Fig. 2. A heat supply circuit: (a) design and (b) redundant.

Table 2

Correspondence between branches and nodes

	Branch 1	Branch 2	Branch 3	Branch 4	Branch 5	Branch 6	Branch 7	Branch 8
Initial node	1	1	1	2	4	5	5	5
End node	2	3	4	3	3	2	3	4

heat supply circuit in Fig. 2*a*, the first Kirchhoff law takes the form

$$x_{\tau 1} + x_{\tau 2} + x_{\tau 3} = Q_{\tau 1}^{G} - Q_{\tau 1}^{DIN},$$

$$-x_{\tau 1} + x_{\tau 4} - x_{\tau 6} = -Q_{\tau 2}^{DH},$$

$$-x_{\tau 2} - x_{\tau 4} - x_{\tau 5} - x_{7} = -Q_{\tau 3}^{DIG},$$

$$-x_{\tau 3} + x_{\tau 5} - x_{\tau 8} = -Q_{\tau 4}^{DH},$$

$$x_{\tau 6} + x_{\tau 7} + x_{\tau 8} = Q_{\tau 5}^{G}.$$

Let the heat volumes $Q_{\tau 1}{}^{G}$ and $Q_{\tau 5}{}^{G}$ and the demands $Q_{\tau 2}{}^{DH}$ and $Q_{\tau 4}{}^{DH}$ be known. Then the total demand from industrial consumers is given by

$$Q_{\tau}^{TSI} = Q_{\tau 1}^{DIN} + Q_{\tau 3}^{DIG} = Q_{\tau 1}^{G} + Q_{\tau 5}^{G} - Q_{\tau 2}^{DH} - Q_{\tau 4}^{DH}.$$

Assume that $Q_{\tau}{}^{TS} > 0$. (Otherwise, the situation becomes trivial.) The scheme in Fig. 2*a* is represented by two industrial consumers, $J_{DIN} \cup J_{DIG} = \{1, 3\}$, $r = 2$. We introduce the dummy node 6, the dummy branch 9 connecting nodes 1 and 6, and the dummy branch 10 connecting nodes 3 and 6, with $\eta(1) = 9$ and $\eta(3) = 10$ (Fig. 2*b*). In this case, the first Kirchhoff law for the redundant circuit is the system of equations

$$x_{\tau 1} + x_{\tau 2} + x_{\tau 3} + x_{\tau 9} = Q_{\tau 1}^{G},$$

$$-x_{\tau 1} + x_{\tau 4} - x_{\tau 6} = -Q_{\tau 2}^{DH},$$

$$-x_{\tau 2} - x_{\tau 4} - x_{\tau 5} - x_{7} + x_{10} = 0,$$

$$-x_{\tau 3} + x_{\tau 5} - x_{\tau 8} = -Q_{\tau 4}^{DH},$$

$$x_{\tau 6} + x_{\tau 7} + x_{\tau 8} = Q_{\tau 5}^{G}.$$

$$-x_{\tau 9} - x_{\tau 10} = -Q_{\tau 6}^{TSI}.$$

We solve the system of equations of the optimal flow distribution in the heating network. Its solution—the flow vector $x_{\tau}^{*}(w_{\tau j}^{HE})$—corresponds to the minimum heat transport costs in the network [18]. Then, the heat volume supplied to industrial consumers (see the expression (7)) is determined by minimizing the network costs under the optimal flow distribution in the heating network. For details, see the book [18]. Industrial consumers will pay for their heat volumes at the prices determined by the inverse demand functions $w_{\tau j}^{DIN} = \Phi^{-1}(Q_{\tau j}^{DIN})$ and $w_{\tau j}^{DIG} = \Phi^{-1}(Q_{\tau j}^{DIG})$.

The network costs are calculated using an original formula from [17] after finding the steady-state flow distribution.

1.5. Management model for heat supply to consumers

When constructing a management model for heat supply on a regulated monopoly district heating market, we have to formulate the problem statement. It is required to find a state of the regulated monopoly district heating market in which heat sources collectively produce a total heat volume to cover the demand from consumers and simultaneously obtain the maximum profit under the available capacities of heat sources and the physical and technical limitations of heating networks. In addition, the price of heat for industrial consumers is determined using their inverse demand functions, and the regulator adjusts a fair heat energy tariff for HCS consumers.

To model the regulator's behavior, we formalize its objective. Assume that the regulator defends the interests of HCS consumers, seeking to minimize their heat energy tariff w_τ^{DH}. The economic balance of the DHS is written as

$$\sum_{j\in J_G}\left(w_{\tau j}^{HE}Q_{\tau j}^{G}+w_j^{P}\overline{Q}_j^{G}\right)+Z_\tau^{NET}(x_\tau)=$$
$$w_\tau^{DH}\sum_{j\in J_{DH}}Q_{\tau j}^{DH}+\sum_{j\in J_{DIG}}w_{\tau j}^{DIG}Q_{\tau j}^{DIG}+\sum_{j\in J_{DIN}}w_{\tau j}^{DIN}Q_{\tau j}^{DIN}, \quad (20)$$

where $Z_\tau^{NET}(x_\tau)$ denotes the costs of the heating network (in rubles); see the formula in [17].

We determine the total demand from the balance relations (20). The price of heat for HCS consumers is expressed as an affine function of the other variables:

$$w_\tau^{DH}=f_\tau\left(w_\tau^{HE},\, Q_\tau^{G},\, Z_\tau^{NET},\, w_\tau^{DIG},\, Q_\tau^{DIG},\, w_\tau^{DIN},\, Q_\tau^{DIN}\right)=$$
$$\frac{1}{\sum_{j\in J_{DH}}Q_{\tau j}^{DH}}\left[\sum_{j\in J_G}\left(w_{\tau j}^{HE}Q_{\tau j}^{G}+w_j^{P}\overline{Q}_j^{G}\right)+Z_\tau^{NET}(x_\tau)-\right.$$
$$\left.\sum_{j\in J_{DIG}}w_{\tau j}^{DIG}Q_{\tau j}^{DIG}-\sum_{j\in J_{DIN}}w_{\tau j}^{DIN}Q_{\tau j}^{DIN}\right]. \quad (21)$$

The heat sources, heating network, and industrial consumers are guided by their economic interests and are not directly subordinate to the regulator. As described above, the heat sources maximize their profits; the heating network optimizes the heat transport costs from sources to consumers by calculating the optimal flow distribution in it; the industrial consumers behave in accordance with their inverse demand functions. The price vector w_τ^{DH} defines the behavior of heat sources, heating networks, and industrial consumers. Therefore, by varying the price vector w_τ^{HE}, the regulator can minimize the heat energy tariff w_τ^{DH} in (21) considering the interests of heat sources, heating networks, and industrial consumers.

These considerations lead to the following two-level model. For each instant $\tau \in T$, the regulator (the upper level) solves the problem

$$\begin{aligned} w_\tau^{DH} &= f_\tau\left(w_\tau^{HE}, Q_\tau^{G}, Z_\tau^{NET},\right.\\ &\left. w_\tau^{DIG}, Q_\tau^{DIG}, w_\tau^{DIN}, Q_\tau^{DIN}\right)\to \min, \end{aligned} \quad (22)$$

$$\underline{w}_\tau^{HE}\le w_\tau^{HE}\le \overline{w}_\tau^{HE},\ j\in J_G. \quad (23)$$

The regulator transmits the price vector w_τ^{HE} to the lower level consisting of heat sources, heating networks, and consumers. Heat sources $j \in J_G$ maximize their profits (1), (2), producing the heat volume $Q^{G}_{\tau j} = Q^{G*}_{\tau j}(w^{HE}_{\tau j})$ (3). The demand for heat energy from the HCS sector is constant and independent of w_τ^{HE}. Given the heat supply from sources and the demand from the HCS sector, the heating network uses the method of redundant circuits to find the optimal flow distribution, determining the network costs and heat volumes

$$Q_\tau^{DIG}=Q_{\tau j}^{DIG}(w_{\tau j}^{HE}),\ Q_\tau^{DIN}=Q_{\tau j}^{DIN}(w_{\tau j}^{HE}),$$

supplied to the industrial consumers (19). Given these volumes, the industrial consumers set the prices in accordance with their reverse demand functions. After that, all variables of the regulator's function f in (22) become known; the regulator determines the heat energy tariff for HCS consumers corresponding to the vector w_τ^{HE}. The interaction between the upper and lower levels of the system is shown in Fig. 3.

We present **the algorithm *FV* (w_τ^{HE})** for calculating the upper-level objective function.

The input is the price vector w_τ^{HE}.

The output is the value $FV(w_\tau^{HE})$.

Step FV.1. Calculating the heat volume $Q_{\tau j}^{G}=Q_{\tau j}^{G,*}(w_{\tau j}^{HE}),\ j\in J_G$, by formula (3).

Step FV.2. Calculating the heat volume $Q_\tau^{TSI}(w_{\tau j}^{HE,k})$ by formula (7).

Step FV.3. Determining the flows $x_\tau^k=x_\tau^*(w_\tau^{HE,k})$ and $\tilde{x}_\tau^k=\tilde{x}_\tau^*(w_\tau^{HE,k})$ from the system of equations (8)–(18).

Step FV.4. Calculating the network costs Z_τ^{NET} as described in [17].

Step FV.5. Determining the heat volumes Q_τ^{DIG} and Q_τ^{DIN} by formulas (19).

Step FV.6. Calculating the prices w_τ^{DIG} and w_τ^{DIN} using the inverse demand functions.

Step FV.7. Calculating the value $FV(w_\tau^{HE})=f_\tau\left(w_\tau^{HE}, Q_\tau^{G}, Z_\tau^{NET}, w_\tau^{DIG}, Q_\tau^{DIG}, w_\tau^{DIN}, Q_\tau^{DIN}\right)$ by formula (22).

With the function *FV*, calculated by this algorithm, problem (22), (23) can be reformulated as follows:

$$FV(w_\tau^{HE})\to \min,$$
$$\underline{w}_\tau^{HE}\le w_\tau^{HE}\le \overline{w}_\tau^{HE},\ j\in J_G.$$

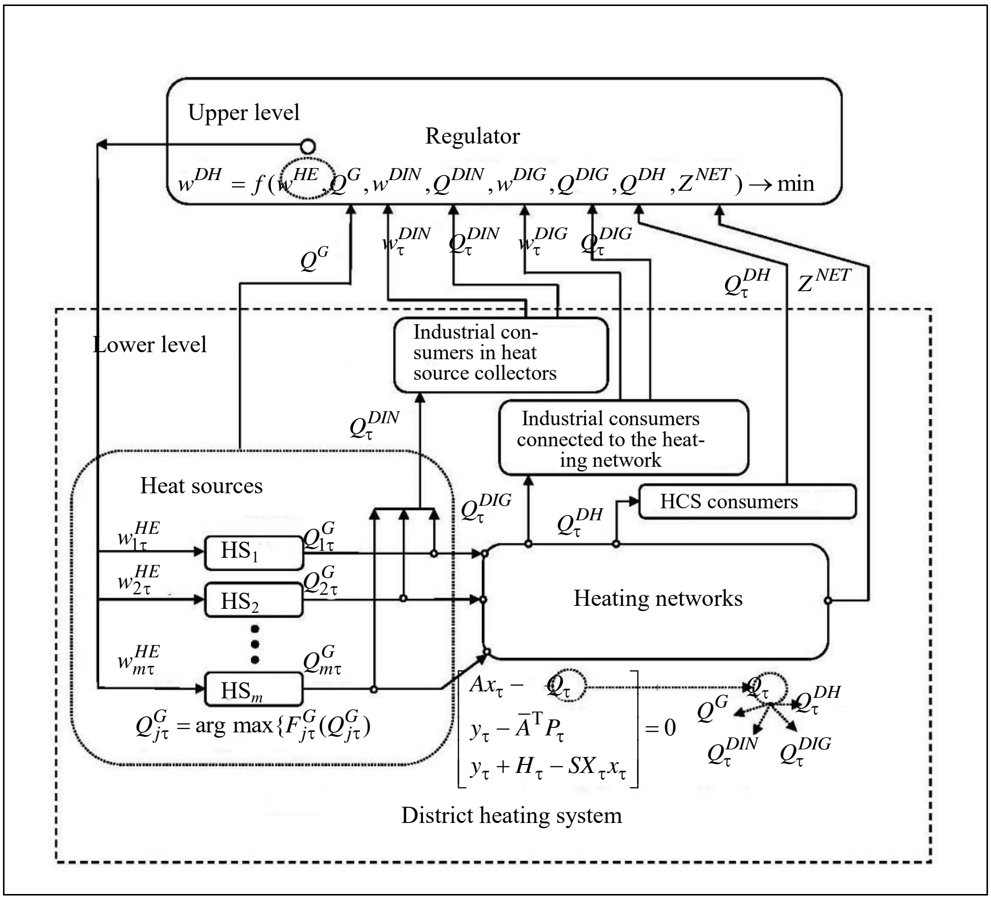

Fig. 3. Input and output parameters in the two-level management heat supply to consumers.

The proposed algorithm for minimizing the objective function (22) is an adaptation of the coordinate descent method to the problem under consideration [19].

A graphical interpretation of this algorithm as a step-by-step computational process is presented in Fig. 4.

In particular, Fig. 4*a* illustrates the change in the heat price function for HCS consumers depending on the heat production price of heat sources in a district heating system with two sources (Fig. 2). Figure 4*b* shows the contour lines of the objective function (the heat price domain for HCS consumers under different values of the heat energy production prices of heat sources).

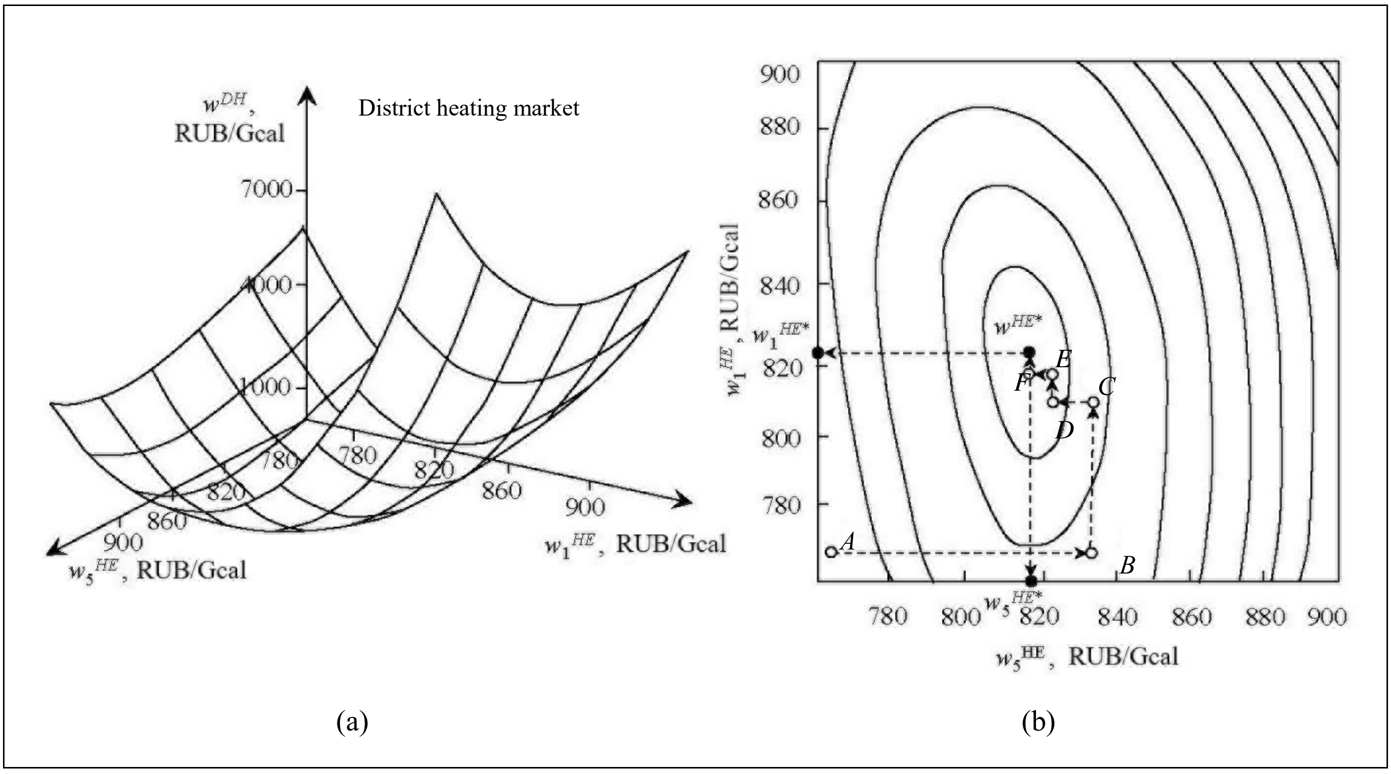

Fig. 4. A graphical interpretation of the computational algorithm.

Each heat source sequentially optimizes its heat production price under a fixed heat production price of the other heat source. The optimal heat production price of heat sources that minimizes the heat price for HCS consumers in this district heating system is achieved at the fifth iteration (see the point $w^{HE}*$). The broken line A-B-C-D-E-F-$w^{HE}*$ is the trajectory of the computational process (Fig. 4*b*).

2. APPLICATION TO A REAL DISTRICT HEATING SYSTEM

The proposed management model was applied to the district heating system of Angarsk. In an enlarged form, this system consists of 1273 edges and 1242 nodes (Fig. 5).

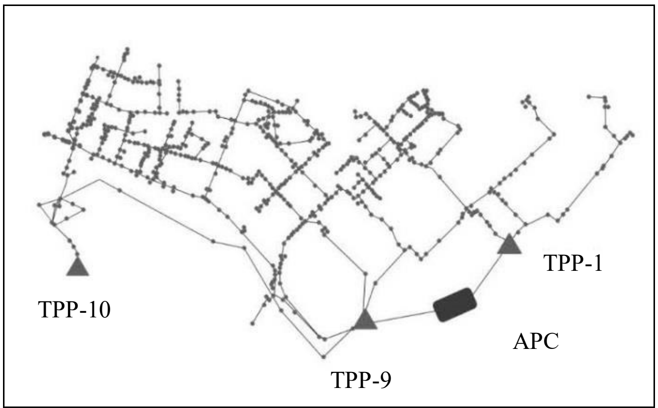

Fig. 5. The heat supply circuit of Angarsk.

The number of generalized consumers is represented by 534 nodes, of which 533 nodes correspond to consumers with fixed heat loads (HCS consumers) and one node to PJSC Angarsk Petrochemical Company (APC), located on the collectors of TPP-1 and TTP-9. We used the following initial data for calculations: the heat supply circuit of Angarsk with the technical characteristics of heating network edges (diameters, lengths, and resistances); the locations of heat sources in the heating system; the cost functions of heat sources; the temperature graph and climatic characteristics of the region; the estimated heat loads of HCS consumers; the heat demand function of PJSC APC.

The calculations were performed for a one-year time interval with a step of one hour. The calculated integral indices of the heating system of Angarsk are combined in Table 3.

According to Table 3, when supplying heat within the UHSO, TPP-9 covers 57.3% of the total heat load, and the shares of TPP-1 and TPP-10 are 22.6% and 20.1%, respectively. The main heat consumer in Angarsk is the HCS sector (69.3% of all heat produced by the district heating system). The remaining 30.7% of heat is consumed by PJSC APC.

Table 3

Estimated integral technical-and-economic indices of the heating system of Angarsk

Indicator	Value
Heat production volume, in million Gcal, including:	**6.85**
TPP-1	1.55
TPP-9	3.92
TPP-10	1.38
Heat production (fuel) costs, in billion RUB, including:	**3.30**
TPP-1	0.78
TPP-9	1.91
TPP-10	0.61
Semi-fixed (operating) costs, in billion RUB, including:	**2.19**
TPP-1	0.43
TPP-9	1.05
TPP-10	0.71
The cost of heat production, in RUB/Gcal:	
TPP-1	647.6
TPP-9	646.8
TPP-10	649.7
The price per unit of heat energy, in RUB/Gcal	
TPP-1	29.4
TPP-9	36.0
TPP-10	42.3
Profit, in billion RUB, including:	**1.02**
TPP-1	0.22
TPP-9	0.62
TPP-10	0.18
Costs of heating networks, in billion RUB	1.22
Heat price for HCS consumers, in RUB/Gcal	862.7
Heat price for PJSC APC, in RUB/Gcal	1350.5
Heat consumption by PJSC APC, in million Gcal	2.10
Heat consumption by HCS consumers, million Gcal	4.75

The average annual minimum tariff for HCS consumers is 862.7 RUB/Gcal (excluding VAT); for PJSC APC, 1350.5 RUB/Gcal. The total sales proceeds of the UHSO amount to 5.65 billion RUB, and the total profit of heat sources for the period is 1.22 billion RUB (0.22 billion RUB for TPP-1, 0.62 billion RUB for TPP-9, and 0.18 billion RUB for TPP-10). Heat transport costs (heating network costs) amount to 1.22 billion RUB (about 178.1 RUB/Gcal).

CONCLUSIONS

This paper has presented:

– new problem statements on developing district heating systems within the organizational regulated monopoly market model;

– requirements to mathematical models and methods used to solve them;

– the conceptual interaction of participants in the process of heat supply to consumers in the form of a hierarchical vertically integrated system.

As a result, a mathematical model of the district heating system within the two-level management system has been developed. In this hierarchical system, the regulator adjusts tariffs for HCS consumers, industrial consumers buy heat in accordance with their demand functions, and heat sources cover a given total demand from consumers to obtain the maximum profit. An optimization criterion has been proposed for the regulator. The two-level management approach has been applied to the real district heating system of Angarsk.

The proposed two-level mathematical model reflects well the real conditions in the local heat energy markets. This model reasonably considers the established "rules of play" in the heat energy market as well as the physical and technical and economic limitations of district heating systems.

REFERENCES

1. Frederiksen, S. and Werner, S., *District Heating and Cooling*, Lund: Studentlitteratur, 2013.
2. Bashmakov, A.I., Analysis of the Main Development Trends of Heating Systems in Russia, *Novosti Teplosnabzheniya*, 2008, no. 2, pp. 6–9. (In Russian.)
3. Wenren, S., District Heating and Cooling in Sweden, *Energy*, 2013, no. 126, pp. 419–429.
4. Belyaev, L., *Electricity Market Reforms: Economics and Policy Challenges*, New York: Springer-Verlag, 2011.
5. Stoft, S., *Power System Economics: Designing Markets for Electricity*, Wiley-IEEE Press, 2002.
6. Paiho, S. and Saastamoinen, H., How to Develop District Heating in Finland?, *Energy Policy*, 2018, no. 122, pp. 668–676.
7. Biggara, D. and Soderberg, M., Empirical Analysis of How Political Ideology and Ownership Influence Price Stability in the Swedish District Heating Market, *Energy Policy*, 2020, no. 145, pp. 1–7.
8. Osman, N., Barriers to District Heating Development in the Netherlands: A Business Model Perspective, *Master Thesis for Business Administration*, Enschede: University of Twente, 2017. URL: https://essay.utwente.nl/73054/1/Osman_BA_BMS.pdf.
9. Wojdyga, K. and Chorzelski, M., Chances for Polish District Heating Systems, *Energy Procedia*, 2017, no. 116, pp. 106–118.
10. *District Heating and Cooling, Combined Heat and Power and Renewable Energy Sources*. URL: http://basrec.net/wp-content/uploads/2014/06/Appendix%20-%20country%20survey.pdf (Accessed February 6, 2021.)
11. Ziemele1, J., Vigants, G., Vitolins, V., Blumberga, D., and Veidenbergs, I., District Heating Systems Performance Analyses. Heat Energy Tariff, *Environmental and Climate Technologies*, 2014, no. 13, pp. 32–43.
12. Sarma, U. and Bazbauers, G., District Heating Regulation: Parameters for the Benchmarking Model, *Energy Procedia*, 2016, no. 95, pp. 401–407.
13. Askeland, K. and Bozhkova, K., District Heating in Norway. An Analysis of Shifting from Individual Electric Heating to District Heating, *Master's Thesis*, Aalborg: Aalborg University, 2017. URL: https://projekter.aau.dk/projekter/files/260084106/Master_thesis_Askeland_Bozhkova.pdf.
14. Šommet, J., Sustainable Development in Estonian Mining, *Environmental and Climate Technologies*, 2013, no. 11, pp. 34–40.
15. Jing, L. and Boqiang, L., Heat Tariff and Subsidy in China Based on Heat Cost Analysis, *Energy Economics*, 2018, no. 71, pp. 411–420.
16. Dempe, S., *Foundations of Bi-level Programming*, Dordrecht: Kluwer Academic Publishers, 2002.
17. Stennikov, V.A., Penkovskii, A.V., and Khamisov, O.V., Search for Cournot Equilibrium on the Heat Energy Market in the Conditions of Heat Sources Competitive Behaviour, *Control Sciences*, 2017, no. 1, pp. 10–18. (In Russian.)
18. Merenkov, A.P., and Khasilev, V.Ya., *Teoriya gidravlicheskikh tsepei* (Theory of Hydraulic Circuits), Moscow: Nauka, 1985. (In Russian.)
19. Shoup, E., *A Practical Guide to Computer Methods for Engineers*, Prentice-Hall, 1979.

This paper was recommended for publication by D.A. Novikov, a member of the Editorial Board.

Received April 9, 2021, and revised April 13, 2022.
Accepted April 13, 2022.

Author information

Stennikov, Valery Alekseevich. Dr. Sci. (Eng.), Melentiev Energy Systems Institute, Siberian Branch, Russian Academy of Sciences, Irkutsk, Russia
✉ SVA@isem.irk.ru

Khamisov, Oleg Valer'evich. Dr. Sci. (Phys.–Math.), Melentiev Energy Systems Institute, Siberian Branch, Russian Academy of Sciences, Irkutsk, Russia
✉ Khamisov@isem.irk.ru

Penkovskii, Andrey Vladimirovich. Cand. Sci. (Eng.), Melentiev Energy Systems Institute, Siberian Branch, Russian Academy of Sciences, Irkutsk, Russia
✉ penkoffsky@isem.irk.ru

Cite this paper

Stennikov, V.A., Khamisov, O.V., Penkovskii, A.V., A Mathematical Model of Managing a Regulated Monopoly District Heating Market. *Control Sciences* **2,** 9–18 (2022). http://doi.org/10.25728/cs.2022.2.2

Translated into English by *Alexander Yu. Mazurov,*
Cand. Sci. (Phys.–Math.),
Trapeznikov Institute of Control Sciences,
Russian Academy of Sciences, Moscow, Russia
✉ alexander.mazurov08@gmail.com

DOI: http://doi.org/10.25728/cs.2022.2.3

REDUNDANCY MANAGEMENT OF ONBOARD EQUIPMENT: AN ARBITRATION APPROACH BASED ON CONFIGURATION SUPERVISORS

A.M. Ageev[1], V.N. Bukov[2], and V.A. Shurman[3]

[1]Zhukovsky and Gagarin Air Force Academy, Voronezh, Russia

[2]Research Institute of Aircraft Equipment, Zhukovsky, Russia

[3]Ramenskoe Instrument-Making Design Bureau (Branch), Zhukovsky, Russia

[1]✉ ageev_bbc@mail.ru, [2]✉ v_bukov@mail.ru, [3]✉ vshurman@rpkb.ru

Abstract. This paper considers the operational reconfiguration of an onboard equipment complex with redundant heterogeneous and non-universal components for achieving fault tolerance and other operational and technical characteristics. An approach to building a redundancy management system (RMS) is formulated as a conceptual solution. This approach uses configuration supervisors according to the number of previously developed competitive configurations of the complex. Each supervisor can be self-sufficient to execute the following functions: monitor the readiness and functional efficiency indicators of the components included in the corresponding configuration; participate in arbitration for the right to implement the corresponding configuration in current conditions; initiate and control the operation of the corresponding configuration. A three-stage algorithm of the RMS is proposed. It contains a sequence of paired arbitrations of computers and configurations. An illustrative example explains RMS operation in different modes of the complex under detectable and undetectable failures of equipment components and computers allocated for redundancy management. The proposed approach unifies and rationalizes the information and logical structure of redundancy management tools, thereby simplifying the creation of an efficient RMS with flexibility, a wide coverage of equipment and configurations, and a significant reduction of bottlenecks in the redundant complex.

Keywords: onboard equipment complex, redundant resource, reconfiguration, redundancy management, configuration supervisor, paired arbitration.

INTRODUCTION

The development of the national aircraft industry must meet the challenges caused by the general global trends in this field as well as the domestic conditions, peculiarities, and goals. Among the main lines of aircraft development [1–4], a prominent place belongs to creating extremely reliable aircraft complexes of onboard equipment (COEs) with wide-range capabilities and high competitiveness in the domestic and world markets. Due to the objectively limited reliability of the element base [5], the single alternative to achieve this effect is creating reconfigurable redundant[1] COEs to ensure fault tolerance using their system properties.

Finding rational ways to design reconfiguration systems for onboard aircraft applications is very topical, as shown by numerous publications in Russia [6–9] and abroad [10–13]. Traditional reconfiguration solutions include FDIR[2] (*Fault-Detection, Fault-Isolation and Recovery Techniques*). These techniques

[1] A redundant technical system has capabilities exceeding those required for its normal operation [5].

[2] URL: http://deacademic.com/dic.nsf/dewiki/422452 (Accessed November 1, 2020.)

imply the redundancy of complete configurations and component-by-component redundancy [7, 8], intended mainly for homogeneous parts of the system (in the sense of their structure and functions). In addition, more advanced approaches are known, e.g., the distribution of functional tasks among the available hardware resources of multiprocessor complexes [9] and the multi-agent interaction of onboard computers based on local dispatchers [14].

However, their common restrictions are narrow orientation toward specific resources managed and a complete focus on fault tolerance. At the same time, users of systems with heterogeneous equipment redundancy are interested in reaching the maximum fault tolerance and, moreover, other operational and technical characteristics preferable in current conditions.

This paper further develops the redundancy management approach [15] to increase its flexibility to heterogeneous and universal equipment components and improve the required integral characteristics of the complex.

1. REDUNDANT RESOURCE CONFIGURATIONS

The redundancy management approach proposed below applies to all components of a redundant COE: informationally separated devices (node, subsystem, system, etc.) or software products (module, application) accessed only through appropriate inputs and outputs. Each component is indivisible in the sense of data flows. The set of components performing the same or interchangeable functions (within the efficiency implemented by a component) forms the corresponding *redundant resource*.

A COE can serve its intended purpose at a given instant only under the following condition: there is a special, minimally sufficient, operable set of its hardware and (or) software components (called a *configuration*). Assume that any operable configuration uses at least one component from each resource.

Component redundancy [7] often has peculiarities; see the following typical cases in Fig. 1:

a) For each resource, there are components with homogeneous interfaces and universal capabilities.

b) Some resource has heterogeneous components. In this case, a barrier arises, insurmountable for intercomponent relations; the replacement of components in one resource requires a corresponding replacement of components in another (adjacent) resource.

c) The non-universal components affect the reconfiguration capabilities of the complex, narrowing the range of admissible alternative configurations.

The fundamental point is that in such circumstances, the choice of alternative (substitute) components can be impossible or inefficient exclusively within a resource. Effective redundancy management may require considering alternatives between resource groups or, in the final analysis, between entire configurations.

2. PROBLEM STATEMENT

Let a redundant COE be represented by a set of K heterogeneous and non-universal informationally separated components. Heterogeneity means different connectivity, whereas non-universality means different functions performed. Due to some conditions, the components are divided into K_i resource groups:

$$k_{i,j} \in K_i, \; K_i \subset K, \; i = \overline{1, N}, \; j = \overline{1, N_i}, \qquad (1)$$

where $k_{i,j}$ denotes the *j*th component in the *i*th group, N is the number of resource groups in the COE, and N_i is the number of components in the *i*th group.

Each component $k_{i,j}$ is assigned a pair of indicators, the readiness index $\mathrm{RI}_{i,j}$ and the functional efficiency indicator $\mathrm{FEI}_{i,j}$. The former's binary value (1–ready, 0–not ready) is an aggregate assessment of the component's readiness (operability and completion of all preparatory procedures) by the monitoring system using built-in check (BiC) means. The latter's real scalar value integrally characterizes the component's potential contribution to the target function of some COE mode. The values of both indicators, $\mathrm{RI}_{i,j}$ and $\mathrm{FEI}_{i,j}$, can be determined a priori or measured for a current instant based on the results of monitoring and external target impact.

In addition, the COE includes an onboard integrated computing environment (OICE), formally not attributed to components, that implements data exchange between them and processes these data by given rules. The computing units of this system, further called computers for simplicity and denoted by v_r, are also redundant:

$$v_r \in V, \qquad r = \overline{1, R}, \qquad (2)$$

where V is the set of computers[3] engaged in the redundancy management of the COE (R in total). Each computer is assigned the index RI_r with a binary value (1–ready, 0–not ready).

[3] Like communication means, computers may have different characteristics. In this case, appropriate modifications are needed for the considerations below.

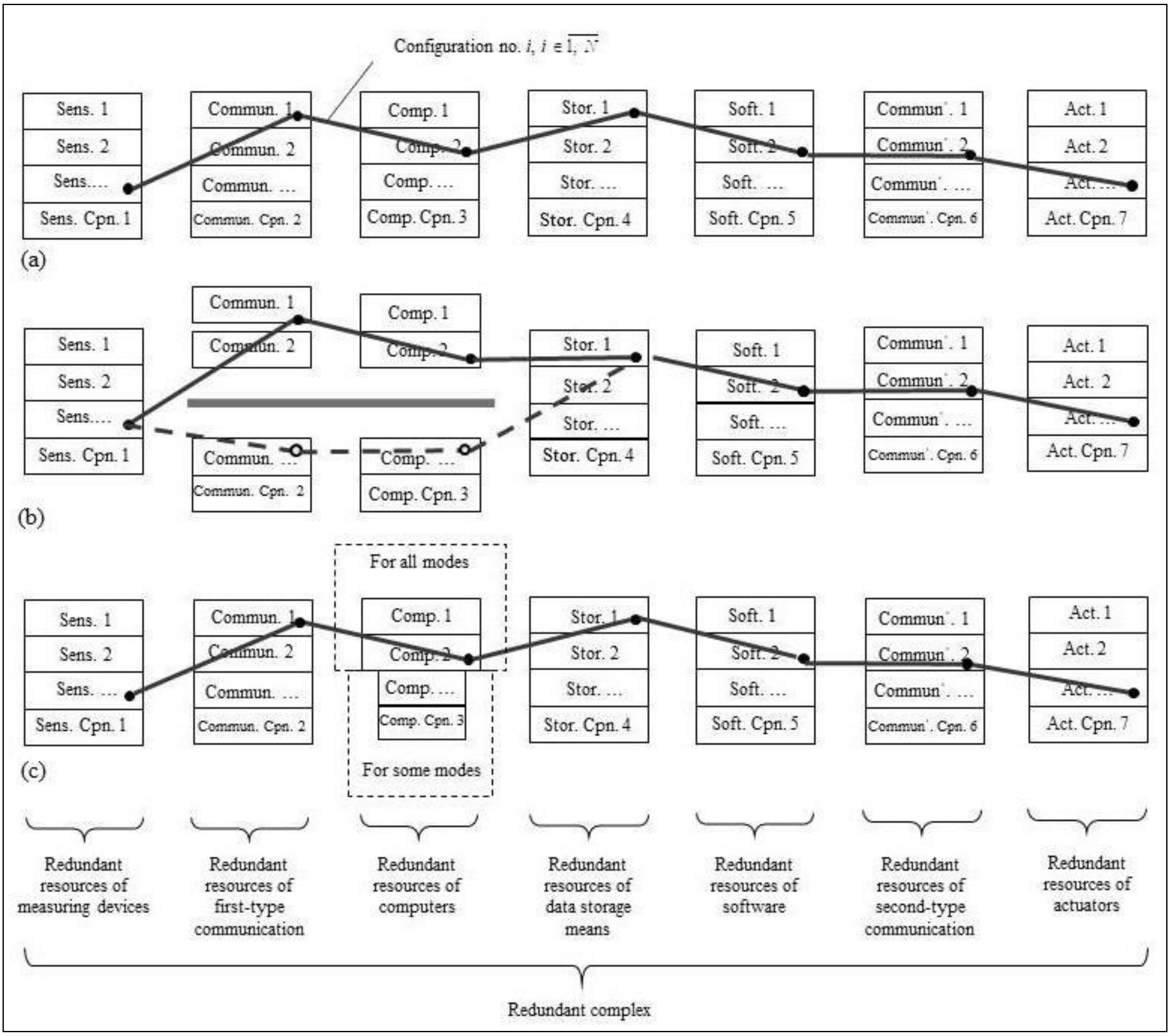

Fig. 1. Typical structure of a redundant complex of equipment: (a) with homogeneous universal components, (b) with heterogeneous components, (c) with non-universal components.

For proper functioning, COE components are combined by OICE means into operable configurations

$$m_q = m_q(k_{1*},\ k_{2*},\ldots,\ k_{N*},\ a_q)\,,$$

$$m_q \in M\ ,\ q = \overline{1,Q}\,, \tag{3}$$

with the following notations: $*$ is the component included in the qth configuration from each resource group; a_q is the aggregate process of data processing by OICE computing means in the qth configuration of the complex; M is a finite set of configurations; Q is the number of different configurations in the set M. Without loss of generality, any COE configuration includes one component from each resource group; the heterogeneity and non-universality of components are overcome by considering all competitive configurations at the RMS design stage.

The corresponding integral indicators are formed for each potential configuration (3) by the following rules:

– The RI of a configuration is the conjunction of the RIs of its components:

$$\mathrm{RI}(m_q) = \mathrm{RI}_{1*} \wedge \mathrm{RI}_{2*} \wedge \ldots \wedge \mathrm{RI}_{N*}\,. \tag{4}$$

– The PFE of a configuration is the result of aggregating the PFEs of its components:

$$\mathrm{FEI}(m_q) = \wp(\mathrm{FEI}_{1*},\ \mathrm{FEI}_{2*},\ldots,\ \mathrm{FEI}_{N*})\,, \tag{5}$$

where $\wp$ is a scalar function. Depending on the characteristics of the components and other conditions, various aggregation methods can be used (summation;

multiplication; choice of typical values; artificial intelligence tools like regularization, normalization, or convolution, including networks with deep learning [16], etc.).

The problem is to develop an approach (a consistent set of algorithms, techniques, and methods) for operationally selecting a preferred element (v_α, m_{opt}) from the direct product $V \times M = \{v_r, m_q\}_{r,q}$ containing:

- a preferred computer v_α (α-computer) for redundancy management, characterized by

$$v_\alpha = \underset{v_r}{\text{opt}} \begin{cases} \text{RI}(v_r) = 1, \\ \text{no preparation errors for configuration } m_{\text{opt}}; \end{cases} \quad (6)$$

- a preferred ready-to-implement configuration m_{opt} with the optimal values of the integral indicators:

$$m_{\text{opt}} = \underset{m_q}{\text{opt}} \begin{cases} \text{RI}(m_q) = 1, \\ \text{FEI}_{\max} = \max\limits_{m_q \in M} \text{FEI}(m_q). \end{cases} \quad (7)$$

According to the second condition in (6), certain attributes are checked to verify that the configuration m_{opt} given by (7) (the corresponding data in the computer v_α) meets the requirements. (In other words, the configuration is validated.)

If there exist several configurations with the same FEI maximum, they are supposed equivalent, and the final choice can be made by some discriminative rules.

The expression (6) and the first equality in (7) mean the fault tolerance of the entire COE; the second equality in (7), its optimal (rational) operational and technical characteristics (FEI). The approach is flexible due to wide possibilities of choosing or constructing the function $\wp$. However, these issues go beyond the scope of the paper.

3. CONFIGURATION SUPERVISORS

The concept of a *configuration supervisor* [15] is central in the proposed approach. Each competitive configuration m_q is assigned a unique configuration supervisor CS_q from the set S bijectively corresponding to the set M:

$$m_q \in M \leftrightarrow s_q \in S.$$

Each supervisor manages its configuration only, performing several tasks:

- It monitors $\text{RI}_{i,j}$ and $\text{FEI}_{i,j}$ (1) for the components within the configuration and all computers (2); it forms $\text{RI}(m_q)$ and $\text{FEI}(m_q)$ for the entire configuration by the expressions (4) and (5).
- It stores information about the configuration: the interconnections of different components, resource allocation, and other implementation data.[4]
- It participates in a periodic competition (arbitration with other supervisors) to identify the preferred configuration m_{opt} by the rule (6); as a result, the corresponding CS acquires the dominant status (DCS): $s_{\text{opt}} \leftrightarrow m_{\text{opt}}$.
- The DSC (winner of the arbitration) performs additional tasks:

– It monitors $\text{RI}(v_r)$ of the computers.

– It executes the arbitration of computers (2) to identify the α-computer v_α among them.

– It validates the configuration m_{opt} corresponding to the DCS s_{opt}.

– It activates and controls the operation of the configuration m_{opt}.

– It coordinates all redundancy management procedures for the COE.

Figure 2 shows the place and role of CSs in the general architecture of the COE.

Thus, each CS is an authorized representative of its configuration. Each CS executes the entire set of functions to assess the readiness of its configuration, store relevant information, represent the configuration in arbitration, and activate it (the winner).

This approach is similar to the multi-agent approach based on a distributed local dispatcher [14]. However, it differs in the following capabilities:

– Achieving fault tolerance not only for computing resources but for all heterogeneous and non-universal equipment of the complex;

– Parrying the errors of configuration readiness monitoring means by the cross-analysis of preference data messages during the paired arbitration[5] of configurations (PAConf);

– Parrying the failures and errors of computers undetectable for the BiC means by the analysis of preference data messages during the paired arbitration[6] of computers (PAComp);

– Achieving various operational and technical indicators simultaneously;

– The self-sufficiency of supervisors (reservation of all redundancy management procedures);

– Unified configuration support tools;

– Unified selection procedures among heterogeneous configurations by structure and composition.

[4] Methods for forming and storing information about configurations were considered in [17].
[5] The paired arbitration of configurations is not described here.
[6] The paired arbitration of computers is not described here.

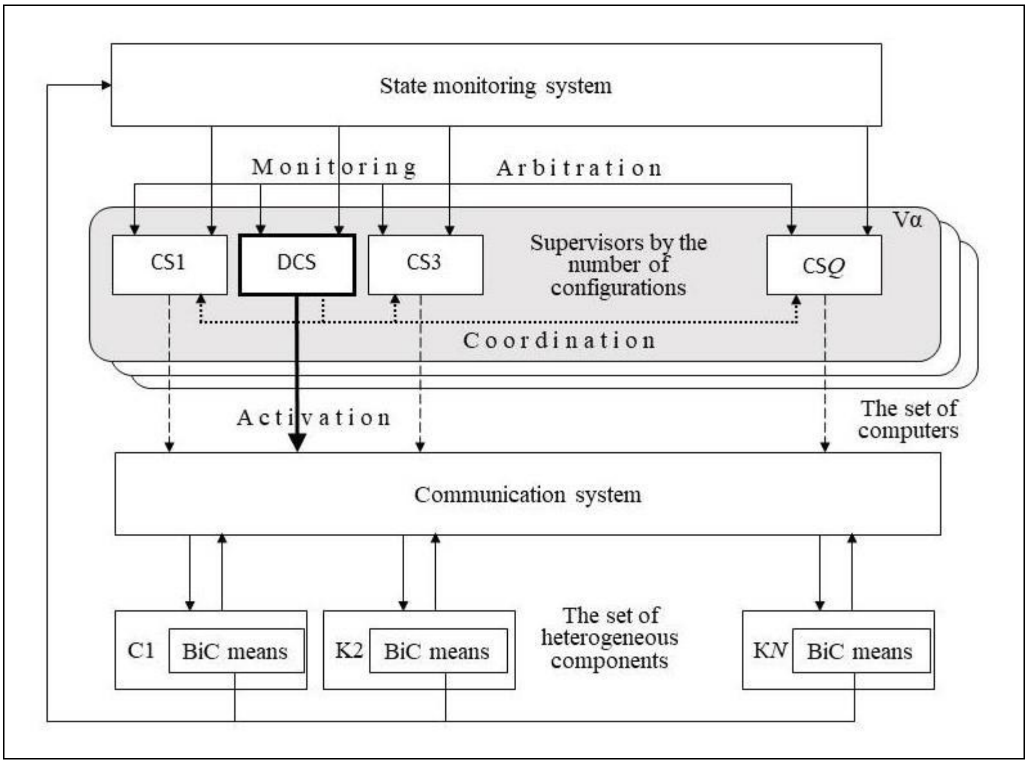

Fig. 2. The place and role of configuration supervisors in the COE architecture.

4. REDUNDANCY MANAGEMENT OF COE

The search domain $V \times K$ contains a finite, possibly significant, number of combinations RQ of computers v_r and configurations m_q. In addition, potential solutions (v_r, m_q) may have different technical characteristics from a large list. Therefore, the refined[7] comparative enumeration of all possible solutions seems to be the most efficient method.

For the scalar criteria $\mathrm{RI}(m_q)$ and $\mathrm{FEI}_{\max}$, the paired arbitration of supervisors has sufficient universality and acceptable complexity. It guarantees the absence of looping and unambiguous determination of the leader (v_α, m_{opt}). The disadvantage of such arbitration—the arrangement of all other solutions (v_r, m_q), $r \neq \alpha$, $q \neq \mathrm{opt}$, without priority—is not critical: the problem statement requires the leader configuration only.

All CSs s_q are placed in each computer v_r, and any computer or configuration has no a priori advantage over the others. Moreover, the rules (6) and (7) are related. Therefore, we propose the following algorithm for the redundancy management system (RMS) of the COE (Fig. 3). Its peculiarity is the sequence of three arbitration stages.

The first stage is the arbitration of computers (phase 1). The α-computer (selected in the previous cycle) or the α-backup (in the case of failure) executes the following functions:

– It forms a group of ready-to-use computers.

– It separates a pair of computers claiming to be α-computers.

– It preliminarily selects a candidate computer (α-candidate) v_{cand} in the separated pair using deterministic, possibly discriminative, rules. The second computer of the pair becomes the α-backup v_{back}.

– It initializes the computer v_{cand}.

The second stage is the arbitration of supervisors. The computer v_{cand} executes the following functions:

– It performs the sequential paired selection of the DCS s_{opt} from the set s_q considering $\mathrm{RI}(m_q)$ and $\mathrm{FEI}(m_q)$ using PAConf procedures.

– It initializes the selected DCS s_{opt} by transferring the control functions.

– It transmits the information about the selected DCS s_{opt} from the computer v_{cand} to the computer v_{back}. This supervisor is renamed to $s_{\mathrm{opt.\,back}}$.

[7] No repetitions or ambiguous steps.

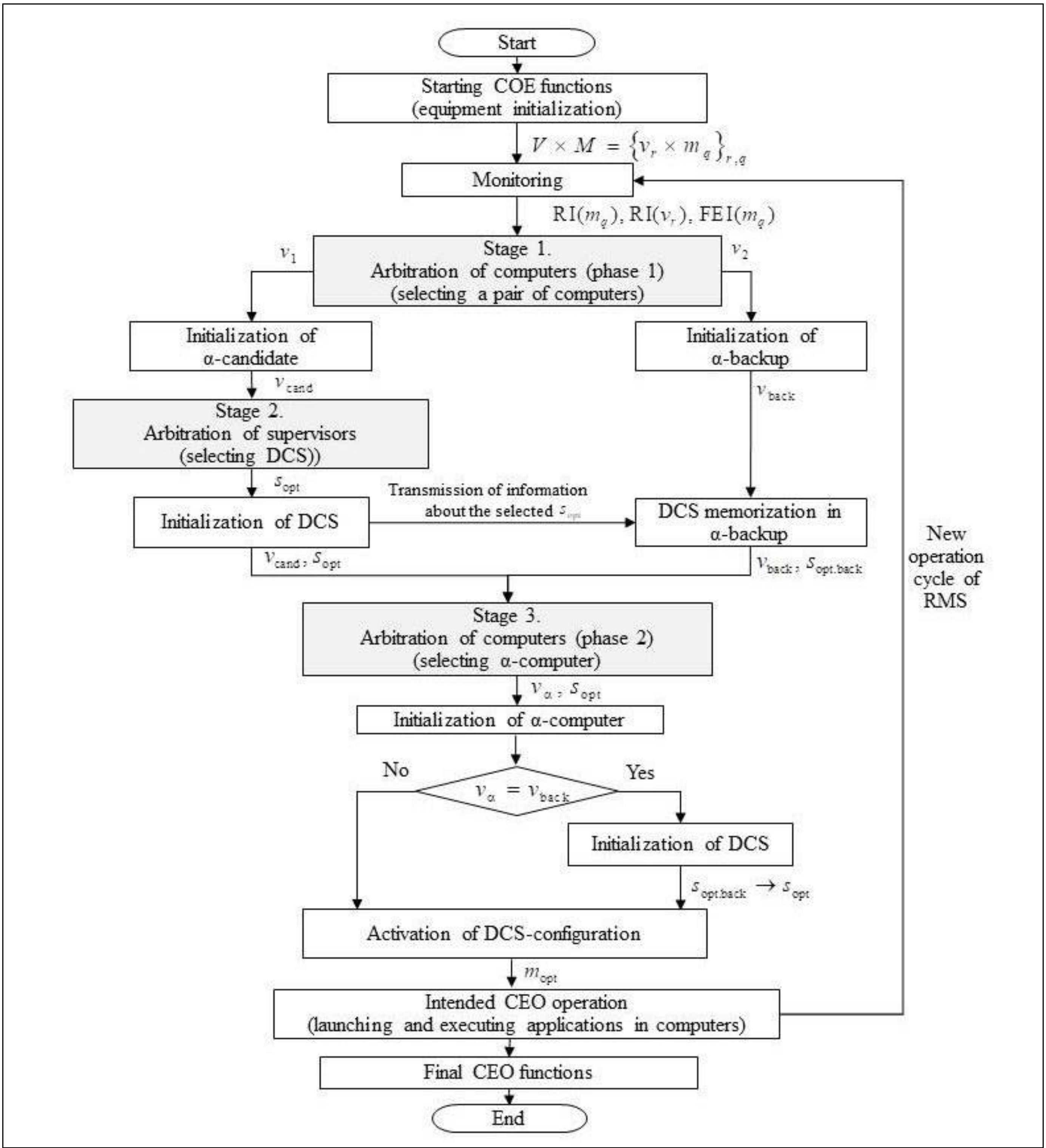

Fig. 3. The algorithm of RMS.

The third stage is the arbitration of computers (phase 2). The computers of the selected pair execute the following functions:

– They perform the final choice of the α-computer v_α based on PAComp procedures: $v_{\text{cand}} \to v_\alpha$ or $v_{\text{back}} \to v_\alpha$.

– In the case $v_{\text{back}} \to v_\alpha$, they initialize the computer v_{back} (transition to the backup computer).

– In the case $v_{\text{back}} \to v_\alpha$, they initialize the DCS $s_{\text{opt. back}} \to s_{\text{opt}}$ in the computer v_{back} (transition to the supervisor in the backup computer).

– They initialize the selected configuration $s_{\text{opt}} \to m_{\text{opt}}$ using the DCS s_{opt} in the α-computer (stop the previous configuration and start the new configuration for the next operation cycle of the COE).

The RMS operates in cycles. On each cycle, arbitration procedures are preceded by monitoring to form RI(m_q) and FEI(m_q) for the configurations in the current operating conditions of the COE. Upon completion of the arbitration and activation of the winning configuration m_{opt}, the COE performs the intended tasks. A new operation cycle of the RMS begins after some time or upon detecting a failure (change of the COE mode).

The loss of a computer or configuration in the current-cycle arbitration (not ready, smaller efficiency) does not mean its elimination from further operation. Hence, the availability of components in another mode or after a failure (reboot) is possible; moreover, a weak configuration can be decisive in a critical situation. The solutions for different failures (including permanent ones) are left to the monitoring and communication system. This issue goes beyond the scope of the paper.

The PAConf and PAComp procedures are based on the mutual cross-analysis of preference data messages (for supervisors and computers, respectively). Therefore, arbitration has required reliability. The failed arbitration objects are determined using the preference matrices generated based on the RI, FEI, and identifiers. These procedures require detailed consideration in a separate paper.

5. AN ILLUSTRATIVE EXAMPLE

Consider a hypothetical COE consisting of six components C1, ..., C6 with three consecutive resource groups. The first two groups have redundancy (two and three components). Figure 4 and the table below demonstrate the corresponding diagram in the five situations explained below.

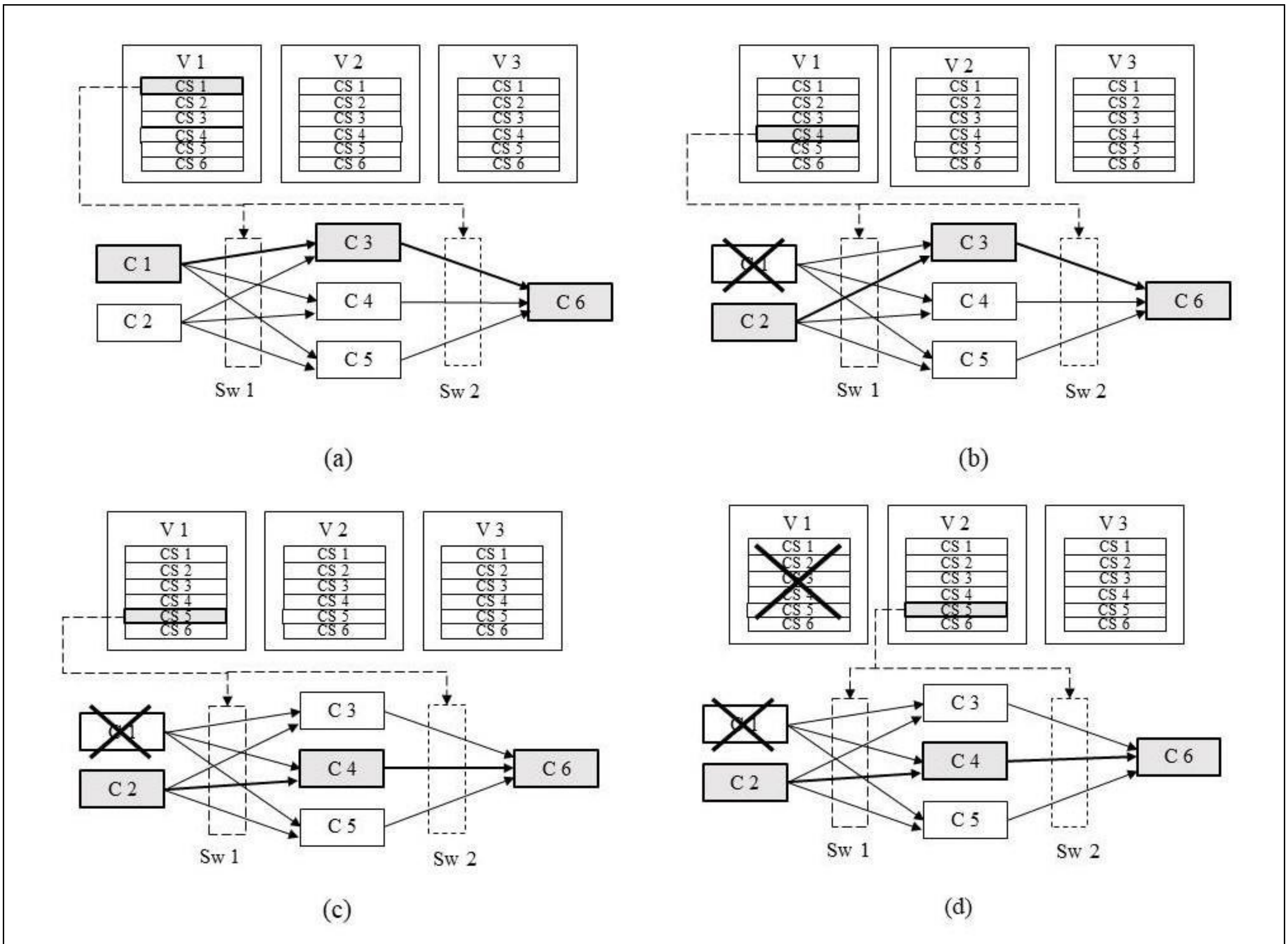

Fig. 4. Structural diagrams of a redundant COE: (a) normal operation in mode 1, (b) failure of component C1, (c) failure of component C1 and transition to mode 2, (d) failure of component C1, transition to mode 2, and failure of computer *V*1 (detectable or undetectable).

COE configurations

Configurations		RI of components in different situations												FEI of configurations	
No.	Connection diagram	Situations *A* and *D*						Situations *B*, *C*, and *D*						Mode 1 (situation *A* and *B*)	Mode 2 (situation *C*, *D*, or *E*)
		C1	C2	C3	C4	C5	C6	C1	C2	C3	C4	C5	C6		
1	C1→ C3→ C6	1	-	1		-	1	0	-	1	-	-	1	6	0.6
2	C1→ C4→ C6	1	-	-	1	-	1	0	-	-	1	-	1	5	5
3	C1→ C5→ C6	1	-	-	-	1	1	0	-	-	-	1	1	4	4
4	C2→ C3→ C6	-	1	1	-	-	1	-	1	1	-	-	1	3	0.3
5	C2→ C4→ C6	-	1	-	1	-	1	-	1	-	1	-	1	2	2
6	C2→ C5→ C6	-	1	-	-	1	1	-	1	-	-	1	1	1	1

The computing means of the OICE include three identical computers V1, V2, and V3, each containing six CSs according to the number of configurations (the table). Two switches Sw1 and Sw2, controlled by supervisors, connect different components in given configurations. In addition, there are two operation modes of the COE, characterized by different FEI values of the configurations.

Typical situations:

A) *Normal operation of the COE*

Mode 1 is implemented, and all components are operable (Fig. 4*a*). In the normal operation of the COE (no failures, fixed mode), the RMS performs the following actions on the first and subsequent cycles:

The first stage of arbitration (in the DCS of the α-computer V1, activated at the RMS start by default):

– All three computers V1, V2, and V3 are included in the group of ready computers.

– The computers V1 and V2 are separated into the pair (by their ordinal numbers).

– The computer V1 is assigned the α-candidate, and the computer V2 is assigned the α-backup (by their ordinal numbers).

– The computer V1 is initialized.

The second stage of arbitration (in the computer V1):

– The supervisor CS1 is selected as the DCS based on the PAConf procedure in the α-candidate by the greater FEI value.

– The supervisor CS1 is initialized.

– The information about the selected supervisor CS1 is transmitted from the computer V1 to the computer V2.

The third stage of arbitration:

– The computer V1 is selected as the α-computer based on the PAComp procedure between the computers V1 and V2.

– The computer *V*1 remains active.

– The supervisor CS1 remains active in the computer V1.

– The supervisor CS1 in the computer *V*1 initializes configuration no. 1.

B) *A failure in a COE component*

Mode 1 is implemented, and the failure of the component C1 is detected by the BCM (Fig. 4*b*). The RMS performs the following actions on the next operation cycle:

The first stage of arbitration (by analogy with situation *A*).

The second stage of arbitration:

– The supervisor CS4 is selected as the DCS based on the PAConf procedure in the α-candidate by the greater FEI value.

– The supervisor CS4 is initialized.

– The information about the selected supervisor CS4 is transmitted from the computer V1 to the computer V2.

The third stage of arbitration:

– The computer V1 is selected as the α-computer based on the PAComp procedure between the computers V1 and V2.

– The computer V1 remains active.

– The supervisor CS4 is initialized in the computer V1.

– The supervisor CS4 in the computer V1 initializes configuration no. 4.

Clearly, the PAConf procedure eliminates the configurations containing the failed component C1 by the condition $RI_1 = 0$. Configuration no. 4, finally selected, parries the failure of component C1 by engaging the backup (alternative) component C2.

C) *COE mode change*

The complex switches from mode 1 to mode 2, and the BiC means detect the failure of component C1 (Fig. 4*c*). The RMS performs the following actions on the next operation cycle:

The first stage of arbitration (by analogy with situation *B*).

The second stage of arbitration:

– The supervisor CS5 is selected as the DCS based on the PAConf procedure in the α-candidate by the greater FEI value.

– The supervisor CS5 is initialized.

– The information about the selected supervisor CS5 is transmitted from the computer *V*1 to the computer V2 and stored as the DCS-backup.

The third stage of arbitration:

– The computer V1 is selected as the α-computer based on the PAComp procedure between the computers V1 and V2.

– The computer V1 remains active.

– The supervisor CS5 is initialized in the computer V1.

– The supervisor CS5 in the computer *V*1 initializes configuration no. 5.

Clearly, the PAConf procedure yields the winning supervisor CS5 since $FEI_{CS5} = 2 > FEI_{CS4} = 0.2$. At the end of the cycle, configuration no. 5 (corresponding to the DCS) is implemented. This configuration changes the COE functionality, adapting it to the mode change.

D) *A detectable failure in a computer*

Mode 2 is implemented, and the failures of the component C1 and the computer V1 are detected by the BiC means (Fig. 4*d*). When the BiC means detect a failure of the computer V1, the redundancy management functions are transferred to the backup computer V2 on the next operation cycle of the RMS. This computer performs the following actions:

The first stage of arbitration:

– The two remaining operable computers V2 and V3 are included in the group of ready computers.

– Both computers V2 and V3 are separated into the pair.

– The computer V2 is assigned the α-candidate, and the computer V3 is assigned the α-backup (by their ordinal numbers).

– The computer V2 is initialized.

The second stage of arbitration (by analogy with situation *C*).

The third stage of arbitration:

– The computer V2 is selected as the α-computer based on the PAComp procedure between the computers V2 and V3.

– The computer V2 is initialized as the α-computer.

– The supervisor CS5 is initialized in the computer V2.

– The supervisor CS5 in the computer V2 initializes configuration no. 5.

E) *An undetectable failure in a computer*

Mode 2 is implemented, the failure of the component C1 is detected by the BCM, and the partial failure of computer V1 is not detected by the BiC means but identified during the PAComp procedure (Fig. 4*d*). The RMS performs the following actions on the next operation cycle:

The first and second stages of arbitration (by analogy with situation *C*).

The third stage of arbitration:

The third stage of arbitration:

– The computer V2 is selected as the α-computer based on the PAComp procedure between the computers V1 and V2.

– The computer V2 is initialized as the α-computer.

– The supervisor CS5 is initialized in the computer V2.

– The supervisor CS5 in the computer V2 initializes configuration no. 5.

Thus, the RMS can be an additional check tool for the state of components: it identifies the failures undetectable for the BiC means through the controlled errors of forming preferred configurations (DCS).

This example has explained the operation of the redundancy management system based on configuration supervisors under controlled and uncontrolled component failures (fault tolerance) and COE mode change (functional reconfiguration).

CONCLUSIONS

This paper has formulated an arbitration approach to manage the redundancy of COEs based on configuration supervisors. The number of supervisors corresponds to the number of competitive COE configurations. Each supervisor summarizes information about the corresponding configuration, participates in arbitration, and performs all procedures to implement the winner configuration.

This approach has several advantages:

– the unified structure of redundancy management tools;

– the flexible coverage of different, including unique, configurations;

– the elimination (or significant reduction) of bottlenecks in the redundancy management structures;

– the selection of operable configurations with preferred operational and technical characteristics.

Potential drawbacks of the approach include an excessive increase in the number of configurations (the number of CSs) and the expected certification difficulties. However, this situation is objective, and its solvability looks optimistic. For example, a close problem—the certification of modern integrated COEs with virtual data channels—is already finding its solution.

Acknowledgments. *The authors are grateful to A.M. Bronnikov, I.F. Gamayunov, and A.S. Popov for their invaluable contribution to some aspects of the proposed approach.*

REFERENCES

1. Aleshin, B. S., Babkin, V. I., Gokhberg, L. M. , et al. *Forsait razvitiya aviatsionnoi nauki i tekhnologii do 2030 goda i na dal'neishuyu perspektivu* (Foresight of Aviation Science and Technology Development until 2030 and Beyond), Moscow: Central Aerohydrodynamic Institute (TsAGI), 2014. (In Russian.)
2. *Digital Avionics Handbook*, 3rd ed. Spitzer, C. R., Ferrell, U., and Ferrell, T., Eds., London-New York: CRC Press, 2015.
3. Fedosov, E. A. The Main Lines of Forming Technological Advance in the Onboard Equipment of Next-Generation Aircraft, *Trudy 4th Mezhdunarodnoi nauchno-prakticheskoi konferentsii "Perspektivnye napravleniya razvitiya bortovogo oborudovaniya grazhdanskikh vozdushnykh sudov"* (Proc. 4th Int. Sci. -Pract. Conf. "Perspective Areas of Developing Civil Aircraft On-Board Equipment"), Moscow, Zhukovsky: State Research Institute of Aviation Systems, 2017, pp. 6–14. (In Russian.)
4. Ezhilarasu, C.M., Zakwan, Skaf Z., and Jennions, I.K., The Arrlication of Reasoning to AeroSpace Integrated Vehicle Health Management (IVHM): Challenges and Orrontinues, *Progress in Aerospace Sciences*, 2019, no. 105, pp. 60–73.
5. Avakyan, A.A., Synthesis of Fault-Tolerant IMA Platforms for Aircraft Equipment Subsystems, *Trudy MAI*, 2013, no. 65, pp. 1–15. URL://http://trudymai.ru/published.php?ID=35845 (Accessed September 13, 2021). (In Russian.)
6. ISO/IEC 19762-1:2008 - Information technology - Automatic identification and data capture (AIDC) techniques - Harmonized vocabulary - Part 1: General terms relating to AIDC.
7. Klepikov, V.I., *Otkazoustoichivost' raspredelennykh sistem upravleniya* (Fault Tolerance of Distributed Control Systems), Moscow: Zolotoe Sechenie, 2014. (In Russian.)
8. Tarasov, A.A., *Funktsional'naya rekonfiguratsiya otkazo-ustoichivykh sistem* (Functional Reconfiguration of Fault-Tolerant Systems), Moscow: Logos, 2012. (In Russian.)
9. Degtyarev, A.R. and Kiselev, S.K., Fault-Tolerant Reconfigurable Complexes of Integrated Modular Avionics, *Electrical and Data Processing Facilities and Systems*, 2016, vol. 12, no. 1, pp. 89–99. (In Russian.)
10. Nicholson, M., Health Monitoring for Reconfigurable Integrated Control Systems, *Proceedings of the 13th Safety-Critical Systems Symposium*, Southampton, UK, London: Springer, 2005, pp. 149–62.
11. Halle, M. and Thielecke, F., Next Generation IMA Configuration Engineering-From Architecture to Application, *2015 IEEE/AIAA 34th Digital Avionics Systems Conference* (*DASC*), Prague, Czech Republic, 2015, pp. 6B2-1–6B2-13.
12. Zhiao, Y., Wang, S., and Jiapan, F., IMA Dynamic Reconfiguration Modelling and Reliability Analysis of Tasks Based on Petri Net, *Complex Systems Design & Management,* 2021, pp. 241–251.
13. Sollock, P., Reconfigurable Redundancy - The Novel Concept Behind the World's First Two-Fault-Tolerant Integrated Avionics System, *Avionics, Navigation, and Instrumentation*, 2019, pp. 243–246.
14. Kalyaev, I. A. and Mel'nik, E.V., *Detsentralizovannye sistemy komp'yuternogo upravleniya* (Decentralized Computer Control Systems), Rostov-on-Don, Southern Research Center RAS, 2011. (In Russian.)
15. Ageev, A.M., Gamayunov, I.F., Bronnikov, A.M., and Bukov, V.N., Supervisory Control Method for Redundant Technical Systems, *Journal of Computer and Systems Sciences International*, 2017, vol. 56, no. 3, pp. 410–419.
16. Nikolenko, S., Kadurin, A., and Arkhangel'skaya, E., *Glubokoe obuchenie. Pogruzhenie v mir neironnykh setei* (Deep Learning. Immersion into the World of Neural Networks), St. Petersburg: Piter, 2021. (In Russian.)
17. Ageev, A.M. Principles of Storing and Monitoring Configuration Information in the Task of On-Board Equipment Complex Redundancy Managing, *Mekhatronika, Avtomatizatsiya, Upravlenie*, 2022, vol. 23, no. 1, pp. 45–55. (In Russian.)

This paper was recommended for publication by V.G. Lebedev, a member of the Editorial Board.

Received February 8, 2022, and revised February 21, 2022.
Accepted February 28, 2022.

Author information

Ageev, Andrey Mikhailovich. Cand. Sci. (Eng.), Zhukovsky and Gagarin Air Force Academy, Voronezh, Russia
✉ ageev_bbc@mail.ru

Bukov, Valentin Nikolaevich. Dr. Sci. (Eng.), Research Institute of Aircraft Equipment, Zhukovsky, Russia
✉ v_bukov@mail.ru

Shurman, Vladimir Aleksandrovich. Ramenskoe Instrument-Making Design Bureau (branch), Zhukovsky, Russia
✉ vshurman@rpkb.ru

Cite this paper
Ageev, A.M., Bukov, V.N., and Shurman, V.A., Redundancy Management of Onboard Equipment: An Arbitration Approach Based on Configuration Supervisors. *Control Sciences* **2,** 19–28 (2022). http://doi.org/10.25728/cs.2022.2.3

Translated into English by *Alexander Yu. Mazurov,*
Cand. Sci. (Phys.-Math.),
Trapeznikov Institute of Control Sciences,
Russian Academy of Sciences, Moscow, Russia
✉ alexander.mazurov08@gmail.com

DOI: http://doi.org/10.25728/cs.2022.2.4

THE HANDLING-COMFORT TRADE-OFF IN A QUARTER-CAR SYSTEM: AUTOMATIC ADAPTIVE MANAGEMENT VIA ACTIVE DISTURBANCE REJECTION CONTROL

M. Alhelou[1], Y. Wassouf[2], V.V. Serebrenny[3], A.I. Gavrilov[4], and E.S. Lobusov[5]

Bauman Moscow State Technical University, Moscow, Russia

[1]✉ alkhelum@student.bmstu.ru [2]✉ thelegend990@gmail.com [3]✉ vsereb@bmstu.ru
[4]✉ alexgavrilov@mail.ru [5]✉ evgeny.lobusov@yandex.ru

Abstract. The effectiveness of a vehicle suspension is often assessed by maximum passenger comfort given continuous contact with the road (road holding). This paper investigates managing the comfort-handling trade-off in a quarter-car suspension system via active disturbance rejection control (ADRC). An adaptive control law is built to manage this trade-off automatically depending on the ADRC parameters. The idea is to use the ADRC-estimated disturbance signal to adjust the system's domain of interest. The effectiveness of the proposed approach is validated: the adaptive control law is tested for a nonlinear hydraulic suspension system. Moreover, the effects of road disturbances amplitudes and road quality on the system performance are studied. Simulation results show the smoothness and simplicity of the adaptive algorithm for managing the comfort-handling trade-off.

Keywords: active disturbance rejection control (ADRC), quarter-car model (QCM), tracking differentiator (TD), nonlinear state error feedback (NLSEF), extended state observer (ESO), disturbance rejection scheme (DRS), proportional-integral-differential (PID) controller, proportional-differential (PD) controller, road handling, ride comfort, road holding, gradient MIT rule (adaptive control method).

INTRODUCTION

A vehicle suspension (suspension system) is a set of parts, components, and mechanisms linking the vehicle body with the road [1]. The suspension performs the following functions:

– It physically connects the wheels or axle assembly to the vehicle's supporting frame.

– It passes to the supporting frame the forces and moments resulting from the interaction of the wheels with the road.

– It ensures the required movement of the wheels relative to the frame and the required smoothness of motion.

When a vehicle moves along a periodic profile with unsprung mass resonances, significant vibrations can occur in its vertical responses to road disturbances. The works [2–6] were devoted to vibrations in vehicle suspension systems. Until recently, when increasing the smoothness of a vehicle, engineers were limited to controlled suspensions to stabilize the position of the vehicle body, mainly its longitudinal (pitch) angle. This contributes to the safe operation of the vehicle by affecting its stability. However, the vehicle's safety in terms of motion instability depends not only on the intensity of vertical, longitudinal-angular, and transverse-angular vibrations of the suspension mass. In the sense of road safety, intense vibrations of the unsprung mass can also be extremely dangerous and therefore undesirable. When moving on periodic bumps under a lateral external force, the vehicle may become unstable due to the weakening of vertical road reactions to the wheels.

A controlled suspension is a type of suspension that adjusts in motion the vertical movement of the wheels relative to the vehicle chassis or body using a special control system. Controlled suspensions can be divided into two classes: active and semi-active. Engineers developed various modifications of automated suspension systems, including active and semi-active spring-based systems [7–10]. Semi-active suspensions change somehow the damping coefficient of the shock absorber to influence the magnitude of the forces generated by it. The corresponding changes in the damping coefficient of the shock absorber depending on the current driving mode of the vehicle are calculated by the control unit based on the information received from the sensor system. Active suspensions include an actuator to control the distance between the wheel center and the chassis.

The smooth ride of a vehicle makes the trip comfortable and minimizes damage to cargo. Moreover, it can minimize the driver's effort during long trips in uncomfortable vehicles [11, 12]. Road handling characterizes how vehicle's wheels respond to the driver's commands and how it moves on the highway or road. This is usually judged by observing the behavior of the vehicle, especially during cornering, acceleration, and braking, as well as by the stability of the vehicle during steady-state motion [13].

Suspension design is often a trade-off between ride comfort and road handling: vehicles with stiff suspension have better regulation of body movements and faster response. Similarly, a low center of gravity is more convenient for road handling, but low ground clearance limits suspension deflection, requiring stiffer springs [8, 14].

In [15], the comfort-handling trade-off was investigated for off-road vehicles on three examples. The authors proposed design criteria for a semi-active suspension system that could significantly reduce or even eliminate the contradiction between ride comfort and road handling. Such a system switches between a stiff spring with high damping mode (for road handling) and a soft spring with low damping mode (for ride comfort). In [10], an automotive active pneumatic system with the main handling and comfort characteristics was presented. The active pneumatic suspension system uses a set of equations for the quarter-car model, a pneumatic valve, and a pneumatic spring. A nonlinear control algorithm based on the *backstepping* principle was adopted. In [16], controllers for improving road holding and passengers' comfort were constructed and analyzed. According to the results, both controllers demonstrate good performance, but the controller proposed below has better performance and reliability.

The contributions of this paper are as follows.

– We propose a new simple and easily adjustable adaptive control system based on *active disturbance rejection control* (ADRC). This control system changes the control domain depending on road disturbances.

– The control strategy is scaled with a single planning parameter and automatically manages the handling-comfort trade-off in the quarter-car model.

– We apply this control algorithm to a hydraulically actuated suspension system.

– We illustrate the effectiveness of the proposed approach using some simulation examples.

1. THE QUARTER-CAR MODEL

The linear active suspension system is shown in Fig. 1. The system dynamics are described by the equations

$$m_s \ddot{z}_s = -K_s(z_s - z_{us}) - C_s(\dot{z}_s - \dot{z}_{us}) + u,$$
$$m_{us}\ddot{z}_{us} = K_s(z_s - z_{us}) + C_s(\dot{z}_s - \dot{z}_{us}) -$$
$$K_{us}(z_{us} - z_r) - C_{us}(\dot{z}_{us} - \dot{z}_r) - u,$$

with the following notations: z_s is the sprung mass displacement; z_{us} is the unsprung mass displacement; z_r is the pavement defect; K_s is the spring stiffness; K_{us} is the tire stiffness; C_s is the damper damping factor; $C_{us} \approx 0$ is the tire damping factor; finally, u is a control signal.

Following [17], we use the suspension parameters $K_s = 17\,765$ N/m, $K_{us} = 190125$ N/m, $C_s = 535$ N·s/m, $m_s = 285$ kg, and $m_{us} = 41$ kg. The value $(z_s - z_{us})$ characterizes the suspension deflection, and $\ddot{z}_s$ is the vertical acceleration of the vehicle body.

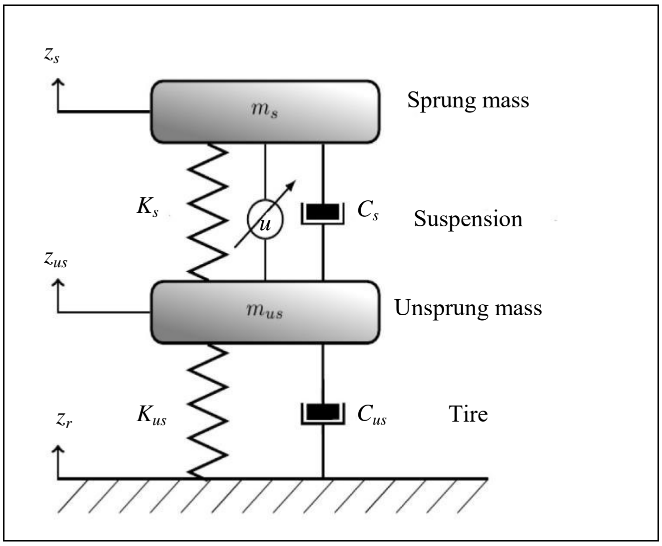

Fig. 1. The linear active quarter-car model.

Ride comfort is characterized by the *root mean square* (RMS) value of the vertical vehicle acceleration. The lower this value is, the higher the comfort level will be. On the other hand, road handling is characterized by the duration of the wheel's contact with the road surface. The greater the RMS value of the suspension deflection is, the lower the road handling level will be.

There is an unavoidable contradiction between ride comfort and road handling since the wheel position approximates the road profile at low frequencies (< 5 rad/s): any decrease in the body travel (vertical position of the sprung mass) at these frequencies will increase the suspension deflection [18]. To resolve this contradiction, this paper considers an adaptive ADRC-based control strategy.

2. ADAPTIVE CONTROL OF ACTIVE DISTURBANCE SUPPRESSION

As stated in [10], both road handling and ride comfort depend on the suspension travel z_s. According to the authors, the suspension travel value should be assigned by prioritizing one of the two aspects. Moreover, a possible way to improve suspension performance is to increase passenger comfort when the relative displacement between the sprung and unsprung masses is far enough from the suspension limits. On the other hand, the control unit must provide safe handling by limiting the suspension travel.

To simplify the idea, consider the control scheme in Fig. 2. Assume that the controller's feedback is given by $(z_s - \alpha z_{us})$, where $\alpha \in [0, 1]$ is the tuning parameter. For $\alpha = 0$, the feedback input is z_s. (In other words, the algorithm aims to minimize the vertical displacement of the sprung mass.) As a consequence, vertical acceleration will be minimized, providing the necessary comfort. When α increases, the suspension travel becomes greater, and the controller gradually gives priority to reducing the suspension travel.

In this paper, we apply ADRC to utilize its adaptive capabilities. In this case, the focus is on an appropriate variation law of the parameter α when changing the suspension deflection.

3. DESCRIPTION OF ADRC

ADRC is an inheritor of the proportional-integral-differential (PID) controller and can be considered a reliable control method: it represents all the unknown dynamics not included in the mathematical model of the controlled system and compensates model uncertainties and exogenous disturbances in real time [19].

As a result, the controlled system behaves like an nth-order integrator ($1/s^n$, where s denotes the Laplace variable and n expresses the system order).

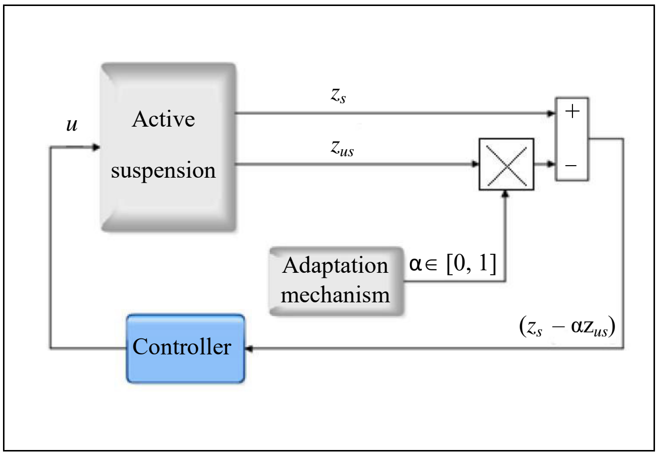

Fig. 2. The control scheme proposed for the handling-comfort trade-off.

Therefore, it is easily controlled by a proportional-differential (PD) controller, even in the case of a nonlinear and time-invariant plant. Figure 3 shows the nonlinear structure of ADRC with four main blocks: the controller (nonlinear state error feedback, NLSEF), the extended state observer (ESO), the tracking differentiator (TD), and the disturbance rejection scheme (DRS).

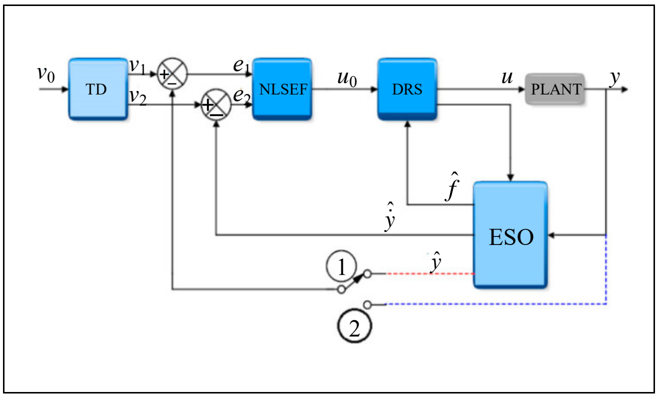

Fig. 3. The diagram of ADRC.

ADRC does not require a precise system model. By assumption, system dynamics can be expressed in the general form

$$\ddot{y} = b_0 u + f, \tag{1}$$

with the following notations: y is the output signal; u is the control input; f is the total disturbances containing the exogenous and endogenous ones; b_0 is the plant's gain.

The reference signal is smoothed by the TD, and the output signals are generated to track the reference signal and its differential. The algorithm is written as

$$\dot{v}_1 = v_2,\ \dot{v}_2 = f_{td}(v_1 - v_0, v_2, r_1) = -r\,\mathrm{sign}\left(v_1 - v_0 + \frac{v_2|v_2|}{2r_1}\right),$$

with the following notations: v_0 is the desired input; v_1 is the tracking signal of the system; v_2 is the differential signal of the system; r_1 is the parameter determining the tracking rate. As proposed in [19], the nonlinear function f_{td} ensures the fastest possible tracking of the reference signal and its derivative considering the acceleration limit r_1. The parameter r_1 depends on the application and is tuned accordingly to speed up or slow down the transients.

A conventional ESO is used to estimate system dynamics. This observer yields the estimates $\hat{y} \approx y$, $\hat{\dot{y}} \approx \dot{y}$, and $\hat{f} \approx f$ (the exogenous disturbances and endogenous dynamic uncertainties). The ESO monitors the performance and forecasts the plant's state in real time. This process was described in the paper [19]:

$$\hat{e} = z_1 - y,\ \dot{z}_1 = z_2 + \alpha_1 \hat{e},$$

$$\dot{z}_2 = z_3 + \alpha_2 |\hat{e}|^{1/2}\,\mathrm{sign}(\hat{e}) + \hat{b}_0 u,$$

$$\dot{z}_3 = \alpha_3 |\hat{e}|^{1/4}\,\mathrm{sign}(\hat{e}),$$

where: y is the system output; z_1 is the tracking signal for y; $\hat{e}$ is the estimated error; z_2 is the differential signal for z_1; z_3 is the tracking signal for the total disturbances; α_1, α_2, and α_3 are the estimation coefficients; u is the control input; $\hat{b}_0$ is the system coefficient (the estimated value of the gain b_0).

The nonlinear state error feedback is a nonlinear control strategy that can improve the accuracy of the control system. As described in [19], it has the form

$$\hat{e}_1 = v_1 - z_1,\ \hat{e}_2 = v_2 - z_2,$$

$$u_0 = \beta_1 f_{nl}(\hat{e}_1, \gamma_1, \eta) + \beta_2 f_{nl}(\hat{e}_2, \gamma_2, \eta), \qquad (2)$$

$$u = \frac{u_0 - z_3}{b_0},$$

with the following notations: $\hat{e}_1$ is the estimated system error; $\hat{e}_2$ is the estimated system error differential; β_1 and β_2 are the gains; the function f_{nl} should provide good control efficiency and high-frequency switching between the modes. Following [19], it can be chosen as

$$f_{nl}(e, \gamma, h) = \begin{cases} \dfrac{e}{\eta^{\gamma-1}}, & |e| \le \eta, \\ |e|^{\gamma}\,\mathrm{sign}(e), & |e| > \eta. \end{cases}$$

Note that the disturbance rejection scheme (DRS) is the last part of equation (2):

$$u = \frac{u_0 - z_3}{b_0},$$

where the estimated disturbance $z_3 = \hat{f}$ is eliminated by subtracting it from the control signal u_0.

Figure 3 shows a switch with two different modes (1 and 2). Thus, the output signal y can be directly passed to the NLSEF controller instead of its estimate $\hat{y}$.

4. AUTOMATIC TUNING OF THE PARAMETER α

We return to the system dynamics (1). Due to the input of the control system (system output) $y = z_s - \alpha z_{us}$, the following fact seems clear: if the closed-loop system is stable, the drift value f will change almost linearly with changing the amplitude of the road disturbances. See Fig. 4, where the values of f are compared with the amplitude of the road disturbances for the closed loop of the sprung mass displacement in the quarter-car model stabilized by ADRC. At the same time, the suspension deflection is inversely proportional to the values of the function *f*.

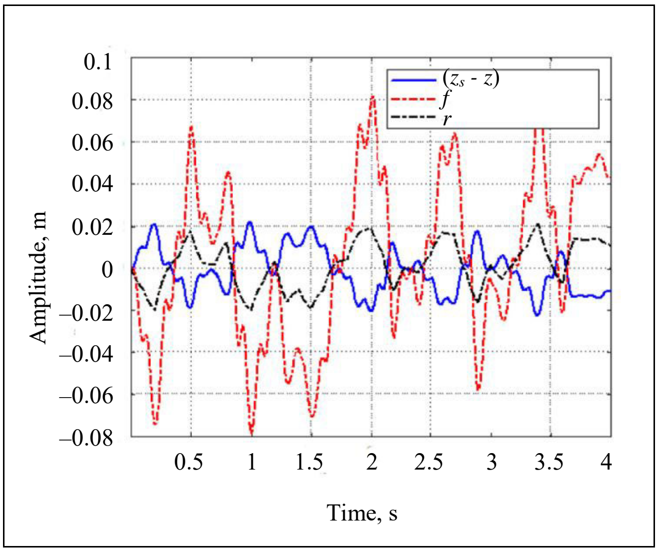

Fig. 4. Values of the function *f* depending on road disturbances.

The parameter α should grow significantly as the suspension deflection increases to balance between the suspension deflection and the passenger comfort.

Therefore, we apply *the MIT rule* with a positive gradient to obtain the variation law of α. In this case, the suspension deflection can be considered a linear function of *f*:

$$z_s - z_{us} = T_{K,w}(f),$$

where $T_{K,w}$ is a *low pass filter* (LPF) with a gain K and a cut-off frequency w. The filter gain and pass frequency are tuning parameters. The performance index α will depend on the absolute value of the suspension deflection, which can be chosen as follows:

$$J(\alpha) = \frac{1}{2}\left(z_s(\alpha) - z_{us}(\alpha)\right)^2.$$

Applying the inversion of the MIT gradient rule to increase the quadratic performance index as the suspension deflection grows, we obtain

$$\frac{\partial J}{\partial\left(z_s(\alpha) - z_{us}(\alpha)\right)} = z_s(\alpha) - z_{us}(\alpha) = T_{K,w}(f),$$

$$\frac{d\alpha}{dt} = \gamma\frac{\partial J}{\partial\alpha} = \gamma\frac{\partial J}{\partial\left(z_s(\alpha) - z_{us}(\alpha)\right)} \times$$

$$\frac{\partial\left(z_s(\alpha) - z_{us}(\alpha)\right)}{\partial\alpha} = \gamma T_{K,w}(f)\frac{\dot{T}_{K,w}(f)}{\dot{\alpha}}$$

$$\Downarrow$$

$$\left(\dot{\alpha}\right)^2 = \gamma T_{K,w}(f)\dot{T}_{K,w}(f) \Rightarrow \dot{\alpha} = \rho\sqrt{T_{K,w}(f)\dot{T}_{K,w}(f)},$$

where $\gamma > 0$ is the tuning parameter, and $\rho = \sqrt{\gamma}$. To simplify the problem, we assume that $\dot{T}_{K,w}(f) = T_{K,w}(\dot{f})$. Hence, the variation rule of the parameter α takes the form

$$\dot{\alpha} = \rho\sqrt{T_{K,w}(f)T_{K,w}(\dot{f})},$$

where ρ is the tuning constant.

Permanent integration of the parameter α will lead to saturation over time, and integration here can be treated as finding the envelope. This is true because the values of α should decrease if the suspension deflection grows over time and increase otherwise.

In applications, this can easily be achieved by adding a simple envelope calculation scheme or equivalent mathematical equations to filter the estimated value of α.

Figure 5 shows the complete control scheme of the system. The Kalman filter extracts the displacements z_s and z_{us} separately using the sensor outputs (a vertical acceleration sensor for the sprung mass, a vertical acceleration sensor for the unsprung mass, and a potentiometer for the suspension deflection). Let the system be equipped with two types of sensors: (1) two acceleration sensors to measure the acceleration of the sprung and unsprung mass, respectively, and (2) a potentiometer to measure the suspension deflection.

The linear Kalman filter is based on the equations

$$\begin{aligned}\dot{x}_1 &= x_2,\\ \dot{x}_2 &= m_1,\\ \dot{x}_3 &= x_4,\\ \dot{x}_4 &= m_2,\end{aligned} \quad y_e = \begin{pmatrix} x_1 \\ x_3 \\ x_1 - x_3 \end{pmatrix}, \quad y_m = \begin{pmatrix} \iint m_1 \\ \iint m_2 \\ m_3 \end{pmatrix}, \tag{3}$$

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, B = \begin{pmatrix} 0 \\ m_1 \\ 0 \\ m_2 \end{pmatrix}, C = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 \end{pmatrix}, \tag{4}$$

with the following notations: *A*, *B*, and *C* are the state, control, and output matrices, respectively; m_1 and m_2 are the measured accelerations of the sprung and unsprung mass, respectively; m_3 is the measured suspension deflection; y_e and y_m are the estimation and measurement vectors, respectively; finally, $\left(x_1 \quad x_2 \quad x_3 \quad x_4\right)^{\mathrm{T}} = \left(z_s \quad \dot{z}_s \quad z_{us} \quad \dot{z}_{us}\right)^{\mathrm{T}}$.

Formulas (3) and (4) are the state-space representation of the two second-order integrators. These integrators yield the positions z_s and z_{us} from the measured accelerations $\ddot{z}_s$ and $\ddot{z}_{us}$. The measured suspension deflection $(z_s - z_{us})$ is used to correct the estimated positions at each time instant, being added to the measurement vector y_m. Without including the measured suspension deflection in the measurement vector, the positions z_s and z_{us} will have constant deflection deviations from the real values.

The covariance matrices for the process and observation noises can be chosen empirically. The initial value of the estimation covariance is chosen large enough, and the initial state estimate is zero.

The values of the parameter $\dot{\alpha}$ are passed to the envelope detector and are limited to the range [0, 1]; see Fig. 5. Here, v_0 is the desired input (zero in the presented approach). Therefore, the TD is irrelevant in this figure: it appears only to preserve the overall ADRC form.

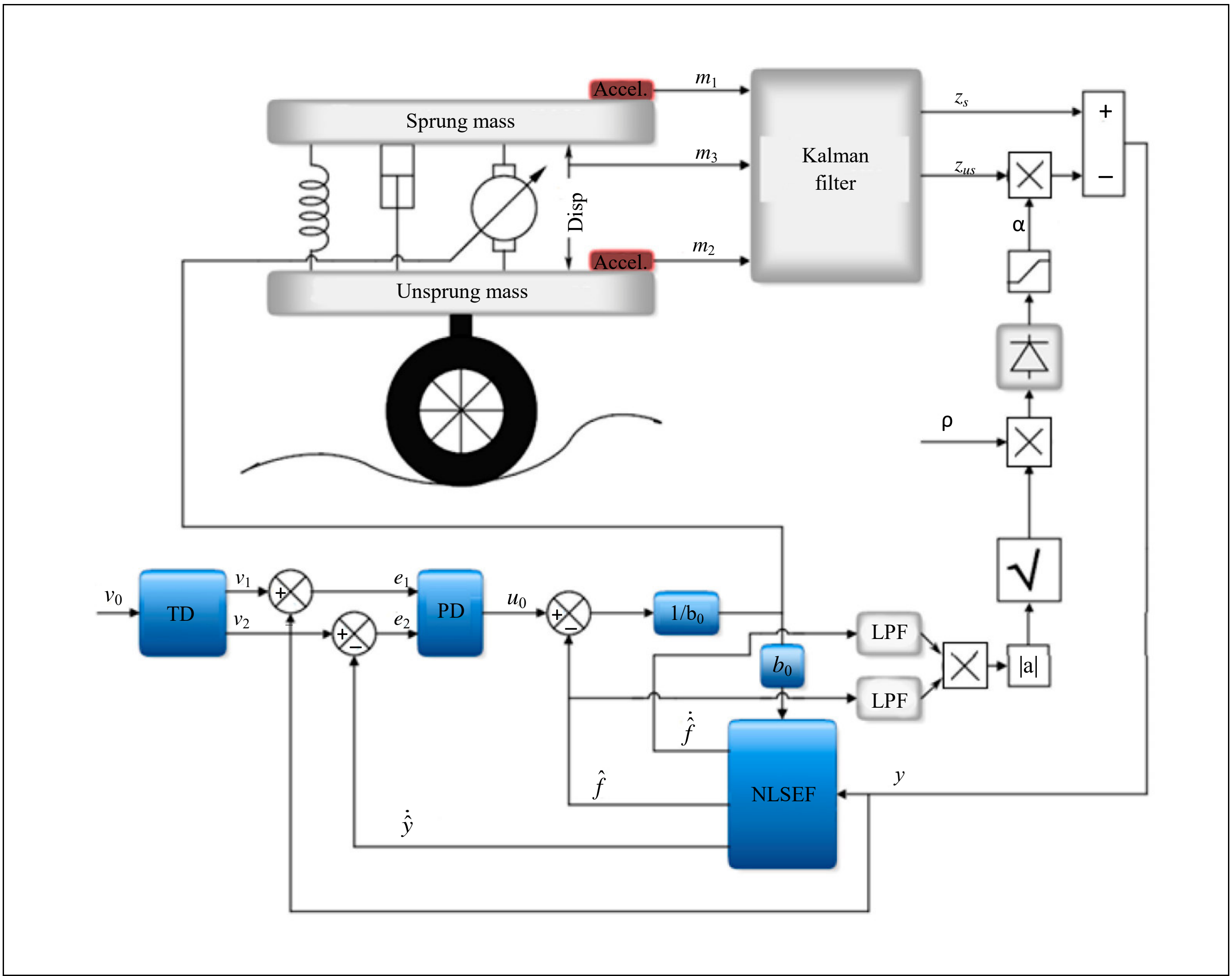

Fig. 5. Adaptive control system (complete diagram).

To demonstrate the effectiveness of this adaptive control law, we estimate the system performance in two cases. In the first case, the road disturbances are random (the ISO-8608 standard), and the system focuses on passenger comfort. In the second case, a sudden shock is treated as a roadblock: when it occurs, the driver is expected to focus on road handling (not to lose control of the vehicle).

To make the simulation more realistic, we choose a hydraulic active system of the quarter-car type with the nonlinear hydraulic actuator model.

5. SIMULATION

This section describes the hydraulic active system of the quarter-car type, the road disturbances, and the simulation results of the adaptive controller.

5.1 Hydraulic active system of the quarter-car type

The active suspension system uses a hydraulic actuator to reduce the external power required to achieve the desired performance. This system exerts an independent impact on the suspension to improve ride quality. The vehicle's active suspension system is presented in Fig. 6.

The hydraulic actuator consists of a spool (servo) valve and a hydraulic cylinder. Figure 6 has the following notations: P_s and P_r are the pressure of the hydraulic fluid entering and leaving the spool valve, respectively; x_{sp} is the position of the spool valve; P_u and P_l are the oil pressure in the upper and lower cylinder chambers, respectively. When the spool valve moves upwards (positive value), the upper chamber of the cylinder is connected to the supply line, and its

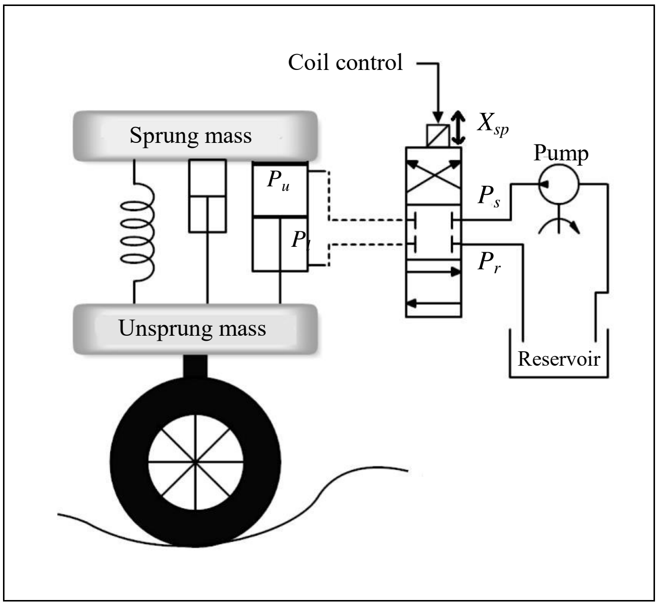

Fig. 6. Active hydraulic suspension system.

pressure increases. Meanwhile, the lower chamber is connected to the return valve, and its pressure decreases. This pressure difference expands the hydraulic cylinder.

For a mechanical movement of the valve spool, an electric current is applied to the coil connected to the servo valve. The powered actuator drives the spool to the desired position.

The actuator is described by the equation

$$v_c = L_c \frac{di_{sv}}{dt} + R_c i_{sv} \Rightarrow \frac{I_{sv}}{V_c} = \frac{1}{L_c s + R_c},$$

with the following notations: R_c and L_c are the coil resistance and inductance, respectively; v_c and i_{sv} are the motor rotor voltage and current, respectively. In the expressions below, the transition from lowercase to uppercase symbols means the variables in the Laplace space.

Let the spool valve displacement x_{sp} be related to the servo valve current i_{sp} through the linear transfer function

$$\frac{X_{sp}}{I_{sv}} = \frac{w_n^2}{s^2 + 2\xi w_n s + w_n^2}, \qquad (5)$$

where ξ is the damping coefficient, and w_n is the natural frequency of the servo valve. By a common assumption, the dynamics (5) are very fast, and we write

$$\frac{X_{sp}}{V_c} \approx \frac{K_v}{\tau s + 1},$$

where τ is the time constant of the servo valve, and K_v is the gain constant.

We introduce several assumptions:

– The valve area is linearly related to the spool displacement.

– The piston area is much larger than the bore.

– The fluid is incompressible.

– The inertia of the piston is negligible.

– The pressure changes in the two chambers are approximately equal, i.e.,

$$\Delta P_u \approx -\Delta P_l = \Delta P.$$

Then the impact is defined as

$$F_a = A_p \Delta P,$$

where A_p is the average piston area, and ΔP is the pressure difference in the valve pipelines. Following [20], this difference can be calculated by

$$\frac{V_t}{4\beta_e}\Delta\dot{P} = Q - C_{tp}\Delta P - A_p(\dot{z}_s - \dot{z}_{us}),$$

with the following notations: V_t is the total cylinder volume; β_e is the effective volume modulus; Q is the hydraulic load flow; C_{tp} is *the total piston leakage rate*. The flow equation with the controlled servo valve load has the form

$$Q = \mathrm{sign}[P_s - \mathrm{sign}(x_{sp})\Delta P] C_d w_g \times x_{sp}\sqrt{\frac{1}{\rho_1}(P_s - \mathrm{sign}(x_{sp})\Delta P)}, \qquad (6)$$

where C_d is the *discharge coefficient*, w_g is the spool valve area gradient, and ρ_1 is the hydraulic fluid density. With the variables

$$\alpha_1 = 4\beta_e/V_t,\ \beta = \alpha_1 C_{tp},\ \gamma = \alpha_1 C_d w_g \sqrt{1/\rho_1},$$

the system dynamics can be represented in the state space, as in equation (6):

$$X_1 = z_s,\ X_2 = \dot{z}_s,\ X_3 = z_{us},\ X_4 = \dot{z}_{us},$$

$$X_5 = \Delta P,\ X_6 = x_{sp},\ \dot{X}_1 = X_2,$$

$$\dot{X}_2 = -\frac{1}{M_s}\big(K_s(X_1 - X_3) + C_s(X_2 - X_4) - A_p X_5\big),$$

$$\dot{X}_3 = X_4,$$

$$\dot{X}_4 = \frac{1}{M_{us}}\big(K_s(X_1 - X_3) + C_s(X_2 - X_4) - K_{us}(X_3 - z_r) - C_{us}(X_4 - \dot{z}_r) - A_p X_5\big),$$

$$\dot{X}_5 = -\beta X_5 - \alpha_1 A_p (X_2 - X_4 + \gamma X_6 \nu),$$

$$\dot{X}_6 = \frac{1}{\tau}(-X_6 + K_v u), \; u = V_c,$$

$$\nu = \mathrm{sign}[P_s - \mathrm{sign}(X_6)X_5]\sqrt{P_s - \mathrm{sign}(X_6)X_5}.$$

We chose the following parameters to simulate the nonlinear model of the electrohydraulic actuator system:

$$\alpha = 4.515\times10^{13}\ \mathrm{N/m^5},\ \beta = 1\ s^{-1},$$

$$\gamma = 1.545\times10^{9}\ \mathrm{N/m^{5/2}kg^{1/2}},$$

$$\tau = \frac{1}{30}\mathrm{s},\ K_v = 4\times10^{-3},\ A_p = 3.35\times10^{-4}\ \mathrm{m^2}.$$

The source and external pressures were set to $P_s = 10$ bar and $P_r = 1$ bar, respectively.

5.2 Models of road disturbances

Following the ISO-8608 standard [9], a random filtered disturbance in the time domain was used as road disturbances to test the comfort level. The corresponding differential equation has the form

$$\dot{q}(t) = -2\pi f_0 q(t) + 2\pi n_0 \sqrt{G_q(n_0)v}\, w_d(t),$$

with the following notations: $q(t)$ is the random input signal; f_0 is a filter with a lower cutoff frequency; $G_q(n_0)$ is the road roughness coefficient; $w_d(t)$ is the Gaussian white noise. The vehicle speed was set to $v = 54$ km/h. A C class road was considered: $G_q(n_0) = 512\times10^{-6}$ and $n_0 = 0.1$.

For the handling test, the road shock was described by

$$w_d(t) = \begin{cases} 0.5h(1-\cos(w_b t)), & t_1 \le t \le t_2, \\ 0, & \text{otherwise}, \end{cases}$$

where: $h = 0.1$ m and w_b are the height and frequency of shocks, respectively; t_1 and t_2 are the lower and upper time limits of the function. The impact frequency is given by $w_b = 2\pi/(t_2 - t_1)$.

5.3 Simulation results

The simulation results are divided into three parts. The first part focuses on ride comfort when the road disturbances are a random signal. The second part concerns vehicle handling when the road disturbances are several consecutive shocks with a given frequency. Finally, the third part considers the two tasks together when the road disturbances of the above types are mixed.

The values of $\dot{\alpha}$ are always positive. Hence, it suffices adding a Butterworth low-pass filter to get the signal envelope. The setting constant for the parameter ρ was set to 200. The low-pass filter was given by $T_{K,w} = 0.1/(0.1s+1)$. The ADRC parameters were assigned as follows:

$$w = 0.5\ \mathrm{rad/s},\ \alpha_1 = 30w,\ \alpha_2 = 15w^2,$$

$$\alpha_3 = 85w^3,\ b_0 = 0.15,\ K_p = w^2,\ K_d = 2w.$$

Figure 7 shows the closed loop response when the road disturbances are random. Clearly, the parameter α has a small value throughout the process: the displacement between the sprung and unsprung masses is still far from its limits. However, the engineer can introduce an additional condition: set α to zero if its value is below some threshold. In this case, the control system will concentrate on ride comfort.

Figure 8 demonstrates the closed loop response when the road disturbances are a sequence of sudden shocks.

Clearly, the parameter α keeps zero values in the absence of exogenous disturbances. Meanwhile, its value gradually grows as the disturbance amplitude increases. Also, the value of α slowly decreases after the shock is over. This situation corresponds to the vehicle's real response to disturbances: the wheels need contact with the road for a short time after the impact is over to ensure greater safety.

For adaptivity tests, we applied this control method to the hydraulic suspension system under the following conditions: a 54 km/h vehicle moving on a C class road (ISO-8608) suddenly encounters several 0.1 m obstacles (in height) in front of it.

Figure 9 shows the closed loop response to these road disturbances. Clearly, the system focuses on ride comfort in the first stage, when the disturbance amplitude is relatively low; it switches to vehicle handling when riding into road bumps. After the high-amplitude disturbances caused by the bumps disappear, the system returns to ride comfort. Note that the proposed control law demonstrates a high degree of flexibility for the disturbances affecting the vehicle. This yields a trade-off between ride comfort and handling using an easily adjusted algorithm requiring no precise knowledge of the system dynamics.

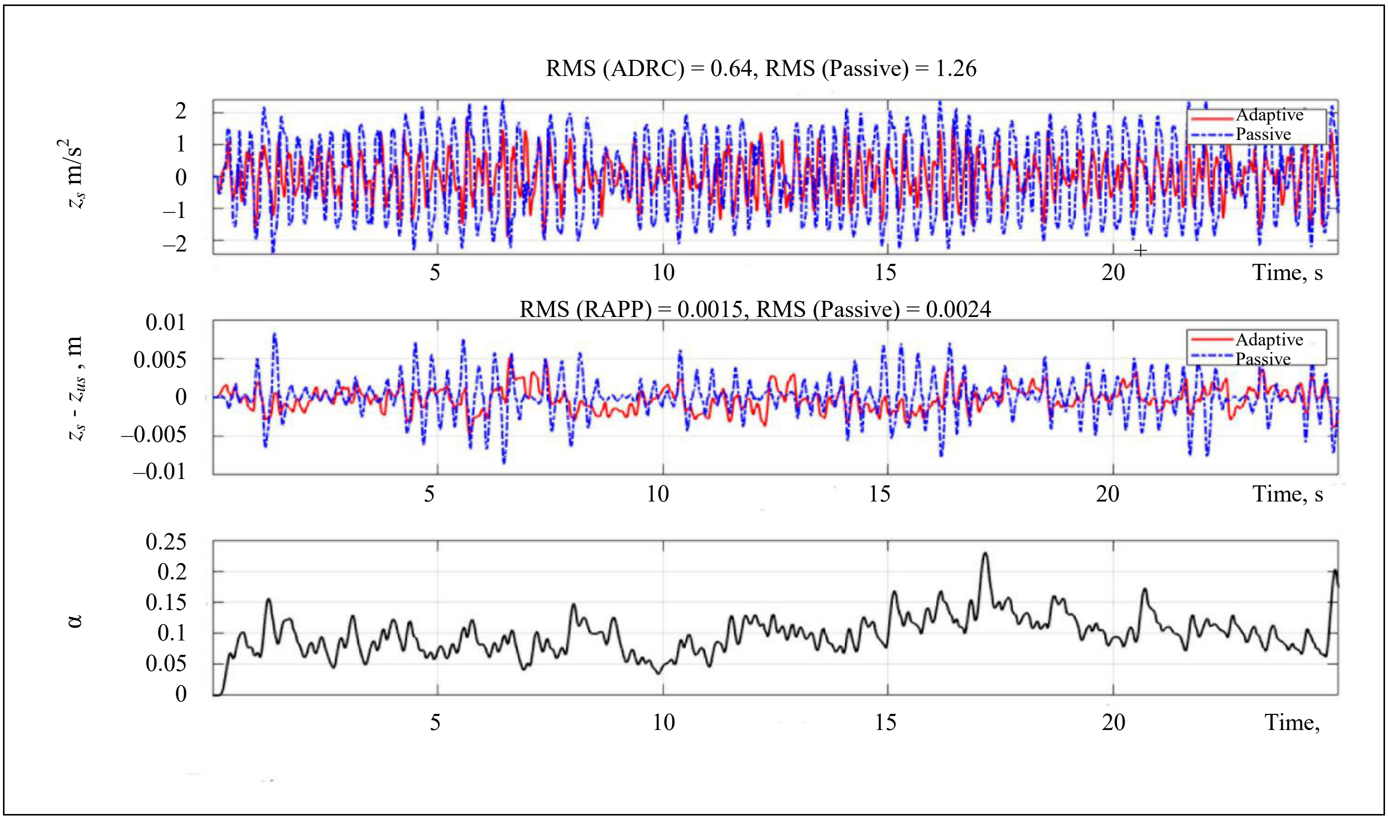

Fig. 7. Simulation results in the case of random road disturbances. The blue lines correspond to the passive system, and the red lines to the adaptive ADRC algorithm.

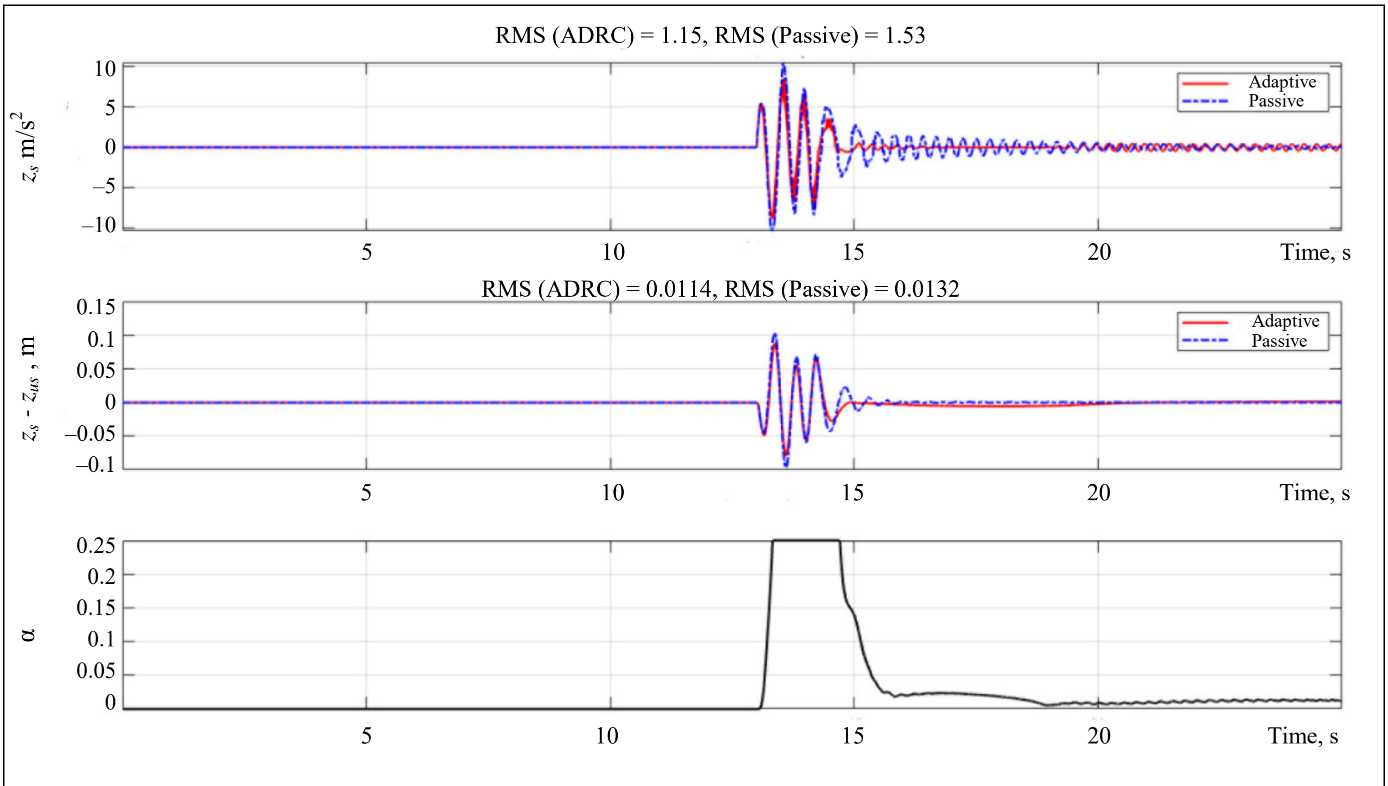

Fig. 8. Simulation results in the case of suddenly riding into road bumps. The blue lines correspond to the passive system, and the red lines correspond to the adaptive ADRC algorithm.

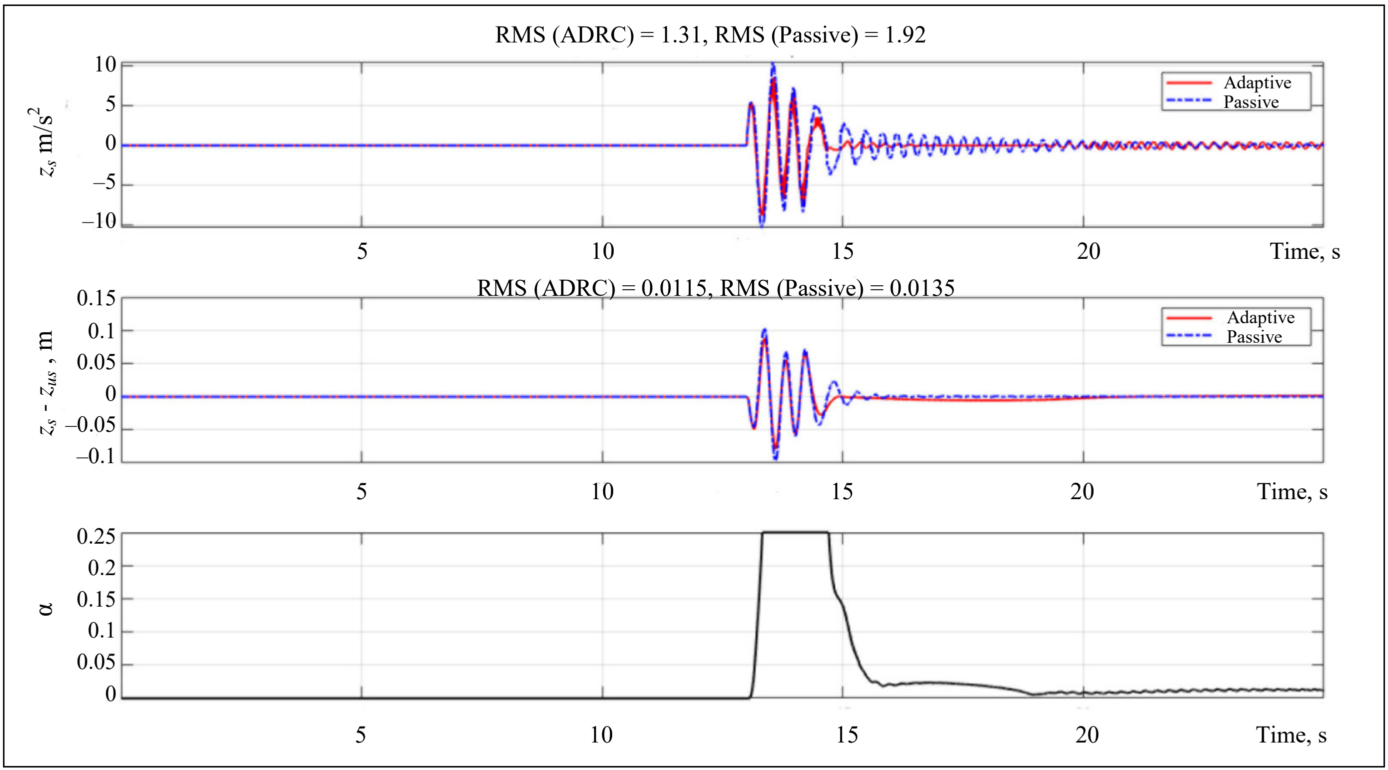

Fig. 9. Closed loop response of the nonlinear quarter-car system to hybrid road disturbances. The first line corresponds to the acceleration of the suspension mass, the second line corresponds to the suspension deflection, and the third line corresponds to the parameter α.

CONCLUSIONS AND FURTHER RESEARCH

This paper has presented a simple approach to managing the trade-off between the vertical acceleration of the vehicle chassis and its position relative to the road surface. Due to the increasing relevance of auto-driving, an auto-tuning ADRC mechanism has been proposed. The basic idea is to use the filtered values of the total disturbances estimated by ADRC to switch the system's operating mode between ride comfort and road handling. This is achieved by embedding the filtered value of the estimated disturbances into the control loop (multiplying this value by the unsprung mass displacement and subtracting it from the sprung mass displacement). The closed loop system has been tested under the following conditions: an average speed vehicle moving on a low-quality road suddenly encounters several obstacles of relatively high height in front of it.

The simulation results have shown the algorithm's ability to adapt and automatically shift focus between ride comfort and road handling. Moreover, this algorithm can be easily tuned for the nonlinear model.

As promising areas of further research, we mention the influence of Kalman filter settings and the change in the road disturbance model on the system performance. In addition, the effect of this controller on the entire vehicle can be studied when maneuvering, turning, and following a given trajectory.

REFERENCES

1. Goodarzi, A. and Khajepour, A., *Vehicle Suspension System Technology and Design*, Kentfield, CA, 2017.
2. Pinkaew, T. and Fujino, Y., Effectiveness of Semi-active Tuned Mass Dampers under Harmonic Excitation, *Engineering Structures*, 2001, vol. 23, no. 7, pp. 850–856.
3. Chaves, M., Maia, J., and Esteves, J., Analysis of an Electromagnetic Automobile Suspension System, *International Conference on Electrical Machines*, Vilamoura, Portugal, 2008.
4. Mihai, I. and Andronic, F., Behavior of a Semi-active Suspension System versus a Passive Suspension System on an Uneven Road Surface, *Mechanics*, 2014, vol. 20, no. 1, pp. 64–69.
5. Zhileykin, M.M., Kotiev, G.O., and Nagatsev, M.V., Synthesis of the Adaptive Continuous System for the Multi-axle Wheeled Vehicle Body Oscillation Damping, *IOP Conference Series: Materials Science and Engineering*, 2018, vol. 315, no. 1.
6. Ovsyannikov, S., Kalinin, E., and Koliesnik, I., Oscillation Process of Multi-support Machines When Driving over Irregularities, in *Energy Management of Municipal Transportation Facilities and Transportation*, Springer, Cham, 2018.
7. Yao, G.Z., Yap, F.F., Chen, G., et al., MR Damper and Its Application for Semi-active Control of Vehicle Suspension System, *Mechatronics*, 2002, vol. 12, no. 7, pp. 963–973.

8. Lajqi, S. and Pehan, S., Designs and Optimizations of Active and Semi-active Non-linear Suspension Systems for a Terrain Vehicle, *Strojniški Vestnik-Journal of Mechanical Engineering*, 2012, vol. 58, no. 12, pp. 732–743.
9. Zhou, Q., Research and Simulation on New Active Suspension Control System, *Master's Thesis*, Lehigh University, 2013.
10. Maizza, G. and Franz, D., *Simulink Control Model of an Active Pneumatic Suspension System in Passenger Cars*, Politecnico di Torino, 2019.
11. Ryu, S., Park, Y., and Suh, M., Ride Quality Analysis of a Tracked Vehicle Suspension with a Preview Control, *Journal of Terramechanics*, 2011, vol. 48, no. 6, pp. 409–417.
12. Tan, B., Wu, Y., Zhang, N., et al., Improvement of Ride Quality for Patient Lying in Ambulance with a New Hydro-Pneumatic Suspension, *Advances in Mechanical Engineering*, 2019, vol. 11, no. 4, pp. 1–20.
13. Abe, M., *Vehicle Handling Dynamics: Theory and Application*, Oxford: Butterworth-Heinemann, 2015.
14. Savsani, V., Patel, V.K., Gadhvi, B., and Tawhid, M., Pareto Optimization of a Half Car Passive Suspension Model Using a Novel Multiobjective Heat Transfer Search Algorithm, *Modelling and Simulation in Engineering*, 2017, vol. 2017, art. ID 2034907.
15. Els, P.S., Theron, N., Uys, P.E., and Thoresson, M., The Ride Comfort vs. Handling Compromise for Off-road Vehicles, *Journal of Terramechanics*, 2007, vol. 44, no. 4, pp. 303–317.
16. Singh, N., Chhabra, H., and Bhangal, K., Robust Control of Vehicle Active Suspension System, *International Journal of Control and Automation*, 2016, vol. 9, no. 4, pp. 149–160.
17. Hasbullah, F., Faris, W., Darsivan, J., and Abdelrahman, M., Ride Comfort Performance of a Vehicle Using Active Suspension System with Active Disturbance Rejection Control, *International Journal of Vehicle Noise and Vibration*, 2015, vol. 11, no. 1, pp. 78–101.
18. de Jesús Lozoya-Santos, J., Tudón-Martínez, J., Morales-Menéndez, R., and Ramírez-Mendoza, R., Comparison of On-off Control Strategies for a Semi-active Automotive Suspension Using HiL, *IEEE Latin America Transactions*, 2012, vol. 10, no. 5, pp. 2045–2052.
19. Han, J., From PID to Active Disturbance Rejection Control, *IEEE Transactions on Industrial Electronics*, 2009, vol. 56, no. 3, pp. 900–906.
20. Pedro, J.O., Dangor, O., Dahunsi, O.A., and Ali, M., CRS and PS-Optimised PID Controller for Nonlinear, Electrohydraulic Suspension Systems, The 9th Asian Control Conference (ASCC), Istanbul, 2013.

This paper was recommended for publication by S.A. Krasnova, a member of the Editorial Board.

Received October 15, 2021, and revised February 15, 2022.
Accepted February 28, 2022.

Author information

Alhelou, Muhammed. Postgraduate, Bauman Moscow State Technical University, Moscow, Russia,
✉ alkhelum@student.bmstu.ru

Wassouf, Yazan. Postgraduate, Bauman Moscow State Technical University, Moscow, Russia, ✉ thelegend990@gmail.com

Serebrenny, Vladimir Valer'evich. Cand. Sci. (Eng.), Bauman Moscow State Technical University, Moscow, Russia,
✉ vsereb@bmstu.ru

Gavrilov, Alexander Igorevich. Cand. Sci. (Eng.), Bauman Moscow State Technical University, Moscow, Russia,
✉ alexgavrilov@mail.ru

Lobusov Evgenii Sergeevich. Cand. Sci. (Eng.), Bauman Moscow State Technical University, Moscow, Russia,
✉ evgeny.lobusov@yandex.ru

Cite this paper
Alhelou, M., Wassouf, Y., Serebrenny, V.V., Gavrilov, A.I., and Lobusov, E.S., The Handling-Comfort Trade-Off in a Quarter-Car System: Automatic Adaptive Management via Active Disturbance Rejection Control. *Control Sciences* **2,** 29–39 (2022). http://doi.org/10.25728/cs.2022.2.4

Translated into English by *Alexander Yu. Mazurov,*
Cand. Sci. (Phys.-Math.),
Trapeznikov Institute of Control Sciences,
Russian Academy of Sciences, Moscow, Russia
✉ alexander.mazurov08@gmail.com