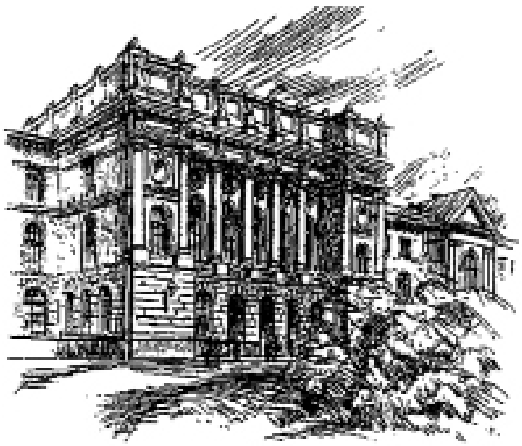
THE MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE RUSSIAN FEDERATION



ISSN 2687-0517

Computing, Telecommunications and Control

Vol. 17, No. 4
2024

Peter the Great St. Petersburg
Polytechnic University
2024

COMPUTING, TELECOMMUNICATIONS AND CONTROL

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

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

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

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

Editorial office

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

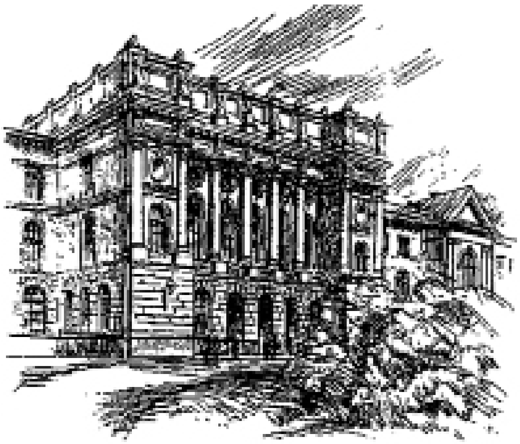
Ph.Ch.S. Bastian – literary editor, proofreader; G.A. Pyshkina – editorial manager; A.A. Kononova – computer layout; I.E. Lebedeva – English translation.

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

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

Release date: 30.12.2024

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ



ISSN 2687-0517

Информатика, телекоммуникации и управление

Том 17, № 4
2024

Санкт-Петербургский политехнический
университет Петра Великого
2024

ИНФОРМАТИКА, ТЕЛЕКОММУНИКАЦИИ И УПРАВЛЕНИЕ

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

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

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

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

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

Редакция журнала

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

Ф.К.С. Бастиан – литературный редактор, корректор; Г.А. Пышкина – ответственный секретарь, выпускающий редактор;

А.А. Кононова – компьютерная вёрстка; И.Е. Лебедева – перевод на английский язык.

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

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

Дата выхода: 30.12.2024

Contents

Содержание

Computing, Telecommunication and Control, 2024, Vol. 17, No. 4, Pp. 7–15.
Информатика, телекоммуникации и управление. 2024. Том 17, № 4. С. 7–15.

Information Technologies
Информационные технологии

Research article
DOI: https://doi.org/10.18721/JCSTCS.17401
UDC 004.422

EVALUATING THE PERFORMANCE OF JAVA VECTOR API IN VECTOR EMBEDDING OPERATIONS

N.A. Tomilov, V.P. Turov ✉
ITMO University, St. Petersburg, Russian Federation
✉ firemoon@icloud.com

Abstract. Hardware vector instructions are widely used to improve the performance of computations. The Java Vector API introduced in Java 16 allows using them portably on any platform supported by the Java Virtual Machine (JVM). In this paper, we evaluate performance benefits from rewriting typical vector search operations, such as computing distance between two vector embeddings, using the Java Vector API. We compare the performance of these vectorized implementations with semantically equivalent scalar code. Furthermore, we compare the Java Vector API with native C++ implementations, called from Java code via different Java-to-native interfaces, namely Java JNI, Project Panama (Foreign Function and Memory API), and manipulating Java JIT compiler via JVM CI and Nalim library. Benchmarking results suggest that in certain situations using Vector API can produce a measurable increase in performance of low-level operations, which can be translated into speedup of high-level algorithms such as Product Quantization. However, under certain scenarios, using Vector API is slower than relying on automatic vectorization provided by JVM, and most benchmarks suggest that invoking calculations implemented in C++ is faster even with all performance penalties incurred by native code invocations. Using techniques to lower these penalties, for example, by avoiding memory copy operations, can decrease the execution time by five times compared to Vector API and by ten times compared to plain Java code. However, in cases where using native code is prohibited, Vector API can still demonstrate a noticeable performance uplift, which can be beneficial for vector-related calculations in Java applications.

Keywords: vector embeddings, Java JVM, Java JNI, Java JVM CI, Project Panama, native code invocation

Citation: Tomilov N.A., Turov V.P. Evaluating the performance of Java Vector API in vector embedding operations. Computing, Telecommunications and Control, 2024, Vol. 17, No. 4, Pp. 7–15. DOI: 10.18721/JCSTCS.17401

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.17401
УДК 004.422

ИЗМЕРЕНИЕ ПРОИЗВОДИТЕЛЬНОСТИ JAVA VECTOR API ПРИ ЕГО ИСПОЛЬЗОВАНИИ В ОПЕРАЦИЯХ НАД ВЕКТОРНЫМИ ПРЕДСТАВЛЕНИЯМИ

Н.А. Томилов, В.П. Туров ✉

Национальный исследовательский университет ИТМО,
Санкт-Петербург, Российская Федерация

✉ firemoon@icloud.com

Аннотация. Аппаратные векторные инструкции широко используются для повышения производительности вычислений. Java Vector API, представленный в Java 16, позволяет переносимо использовать их на любой платформе, поддерживаемой виртуальной машиной Java. В данной работе выполняется оценка производительности при реализации типичных операций поиска по векторным представлениям, таких как вычисление расстояния между двумя векторными представлениями, с использованием Java Vector API. Производительность векторизованных реализаций этих операций сравнивается с семантически эквивалентным скалярным кодом. Кроме того, производится сравнение Java Vector API с нативными реализациями на C++, вызываемыми из Java-кода через различные интерфейсы взаимодействия Java с нативным кодом, а именно Java JNI, Project Panama (Foreign Function and Memory API) и управление JIT-компилятором через JVM CI и библиотеку Nalim. Результаты тестирования показывают, что в определенных ситуациях использование Vector API может привести к заметному увеличению производительности низкоуровневых операций, что может выражаться в ускорении высокоуровневых алгоритмов, таких как Product Quantization. Однако в некоторых сценариях использование Vector API оказывается медленнее по сравнению с автоматической векторизацией, предоставляемой JVM, и большинство тестов показывают, что вызов вычислений, реализованных на C++, занимает меньше времени по сравнению с Vector API, даже с учетом всех накладных расходов, возникающих при вызовах нативного кода. Используя методы для снижения этих накладных расходов, например, избегая операций копирования памяти, можно уменьшить время выполнения в пять раз по сравнению с Vector API и в десять раз по сравнению с обычным Java-кодом. Тем не менее, в случаях, когда использование нативного кода запрещено, Vector API все еще может демонстрировать заметное повышение производительности, что может быть полезно для вычислений, связанных с векторными представлениями, в Java-приложениях.

Ключевые слова: векторные представления, Java JVM, Java JNI, Java JVM CI, Project Panama, вызов нативного кода

Для цитирования: Tomilov N.A., Turov V.P. Evaluating the performance of Java Vector API in vector embedding operations // Computing, Telecommunications and Control. 2024. Т. 17, № 4. С. 7–15. DOI: 10.18721/JCSTCS.17401

Introduction

Modern processors have instruction sets that enable the simultaneous execution of certain operations on arrays of data. These instructions are referred to as SIMD instructions, where SIMD stands for Single Instruction Multiple Data. The use of such instructions significantly accelerates computations over numerical arrays and matrices [1, 2]. The process of transforming a loop that performs calculations on single data elements into a loop that operates on data blocks using SIMD instructions is called vectorization.

Vectorization can significantly accelerate any data array processing. One example of such data processing is the organization of search operations in large document databases, where machine-learning techniques are used to transform the data into representations that reflect the semantic structure of textual [3] and multimodal documents in the form of vector embeddings (representations), which are arrays of floating-point numbers. In this approach, an efficient search can be organized by constructing an index of these embeddings, converting the search query into its own embedding, and then performing the search for the nearest vector embeddings using the constructed index [4]. Vectorizing the computation of the distance between two vectors significantly reduces the time required for vector search [5]. In general, vector embeddings can have any dimensionality, but in vector search operations the dimensionality of each vector is typically small and each vector is often represented as an array of floating-point numbers with up to 1000 elements. Another key point of vector search operations is having a large number of relatively small operations (e.g., distance calculation or computing an average vector) [6].

High-level programming languages often have their own mechanisms for automatic vectorization, either during runtime, as in Java JVM, or at the compilation stage, as in C++. In case of Java, the JVM can automatically vectorize only a small set of operations[1]; in other cases, vectorization must be implemented manually, either through certain intrinsic functions of the language, such as FMA (Fused-Multiply-Add), or by implementing computational operations in a low-level language with access to assembly-level SIMD instructions, and then invoking the implemented functions from Java through various mechanisms for executing platform-dependent code. Java 16 introduced an additional mechanism for working with vector instructions – Java Vector API – which enables convenient use of vector instructions without the need for platform-dependent code. This mechanism offers a significant performance boost compared to Java's automatic vectorization [7], and despite its experimental status, it is already being adopted in some software products, such as Apache Lucene[2].

In this paper, we test the performance of the Java Vector API in the scope of operations on vector embeddings that are used in vector search. We compare the usage of this API to the implementation of the same operations in Java without vector instructions, as well as to the implementations in C++, where we call the corresponding functions from Java using mechanisms, such as Java JNI, Java JVM CI, and the Project Panama (Foreign Function and Memory API).

Methods of using vector instructions in Java

As mentioned above, there are three methods of working with vector instructions in Java: using intrinsics, invoking a shared library that was built in some other language with support for vector instructions, and using Vector API.

The first method involves using intrinsics, such as the FMA operation, which are optimized functions provided directly by the JDK developers. Unfortunately, the existing set of intrinsics is mostly limited to cryptographic operations, and it is not practical to use them when implementing operations applied in vector search.

The second method involves implementing vector computational operations in a low-level language, such as C or C++, and subsequently invoking these operations from Java JVM [8]. This approach allows us to optimize computational functions more precisely and take full advantage of any vector or other assembly instructions, automatic code vectorization provided by the low-level compiler [9], or even already implemented and highly optimized shared libraries, resulting in higher performance compared to plain Java [10]. However, this approach comes with several significant drawbacks. The most obvious are the need to write platform-dependent code and the need to modify it for each new architecture or even processor generation to achieve maximum performance. Additionally, this method comes with the complexity of working directly with the required vector and assembly instructions, direct memory handling, and potential security risks [11].

---

[1] Vladimir Ivanov. Vectorization in HotSpot JVM. Available: https://cr.openjdk.java.net/~vlivanov/talks/2017_Vectorization_in_HotSpot_JVM.pdf (Accessed: 29.09.2024)

[2] Available: https://github.com/apache/lucene/pull/12311 (Accessed: 06.12.2024)

A less obvious drawback of this approach is the overhead associated with data copying. When transferring control to platform-dependent code, the JVM typically does not allow direct access to memory managed by the JVM. If the data for computations resides in JVM memory, we must copy it to memory outside the JVM, and then we need to free that memory ourselves. This copying is necessary for each data array and, in addition to the time spent on the copying itself, incurs context switching from managed code to the JVM and other overheads [12]. When we need to pass matrices, being arrays of arrays, to managed code, the context-switching overhead can become so significant that it is more efficient to copy the data into a single large-dimensional array and transfer it, rather than copying each row of the matrix as a separate array. Alternatively, we can store all data in unmanaged memory beforehand, which complicates access to that data from the JVM.

Although the problems associated with executing unmanaged code in a managed environment are generally insurmountable, modern Java provides several mechanisms to reduce overhead from memory copying and context switching between managed and unmanaged code. One of the most effective methods for reducing overhead is through altering the just-in-time (JIT) compilation within the JVM. This allows us to completely replace the function body generated by the JIT compiler, specifically substituting the proper call to platform-dependent code, which involves memory copying and context switching, with a direct call to the platform-dependent function [13, 14]. However, this method has significant drawbacks, including the lack of exception handling, complete blocking of the thread invoking unmanaged code, and having access to the JVM from unmanaged code. Consequently, we cannot copy objects or complex data structures, such as arrays of arrays, into unmanaged code; we can only work with primitive types or one-dimensional arrays. When we need to pass two-dimensional arrays, such as arrays of vectors, we must manually copy the data into a large one-dimensional array. However, despite these limitations, this method remains popular and is used in frameworks for heterogeneous computing, such as Tornado VM [15], and in the Nalim library[3], which utilizes Java JVM CI capabilities. Another method for reducing overhead, aside from managing the JIT compiler, is the Foreign Function and Memory API, or Project Panama. This API, which recently achieved Released status in Java 22, offers a modern alternative to JNI and allows, under certain conditions, direct access to JVM memory from unmanaged code, including complex data structures like objects or arrays of arrays. This means that while the overhead from switching thread states between managed and unmanaged code still exists, we can entirely eliminate memory copying for complex data structures, potentially accelerating computations for arrays of vectors even more than through JIT compilation management.

The third method involves using the recently introduced Java Vector API. This framework provides a high-level abstraction over various vector instructions available on different architectures supported by the Java JVM. When using this framework, developers do not need to consider the target architecture or which specific vector instructions to utilize; they only need to prepare the data correctly and invoke a certain function, for example, addition. This function will be translated into the appropriate vector instruction, often the most suitable one, depending on the processor architecture and its supported instruction set. If the processor lacks support for vector instructions, the framework transparently falls back to using standard scalar instructions. Apart from ease of use, the Java Vector API offers several advantages, including the absence of manual memory copying to and from unmanaged code, and full support from the JVM, including proper exception handling and effective organization of computations in a multithreaded environment. Moreover, not having to deal with unmanaged code allows developers to use vector instructions for speeding up computations in contexts where using unmanaged code is prohibited or requires extensive precautions, such as in the financial sector.

However, a significant drawback is the need for manual organization of computations using vector instructions, which, albeit not directly related to assembly instructions, still closely resembles the use of vector instructions on the processor. Furthermore, there are no optimizations available if the instructions are

[3] Available: https://github.com/apangin/nalim (Accessed: 06.12.2024)

used incorrectly or inefficiently. For instance, when computing the sum of two vectors with a dimension of 390, the developer must perform a summation of 48 parts of vectors of length 8 in a loop, and then separately sum the remaining 6 elements manually, assuming the processor supports a maximum vector width of 256, meaning it can handle operations on up to 8 32-bit floating-point numbers at once. Any mishandling of this vector alignment will cause performance degradation. In addition, it is important to note that as of the time of writing, the Java Vector API remains in Incubator Preview status, and there is no information on when it will become fully supported and available for use in production environments.

Experiment setup

To test the performance of the Vector API, we created an interface that describes various operations for vector embeddings. We then implemented the following versions of this interface:

1. Standard Java;
2. Java with Vector API;
3. Java with Calls to Platform-Dependent Code Implemented in C++:
   - Using Java JNI, which includes all potential overheads;
   - Using the Nalim library, which leverages Java JVM CI to manage JIT compilation, eliminating context-switching overhead, but requiring memory copying for two-dimensional arrays;
   - Using the Project Panama (Foreign Function and Memory API), which incurs overhead from switching to native code, but does not require memory copying for one-dimensional and two-dimensional arrays.

The implemented operations can be divided into several sets.

The first set includes operations on a single vector or a pair of vectors of dimensionality D and consists of the following operations:

- Calculation of the Average value of the vector elements;
- Calculation of the Variance, also known as Dispersion, of the vector element values;
- Calculation of the Angular, also known as Cosine, and Euclidean Distances between two vectors [16].

The second set includes operations performed on multiple vectors simultaneously and uses an array of vectors (a two-dimensional array) as one of the input parameters:

- Calculation of the Average Vector from N source vectors;
- Calculation of N Angular and Euclidean Distances between one vector and N other vectors.

The third set includes complex computations that utilize operations from the first two sets:

- Clustering using the k-Nearest Neighbors (kNN) method from N source vectors into K clusters, which involves calculating N distances and calculation of the Average Vector [17];
- Product Quantization of N source vectors [18] to M subvectors with a quantization depth of nBits, which uses kNN-clustering.

The purpose of this set is to measure the extent of the overhead incurred due to the need to convert sets of vectors, which are arrays of arrays, into a one-dimensional array for passing to unmanaged code.

The sets of vectors were randomly generated and consisted of N arrays of 32-bit floating-point numbers (FP32) of size D. The following parameters were used:

- For the first two sets of operations:
  - D = 256, 384, 768, N = 100, 250, 500;
- For kNN-clustering:
  - D = 256, N = 5000, 10000;
  - K = 20, 50;
- For PQ:
  - D = 256, N = 10000;
  - M = 4, 8;
  - nBits = 8, 12, 16.

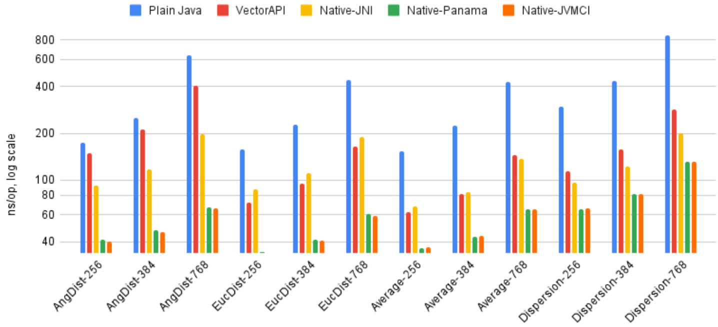


Fig. 1. Results for Operations on a single vector set

The test setup has the following specifications: AMD Ryzen 7 7700X (8C16T); 32GB RAM; Operating System: Ubuntu 22.04; a framework of comparing vector search algorithms, implemented in previous research [19], that uses the Java Microbenchmark Harness (JMH) [20]. The JMH was configured with a warm-up mode to allow the JIT compiler to perform necessary optimizations, including automatic vectorization. The C++ code was compiled using GCC 11.4.0. For all computations except JVMCI, JDK 22 version 22.0.2+9 was used. For JVMCI, JDK 17 version 17.0.10+7 was selected, because it was the most recent long-term support version of the JVM that the Nalim library works correctly with. In newer JVM versions, JVM CI was modified in a way that caused the library to malfunction, and modifying it is beyond the scope of this article. The source code for the benchmark and all implemented vector operations is available on GitHub[4].

Experiment results

The results for the first set of operations are presented in Fig. 1. The measured execution time for a single operation is in nanoseconds, where lower values indicate better performance.

It can be observed that using the Vector API significantly reduces the execution time for a single operation compared to standard scalar Java. However, the execution time for calculations using the Vector API is nearly indistinguishable (sometimes longer, sometimes shorter) from that of calculations implemented in C++ with an optimizing compiler and function calls via JNI. The execution time for calculations implemented in C++ and invoked through JVM CI or Project Panama is substantially lower than that using JNI, and is two to four times less than the execution time for calculations using the Vector API. This indicates significant overhead from both using JNI and the Vector API. Nevertheless, the speedup from the Vector API compared to regular Java computations is notable and directly depends on the number of vector operations: the more operations are necessary to perform the computation, the less pronounced the speedup becomes, as seen when comparing computation time of angular distance (which needs three operations in total) and Euclidean distance (which needs only one operation) for the same number of dimensions. Such difference depending on the number of operations also demonstrates the overhead involved when invoking the Vector API.

The results for the second set of operations are presented in Fig. 2. Fig. 2a shows the computation time for the average of N vectors of dimensionality D, Fig. 2b displays the computation time for N Euclidean distances between a given vector and N specified vectors, and Fig. 2c illustrates the similar time for angular distances.

When using the Vector API, the computation time for the average vector is approximately equal to that of standard Java and significantly less than the time required for calling C++ code via JNI or JVM CI. The performance degradation of the C++ code is attributed to the overhead of converting N arrays

[4] Available: https://github.com/nikita-tomilov/vector-vapi-jni-jvmci-panama (Accessed: 06.12.2024)

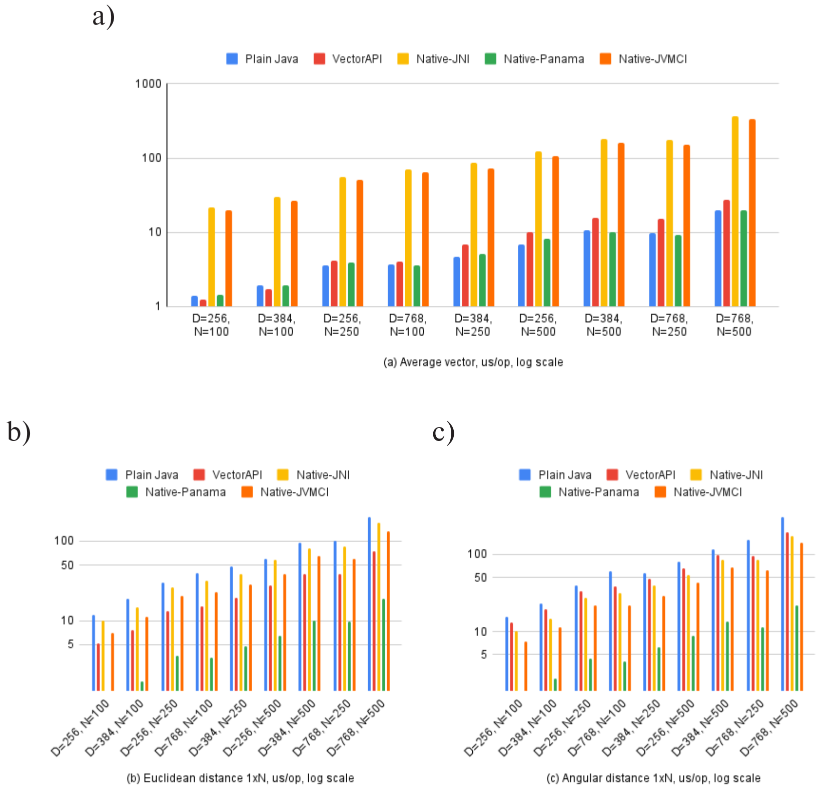


Fig. 2. Results for the multi vector operations

of dimensionality D into a one-dimensional array of size N×D for passing to unmanaged code. A similar conversion is necessary for calculating N distances; however, in this case, the slow data transfer to unmanaged code is offset by the more optimized code generated by the optimizing compiler, resulting in the use of C++ code in conjunction with JVM CI yielding the greatest speedup for distance calculations. Nevertheless, although in both cases the computation of N distances using the Vector API is significantly faster than using standard Java, and in the case of Euclidean distance, it is somewhat faster than using C++ and JNI, the combination of C++ and Project Panama remains an order of magnitude faster, thanks to the advantages provided by the optimizing compiler and the absence of memory copying overhead.

The results for the third set of operations are presented in Fig. 3. Fig. 3a shows the time for clustering using the kNN method, while Fig. 3b demonstrates the execution time for Product Quantization.

The results for the third set of operations partially replicate those of the second set: the parity between scalar Java and the Vector API remains for the measurement of Product Quantization time, with both benchmarks showing significant time overhead for JNI and JVM CI due to memory copying. In both benchmarks, Project Panama is the fastest option, benefiting from the absence of memory copying overhead. However, for kNNs, the Vector API is faster than scalar Java in all cases, achieving up to a twofold increase in performance at best.

Conclusion

The use of the Java Vector API significantly accelerates vector operations in applications written in Java or running in the JVM. However, computations implemented using the Vector API and manual vectorization generally lag behind calls to functions implemented in C++ that utilize automatic vectorization and mechanisms to reduce overhead associated with calling unmanaged code, such as the Project Panama

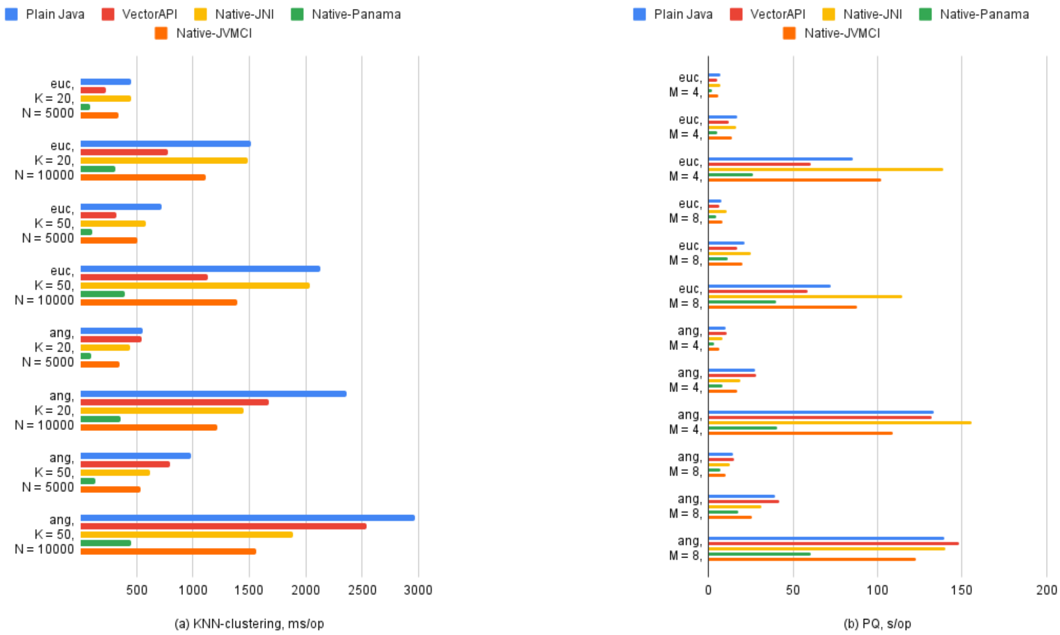


Fig. 3. Results for the complex vector operations

(Foreign Function and Memory API). Nevertheless, in scenarios where the use of unmanaged code is not feasible, as well as in cases where the conversion and transfer of data to managed code incur significant overhead despite the use of optimized methods for calling unmanaged code, the Vector API serves as a good alternative to traditional Java computations.

REFERENCES

1. **Muła W., Kurz N., Lemire D.** Faster population counts using AVX2 instructions. The Computer Journal, 2017, Vol. 61, No. 1, Pp. 111–120. DOI: 10.1093/comjnl/bxx046

2. **Cree M.J.** An exploration of using the Intel AVX2 gather load instructions for vectorised image processing. 2018 International Conference on Image and Vision Computing New Zealand (IVCNZ), 2018, Pp. 1–9. DOI: 10.1109/IVCNZ.2018.8634707

3. **Grbovic M., Cheng H.** Real-time personalization using embeddings for search ranking at Airbnb. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, Pp. 311–320. DOI: 10.1145/3219819.3219885

4. **Berry M.W., Drmac Z., Jessup E.R.** Matrices, vector spaces, and information retrieval. SIAM Review, 1999, Vol. 41, No. 2, Pp. 335–362. DOI: 10.1137/S0036144598347035

5. **Douze M., Guzhva A., Deng C., Johnson J., Szilvasy G., Mazaré P., Lomeli M., Hosseini L., Jégou H.** The Faiss library, 2024. DOI: 10.48550/arXiv.2401.08281

6. **Sanca V.** Ailamaki A. Efficient data access paths for mixed vector-relational search. Proceedings of the 20th International Workshop on Data Management on New Hardware, 2024, Pp. 1–9. DOI: 10.1145/36620-10.3663448

7. **Basso M., Rosà A., Omini L., Binder W.** Java Vector API: Benchmarking and performance analysis. Proceedings of the 32nd ACM SIGPLAN International Conference on Compiler Construction, 2023, Pp. 1–12. DOI: 10.1145/3578360.3580265

8. **Stojanov A., Toskov I., Rompf T., Püschel M.** SIMD intrinsics on managed language runtimes. Proceedings of the 2018 International Symposium on Code Generation and Optimization, 2018, Pp. 2–15. DOI: 10.1145/3168810

9. **Naishlos D.** Autovectorization in GCC. GCC Developers' Summit, 2004, Pp. 105–118.

10. **Vivanco R., Pizzi N.** Computational performance of Java and C++ in processing large biomedical datasets. IEEE CCECE2002. Canadian Conference on Electrical and Computer Engineering. Conference Proceedings (Cat. No. 02CH37373), 2002, Vol. 2, Pp. 691–696. DOI: 10.1109/CCECE.2002.1013025

11. **Grichi M., Abidi M., Jaafar F., Eghan E.E., Adams B.** On the impact of interlanguage dependencies in multilanguage systems empirical case study on Java Native Interface applications (JNI). IEEE Transactions on Reliability, 2021, Vol. 70, No. 1, Pp. 428–440. DOI: 10.1109/TR.2020.3024873

12. **Halli N.A., Charles H.-P., Mehaut J.-F.** Performance comparison between Java and JNI for optimal implementation of computational micro-kernels, 2015. DOI: 10.48550/arXiv.1412.6765

13. **Stepanian L., Demke Brown A., Kielstra A., Koblents G., Stoodley K.** Inlining Java native calls at runtime. Proceedings of the 1st ACM/USENIX International Conference on Virtual Execution Environments, 2005, Pp. 121–131. DOI: 10.1145/1064979.1064997

14. **Grimmer M., Rigger M., Stadler L., Schatz R., Mössenböck H.** An efficient native function interface for Java. Proceedings of the 2013 International Conference on Principles and Practices of Programming on the Java Platform: Virtual Machines, Languages, and Tools, 2013, Pp. 35–44. DOI: 10.1145/2500828.2500832

15. **Fumero J., Papadimitriou M., Zakkak F.S., Xekalaki M., Clarkson J., Kotselidis C.** Dynamic application reconfiguration on heterogeneous hardware. Proceedings of the 15th ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments, 2019, Pp. 165–178. DOI: 10.1145/3313808.3313819

16. **Abu Alfeilat H.A., Hassanat A.B.A., Lasassmeh O., Tarawneh A.S., Alhasanat M.B., Eyal Salman H.S., Surya Prasath V.B.** Effects of distance measure choice on K-nearest neighbor classifier performance: A Review. Big Data, 2019, Vol. 7, No. 4, Pp. 221–248. DOI: 10.1089/big.2018.0175

17. **Seidl T., Kriegel H.-P.** Optimal multi-step K-nearest neighbor search. ACM SIGMOD Record, 1998, Vol. 27, No. 2, Pp. 154–165. DOI: 10.1145/276305.276319

18. **Jégou H., Douze M., Schmid C.** Product quantization for nearest neighbor search. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, Vol. 33, No. 1, Pp. 117–128. DOI: 10.1109/TPAMI.2010.57

19. **Turov V.P., Tomilov N.A., Babaiants A.A.** Razrabotka instrumenta sravneniia algoritmov vektornogo poiska [Development of a tool for comparing vector search algorithms]. XI Kongress Molodykh Uchenykh [XI Congress of Young Scientists], 2022, Vol. 1, Pp. 446–450.

20. **Laaber C., Leitner P.** An evaluation of open-source software microbenchmark suites for continuous performance assessment. Proceedings of the 15th International Conference on Mining Software Repositories, 2018, Pp. 119–130. DOI: 10.1145/3196398.3196407

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Tomilov Nikita A.
Томилов Никита Андреевич
E-mail: programmer174@icloud.com

Turov Vladimir P.
Туров Владимир Павлович
E-mail: firemoon@icloud.com

Submitted: 30.09.2024; Approved: 09.11.2024; Accepted: 18.11.2024.
Поступила: 30.09.2024; Одобрена: 09.11.2024; Принята: 18.11.2024.

Review article
DOI: https://doi.org/10.18721/JCSTCS.17402
UDC 004.03

NEW TYPES OF THREATS AND ASSESSMENT OF QUANTUM STABILITY OF INFORMATION SYSTEMS IN THE FIELD OF FOREIGN TRADE ACTIVITY

V.Yu. Skiba[1,2] ✉, *S.A. Petrenko*[1], *A.A. Murzina*[1], *K.R. Popova*[1]

[1] Sirius University of Science and Technology, Sirius Federal Territory, Krasnodar region, Russian Federation;

[2] Bauman Moscow State Technical University, Moscow, Russian Federation

✉ vskiba69@mail.ru

Abstract. The introduction of technologies for performing customs operations through information systems without the direct participation of customs officials, as well as the evolving political situation and the adoption of sanctions against the Russian Federation by the United Stated and its allied states, significantly increase the importance of ensuring information security of customs authorities. The results obtained in the field of quantum informatics clearly demonstrate the high technological potential of quantum technologies. A cryptanalytically relevant or significant quantum computer can threaten civil and military communication systems, including information systems of customs authorities and individual participants in foreign trade activity. In this situation, there is a growing need to prepare in advance for possible collisions and take all necessary measures to protect against the mentioned quantum threat, including developing a plan for relevant priority measures.

Keywords: foreign trade activity, information security, quantum security threat, quantum stability, critical information infrastructure

Acknowledgements: The project "Technologies for countering previously unknown quantum cyber threats" was selected for support within the framework of event 2.3 of the state program of the federal territory Sirius "Scientific and technological development of the federal territory Sirius" (application registration No. FCS-2024-2.3-VY-1160-5744, leading scientist – Petrenko S.A., PhD).

Citation: Skiba V.Yu., Petrenko S.A., Murzina A.A., Popova K.R. New types of threats and assessment of quantum stability of information systems in the field of foreign trade activity. Computing, Telecommunications and Control, 2024, Vol. 17, No. 4, Pp. 16–34. DOI: 10.18721/JCSTCS.17402

Обзорная статья
DOI: https://doi.org/10.18721/JCSTCS.17402
УДК 004.03

НОВЫЕ ВИДЫ УГРОЗ И ОЦЕНКА КВАНТОВОЙ УСТОЙЧИВОСТИ ИНФОРМАЦИОННЫХ СИСТЕМ В СФЕРЕ ВНЕШНЕЭКОНОМИЧЕСКОЙ ДЕЯТЕЛЬНОСТИ

В.Ю. Скиба[1,2] ✉ , *С.А. Петренко*[1],
А.А. Мурзина[1], *К.Р. Попова*[1]

[1] Научно-технологический университет «Сириус», федеральная территория «Сириус», Краснодарский край, Российская Федерация;

[2] Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет), Москва, Российская Федерация

✉ vskiba69@mail.ru

Аннотация. Внедрение технологий совершения таможенных операций посредством информационной системы без непосредственного участия должностных лиц таможенных органов, а также складывающаяся геополитическая обстановка и принятие США и примкнувшими к ним государствами санкций в отношении Российской Федерации существенным образом повышают значимость обеспечения информационной безопасности таможенных органов. Полученные результаты в области квантовой информатики наглядно показывают высокий технологический потенциал квантовых технологий. Криптоаналитически релевантный или значимый квантовый компьютер может поставить под угрозу системы гражданской и военной связи, в том числе информационных систем таможенных органов и отдельных участников внешнеэкономической деятельности. В этой ситуации назревает необходимость заранее готовиться к возможным коллизиям и выполнять все необходимые мероприятия по защите от упомянутой квантовой угрозы, в том числе разработать план соответствующих первоочередных мероприятий.

Ключевые слова: внешнеэкономическая деятельность, информационная безопасность, квантовая угроза безопасности, квантовая устойчивость, критическая информационная инфраструктура

Финансирование: Проект «Технологии противодействия ранее неизвестным квантовым киберугрозам» был отобран для поддержки в рамках мероприятия 2.3 государственной программы федеральной территории «Сириус» «Научно-технологическое развитие федеральной территории “Сириус”» (регистрационный номер заявки ФТС-2024-2.3-VY-1160-5744, ведущий ученый – д.т.н. Петренко С.А.).

Для цитирования: Skiba V.Yu., Petrenko S.A., Murzina A.A., Popova K.R. New types of threats and assessment of quantum stability of information systems in the field of foreign trade activity // Computing, Telecommunications and Control. 2024. Т. 17, № 4. С. 16–34. DOI: 10.18721/JCSTCS.17402

Introduction

One of the key goals of the creation of the Eurasian Economic Union (EAEU) is “comprehensive modernisation and improving competitiveness of national economies within the framework of the global economy”[1]. The creation of competitive conditions directly depends on the reduction of time costs on the route of goods from the manufacturer to the end consumer [1].

The rate of technological change is constantly increasing, which leads to the need to master new skills and knowledge that allow taking into account such factors as instability, uncertainty, complexity

[1] Treaty on the Eurasian Economic Union, Available: https://docs.eaeunion.org/ru-ru/Pages/DisplayDocument.aspx?s=bef9c798-3978-42f3-9ef2-d0fb3d53b75f&w=632c7868-4ee2-4b21-bc64-1995328e6ef3&l=540294ae-c3c9-4511-9bf8-aaf5d6e0d169&EntityID=3610 (Accessed: 28.01.2025)

and ambiguity in the economic models used. The concept of economic development in the context of VUCA (volatility, uncertainty, complexity and ambiguity) conditions is becoming increasingly popular [2], which has a direct impact on the processes of foreign trade activities (FTA) management.

In these conditions, FTA acquires new specifics, which was proposed in [3] to consider from the standpoint of reducing the time costs of information and logistics operations by eliminating “false disagreements” by creating services for implementing trade operations in the format of smart contracts; for forming indicators that directly characterize the integrative aspects of trade transactions within FTA.

Information systems that directly support FTA have begun to actively develop in the direction of creating platform solutions that involve the implementation of smart contracts and blockchain technologies.

In the context of growing volumes of cross-border trade, the Federal Customs Service of Russian Federation was one of the first customs services in the world to take a course on reducing the time it takes to complete customs formalities through the digitalization of its activities and the introduction of innovative technologies for interaction with all participants involved in the process of international movement of goods [4]. In the current conditions, there was an urgent need to develop and implement fundamentally new approaches, technologies and means of performing customs operations through information systems and information and communication technologies without the participation of customs officials, i.e. in automatic mode.

At the same time, the global information space (GIS) and its main component – the Internet – is used and will be used in the foreseeable future in the interests of the functioning of critical infrastructure facilities of customs authorities and individual participants in FTA, expanding citizens’ access to information. On the other hand, it is already being used by criminal structures, international terrorist organizations and unfriendly states to disrupt the functioning of these facilities and create centers of social tension [5].

The results obtained in the field of quantum information science clearly demonstrate the high technological potential of quantum technologies. At the same time, it becomes clear that a cryptanalytically relevant or significant quantum computer could threaten civil and military communications systems and undermine the combat capability of strategic control and management systems of the Russian Federation's critical information infrastructure [6], including critical information infrastructure facilities of customs authorities and individual participants in FTA.

In these conditions, the relevance of conducting research in the field of quantum technologies and information security, the creation of safe quantum-resistant ecosystems and platforms for conducting FTA is beyond doubt.

GIS and modern threats to international information security

It is necessary to conduct a permanent detailed analysis of the current situation in the GIS, primarily an analysis of the quantity and quality of threats from states, criminals and terrorists using information and communication technologies for destructive purposes. It is also necessary to constantly clarify and classify threats to international information security (IIS), make a detailed assessment of these threats, clarify plans and the format of activities to counter threats to the IIS for the future.

The main basic concepts of IIS, the classification of IIS threats, their types and mechanisms (or channels) for their implementation are defined in [6] and subsequently clarified[2,3] in a series of works devoted to modern IIS threats. It should be noted that with the adoption of sanctions against the Russian Federation by the United States and its allied states, the number of attempts to violate the information security of various critical information infrastructure facilities, including customs authorities, has increased significantly.

---

[2] Updated concept of the convention of the United Nations on ensuring international information security, Available: http://www.scrf.gov.ru/security/information/Inf_conc/ (Accessed: 29.01.2025)

[3] Decree of the President of the Russian Federation on approval of the Fundamentals of the state policy of the Russian Federation in the field of international information security, Available: http://pravo.gov.ru/proxy/ips/?docbody=&firstDoc=1&lastDoc=1&nd=602148302 (Accessed: 29.01.2025)

Over the past decades, the scale of IIS threats has increased significantly under the influence of such a complex and contradictory phenomenon as globalization. On the one hand, in the context of globalization, the interdependence of states has increased sharply and conflicts in the GIS seriously threaten global security and stability. On the other hand, by deepening the unevenness of economic development of states, globalization creates a fertile ground for the accumulation of crisis potential in many countries of the world. It is on this basis that new types of IIS threats arise and grow, various new actors of the GIS appear, who have made violence and lawlessness their weapon in it [5].

Over the past three decades, so many threats to information security have emerged that it would seem that there is no room for new ones. Examples of the ingenuity and resourcefulness of manufacturers and developers of new means of conducting cyberattacks and hostile use of content (new mechanisms of information-technical and/or information-humanitarian influence) and, accordingly, the emergence of new threats [5].

Since 2022, many Russian services have been subjected to cyberattacks: Gosuslugi – the Unified Portal of State and Municipal Services (Functions), websites of Russian banks, courts, media, federal companies, electronic document management systems, as well as Android-based equipment. Over the past two years, there have been large-scale leaks of personal data of Russians. In the spring of 2024, on the eve of the admissions campaign, university websites were attacked, as well as TV channels, where the broadcast of the Victory Parade was interrupted.

It is now clear that cyber operations against transport infrastructure, power grids, dams, chemical plants, nuclear power plants, customs authorities, and other critical infrastructures are technically possible. Such operations can have far-reaching consequences, causing significant damage and large numbers of civilian casualties [5].

In accordance with the passport of the national program "Digital Economy of the Russian Federation"[4], ensuring the integral, sustainable and secure functioning of critical information infrastructure and services for the transmission, processing and storage of large volumes of data in the context of the growth of both classical and previously unknown (and, accordingly, poorly studied) security threats is one of the key goals of the "Information Security" project.

In the current geopolitical situation and in accordance with the Concept of ensuring information security of customs authorities, the priority areas for ensuring information security and technical protection of information of customs authorities are:

- ensuring resistance to deliberate destructive impacts on the information and telecommunications infrastructure of customs authorities;
- ensuring trust in the processes implemented (including without human participation) by automated information systems of customs authorities and the generated electronic documents;
- unconditional compliance with the requirements for ensuring information security at all stages of the creation, development, operation, use and decommissioning of components of the information and telecommunications infrastructure of customs authorities.

At the same time, the destructive impact is exerted not only on the information systems of customs authorities, but also on the information systems of participants in FTA [7] and other interested parties, information interaction with whom is ensured in accordance with the legislation of the EAEU and the Russian Federation.

There are a number of IIS threats arising within the framework of the activities of international (regional) economic organizations. Thus, in the scheme of exchanging electronic documents during cross-border interaction of government bodies of the EAEU member states with each other and with the Eurasian Economic Commission with the participation of a trusted third party, collisions arise. They are associated with the emergence of a situation in which the subject of interaction (legal entity or

[4] Passport of the national program "Digital Economy of the Russian Federation", Available: https://digital.gov.ru/uploaded/files/tsifrovaya-ekonomika-rossijskoj-federatsii.pdf (Accessed: 29.01.2025)

individual) can potentially submit to the government body an electronic document that it received from another subject located in a jurisdiction different from the government body of submission. In this case, the government body will not be able to correctly verify such an electronic document.

Another type of threat in case of violation of the cross-border transfer of electronic documents is associated with an electronic document with uncertain legal force that has already entered the recipient's jurisdiction, but its legal force has not yet been confirmed using receipts from a trusted third party. For example, it is possible that an electronic document was signed with an electronic signature and sent to the recipient, but before the recipient initiated the procedure for generating receipts, the public key certificate for verifying the signature was compromised. Obviously, such an electronic document will have no legal force in the recipient's jurisdiction. At the same time, at the time of the electronic document's formation and during its transfer from the sender to the recipient, the document had all the properties that allow it to be considered legally significant [5].

The achievements of IBM, as well as a number of other high-tech manufacturers of quantum computers, convincingly demonstrate the reality of the so-called “quantum threat”. For this reason, a number of technological countries in the world have already begun preparations to counter future quantum cyberattacks. For example, the administration of former US President Joe Biden has issued two new directives to prepare the state and business for future quantum cyberattacks.

Thus, it should be considered that computer attacks and impacts using specially prepared content (information-technical and information-humanitarian impacts) would constantly develop, and their number would grow. Therefore, it is necessary to systematically update the existing lists of information security threats and conduct predictive research in this area in order to counter them in the near and medium term.

Automated information systems in the customs sphere

Currently, the customs sphere not only ensures 100% electronic declaration[5] [1, 4], but also various customs technologies for carrying out customs operations through information systems without the participation of customs officials are actively used.

Thus, for the first time, self-regulating mathematical methods, algorithms and software for format-logical control, interdepartmental exchange and verification of permits, as well as decision-making by information systems without the participation of customs officials have been developed and implemented.

In this case, electronic signature mechanisms are used to ensure the legal significance of decisions made by information systems. The proposed system-technical and information-technological solutions ensured the implementation of technologies for automatic registration of customs declarations, automatic verification of the risks of violation of the customs law of the EAEU and automatic release of goods in accordance with the declared customs procedure, as well as writing off customs duties and payments from the single personal accounts of participants in FEA [1].

Information systems of the customs authorities of the Russian Federation ensure the implementation of information and communication technologies used in the performance of customs operations and customs control of goods and vehicles, the use of the risk management system, accounting and control of the completeness and timeliness of receipt of customs payments and their payment, the maintenance and analysis of customs statistics of foreign trade, currency control, the analysis and assessment of the effectiveness of the activities of customs authorities, the implementation of other functions assigned to customs authorities in the field of customs affairs in accordance with the current EAEU Customs Code.

The basis for the implementation of various customs technologies was the creation, since 1998, of information systems of customs authorities in a secure design [1], first of all, Unified automated information

[5] The first electronic declaration was submitted to customs authorities on 25.11.2002.

system of customs authorities (UAIS CA)[6]. This system is designed to ensure automation of the activities of customs authorities and the implementation of information customs technologies in accordance with the legislation of the EAEU, the legislation of the Russian Federation in the customs sphere, as well as other relevant regulatory legal and acts of the Russian Federation and international treaties. The UAIS CA is a hierarchical multi-level information system corresponding to the organizational and staff structure of the customs authorities of the Russian Federation.

The set of components (objects) of the UAIS CA is operated in the central office of the Federal Customs Service of Russian Federation, specialized and regional customs departments, customs offices, at customs posts and checkpoints on the external border of the EAEU.

Information interaction between customs authorities is carried out using the Departmental Integrated Telecommunications Network (DITN) of the Federal Customs Service of Russian Federation, which is a system of telecommunications nodes of customs authorities that are interconnected according to the hierarchical principle.

Since 2002, when the first departmental certification center of the State Customs Committee of Russian Federation was created and the first electronic declaration of goods was filed, customs authorities have consistently created, put into operation and developed a system of departmental certification centers of customs authorities and automated information system for external access of customs authorities[7].

The automated information system for external access of customs authorities ensures secure information interaction between the information and software systems of the UAIS CA and external information systems (information systems of FTA participants, federal executive authorities, customs authorities of foreign states, including members of the EAEU, international organizations, etc.).

The system of departmental certification centers of customs authorities ensures the issuance of qualified certificates of electronic signature verification keys to customs officials and server components of the UAIS CA. Qualified certificates of the electronic signature key are required to create electronic documents within the framework of the procedures for automatic registration of declarations and automatic release of goods, as well as to check the status of a document at various stages of customs clearance and customs control, including after the release of goods within five years.

Since 2002, information interaction with the information systems of FTA participants and other interested parties has developed in the direction of a complete transition to the use of legally significant electronic documents.

Since 2018, the implementation of customs technologies has begun, which provide for the adoption of legally significant decisions when carrying out customs operations through information systems without the participation of customs officials.

Since November 2021, the process of automatic processing of goods declarations has been carried out by the customs authorities' information system around the clock. If a decision on the automatic release of goods in accordance with the declared customs procedure is not made by the information system, then such a declaration is automatically sent to the customs authority (electronic declaration center), whose officials make the final decision on the submitted declaration “manually”. Simultaneously with the goods declaration, the official receives the results of inspections carried out by the information system, which he uses to make decisions. In this case, the decision is made by officials during the working hours of a specific customs authority where the declaration was received (most electronic declaration centers work daily in a 12-hour mode).

Along with this, work was carried out to develop the information and software tools of the UAIS CA, implementing in automatic mode the verification of format-logical control, algorithms for automatic registration of customs declarations and algorithms for the automatic release of goods.

---

[6] Order of the Federal Customs Service of Russia dated 17.06.2010 No. 1154 “On approval of the Regulation on the Unified Automated Information System of Customs Authorities”, Available: https://www.alta.ru/tamdoc/10pr1154/?ysclid=m6izdrt2k5240885086 (Accessed: 30.01.2025)

[7] In 2002, the Departmental Certification Center of the State Customs Committee of Russia began its work, on the basis of which a system of departmental certification centers of customs authorities was subsequently created, which was put into operation on 11.03.2005. On 06.04.2009, an automated information system for external access of customs authorities was created and put into operation.

The digitalization of customs authorities' activities was accompanied by changes in international and national legislation. For example, the Customs Code of the Customs Union established the priority of using paper documents, and all customs operations were provided for only by customs officials. Since 2014, the electronic form of declaration has become mandatory in the Russian Federation, with the exception of cases determined by the Government of the Russian Federation or supranational legislation. Since 01.01.2018, the Customs Code of the EAEU has already entered into force, which established the priority of the electronic form of documents and the possibility of carrying out customs operations through an information system without the direct participation of customs officials.

Thus, currently developed technologies for performing customs operations through the information systems without the participation of customs officials (in automatic mode) involve the use, together with the UAIS CA, of the System of Departmental Certification Centers of Customs Authorities and the Automated System of External Access of Customs Authorities, ensuring interaction with the information systems of FTA participants and other interested parties, as well as the System of Interdepartmental Electronic Interaction, ensuring interaction with the information systems of more than 40 other federal executive authorities).

The transition of the customs authorities of the Russian Federation to the use of digital technologies served as the basis for the use of electronic documents by other federal executive authorities whose powers include the issuance of permits necessary for the movement of goods and vehicles across the customs border of the EAEU.

The introduction of an electronic form for submitting documents, the organization of obtaining permits in electronic form directly from the agency that issued them, and a number of other measures related to the organization of electronic document flow between participants in the cross-border movement of goods, made it possible for FTA participants to submit an electronic declaration for goods from any point in the Russian Federation and, as a result, eliminate the need for the personal presence of a representative of a participant in FTA at the customs authority when declaring goods.

The development of information systems of FTA participants took place, although the development and implementation of the CUCA interaction model scheme [3], which ensures increased management efficiency based on the processing of data on goods and thereby minimizing the impact of the so-called VUCA processes, had a greater influence on the development of information systems in the sphere of FTA.

When creating a digital information and logistics platform under VUCA conditions, it becomes possible to control all routine operations: from concluding a contract to successfully closing a deal, subsequent warranty service and compliance with the rights to the intellectual property used, processing and storage in distributed ledgers in the format of smart contracts. Automatic execution of smart contracts reduces the risk of logistical errors, unifies the document flow process, visualizes in real time the processes of movement of goods and the status of customs documents [3].

This model scheme consists of coordinating operations: data transfer during the implementation of business processes, defining the goal, in terms of knowledge of laws and regulations of FTA, observability of the current state of resources and controlled distribution of decision-making powers (Fig. 1).

The CUCA model scheme allows for a radical increase in the reliability and security of the chain of foreign trade transactions, the level of automation of accounting and control systems, a reduction in the time and financial costs of exporting enterprises, and the introduction of the principle of one-time provision of data (the “single window” principle) [8, 9].

The use of CUCA significantly reduces the potential for fraud, violation of liability regulations, certification rules and the use of false information. Thanks to the peer-to-peer architecture of the CUCA database, the FTA information space is endowed with new functions, which allows the exporter and importer to directly exchange legally significant documents organized in the form of a cryptographically protected chain of blocks (blockchain). The correctness of the information update is confirmed by all nodes of the blockchain network, and specially organized chains of records (sidechain) allow customs

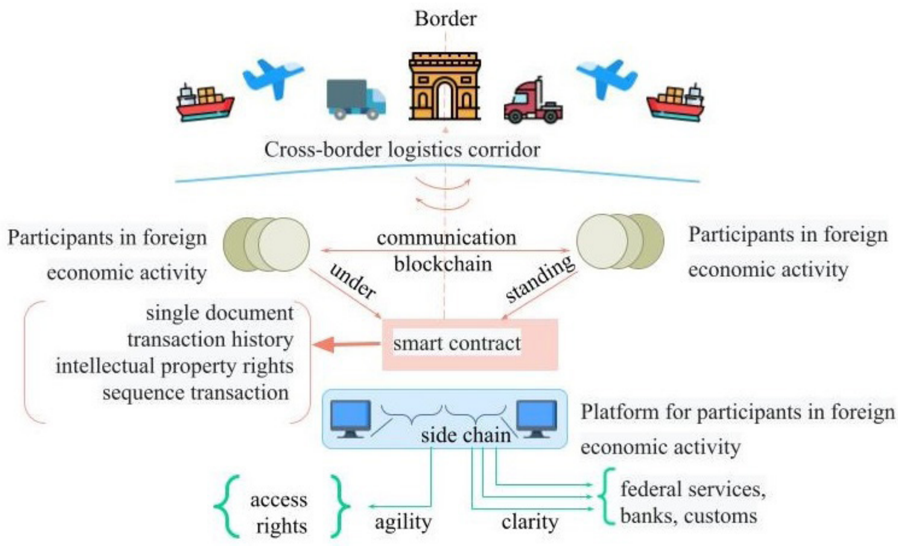


Fig. 1. CUCA model scheme

and currency control authorities to automatically receive the necessary information, which reduces administrative costs and speeds up all FTA processes.

Considering that distributed ledgers are decentralized network-centric technology for storing and processing data that includes meta-information about previous operations, a new addition operation using ledgers is executed taking into account the context of previous operations. This ensures verification of new operations in the future.

Each result of a successful operation is called “smart contract” and stored in a distributed ledger and is duplicated multiple times in the system, which ensures automated control of its correctness, radically reducing the number of violations in the preparation of documentation and customs clearance, reducing logistics costs due to:

- processing of all transactions in a common cryptographically protected information space of manufacturing enterprises, licensed exporting enterprises, customers of high-tech products, financial organizations and state customs control authorities;
- creating a distributed cross-border digital infrastructure for operational management and planning of FTA;
- integrating production and logistics capabilities for export-import operations, typical for enterprises engaged in high-tech production.

In the context under consideration, the essence of digital transformation of FTA comes down to using the capabilities of end-to-end information and computing systems and IoT networks to stimulate the activities of industrial enterprises to dramatically increase the volume of exports of high-tech products, including changes in business models of relationships with partners and competitors through the introduction of smart contracts.

An information and logistics model of FTA (using the example of transportation between Russia and China) with transaction control technology using distributed ledgers and smart contracts, is presented in Fig. 2.

All regulatory financial and logistical operations, up to the successful closing of a deal and subsequent warranty service, are proposed to be recorded and stored in distributed secure ledgers. In this case, the basis of FTA is smart contracts (application, description, letter of credit etc.), which automatically guarantee the regulatory correct settlement of requirements in accordance with the model of FTA adopted by the company and approved by the customs authorities (smart contract FTA).

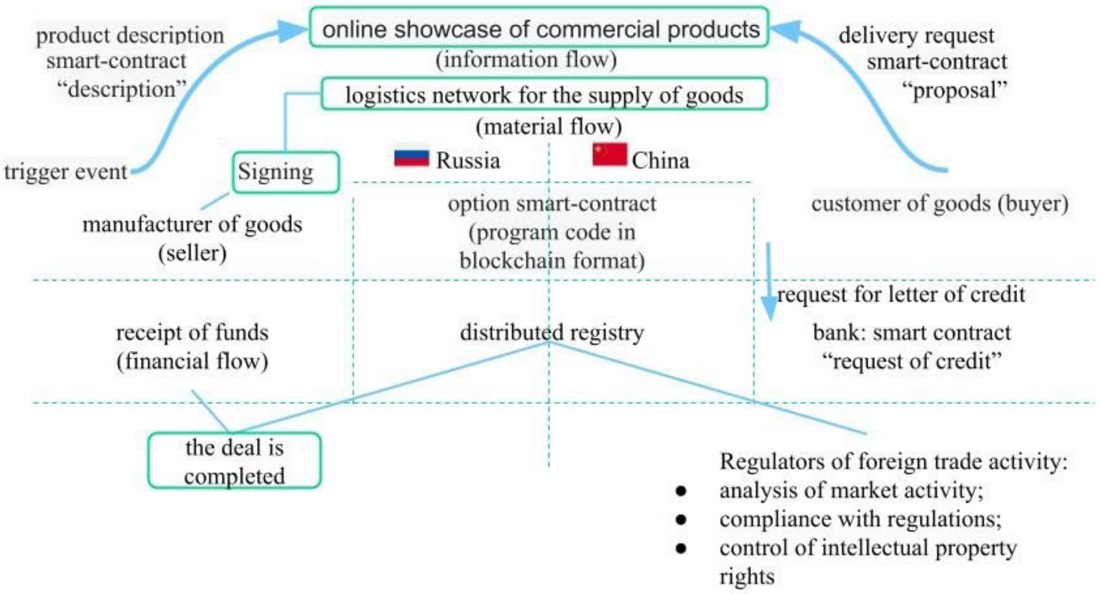


Fig. 2. Information and logistics model

It should be noted that a smart contract implements an algorithm executed in a local or cloud computing environment, which describes the sequence of fulfillment of contractual obligations agreed upon by the supplier and customer. When creating an information platform for the implementation of Russian-Chinese FTA, it must be taken into account that the processed data used, for example, to certify goods in accordance with national quality standards, are in "constant motion" [10], the trajectory of which can be clearly displayed on a multi-layer telematics map.

In modern conditions, the formation of a register of "obligations plus regulations", which characterizes the time sequence of fulfillment of contractual obligations, using traditional paper declarations significantly reduces the speed of trade transactions and at the same time increases the amount of inaccurate information associated with the influence of the "human factor". That is why the considered information and logistics model of foreign trade operations is based on digital technologies that allow the creation of a new class of network infrastructure for the implementation of FTA or the "Internet" of economically significant values. Such infrastructure can be created on the basis of an intelligent digital information and logistics platform that implements and controls material and information flows formed during the implementation of export contractual obligations and customs control regulations.

As noted above, business processes occurring during the implementation of FTA can be used to design smart contracts at the normative-algorithmic level regulating the sequence of both preparatory and customs-logistics operations [11, 12]. In this case, preparatory operations include the processes of selling and purchasing finished products or materials, drawing up relevant foreign trade contracts and carrying out banking transactions. Customs and logistics processes include the processes of obtaining permits (certificates of conformity, etc.), passing customs formalities, handling cargo along the entire route or transferring goods using information carriers through computer networks.

In the digital information and logistics platform, the distributed ledger actually functions as an information integrator, which receives information from both participants in export transactions and from equipment that forms the IoT class transport infrastructure used by the logistics operator or IaaS service provider. RFID sensors, digital video cameras, temperature and humidity sensors, GPS navigators and other telematic devices that transmit information about the current state, route, and location in the warehouse or other logistic attributes of the goods being transported can be used as sources of data processed during customs support of the smart contract.

At the same time, the technology of distributed ledgers at the algorithmic level fixes the list of requirements for the goods, their transportation and storage, which is mandatory for the supplier and the logistics operator company. Subsequently, the smart contract automatically prepares financial flows and forms a neutral account, where the recipient of the goods transfers payment for the completed goods transaction. At the same time, the funds are in this account until the goods cross the customs border and become the object of control of the smart contract.

An analysis of the structure of the digital information and logistics platform shows that the information used in it, being a carrier of events and facts, forms a distributed ledger or a set of data ordered according to certain rules, which are endowed with attributes that characterize how both rights and fact their ownership. These attributes ensure “high accuracy” of data accounting and management, which opens up fundamentally new possibilities for the use of distributed ledgers for the implementation and control of foreign trade transactions. Increasing the accuracy of data accounting reduces uncertainty in the planning process and economic risks of logistics transactions, which creates the preconditions for a radical change in business models based on the fulfillment of mutual obligations. This is especially important in the implementation of FTA, as it allows for the automation of the processes of control over the fulfillment of obligations, analyzing the ledgers of financial resources of suppliers, clients, intellectual property and applicable standards based on agreed algorithms. Considering that contract ledgers or smart contracts can be complete (take into account any possible situations) or open (can be supplemented or changed depending on the situation), the operational management of economic processes within the framework of FTA can be considered as a collective and non-authoritarian technology that allows for immediate and public confirmation of the accuracy and authenticity of the data provided.

The CUCA model scheme that forms the basis of the digital information and logistics platform allows for increasing the efficiency of operational management of logistics operations and automating the processes of FTA control by customs authorities. At the same time, enterprises exporting high-tech products have the opportunity to use new business models of FTA that increase the reliability, security and speed of trade operations.

Using similar approaches, the information system of the Russian Export Center is being developed. The digital platform “My Export” (state information system “Single Window”) and the National digital transport and logistics platform (DTLP) are being created.

My Export platform was created within the framework of the national project “International cooperation and export”, which is aimed at increasing the export of non-resource non-energy goods. Eleven relevant ministries, federal executive bodies and business associations worked together with the Russian Export Center to create the platform, including the Ministry of Industry and Trade of Russian Federation, the Ministry of Agriculture of Russian Federation, Rosselkhoznadzor, the Federal Customs Service of Russian Federation and the Federal Tax Service of Russian Federation. The digital platform “My Export” provides online access to government and business services that support companies' export activities. The platform's services provide solutions to key tasks at each stage of the export cycle[8].

The National DTLP is being created as a state information system that defines uniform standards of digital interaction for all participants in transport and logistics activities and government agencies. It is designed to unite digital logistics services on one platform and become a kind of a “single window” for interaction between the state and carriers. One of the goals is to implement electronic document management at all stages of cargo transportation by road, sea, river, rail and air[9].

[8] Over 115 thousand services rendered and over 23 thousand users. Digital platform "My export" is three years old. News of Russian export. Available: https://www.exportcenter.ru/press_center/svyshe-115-tysyach-okazannykh-uslug-i-bolee-23-tysyach-polzovateley-tsifrovoy-platforme-moy-eksport-/?ysclid=m6jdtzaob1137958268 (Accessed: 30.01.2025)

[9] National Digital Transport and Logistics Platform | Ministry of Transport of the Russian Federation, Available: https://mintrans.gov.ru/activities/297/367?ysclid=m6je3lweyh694738446 (Accessed: 30.01.2025)

Threats to the information infrastructure of customs authorities associated with the development of quantum computing technologies

The creation of secure quantum-resistant ecosystems and platforms for the digital economy of Russia, including the FTA sphere, is a long-term challenge due to the current lack of unified scientific, methodological and technical base for creating the aforementioned systems. In particular, there are no technologies to counter the new quantum threat to cybersecurity. This is a fundamental scientific problem, without solving which it is impossible to talk about achieving the goals of the national project "Data Economy".

According to [6], today we are in the so-called era of noisy intermediate-size quantum devices (NISQ) [13–19].

Quantum computers are capable of solving some computational problems much more efficiently than any modern classical computer (fifth-generation supercomputer) of the von Neumann architecture [20, 21].

The components of a quantum computer that can be implemented in practice are imperfect in terms of accuracy and are highly susceptible to interference and errors. However, if these components are used in combination with classical computers and fifth-generation supercomputers, it is possible to achieve significant acceleration in calculations in the field of solving a wide class of multidimensional optimization problems. It should be noted that these problems include the problems of ensuring information security for critical infrastructure and problems associated with information impact (primarily unfriendly) on this infrastructure.

Here the question of "price vs quality" arises: how much more expensive is such a quantum component of a computing system than a classical one, capable of doing the same work, but in a longer time?

The results obtained in the field of quantum information scince clearly demonstrate the high technological potential of quantum technologies. At the same time, it becomes clear that a cryptanalytically relevant or significant quantum computer can threaten civil and military communication systems and undermine the combat capability of strategic control and management systems of critical information infrastructure [6], including critical information infrastructure facilities of customs authorities and individual participants in FTA.

In this situation, there is a growing need to carry out all necessary measures to protect against the aforementioned quantum threat, including developing a plan of relevant priority measures at the state and military levels.

Touching upon the task of ensuring the security of critical information infrastructure in the context of the introduction (full-scale or limited) of quantum computing systems and quantum computing based on them, one can try to identify two main "scenarios" for the development of events [21].

The first scenario assumes the direct use of quantum technologies in computing facilities that provide control, processing of all types of information and modeling of all levels in the elements of the critical information infrastructure facility. It is obvious that quantum computing will be used in this case in parallel with calculations and control operations implemented on classical computing facilities (modern supercomputers in this approach will also be considered classical facilities).

The second scenario assumes that quantum technologies in the near future will not find wide application in critical information infrastructure due to insufficient reliability and maturity, and will most likely be used outside the real control and information processing circuits to solve individual local computationally intensive tasks. In this case, the processes of forming "hostile" information impacts can be implemented using quantum technologies outside the real control and computing circuits with subsequent "introduction" into structures related to critical information infrastructure facilities that are potential targets of impact.

It should be noted that the second scenario may become feasible in the near future, while the first one is associated with the need to master the serial production of basic elements that are (by analogy with

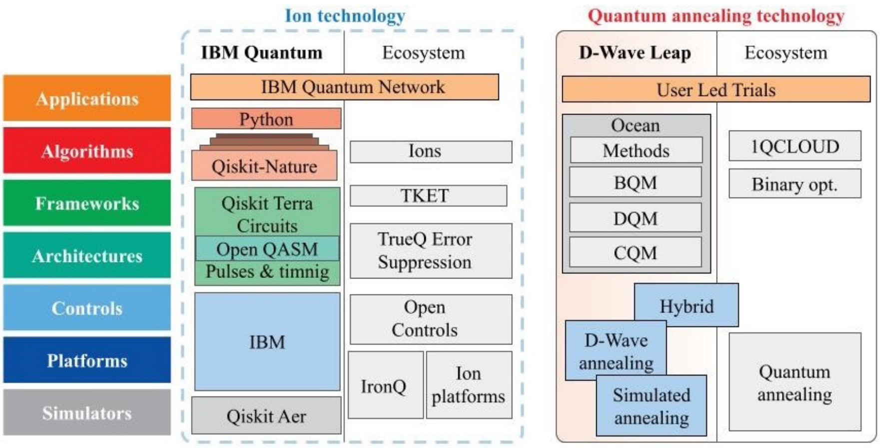


Fig. 3. The emergence of the first quantum computing technologies

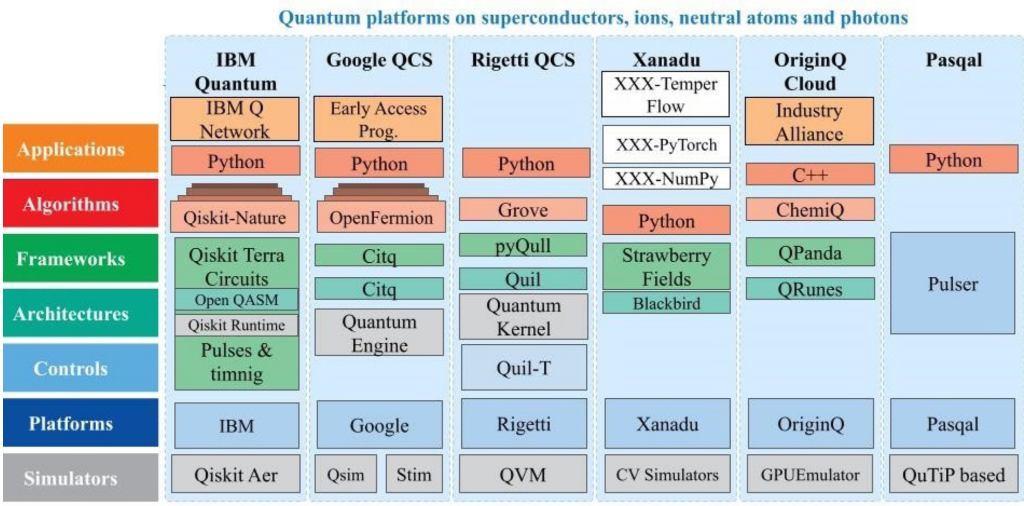


Fig. 4. The first quantum computing systems

microelectronic chips) the basis for hardware solutions for quantum computers (so-called "quantum chips"). The general situation here is that, despite active research in the field of quantum computers, there are still no technologies for the mass production of "quantum chips". Moreover, a physical platform for the creation and production (at least small-scale) of quantum computers has not been selected.

Currently, researchers around the world are working on the creation of quantum computers on four main platforms (Fig. 3 and 4): superconductors, ions, neutral atoms and photons.

The main features of these platforms are studied in [6]. The first prototypes of quantum computers do not differ fundamentally from each other in performance and are practically at the same stage of development. At the same time, if we talk about specific models of quantum computers, then at present, computing devices on superconductors have become more widespread.

The situation in the field of quantum computing is characterized by a kind of "technological race" between leading companies. Periodically, there are reports of achieving "quantum supremacy", i.e. the ability to solve problems that are impossible for classical von Neumann supercomputers.

Thus, in 2019, Google claimed to have achieved quantum supremacy on a 54-qubit array[10].

[10] **Arute F., Arya K., Babbush R. et al.** Quantum supremacy using a programmable superconducting processor. Nature, 2019, Vol. 574, 505–510. DOI: 10.1038/s41586-019-1666-5

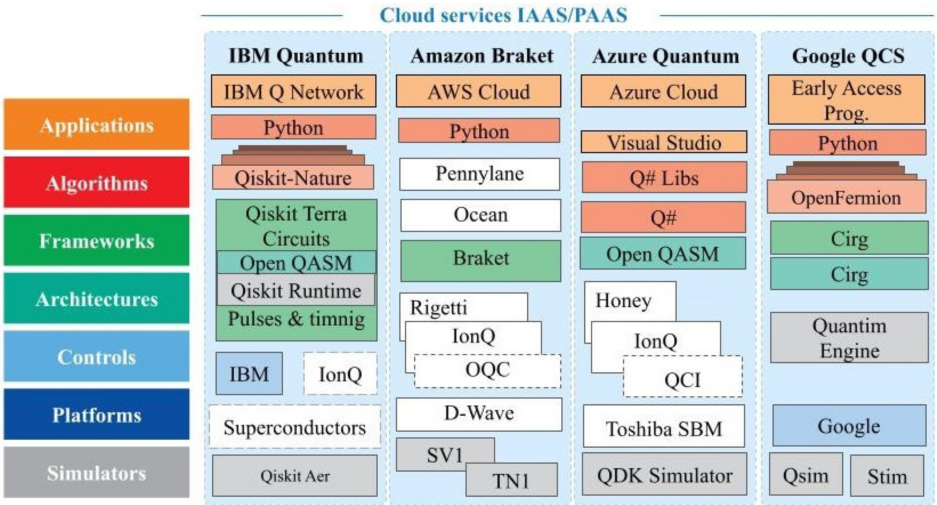


Fig. 5. First quantum computing services

In 2021, Chinese scientists reported[11] achieving quantum supremacy with the Jiuzhang 2 quantum computer on photons (the problem of sampling Gaussian bosons was solved 107 times faster than on classical supercomputers).

At the IBM Quantum Summit 2022[12], the Osprey quantum processor was presented, consisting of 433 qubits. Google, IBM, and D-Wave have provided access to their prototypes of cloud quantum computers under the Iaas and PaaS models. IBM plans to develop a quantum system with more than 4000 qubits by 2025. Google plans to introduce a cloud quantum computer with 1 million qubits by 2029. For comparison, the current leader among cloud computers is the Canadian company D-Wave with a 7000-qubit D-Wave Advantage 2 processor based on quantum annealing technology. An open cloud service Leap has been developed to work with this computer, which allows you to create and run various quantum applications (Fig. 5 and 6).

In Russian Federation, scientific research is also being conducted aimed at creating the first domestic quantum computers. For example, scientists from the Russian Quantum Center and the P.N. Lebedev Physical Institute of the Russian Academy of Sciences have developed a prototype of a quantum computer based on ytterbium ions[13].

The emergence of a relevant quantum computer capable of cracking traditional cryptographic algorithms is expected in the period 2026–2030.

In the context of the emergence of a new quantum security threat, it is necessary to set and solve the problem of ensuring the stability of customs authorities' information systems in such a way that quantum stability is ensured for technologies for performing customs operations through information systems without the participation of customs officials when making the following decisions:

- registration of goods declarations;
- release of goods and vehicles;
- registration of transit declaration;
- issue of transit declarations;
- risk level category of FTA participants;
- results of format-logical control of goods declarations;

[11] The Jiuzhang 2.0 Photonic Quantum Computer, Available: https://www.youtube.com/watch?v=R57M0SmTPHI (Accessed: 31.01.2025)
[12] The Next Wave – IBM Quantum Summit 2022 Keynote, Available: https://www.youtube.com/watch?v=8ySjHqfioJM (Accessed: 31.01.2025)
[13] **Dorohova I.** Sozdan prototip kvantovogo komp'yutera na ionah itterbiya. 2022, Available: https://strana-rosatom.ru/2022/02/25/sozdan-prototip-kvantovogo-kompjute/ (Accessed: 31.01.2025)

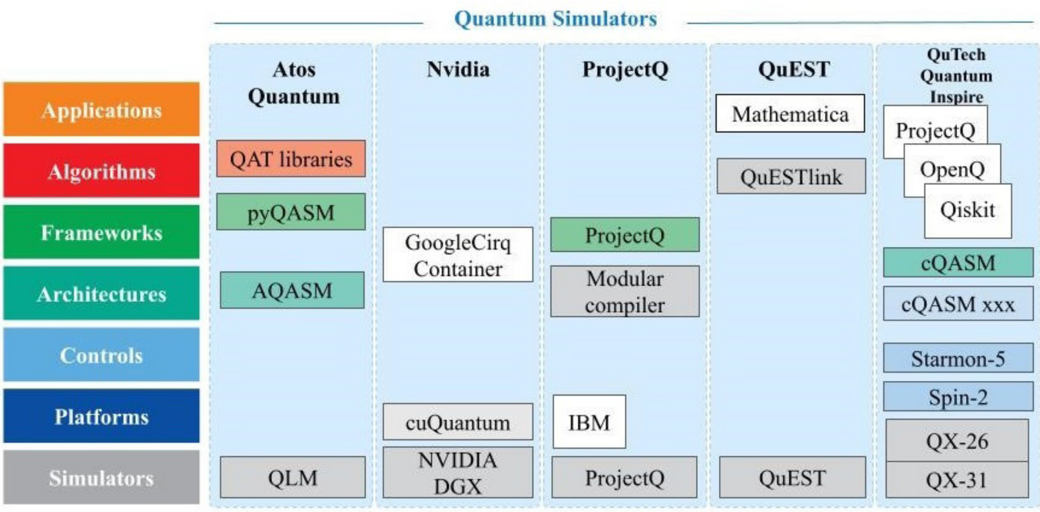


Fig. 6. First models of quantum simulators

- results of reconciliation of permits with the information declared in goods declarations;
- accrual and write-off of customs duties and fees.

At the same time, the main problematic issues of ensuring quantum stability of critical information infrastructure facilities of customs authorities include:

- insufficient level of readiness for the growth of quantum cyberattacks by intruders;
- growing complexity of the structure and behavior of critical information infrastructure facilities of customs authorities in the context of unfinished import substitution and technological security;
- difficulty of identifying quantitative patterns that allow us to study the cyber-resilience of customs authorities' critical information infrastructure facilities in the context of classical and quantum cyber-attacks by intruders.

Ignorance of the above-mentioned problematic issues leads to a decrease in the efficiency of the functioning of the critical information infrastructure facilities of customs authorities.

Moreover, this problem is significantly aggravated by the growth in the number of classical and quantum cyberattacks by intruders. Of particular concern are the so-called quantum attacks or attacks using a quantum computer. The fact is that most of the crypto primitives used in modern information systems (including hash functions, electronic signatures, asymmetric cryptographic algorithms and corresponding protocols) are no longer resistant to such attacks.

Today, effective quantum algorithms are known, in particular, Shor's algorithm for factorization and discrete logarithm, which can be successfully used to hack the listed crypto primitives [22–24].

The functioning of critical information infrastructure facilities are also strongly influenced by factors of the external and internal environment, which are either fundamentally impossible to manage, or can be controlled with an unacceptable delay. In addition, the external and internal environments have incomplete certainty of their possible states in the future.

That is, the factors influencing the behavior of critical information infrastructure facilities of the Russian Federation undergo changes over time that can radically change their functioning algorithms or make the set goals unattainable. Changes in the external and internal environment occur regularly and randomly, therefore generally they cannot be accurately predicted, resulting in uncertainty in their values. These facilities have a certain “safety margin” – features that allow achieving the set goals with certain deviations in the influencing factors of the external and internal environment.

Until recently, two main approaches were used to identify the above-mentioned patterns of functioning of critical information infrastructure facilities: experimental (for example, methods of mathematical statistics and experimental design) and analytical (for example, analytical verification methods). Unlike experimental methods, which make it possible to study the individual behavior of a critical information infrastructure facility, analytical methods allow us to consider the most general properties of the behavior of this facility, characteristic of the class of functioning processes as a whole. These approaches have significant shortcomings.

For experimental methods, this is the impossibility of extending the results obtained during the experiment to other behavior of a critical information infrastructure facility that differs from the one studied, and for analytical verification methods, this is the difficulty of moving from a class of processes of functioning of a critical information infrastructure facility, characterized by the derivation of generally significant properties, to a single process that is additionally characterized by corresponding conditions of functioning (in particular, specific values of the behavioral parameters of this facility under conditions of classical and quantum cyberattacks by intruders).

Consequently, each of the named approaches separately is insufficient for effective research of quantum stability of customs authorities' critical information infrastructure facilities. The necessary mathematical apparatus for identifying the required quantitative patterns of behavior and ensuring quantum stability of critical information infrastructure facilities can be obtained only by using the strengths of both approaches and combining them.

Thus, the practice of operating and maintaining critical information infrastructure facilities both in customs authorities and in other areas indicates the following. The conditions of modern confrontation in cyberspace impart to the mentioned facilities features that exclude the possibility of creating quantum-resistant facilities of critical information infrastructure of customs authorities by traditional methods.

At present, three main directions can be identified for resolving the scientific problem of ensuring quantum stability of information systems.

The first direction is the justification and preparation for the transition to the emerging domestic post-quantum crypto-primitives of blockchain and electronic signature. For example, to the post-quantum electronic signature “Rosehip” (2022), the stability of which is based on the mathematical problem of decoding a random linear code, which is computationally complex, and to the protocol for generating a common key based on the apparatus of isogenies of supersingular elliptic curves “Forsythia” (2022).

The second direction is the justification of the application of the first quantum-resistant solutions of quantum cryptography on physical principles and laws of quantum mechanics with mathematically provable stability. Including data transfer protocols that cannot be intercepted and decrypted unnoticed, quantum key distribution systems, quantum generators of truly random numbers, etc.

The third direction is the creation of a fully quantum model of customs authorities' information systems. It is clear that such an approach will require the creation of a full-fledged quantum (physical) infrastructure of customs authorities, which is a distant prospect.

Note that before this, information theory and information security dealt exclusively with attacks by intruders using traditional von Neumann computers. For example, using the recommendations and corresponding Threat Models of domestic regulators[14].

In practice, a number of interesting results have already been obtained. For example, high-speed hardware symmetric key encoders and quantum key distribution devices, which provided these encoders with secret keys, were integrated into highly loaded communication channels. A dedicated fiber optic core was used to transmit single photons. Since networks are often overloaded, scientists are working to ensure that quantum and classical signals can coexist in one fiber optic at different wavelengths. Thus, a pilot project of the Rosatom state corporation connected two offices of the organization in Moscow.

[14] Methodology paper dated February 5, 2021, Available: https://fstec.ru/dokumenty/vse-dokumenty/spetsialnye-normativnye-dokumenty/metodicheskij-dokument-ot-5-fevralya-2021-g?ysclid=m6kmwynfoo336914156 (Accessed: 31.01.2025); MITRE ATT&CK®, Available: https://attack.mitre.org (Accessed: 31.01.2025)

This project is a pilot fiber-optic communication line using quantum key distribution technology and was implemented with the support of Rostelecom PJSC on equipment from the QRate research and production company.

It is interesting that during the testing, a quantum channel rupture was simulated, during which the secret key storage buffer worked. Successful testing confirmed the required level of reliability for the implementation of the current solution in the network infrastructure. Another project is being implemented by Russian Railways, which is responsible for the implementation of the roadmap for the development of the high-tech field of Quantum Communications in the Russian Federation. The project involves the creation of an 800 km long quantum network, based on domestic solutions ViPNet Quantum Trusted System from InfoTeKS. At present, the largest backbone quantum network in Europe from Moscow to St. Petersburg has already been created.

At the same time, the following must be taken into account. Firstly:

- relative youth of the field, and therefore insufficient study and trust in post-quantum crypto primitives (performance and security issues). For example, in 2023, researchers from the Royal Institute of Technology in Sweden discovered a vulnerability in the post-quantum algorithm CRYSTALS-Kyber, one of the finalists of the well-known NIS competition;
- emergence of quantum algorithms that effectively solve "new" mathematical complex problems, i.e. "new" post-quantum crypto-primitives immediately become unstable. The sensational story of a new algorithm by Chinese scientists based on the Schnorr method, which used quantum acceleration to obtain approximate results for one of its stages – solving the problem of finding a short vector in a lattice of small dimension;
- emergence of a large number of open and commercial libraries for developers of digital platforms, SDKs implementing new cryptographic schemes, and, consequently, a high probability of the presence of so-called undeclared capabilities and software backdoors (up to 95% of the software code of open libraries contains the aforementioned backdoors);
- lower efficiency of post-quantum cryptographic schemes compared to classical ones: large sizes of keys, ciphertexts and signatures, low productivity etc.;
- practical complexity of a mass transition to post-quantum schemes and the unclear timeframe for implementing such a transition, etc.

Secondly, the creation and transition to quantum-resistant solutions and components of the RF CII facilities based on quantum cryptography with mathematically provable resistance. Including data transfer protocols that cannot be intercepted and decrypted unnoticed, quantum key distribution systems, quantum generators of truly random numbers.

Well-known Russian mathematicians have made significant contributions to this area of knowledge, for example, employees of the Steklov Mathematical Institute of the Russian Academy of Sciences:

- Volovich I.V., PhD, – head of the Department of Mathematical Physics, corresponding member of the Russian Academy of Sciences;
- Kholevo A.S., PhD, – head of the Department of Probability Theory and Mathematical Statistics, laureate of the Claude E. Shannon Prize for outstanding achievements in information theory, academician of the Russian Academy of Sciences.

Conclusion

An analysis of the information and software tools, technologies and information resources of the customs authorities of the Russian Federation has shown that they are currently the only complex in the Russian Federation that ensures the performance of legally significant actions electronically around the clock without human participation (registration of customs declarations, automatic verification of risks of violation of the customs legislation of the EAEU and automatic release of goods in accordance with the declared customs procedure).

In the future, the implementation of automatic customs operations will be transformed into a new vector – automatic business processes, for example, automatic intelligent control at checkpoints (using elements of artificial intelligence). One of the elements of such an intelligent business process is already a qualitatively new project of the Federal Customs Service – analysis of images of inspection and screening complexes using machine learning technology.

The results obtained in the field of quantum information science clearly demonstrate the high technological potential of quantum technologies. A cryptanalytically relevant or significant quantum computer can threaten civil and military communication systems and undermine the combat capability of strategic control and management systems of critical information infrastructure.

It has been shown that critical information infrastructure facilities, including those of customs authorities, do not have the required stability for their intended functioning in the face of previously unknown quantum attacks by intruders.

In this situation, there is a growing need to prepare in advance for possible collisions and carry out all necessary measures to protect against the aforementioned quantum threat, including developing a plan for relevant priority measures at the level of the Federal Customs Service of Russian Federation and, possibly, at the level of FTA participants who ensure the largest volume of customs operations or carry them out in the interests of critical information infrastructure facilities in other areas.

REFERENCES

1. **Skiba V.Yu., Pozdnyakova K.E.** Modern automated information systems for customs operations without the participation of customs officials. Vestnic of Russian Customs Academy, 2022, Vol. 59, no. 2, pp. 19–33. DOI: 10.54048/20727240_2022_02_19

2. **Giles S.** How VUCA is reshaping the business environment, and what it means for innovation. Forbes, 2018. URL: https://www.forbes.com/sites/sunniegiles/2018/05/09/how-vuca-is-reshaping-the-business-environment-and-what-it-means-for-innovation/ (Accessed: 10.12.2024)

3. **Zaborovsky V.S., Lei Dzhan, Skiba V.Yu., Strekalov S.V.** Digital information and logistic platform for operational management of foreign trade activities of high-tech products suppliers. St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunications and Control Systems, 2018, Vol. 11, no. 4, pp. 7–20. DOI: 10.18721/JCSTCS.11401

4. **Bulavin V.I., Vasil'ev D.A., Skiba V.Yu., Tereshchenko D.V. et al.** 30 let avtomatizatsii tamozhennykh organov. Ot zamysla sozdaniia avtomatizirovannykh sistem obrabotki informatsii v organakh gosudarstvennogo tamozhennogo kontrolia do Strategii razvitiia tamozhennoi sluzhby Rossiiskoi Federatsii do 2030 goda [30 years of automation of customs authorities. From the idea of creating automated information processing systems in state customs control bodies to the Strategy for the Development of the Customs Service of the Russian Federation until 2030]: monograph. Moscow: TsBT, 2021. 304 p.

5. **Kazarin O.V., Skiba V.Yu., Sharyapov R.A.** Novye raznovidnosti ugroz mezhdunarodnoi informatsionnoi bezopasnosti [New types of threats to international information security]. History and Archives, 2016, Vol. 3, no. 1, pp. 54–72.

6. **Stupin D.D., Petrenko A.S., Petrenko S.A.** Razvitie tekhnologii kvantovykh vychislenii i sviazannye s nim ugrozy dlia kriticheskoi informatsionnoi infrastruktury rossiiskoi federatsii [Development of quantum computing technologies and associated threats to the critical information infrastructure of the Russian Federation]. XVI Vserossiiskaia Mul'tikonferentsiia po Problemam Upravleniia [All-Russian Multi-Conference on Management Problems] (MKPU-2023), 2023. pp. 168–172.

7. **Skiba V.Yu.** Rol informatsionnoy bezopasnosti vo vneshneekonomicheskoy deyatelnosti [The role of information security in foreign economic activity], Mezhotraslevoy forum direktorov po informatsionnoy bezopasnosti [Abstracts of reports of the Inter-industry Forum of Directors of Information Security] (Moscow, November 17 – 182008.). Inform-media Russia, 2008.

8. Ot evraziiskoi integratsii k dal'nevostochnomu vektoru politiko-ekonomicheskikh interesov Rossii: sbornik nauchnykh trudov [From Eurasian integration to the Far Eastern vector of Russia's political and economic interests: a collection of scientific papers]. Vladivostok: Vladivostokskii filial Rossiiskoi tamozhennoi akademii, 2015. 265 p.

9. **Skiba V.Yu., Strekalov S.V.** Implementing the "Single Window" Concept in the European Union and Asian-Pacific Region: Organizational and Financial Aspects. Tamozhennaia politika Rossii na Dal'nem Vostoke [Customs policy of Russia in the Far East], 2016. Vol. 76, no. 3, pp. 32–44.

10. **Zubakov G.V., Strekalov S.V.** Current issues of information's interaction of participants in logistics chains while the cross-border air cargo carriage executing. Academic Vestnic of the Rostov Branch of the Russian Customs Academy, 2016, Vol. 24, no. 3, pp. 21–24.

11. **Motorin D.E., Popov S.G.** Algoritm mnogokriterial'nogo poiska traektorii dvizheniia robota na mnogosloinoi karte [Algorithm for multi-criteria search of robot movement trajectory on multi-layer map]. Information and Control Systems, 2018, no. 3, pp. 45–53. DOI: 10.15217/issn1684-8853.2018.3.45

12. **Motorin D.E., Popov S.G., Chuvatov M.V., Kurochkin M.A., Kurochkin L.M.** A study of the evaluation function for the cost of transport operations in distribution of purpose in a group of robots. 2017 XX IEEE International Conference on Soft Computing and Measurements (SCM), 2017, pp. 536–538. DOI: 10.1109/SCM.2017.7970642

13. **Valiev K.A.** Quantum computers and quantum computing. Physics-Uspekhi, 2005, Vol. 48, no. 1, pp. 1–36. DOI: 10.1070/pu2005v048n01abeh002024

14. **Kitaev A., Shen' A., Vialyi M.** Klassicheskie i kvantovye vychisleniia [Classical and quantum computing]. Moscow: MCNO, CheRo Publishing House, 1999. 192 p.

15. **Nielsen M.A., Chang I.L.** Quantum computing and quantum information. Cambridge: Cambridge University Press, 2011. 702 p.

16. **Petrenko A.S.** Kvantovaia ugroza bezopasnosti tekhnologii blokchein [Quantum Threat to Blockchain Security]. St. Petersburg: Athena, 2022. 105 p.

17. **Kholevo A.S.** Vvedenie v kvantovuiu teoriiu informatsii [Introduction to Quantum Information Theory]. Moscow: MTsNMO, 2002. 128 p.

18. **Kholevo A.S.** Veroiatnostnye i statisticheskie aspekty kvantovoi teorii [Probabilistic and statistical aspects of quantum theory], 2nd ed. Moscow, Izhevsk: Institut Komp'iuternykh Issledovanii, 2002. 128 p.

19. **Bennett C.H., Shor P.W.** Quantum information theory. IEEE Transactions on Information Theory, 1998, Vol. 44, no. 6, pp. 2724–2742. DOI: 10.1109/18.720553

20. **Petrenko S.A., Petrenko A.S., Kostyukov A.D.** Countermeasures technologies previously unknown quantum cyber threats. Zaŝita informacii. Inside, 2024, Vol. 118, no. 4, pp. 66–76.

21. **Petrenko A.S., Petrenko S.A.** Quantum resilience estimation method blockchain. Cybersecurity Issues, 2022, Vol. 49, no. 3, pp. 2–22. DOI: 10.21681/2311-3456-2022-3-2-22

22. **Simon D.R.** On the power of quantum computation. Proceedings 35th Annual Symposium on Foundations of Computer Science, 1994, pp. 116–123. DOI: 10.1109/SFCS.1994.365701

23. **Shor P.W.** Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM Journal of Computing, 1997, Vol. 26, no. 5, pp. 1484–1509. DOI: 10.1137/S0097539795293172

24. **Shor P.W.** Algorithms for quantum computation: discrete logarithms and factoring. Proceedings 35th Annual Symposium on Foundations of Computer Science, 1994, pp. 124–134. DOI: 10.1109/SFCS.1994.365700

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Skiba Vladimir Yu.
Скиба Владимир Юрьевич
E-mail: vskiba69@mail.ru

Petrenko Sergei A.
Петренко Сергей Анатольевич
E-mail: petrenko.sa@talantiuspeh.ru

Murzina Anastasiya A.
Мурзина Анастасия Алексеевна
E-mail: myrzina.aa@talantiuspeh.ru

Popova Kristina R.
Попова Кристина Романовна
E-mail: popova.kr@talantiuspeh.ru

Submitted: 11.11.2024; Approved: 23.12.2024; Accepted: 27.12.2024.
Поступила: 11.11.2024; Одобрена: 23.12.2024; Принята: 27.12.2024.

Circuits and Systems for Receiving, Transmitting and Signal Processing

Устройства и системы передачи, приема и обработки сигналов

Research article
DOI: https://doi.org/10.18721/JCSTCS.17403
UDC 621

ON THE POSSIBILITIES OF HUMAN EMISSION AND RECEPTION OF HIGH-FREQUENCY ULTRASOUND

V.H. Avetisyan ✉

Russian-Armenian University, Yerevan, Republic of Armenia

✉ avahan@mail.ru

Abstract. This article suggests the possibility of the existence of an ultrasonic channel of information exchange in humans, similar to bats, dolphins, whales, etc. The possible frequency band of such a channel is estimated, resulting in frequencies from several hundred kilohertz to several hundred megahertz. In order to verify the assumption made and to minimize the efforts to find the frequency band of the supposed communication channel in such a wide frequency range, the corresponding experiment is considered and the results of previously performed scientific researches are discussed. The consequences of the presence of such a possibility in humans are considered and, on their basis, an attempt is made to explain some unusual phenomena inherent in humans.

Keywords: frequency band, human unusual phenomena, information exchange channel, ultrasound, emission

Citation: Avetisyan V.H. On the possibilities of human emission and reception of high-frequency ultrasound. Computing, Telecommunications and Control, 2024, Vol. 17, No. 4, Pp. 35–45. DOI: 10.18721/JCSTCS.17403

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.17403
УДК 621

О ВОЗМОЖНОСТЯХ ИЗЛУЧЕНИЯ И ПРИЕМА ЧЕЛОВЕКОМ УЛЬТРАЗВУКА ВЫСОКИХ ЧАСТОТ*

В.Г. Аветисян ✉

Российско-Армянский университет,
Ереван, Республика Армения

✉ avahan@mail.ru

Аннотация. В данной статье предполагается возможность существования у человека ультразвукового канала обмена информацией, подобно летучим мышам, дельфинам, китам и т.д. Оценена возможная частотная область такого канала, в результате чего получены частоты от нескольких сот килогерц до нескольких сот мегагерц. С целью проверки сделанного предположения и снижения усилий по поиску частотной полосы предполагаемого канала общения в столь широкой частотной области рассматривается соответствующий эксперимент и обсуждаются результаты ранее выполненных научных исследований. Рассмотрены также следствия наличия такой возможности у человека, и на их основе сделана попытка объяснения присущих человеку некоторых необычных явлений.

Ключевые слова: частотный диапазон, необычные способности человека, информационный канал обмена, ультразвук, излучение

Для цитирования: Avetisyan V.H. On the possibilities of human emission and reception of high-frequency ultrasound // Computing, Telecommunications and Control. 2024. Т. 17, № 4. С. 35–45. DOI: 10.18721/JCSTCS.17403

* Статья печатается в порядке обсуждения

Introduction

The question of the existence of “unusual abilities” in sertain representatives of humanity has divided people and the entire scientific world into two camps, whose opinions are opposite. The first camp, the camp of deniers, declared these abilities untenable, and scientists of this camp declared research in this area as pseudoscience. The second camp, relying on the practice of famous individuals, including the masters of the Chinese method of “qigong”, believes that there are such “unusual abilities” in humans. However, the scientists of the second camp still have not yet presented any convincing researches and explanations of these abilities, which has become the basis for creation of various legends and occultism.

Let us explain what is meant by the term “unusual human abilities”. The perception of information by an average person is based on the reaction of his well-known sense organs – smell, taste, touch, sight, hearing and the vestibular apparatus. The transmission of information is carried out by influencing the said corresponding organs of another person with the help of conditional gestures of body parts (limbs and head), speech and artificial systems created by man. The latter use various devices for converting propagating signals of electromagnetic waves (including light) and elastic waves. However, those human abilities, which go beyond the ordinary just in the process of exchange information without artificial systems and lie outside the sphere of occult conceptions, we will attribute as “unusual abilities”.

The author supports the opinion about the possible existence of such abilities, which was supported by numerous studies of the abilities of individual representatives of the terrestial fauna. On this basis, the author tried to build a chain of assumptions and considerations leading to the possibility of the existence of unusual abilities in humans and their explanation with the help of ultrasonic communication channel.

The purpose of the article is an attempt, based on the supposed ultrasonic impact of a person and his reaction to it, firstly, to present ways to check the presence of such a channel of influence and to identify its frequency band; secondly, to compare the proposed facts with the available facts of modern studies of the fauna of our planet; and thirdly – on the basis of the consequences of the supposed attempt to present the physical essence of some unusual phenomena inherent in humans.

Basic assumptions, early research and discussion of results

Below are the results of works carried out by representative research teams in the field of studying unusual human abilities. Some author's considerations concerning the results of these studies are given and the ways of experimental verification of the assumptions made are shown.

It is known that the human skin and internal organs are dotted with nerve endings, with the help of which a person senses acoustic waves starting from zero frequencies, in contrast to the human hearing organs. They perceive only sound oscillations in the maximum range of ΔF_a = 16 Hz...20 kHz frequencies, according to the wavelengths in the interval of $\Lambda_a \approx$ 20.6 m...16.5 mm, taking into account the speed of $v_a \approx$ 330 m/s of propagation of acoustic waves in the air. Below these frequencies lies infrasound, and above – ultrasound [1, 2].

The essence of the author's basic assumption lies in the fact that humans have such areas of the body, which, with their corresponding nerve endings, can be likened to a kind of acoustic transducer of ultrasonic frequencies. The latter converts the ultrasound affecting it into a certain signal, which is perceived by the human nervous system, delivering certain information to the brain. Moreover, it is assumed that this acoustic transducer is reciprocal, i.e. the human brain, through its nervous system, is able to influence this acoustic transducer in such a way that it generates ultrasound. The set of these acoustic transducers, i.e. the human nervous system and the brain, capable of receiving and processing the influencing ultrasound, as well as to produce it in accordance with the causal activity of a person, we will conventionally call the transceiving center.

Thus, the human body is similar to a system consisting of a set of reciprocal acoustic transducers – ultrasonic antennas, the input signals of which are received by the transceiving center, which controls and directs the work of these antennas. Then such a modeled system, according to the theories of antennas and general communication, is capable of the following [3, 4]:

- to emit ultrasound coherently and, therefore, directionally;
- the parameters of the emitted ultrasonic wave can be modulated by the brain (in other words, the ultrasonic emission can be informative);
- when receiving ultrasonic energy, the system is able to determine the direction of its inflow;
- to exchange information with a similar system.

In other words, a person is capable to exert an ultrasonic influence on someone or something at a distance and in a certain direction, and to determine the direction of arrival of such an influence acting on oneself. Moreover, he is capable to establish communication, contact with another person at a distance without the participation of the usual organs of sight, hearing and speech.

Why is an ultrasonic, and not an infrasonic channel of communication assumed? The answer to this question implies the effectiveness of the channel with which a person may be gifted. There are three significant circumstances, when comparing these channels. The minimum wavelength of infrasound is $\Lambda_{a\min}^{infr} \approx 20.6$ m (corresponding to $f_{a\max}^{infr} = 16$ Hz) about 10 times greater than the natural size of a person and three orders of magnitude greater than the maximum wavelength of ultrasound $\Lambda_{a\max}^{ultr} \approx 16.5$ mm ($f_{a\min}^{ultr} = 20$ kHz). Then, according to the antenna theory [3], the human body is practically unable to provide either directional transmission of influence or directional transmission of information in the assumed way via the infrasonic channel, in contrast to such a possibility via the ultrasonic channel. This is the first circumstance.

The second circumstance is the obvious smallness of the level of natural interference of the ultrasonic channel compared to the infrasound one.

The third circumstance is that from the point of view of the information capacity of the communication channel, it is much larger for a higher-frequency ultrasonic channel compared to a low-frequency infrasonic channel [4].

In the 1970–80s, the intensive scientific research has been carried out in the USSR on people with unusual abilities. The research was conducted at the Leningrad Institute of Fine Mechanics and Optics (LIFMO; now ITMO University) under the leadership of academician G.N. Dulnev and at the Moscow Institute of Radio-Engineering and Electronics by the USSR Academy of Sciences (now IRE RAS) under the leadership of academician Yu.V. Gulyaev. People with unusual abilities were called “operators” by scientific teams, and we will stick to this terminology later in the article.

During the study, the research groups identified the following conclusions [5–8]. Firstly, in the “working mode”, the operator experiences great stress, and each session is hard work for him, which was confirmed by the operators themselves. Secondly, during the session, the operator generates pulses of magnetic and electric fields and there is a characteristic increase in the intensity of infrared and optical emission around his hands. At these moments, acoustic pulses with duration of hundredths and tenths fractions of a second are also recorded. Thirdly, the operator is able to move objects by acting on them with his hands at a distance. If a glass wall opaque to infrared emission is placed between the operator's hand and the object, the effect of the action is not disrupted. Fourthly, in the same mode, the charged microdroplets of sweat from the sweat glands of the operator’s hands were sprayed at a distance of several centimeters. It should be noted that this phenomenon in subsequent studies by Moscow scientists was presented as the basis for the “triboelectric” model of charge separation around the operator. Its essence is that the operator is charged with a charge of a sign opposite to the charges of the microdroplets of sweat that are sprayed out. They “wet” with their charge the surrounding objects. As a result, a Coulomb attraction of the nearby object by the operator occurs [8].

The author assumes that the above-described changes registered in the operator are an external manifestation of physiological restructuring of the operator's body and its work in the mode of generation of marked fields (including registered acoustic pulses, for which the glass barrier is practically transparent). It is physically clear that the operator's body must spend a significant part of its energy from its resources to generate the described processes, which leads to great physical exertion.

The author's next assumption concerns the operators themselves.

Let us think about how people communicated with each other at the dawn of the formation of man as a thinking person. If a person was gifted with the above-mentioned abilities, then, probably, the transmission of some desires, primitive thoughts was realized by means of ultrasonic channel of communication. The similarity of this is clearly expressed, for example, in killer whales and dolphins in mating season or when organizing a hunt, which is repeatedly proved by scientific researches. However, with the evolution of man, his vocal cords, language and speech developed. A person had to use the communication channel with another person less and less often, when significant expenses of his own physical resources were required. Naturally, with the evolution of man, his ability to communicate, to make contact with another person in this way was lost. Nowadays, there are people who, due to hereditary characteristics, have retained this ability to a greater extent than others, and we now call them people with unusual abilities. The masters of "qigong" practice call such people as “gifted” and even developed a method for improving their abilities for the practice of healing people.

How to check whether the supposed ultrasonic channel exists? For this purpose, the author proposes an experiment schematically depicted in Fig. 1.

The operator acts on a object and moves it. The fact that the lightweight object (a scale pan, a matchbox, a cigarette) moves due to the operator’s directed impact on the object is undeniable and is cited in the studies of both Moscow and Leningrad scientists [5–8]. Then we place the cigarette at the bottom of

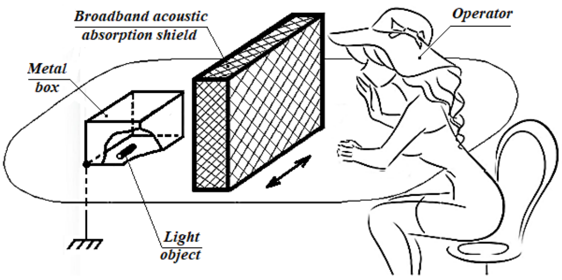


Fig. 1. Scheme of the experiment

hermetically sealed grounded metal box. Clearly, it is transparent to ultrasonic waves, which, according to wave theory [1, 9], create acoustic pressure that repels the object from the operator. However, the box excludes any electromagnetic influences from the operator, as well as the effects of convective air flows and any possible particles. In the case of particles, this also includes the above-mentioned charged microdroplets of sweat sprayed by the operator. In such isolated state, the operator repeats the session of impact on the cigarette. If after the session, when the box is opened, the fact of the cigarette moving in it away from the operator is still recorded, then we can talk about the presence of ultrasonic influence. Then, in the space between the box and the operator, a broadband sound-absorbing screen is introduced, after which the operator again repeats a session of influence. If the fact of the cigarette moving is not recorded, it confirmes the ultrasonic nature of the impact of the operator.

Another experiment was described in the studies of Leningrad scientists. The operator N.S. Kulagina acted on the cup of balanced sensitive analytical scales, again placed under a shielding metal cap. Then a forevacuum (about 10^{-3} Torr) was created in the cap, and the operator's action on the cup was not recorded during her numerous attempts. This led the researchers to the idea of a possible acoustic nature of the operator's action on the cup [5, 6]. However, the same works note that after the description of the performed experiments, the editorial board of the journal "Parapsychology and Psychophysics" in 1992 stated the following. As part of a research program for the study of the ability of an operator to move (in a working mode) nearby objects, the L.L. Vasiliev Parapsychology Foundation was able to record the movement of suspended objects in a vacuum of up to 10^{-2} Torr. The Foundation did not have such outstanding operators as Kulagina, who were able to move heavy objects over long distances. They studied the actions of operators on torsion balance, a much more sensitive device that requires less skill and effort from the operator to move the device. Their conclusion was that the observed movement of the torsion balance in a vacuum casts doubt on the explanation of the phenomenon by any acoustic fields.

The author does not agree with the conclusion of the Foundation, because the performed experiment with such sensitive torsion balance can be interpreted as follows. After the formation of a forevacuum, the possible ultrasonic impact of the operator on the balance ceases, which agrees with the experiment involving Kulagina. However, due to the very high sensitivity of the torsion balance, its movement is caused by an unaccounted phenomenon of a different nature, which is not acoustic.

Given the fact of the existence of ultrasonic channel in humans, the question of the frequency band of this channel remains open. One thing is clear: humans must have such a channel at frequencies higher than the ultrasonic frequencies emitted by dolphins and bats, because humans are insensitive to the ultrasound of these animals.

To localize the frequency band of the human ultrasonic channel, we resort to the phenomenon of the presence of an aura (halo), which an ordinary person can see in special cases and in a special environment around an outstanding person. It should be noted that images of a halo are found in ancient religious pictures and icons around the faces of saints, and the existence of the halo is not denied by anyone for sure.

The author supposes that this pronounced manifestation of an aura around a person with outstanding unusual abilities is the result of the effect of his ultrasonic emission on the surrounding air. Its density is modulated in accordance with the standing ultrasonic wave that somehow arose, forming a kind of diffraction grating in the air. Then the solar white light will diffract on such a formed grating [10]. We consider that the wavelengths of the incident white light are approximately $\lambda \approx 1$ μm. To observe the diffraction of light on such a spatial grating, it is necessary that its period d lies in the interval from lengths comparable to the wavelength of light incident on it, i.e. $d_1 \approx \lambda$ (case of a good diffraction grating), to lengths of the order of several hundred wavelengths of light, i.e. $d_2 \prec 1000\lambda$ (case of a rough diffraction grating). In these cases, for the period of the grating we approximately obtain 10^{-6} m $\approx d_1 \leq d \prec d_2 \approx 10^{-3}$ m. On the other hand, the period d of the diffraction grating formed under the influence of ultrasound will be of the order of the acoustic wavelength Δ_a, i.e $d \sim \Lambda_a$. Then we obtain that the frequency interval $\Delta F_a \approx F_{a1} - F_{a2}$ of acoustic waves emitted by a person is limited by the frequencies:

$$F_{a1} = v_a / \Lambda_{a1} \approx v_a / d_1 \leq \frac{3,3 \cdot 10^2}{10^{-6}} = 330 \text{ MHz}; \tag{1}$$

$$F_{a2} = v_a / \Lambda_{a2} \approx v_a / d_2 \leq \frac{3,3 \cdot 10^2}{10^{-3}} = 330 \text{ kHz}. \tag{2}$$

This is a large range for finding the frequency band of such a communication channel. In order to reduce the effort in attempts to find it, we propose to analyze the experiment on the scattering of laser radiation passing between the operator's palms. This unique effect was discovered by prof. G.B. Altshuller in experiments when the operator Kulagina affected a long glass cuvette with her palms [5]. Its length was 40 cm and the cuvette contained a solution of PGG dye in alcohol, through which a helium-neon laser beam passed. The palms were placed at a distance of 3–50 cm from the cuvette. As noted in this work, the effect of operator’s influence was manifested in visually observed “flashes” the beam scattering on emerging inhomogeneities in the cuvette in the impact zone, as well as in a strong flickering of the laser spot on the screen behind the cuvette. These inhomogeneities in the cuvette had the form of thin thread-like particles or formations. Moreover, it was observed that other operators were failed to achieve similar effects.

The team of Moscow researchers under the leadership of academician Yu.V. Gulyaev explained the effect of laser beam scattering by the influence of the above-mentioned microdroplets of sweat sprayed by the operator's palms on the laser beam. This version is presented in [8] by prof. E.E. Godik, the leading scientist of this team, and it is quite acceptable for the case of the beam passing through the air directly between the operator's palms. However, the proposed version does not explain the laser beam scattering effect in the case of prof. G.B. Altshuller's experiments. In these experiments, the effect of microdroplets of sweat sprayed by the operator's palms on the laser beam in the air in front of or behind the cuvette is excluded. The reason for this is the distance of 3–50 cm of the operator's palms from the cuvette, due to which the sweat microdroplets (which are sprayed at a distance of several centimeters) practically do not reach the ends of the long 40-cm cuvette. On the other hand, the same charges of microdroplets on the cuvette are not capable to cause thread-like inhomogeneities in the cuvette solution, which were observed in the experiment.

Another interesting effect described in [8] should also be noted. It consisted in the registered health-improving effect on sick patients during classes with a master of the Chinese practice of “qigong”. During the health session, the patient was in a compartment, which was located at a distance of 2–3 m from the master. The compartment was not soundproofed, and during the session, the master sat quietly on a chair, unlike the previously studied operators. He did not show any activity, expressed in the form of his sound commands or movement of his body parts. Prof. Godik, trying to explain this healing effect, suggested hypothesis about the possibility of infra- or ultrasonic influence on the patient.

The author suggests the possible nature of the operator's exposure to high-frequency ultrasound. This ultrasonic channel provides the above-mentioned circumstances regarding the pronounced directional properties of communication, its noise immunity and infomativity. Then, in the experiment with the laser beam scattering between the operator's palms, the medium (whether it is a solution of PGG dye in alcohol in a cuvette or in the air – in the absence of a cuvette) is subjected to such possible ultrasonic influence. The resulting standing ultrasonic wave between the operator's palms, and hence in the cuvette or in the air, will produce thread-like scattering inhomogeneities in the medium, through which the laser beam will pass. These inhomogeneities in density will be the greater, the higher the intensity of the ultrasound. If the operator is also able to vary the frequency of the ultrasound, the result will be oscillations in the displacements of inhomogeneities, which will cause the laser spot to flicker on the screen. It is clear that the intensity of the scattering effect of the laser beam will depend on the density of the medium and the intensity of the standing ultrasonic wave in it, as well as on the degree of the operator's ability to vary the frequency of the emitted ultrasound. Then it becomes clear that only Kulagina, who had the most pronounced unusual abilities in comparison with other operators, was able to cause the observed effect of laser beam scattering in Altshuller's experiments.

However, in the case of a possible standing ultrasonic wave in the medium, when the laser beam passes, Raman scattering of laser light caused by this standing ultrasonic wave should also take place [10]. It is known that the results of this phenomenon is the appearance of frequency satellites in the spectrum of scattered laser light, which are removed from the primary light frequency f by an offset value Δf, equal to the frequency F_a of the ultrasonic wave, i.e $\Delta f \approx F_a$. By performing a spectral analysis of the scattered light with a spectral device, it is possible to determine the frequency of the ultrasound used.

Let us estimate the resolution A of the spectral device. Taking into account the frequency $f \succ 3 \cdot 10^{14}$ Hz ($\lambda \prec 1\ \mu\text{m} = 10^{-6}$ m – in the visible light region) of the incident primary laser radiation and the frequency of the proposed generated ultrasound, according to (1) and (2), we obtain that the resolution A of the spectrum analyzer should lie in the interval of values:

$$A \approx f/\Delta f \approx \frac{f}{F_{a1}} \ldots \frac{f}{F_{a2}} \approx \frac{3 \cdot 10^{14}}{3{,}3 \cdot 10^{8}} \ldots \frac{3 \cdot 10^{14}}{3{,}3 \cdot 10^{5}} \approx 10^{6} \ldots 10^{9}. \tag{3}$$

It follows from (3) that to measure the ultrasound frequency F_a, a spectral device of high class will be required.

However, in the absence of such a device, the narrow frequency range of the assumed ultrasonic emission of the operator can also be determined in the experiment proposed by us (see Fig. 1) to verify the existence of such emission. For this purpose, in this experiment, instead of a broadband acoustic absorbing screen, it is necessary to use a set of sequential narrowband absorbing ultrasonic filters.

By measuring the frequency band of ultrasound generated by the operator and constructing a corresponding acoustic receiver, it is possible to directly measure many parameters of human-generated ultrasonic emission, including the degree of informativity of the ultrasonic channel.

Some modern studies on the generation of ultrasound by living organisms and effects of ultrasound on them

As follows from the above, the physical interpretation of human unusual abilities is based on the assumption that ultrasound is generated by humans with frequencies above several hundred kilohertz. Here we note that there was an attempt to explain such abilities using the "theory of torsion or microlepton fields", which, however, were not recognized by the scientific community [11]. The paper [2] presents extensive studies on the search of ultrasonic emission by mice, rats, hamsters, guinea pigs, rabbits, monkeys and humans, as well as on the effects of this emission on them. The studies were carried out in the frequency range up to 100 kHz. In [1], the studies on the effects of ultrasound on human organs

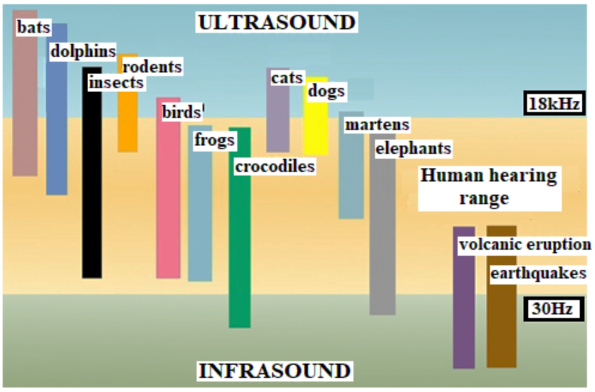


Fig. 2. Frequency ranges heard by different animals from the MoDiVa study, Walloon region, Belgium[1]

and cells are presented, but these studies are limited to a number of discrete frequencies up to 10 MHz. The author is not aware of any open publications of studies directly aimed at searching for and determining the frequencies of ultrasonic emission by humans at frequencies above several hundreds of kilohertz.

To find out how probable the author's assumption about generation of ultrasound of high frequencies by humans and mutual acoustic conversion in their bodies to ensure information exchange, we will give similar examples in nature. From the point of view of the presence of ultrasonic channel of communication, dolphins and bats have been studied well[2] [1, 2, 12–16]. Based on modern studies of fauna representatives, whales, rodents, insects, dogs and cats can also be classified as such "ultrasonic" creatures. The classification of some animals by the frequencies of the sound they produce, is shown in Fig. 2 [18].

Whales use ultrasound to navigate and catch fish in turbid water, and as a weapon to stun the fish. Ultrasonic signals from the whales cause the air-filled swim bladders of fish to resonate so intensely that the vibrations transmitted to body tissues disorient the fish and sometimes even kill them[3].

The bats use ultrasonic waves to detect obstacles, navigate spatially and communicate[4] [1, 9, 12, 16]. Their ability to echolocate is due to the generation of ultrasonic waves in the frequency range of 40–100 kHz. With the help of reflected waves, they navigate in space, determining directions and distances to surrounding objects. The higher the frequency of the sound, the finer the details the bats can distinguish and the more accurately they plot their flight path. In flight, bats use ultrasonic communication to manage flight safety to avoid collisions. Common bats adjust their flight paths and make way for the socially superior bat.

The little brown bat (lat. *Myotis lucifugus*) is an insectivorous bat found throughout North America that can use ultrasound to detect and avoid objects as thin as a human hair during the flight. Other fish-eating bats capture small fish swimming near the surface of the water, navigating only by the ripples in the water created by their movement. The leatherback bat, searching for prey, emits ultrasonic squeaks about five times per second, each squeak lasting 10–15 milliseconds. When the bat detects potential prey, it continuously increases the number of these squeaks and decreases their duration, similar to the emission of radar pulses when a target approachs to radar. The duration of the squeaks decreases to one millisecond when the number of squeaks exceeds 200 per second.

[1] Fedorov A. Kto i kak slyshit ul'trazvuk [Who and how hears ultrasound], Available: https://blog.pribormaster.ru/kto-i-kak-slyshit-ultrazvuk (Accessed 13.01.2025) (In Russian)
[2] Ter-Grigorian I. Letuchie myshi peregovorili vsekh [The bats talked everyone out], Available: http://www.gazeta.ru/science/2007/10/22_a_2257226.shtml (Accessed 13.01.2025) (In Russian)
[3] Kak zhivotnye ispol'zuiut ul'trazvuk [How Animals Use Ultrasound], Available: https://mir-znaniy.com/kak-zhivotnyie-ispolzuyut-ultrazvuk (Accessed 13.01.2025) (In Russian)
[4] Ter-Grigorian I. Letuchie myshi peregovorili vsekh [The bats talked everyone out], Available: http://www.gazeta.ru/science/2007/10/22_a_2257226.shtml (Accessed 13.01.2025) (In Russian)

Other insectivorous bats change the frequency of their squeaks and use overtones when hunting. Leaf-eating bats that hunt in the jungle use echolocation without changing the frequency of their squeaks. To detect prey, they use the Doppler Effect – the difference between the frequency of the ultrasound generated by the source and the frequency of ultrasound reflected from an object moving forward or backward relative to the ultrasound source.

Many species of moths are sensitive to the ultrasonic echolocation signals of bats. To avoid being caught by a bat, which is about 6 m away, these moths suddenly fold their wings and drop down, disappearing from the predator's flight path, or land on something. Some moths perceive the ultrasonic cry of bats with the help of paired organs on the abdomen that resemble the eardrums. Each such organ consists of a thin cuticular membrane, behind which is an air sac, that enables the membrane to vibrate when struck by an acoustic wave. Connected by nerves to the brain, these organs are sensitive to the frequency range of ultrasonic cries produced by insectivorous bats.

A special place is occupied by small long-necked primates (lat. *Tarsius*)[5]. Scientists from Humboldt University (California, USA) found that tarsiers "communicate" using pure ultrasound. For example, Philippine tarsiers hear ultrasound with frequency up to 90 kHz and emit cries at frequency of about 70 kHz. So far, these are the highest frequencies ever recorded in the "speech" of terrestrial mammals.

From the above, it follows that there are numerous examples of the existence of an ultraacoustic channel of communication between the animals of our planet. Humans, according to modern views on biology, are the highest thinking representatives of nature. Then, the author's assumptions about the possibility of humans having their own ultraacoustic communication channel with a higher frequency and greater informativeness do not seem so incredible.

Assessment of the concentration of possibility human ultrasonic emission and its implications

Let us discuss the degree of concentration of emitted ultrasonic energy by the operator and the consequences of this concentration. For this purpose, let us give some estimates for ultrasonic frequencies of several hundred kilohertz, for example $F_a \approx 400$ kHz. The wavelength of such an ultrasound in air will be $\Lambda_a = v_a / F_a \approx 330 / 4{\cdot}10^5 = 8{,}25{\cdot}10^{-4}$ m = 0,825 mm. Let us assume that the circular emitting ultrasound surface, comparable to the human palm, has a diameter of $D \approx 8$ cm. Then the plane angle $\Delta\theta$ of the solid angle of the emission pattern of the main lobe of such a emitting site, according to the antenna theory [3], will have the following value:

$$\Delta\theta \approx \frac{\Lambda_a}{D} \approx 0{,}01\,\text{rad} \approx 0{,}6^0. \quad (4)$$

Thus, from the author's assumption about the coherent emission of high-frequency ultrasound by the operator, from (4) follows a significant directionality of human palm emission. In other words, in the main share of emission it is possible to create a noticeable density of ultrasonic energy at distances of 10 m or more, similar to "ultrasonic" representatives of the fauna. At the same time, according to another author's assumption, there is a physiological restructuring of the operator's organism associated with this process. It is expressed in the intensive work of his sweat glands, revealed in the course of early studies. As a result, the operator's ability to attract (due to triboelectric effect) and repel (due to acoustic pressure) closely located objects is physically explained. Moreover, it is possible to explain the experiments with the scattering of a laser beam (due to the formation of inhomogeneity of the medium) by the operator in the air and in the cuvette at Altshuller's experiments.

The experiment, demonstrated by Kulagina with a ping-pong ball hovering in the air between her palms, may also be clearly explained physically. Due to the triboelectric effect, the charged ball is repelled from

[5] Dolgopiaty govoriat "na chistom ul'trazvuke" [Tarsiers speak "pure ultrasound"], Available: https://zoom.cnews.ru/rnd/news/line/dolgopyaty_govoryat_na_chistom_ultrazvuke (Accessed 13.01.2025) (In Russian)

the similarly charged surface of the table (also “wetted” with sweat droplets) upwards for some distance until the distance-dependent Coulomb repulsion force is balanced by the opposite attractive force of the ball. At the same time, the Coulomb force of attraction between the oppositely charged ball and the operator’s body is compensated by the opposite force of acoustic pressure on the ball, created by the ultrasonic wave emitted by the operator.

The human sensitivity to ultrasound of high frequencies, which, according to the author's assumption, has been preserved in humans during the evolution, and the possibility of directed emission of such concentrated ultrasonic energy by a person explain the following unusual phenomenon. Sometimes a person at a distance feels the attention of another person without seeing or hearing him or her. Turning round, the person quickly finds among many other people the one whose attention is concentrated on him. This is, in fact, a consequence of the acoustic location produced by humans, similar to dolphins, bats, etc. Such detection is known in radio engineering as “direction-finding”, which is used in “passive radar” [17].

A physical basis for interpreting other unusual phenomena appears, which manifest themselves when the operator concentrates his influence on a person nearby and establishes an information connection with him or her. Through such a connection, it is possible to transmit and receive thoughts, images, feelings and unconscious states from one person's brain to another person's brain or body at a distance, without using any known means of communication or manipulation.

In light of the above, it becomes clear that the effectiveness of the above-described phenomena, related to information exchange, will be determined by both the information capacity of ultrasonic communication channel and the process of detecting the ultrasonic signal by a person.

Conclusion

The article attempts to explain the unusual abilities of a person using specific physical and biological concepts and a probable model of information exchange via a channel with ultrasonic waves. Of course, the proposed assumption requires experimental confirmation. This is where one important question arises. Is it worth continuing research in the direction of clarifying the nature of such abilities or denying them immediately? The author is convinced of the need for such research, because if his assumptions are correct, the purpose of further research should be the answers to the following tempting, but difficult questions:

- How is the ultrasound emitted by humans modulated by their brains?
- What parameters of the ultrasound are modulated and by how much?
- How are ultrasound signals detected?
- What is the degree of perception of information embedded in the carrier of ultrasound of such frequencies?

Clarification of these questions, firstly, it will allow to eliminate some legends and occult ideas about unusual human abilities. Secondly, it will allow to take measures towards the urgent problem of protecting a person from the effects of intense ultrasound, which can cause great harm to the physical and mental state of a person.

REFERENCES

1. **Akopian B.V., Ershov Iu.A.** Osnovy vzaimodeistviia ul'trazvuka s biologicheskimi ob"ektami: Ul'trazvuk v meditsine, veterinarii i eksperimental'noi biologii [Fundamentals of interaction of ultrasound with biological objects: Ultrasound in medicine, veterinary science and experimental biology]. Moscow: Izdatel'stvo MGTU im. N.E. Baumana, 2005.

2. **Karkischenko N.N., Fokin Yu.V., Sakharov D.S., Karkischenko V.N., Kapanadze G.D., Chayvanov D.B.** Ultrasonic vocalization and its informative parameters in animals and man. Biomedicine, 2011, Vol. 1, Pp. 4–23.

3. **Somov A.M., Starostin V.V., Kabetov R.V.** Antenno-fidernye ustroistva [Antenna-feeder devices]. Moscow: Goriachaia liniia–Telekom, 2014.

4. **Dar'ian A.V.** Obshchaia teoriia sviazi [General theory of communication]. Yerevan: Rossiisko-Armianskii universitet, 2017.

5. **Dul'nev G.N., Volchenko V.N., Vasil'eva G.N. et al.** Svetloi pamiati N.S. Kulaginoi. Issledovanie K-fenomena [In Blessed Memory of N.S. Kulagina. Research of the K-phenomenon]. Parapsikhologiia i psikhofizika [Parapsychology and psychophysics], 1992, Vol. 7, No. 5, Pp. 35–51.

6. **Dul'nev G.N.** Energoinformatsionnyi obmen v prirode [Energy and information exchange in nature]. St. Petersburg: SPbGITMO, 2000.

7. **Guliaev Iu.V., Godik E.E.** Fizicheskie polia biologicheskikh ob"ektov [Physical fields of biological objects]. Vestnik AN SSSR. Seriia fizicheskaia [Bulletin of the USSR Academy of Sciences. Series: Physics], 1983, Vol. 8, Pp. 118–125.

8. **Godik E.E.** Zagadka ekstrasensov: chto uvideli fiziki. Chelovek v sobstvennom svete [The Riddle of Psychics: What Physicists Saw. Man in His Own Light]. Moscow: AST-Press kniga, 2010.

9. **Airapet'iants E.Sh., Konstantinov A.I.** Ekholokatsiia v prirode [Echolocation in nature]. Leningrad: Nauka, 1974.

10. **Landsberg G.S.** Optika [Optics]. Moscow: FIZMATLIT, 2003.

11. **Mineev V.V.** Vvedenie v istoriiu i filosofiiu nauki: uchebnik dlia vuzov [Introduction to the History and Philosophy of Science: A Textbook for Universities]. Krasnoyarsk: Krasnoiarskii gosudarstvennyi pedagogicheskii universitet im. V.P. Astaf'eva, 2012.

12. **Thomas J.A., Moss C.F., Vater M.** Echolocation in bats and dolphins. Chicago: The Univercity of Chicago Press, 2003.

13. **Au W.W.L.** The sonar of dolphins. New York: Springer New York, 1993. DOI: 10.1007/978-1-4612-4356-4

14. **Ryabov V.A.** The study of acoustic signals and the supposed dolphins' spoken language. St. Petersburg State Polytechnical University Journal. Physics and Mathematics, 2016, Vol. 248, No. 3, Pp. 88–101. DOI: 10.5862/JPM.248.10

15. **Zaytseva K.A., Korolev V.I., Akhi A.V., Krivchenko A.I., Butyrskiy E.Yu.** Dolphin Tursiops Trucatus and its radar capabilities. National Security and Strategic Planning, 2018, Vol. 21, No. 1, Pp. 74–82.

16. **Shchelkanov E.M., Ukolov S.S., Dunaeva M.N., Moskvina T.V., Popov I.A., Belov Yu.A., Kakareka N.N., Ganzevich A.V., Tolkach V.F., Volkov Yu.G., Galkina I.V., Shchelkanov M.Yu.** Echolocation of bats (Chiroptera Blumenbach, 1779) as an element of their ecological plasticity. South of Russia: Ecology, Development, 2020, Vol. 15, No. 4, Pp. 6–20. DOI: 10.18470/1992-1098-2020-4-6-20

17. **Bakulev P.A.** Radiolokatsionnye sistemy [Radar systems]. Moscow: Radiotekhnika, 2004.

INFORMATION ABOUT AUTHOR / СВЕДЕНИЯ ОБ АВТОРЕ

Avetisyan Vahan H.
Аветисян Ваан Генрихович
E-mail: avahan@mail.ru

Submitted: 07.11.2024; Approved: 10.12.2024; Accepted: 27.12.2024.
Поступила: 07.11.2024; Одобрена: 10.12.2024; Принята: 27.12.2024.

Computing, Telecommunication and Control, 2024, Vol. 17, No. 4, Pp. 46–55.
Информатика, телекоммуникации и управление. 2024. Том 17, № 4. С. 46–55.

Research article
DOI: https://doi.org/10.18721/JCSTCS.17404
UDC 621.3.049.774.2

CHANNEL SELECTION AT INPUT OF NARROWBAND DIRECT-CONVERSION RECEIVER BASED ON CURRENT-DRIVEN PASSIVE MIXER

T.D. Tran ✉

Peter the Great St. Petersburg Polytechnic University,
St. Petersburg, Russian Federation

✉ thanhdat140495@gmail.com

Abstract. Some issues when applying a Miller *N*-path filter to narrowband direct-conversion receiver (DCR) based on current-driven passive mixer and solutions to mitigate those issues are presented. The small impedances of the parallel LC circuit of low noise amplifier (LNA) at high frequencies cause undesirable effects: conversion gain reduction, noise figure increase and non-linear effects of the switches. A commutated network with the addition of resistors R_{FB1} and R_{FB2} is proposed to reduce the influence of those undesirable effects. The receiver is designed in the frequency range from 2.4 GHz to 2.6 GHz with 0.18 μm CMOS UMC technology. When applying the Miller *N*-path filter, although the conversion gain decreased by 3 dB and the noise figure increased by 1 dB, the in-band linearity (P_{1dB} and IIP3) of the DRC increased by 6–9 dB compared to the DRC without the Miller *N*-path filter.

Keywords: narrowband direct-conversion receiver, current-driven passive mixer, Miller *N*-path filter, channel selection, low noise amplifier, commutated network

Citation: Tran T.D. Channel selection at input of narrowband direct-conversion receiver based on current-driven passive mixer. Computing, Telecommunications and Control, 2024, Vol. 17, No. 4, Pp. 46–55. DOI: 10.18721/JCSTCS.17404

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.17404
УДК 621.3.049.774.2

КАНАЛЬНАЯ ФИЛЬТРАЦИЯ НА ВХОДЕ УЗКОПОЛОСНОГО ПРИЕМНИКА ПРЯМОГО ПРЕОБРАЗОВАНИЯ НА ОСНОВЕ ПАССИВНОГО СМЕСИТЕЛЯ С УПРАВЛЕНИЕМ ПО ТОКУ

Т.Д. Чан ✉

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

✉ thanhdat140495@gmail.com

Аннотация. В статье рассмотрены некоторые проблемы, возникающие при применении N-канального фильтра Миллера в узкополосном приемнике прямого преобразования (ППП) на основе пассивного смесителя с управлением по току, а также методы устранения этих проблем. Малые импедансы параллельного LC-контура малошумящего усилителя (МШУ) на высоких частотах приводят к таким нежелательным последствиям, как снижение коэффициента передачи, увеличение коэффициента шума и нелинейные эффекты в переключателях. Для уменьшения влияния этих нежелательных последствий предлагается коммутируемая схема с добавлением резисторов R_{FB1} и R_{FB2}. Приемник предназначен для работы в диапазоне частот от 2,4 ГГц до 2,6 ГГц с использованием 0,18 мкм КМОП-технологии компании UMC. При применении N-канального фильтра Миллера, хотя коэффициент передачи снизился на 3 дБ, а коэффициент шума увеличился на 1 дБ, внутриполосная линейность (P_{1dB} и IIP3) ППП увеличилась на 6–9 дБ по сравнению с приемником без применения N-канального фильтра Миллера.

Ключевые слова: узкополосный приемник прямого преобразования, пассивный смеситель с управлением по току, N-канальный фильтр Миллера, канальная фильтрация, малошумящий усилитель, коммутируемая схема

Для цитирования: Tran T.D. Channel selection at input of narrowband direct-conversion receiver based on current-driven passive mixer // Computing, Telecommunications and Control. 2024. Т. 17, № 4. С. 46–55. DOI: 10.18721/JCSTCS.17404

Introduction

Current-driven passive mixer is a suitable choice for modern direct-conversion receivers (DCR), because it has low flicker noise and high linearity [1–8]. For narrowband DRC based on current-driven passive mixer, to achieve the maximum conversion gain of the receiver, the coupling capacitor needs to be nearly resonant with the inductor at the load of the low noise amplifier (LNA) [9–11]. This resonance not only suppresses out-of-band interferers, but also eliminates noise at the harmonics of local oscillator (LO) frequency that are not down-converted to the baseband at the mixer output, significantly reducing the noise figure of the receiver.

However, the large conversion gain of the receiver reduces its linearity. In practice, the 1 dB compression point (P_{1dB}) and input third-order intercept point (IIP3) of the receiver are in the ranges of –26 dBm to –23 dBm and –14 dBm to –10 dBm, respectively. With such a low level of linearity, the conversion gain of the receiver will be drastically reduced when receiving in-band interference of not very high power, such as –15 dBm. To suppress in-band interferences the Miller N-path filter was used as a channel selection filter. In these papers, LNA is designed to work over a wide frequency range [12–16]. Therefore, its output impedance is a resistance. On the other hand, for narrowband DCR based on current-driven passive mixer, the LNA loads a parallel LC circuit. Therefore, when applying a Miller N-path filter to this receiver, some issues appear, such as conversion gain reduction, noise figure increase and non-linear effects of switches.

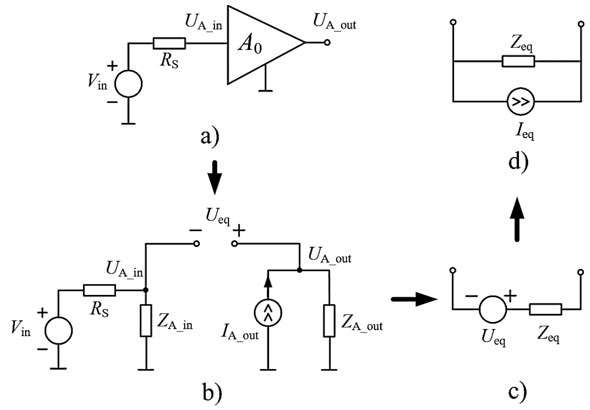


Fig. 1. LNA circuit with input source (a), equivalent circuit of LNA with input source (b), equivalent source between the input and output nodes of the LNA (Thevenin equivalent) (c, d)

This paper studies some issues when applying a Miller N-path filter to narrowband DCR based on current-driven passive mixer and proposes solutions to mitigate those issues.

Some issues applying a Miller *N*-path filter to narrowband DCR based on current-driven passive mixer

When applying a Miller N-path filter to a narrowband DCR, it is necessary to connect a commutated network between the input and output nodes of the LNA as shown in [13]. Using Thevenin equivalent LNA together with input voltage source $V_{in}(\omega)$ (Fig. 1a) and its impedance R_S = 50 Ohm is equivalent to a circuit consisting of a voltage source $U_{eq}(\omega)$ connected in series with impedance $Z_{eq}(j\omega)$ (Fig. 1c) or a current source $I_{eq}(\omega)$ in parallel with impedance $Z_{eq}(j\omega)$ (Fig. 1d) calculated by the following formulas:

$$U_{eq}(\omega)=\frac{Z_{A\_in}(j\omega)V_{in}(\omega)\left[A_0(j\omega)-1\right]}{Z_{A\_in}(j\omega)+R_S},$$

$$Z_{eq}(j\omega)=\frac{Z_{A\_in}(j\omega)R_S\left[1-A_0(j\omega)\right]}{Z_{A\_in}(j\omega)+R_S}+Z_{A\_out}(j\omega), \quad (1)$$

$$I_{eq}(\omega)=\frac{U_{eq}(\omega)}{Z_{eq}(j\omega)},$$

where $Z_{A\_in}(j\omega)$, $Z_{A\_out}(j\omega)$ are input and output impedances of the LNA, respectively; $A_0(j\omega)$ is the conversion gain of the LNA.

Then, when connecting the commutated network to the input and output nodes of the LNA, we obtain the circuit for the small-signal model in Fig. 2. Using the methodology presented in [17], the current $I_{in1}(\omega)$ is calculated. Then, at $0.8\omega_c < \omega < 1.2\omega_c$, where ω_c is the switching angular frequency, the voltage $U_{eq1}(\omega)$ is calculated by the formula:

$$U_{eq1}(\omega)=\left[I_{eq}(\omega)-I_{in1}(\omega)\right]Z_{eq}(j\omega)=$$
$$=\left[R_{SW}+Z_{CL}(j\omega)+\frac{c(\omega)Z_{eq}(j\omega)}{\left(1+g(\omega)\right)Z(j\omega)}\right]\frac{I_{eq}(\omega)Z_{eq}(j\omega)}{Z(j\omega)}\approx\left[R_{SW}+\frac{c(\omega)}{1+g(\omega)}\right]I_{eq}(\omega), \quad (2)$$

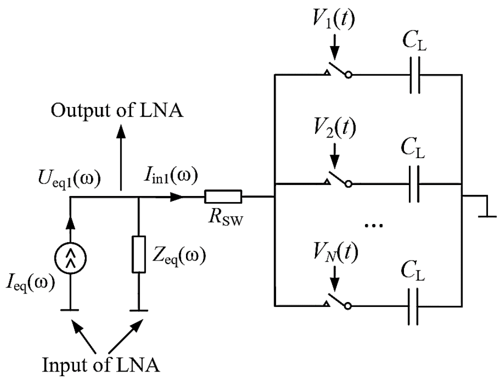


Fig. 2. Equivalent circuit for the small-signal model
when connecting a commutated network to the input and output nodes of the LNA

where $c(\omega)=\dfrac{N\sin\left(\left(\pi-\dfrac{\pi}{N}\right)\dfrac{\omega}{\omega_c}\right)\sin\left(\dfrac{\pi\omega}{N\omega_c}\right)Z_{CL}(j\omega_c)}{2\pi\sin\left(\left(\dfrac{\omega-\omega_c}{\omega_c}\right)\pi\right)}$, $g(\omega)=\sum\limits_{l=-\infty}^{+\infty}\dfrac{c(\omega)}{(1+lN)^2 Z(j\omega+lNj\omega_c)}$,

$$Z(j\omega)=Z_{eq}(j\omega)+R_{SW}+Z_{CL}(j\omega)\approx Z_{eq}(j\omega),\ Z_{CL}(j\omega)=\frac{1}{j\omega C_L}.$$

Firstly, let us consider the conversion gain reduction of the LNA. Now we calculate the output voltage $|U_{eq1}(\omega)|$ of the LNA for the desired signal at the frequency $\omega \approx \omega_c$. In this case, the ratio $(\omega - \omega_c)/\omega_c \approx 0$, therefore $|c(\omega)| >> |Z_{eq}(j\omega)| \approx |Z(j\omega)|$. Then formula (2) will take the form:

$$U_{eq1}(\omega)=\left[R_{SW}+\frac{1}{\sum\limits_{l=-\infty}^{+\infty}\dfrac{1}{(1+lN)^2 Z_{eq}(j\omega+lNj\omega_c)}}\right]I_{eq}(\omega)\approx\frac{I_{eq}(\omega)}{\sum\limits_{l=-\infty}^{+\infty}\dfrac{1}{(1+lN)^2 Z_{eq}(j\omega+lNj\omega_c)}},$$

taking into account the following approximation:

$$R_{SW}<<\left|\frac{1}{\sum\limits_{l=-\infty}^{+\infty}\dfrac{1}{(1+lN)^2 Z_{eq}(j\omega+lNj\omega_c)}}\right|.$$

As mentioned above [12–16], $Z_{eq}(j\omega)$ is a resistor over a wide frequency range. Then we have:

$$U_{eq1}(\omega)=\frac{I_{eq}(\omega)}{\sum\limits_{l=-\infty}^{+\infty}\dfrac{1}{(1+lN)^2 Z_{eq}(j\omega+lNj\omega_c)}}\approx\frac{I_{eq}(\omega)}{\dfrac{1}{Z_{eq}(j\omega)}}=I_{eq}(\omega)Z_{eq}(j\omega)=U_{eq}(\omega).$$

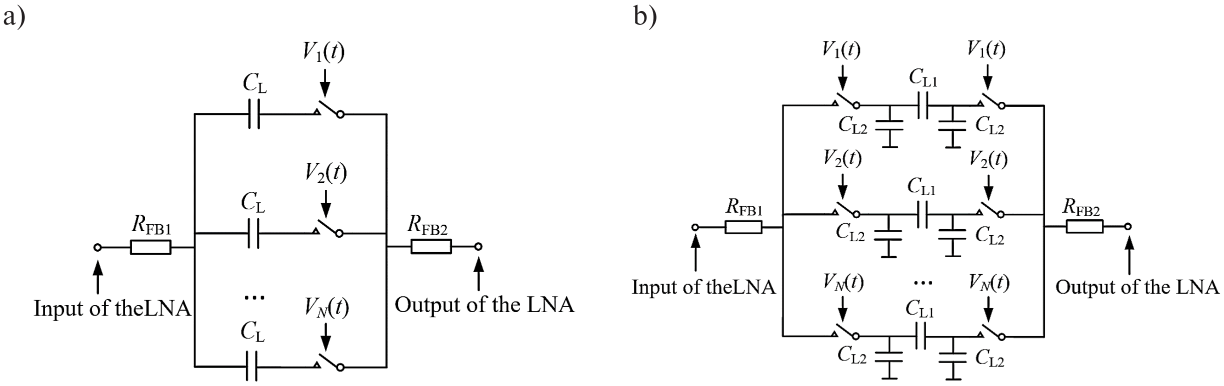


Fig. 3. Commutated network with resistors R_{FB1} and R_{FB2} (a); commutated network with increased filtering efficiency (b)

This means that the output voltage $|U_{eq1}(\omega)|$ of the LNA with the commutated network for the desired signal at frequency ω is approximately equal to the output voltage $|U_{eq}(\omega)|$ of the LNA without the commutated network for the desired signal at frequency ω. On the other hand, in a narrowband DCR based on current-driven passive mixer, $|A_0(j\omega)|$ and $|Z_{A\_out}(j\omega)|$ are large only near the input frequency f_{in} and are very small at high frequencies due to the selectivity of the parallel LC circuit at the output of the LNA. Therefore, from (1) it is clear that $|Z_{eq}(j\omega)|$ is large only near the input frequency f_{in} and is very small at high frequencies. Consequently, it is impossible to apply the following approximation:

$$U_{eq1}(\omega) = \frac{I_{eq}(\omega)}{\sum_{l=-\infty}^{+\infty} \frac{1}{(1+lN)^2 Z_{eq}(j\omega + lNj\omega_c)}} \approx \frac{I_{eq}(\omega)}{\frac{1}{Z_{eq}(\omega)}}.$$

Then, small values of $|Z_{eq}(j\omega + lNj\omega_c)|$ at high frequencies will significantly reduce $|U_{eq1}(\omega)|$. Consequently, the conversion gain of the LNA and receiver decreases.

Secondly, consider the noise figure increase of the LNA and receiver. Due to the conversion gain reduction of the LNA, the noise contribution of the LNA noise sources, the mixer, and subsequent stages will increase significantly, increasing the noise figure of the LNA and receiver.

Thirdly, when connecting transistors (switches) of the commutated network directly between the input and output nodes of the LNA, the voltage level at the transistor terminal is quite high, which causes nonlinear effects of the transistors.

Solutions to mitigate indicated issues

The reason for the conversion gain reduction and the noise figure increase of the narrowband DCR based on current-driven passive mixer is that $|Z_{eq}(j\omega)|$ has very small values at high frequencies. To reduce these undesirable effects, it is necessary to increase $|Z_{eq}(j\omega)|$ at high frequencies. For this purpose, resistors R_{FB1} and R_{FB2} are introduced, as shown in Fig. 3a. To further increase the filtering efficiency, a commutated network is used in [13], which is shown in Fig. 3b with the resistors R_{FB1} and R_{FB2}.

In addition, the resistor R_{FB2} isolates the terminal of the transistors (switches) of the commutated network from the output of the LNA, thereby reducing the voltage level at this terminal of the transistors and not causing nonlinear effects.

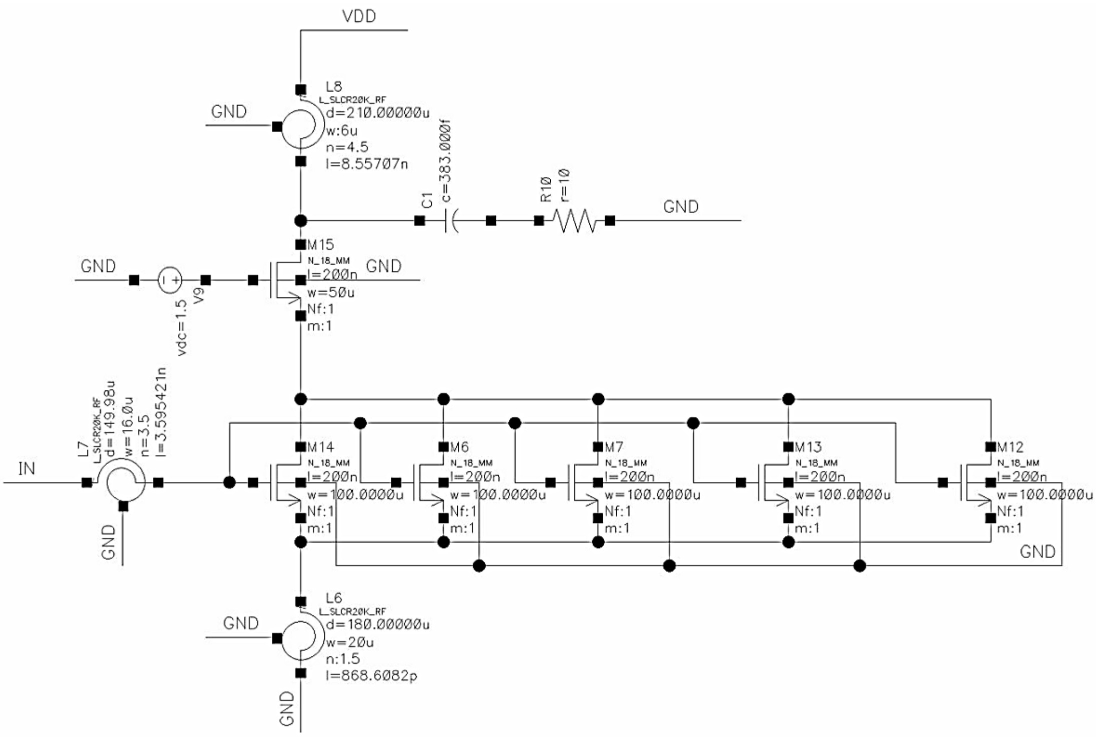


Fig. 4. The schematic of the LNA

Design of narrowband DCR based on current-driven passive mixer

The schematic of the narrowband DCR based on current-driven passive mixer is presented in [9]. The receiver is designed in the frequency range from 2.4 GHz to 2.6 GHz (desired frequency band) with 0.18 μm CMOS UMC technology. Fig. 4 shows the schematic of the LNA based on inductively-degenerated cascode common-source topology [18]. In this circuit, C is the coupling capacitor between the LNA and the mixer and is calculated using the method in [9]. The resistor R = 10 Ohm at the output of the LNA is the input impedance of the subsequent stages: mixer and transimpedance amplifier (TIA). This resistance needs to be small enough to ensure the switches work in current mode and not worsen the quality factor of the inductor L of the LNA. In the desired frequency band, the performances of LNA according to simulation results are: $|A_0|$ = 22.3 dB; noise figure NF = 2.1 dB; transconductance G = = 77.5 mA/V, which is determined by the ratio of the current flowing through the coupling capacitor C to the input voltage V_{in} of the LNA.

The schematic of the TIA is shown in Fig. 5. The TIA is designed using a common gate circuit to ensure its minimum input impedance. According to the simulation results, its input impedance is a parallel RC-circuit with R_{BB} = 82 Ohm and C_{BB} = 20 pF.

The width W_{mix} of the transistors (switches) of the current-driven passive mixer is chosen so that its resistance R_{SW} is small enough and its parasitic capacitance does not significantly reduce the conversion gain of the receiver. The optimal value is W_{mix} = 120 μm with "the number of fingers" being two. Then R_{SW} = 5 Ohm.

Simulation results when applying Miller N-path filter

The LNA circuit with the commutated network in Fig. 3b when $C_{L1} = C_{L2}$ = 2 pF was simulated using the Cadence software platform for the 0.18 μm CMOS UMC technology. For a fixed width W = 10 μm with "the number of fingers" being one of the switches, the dependence of the LNA performances on R_{FB1} and R_{FB2} is presented in Table 1.

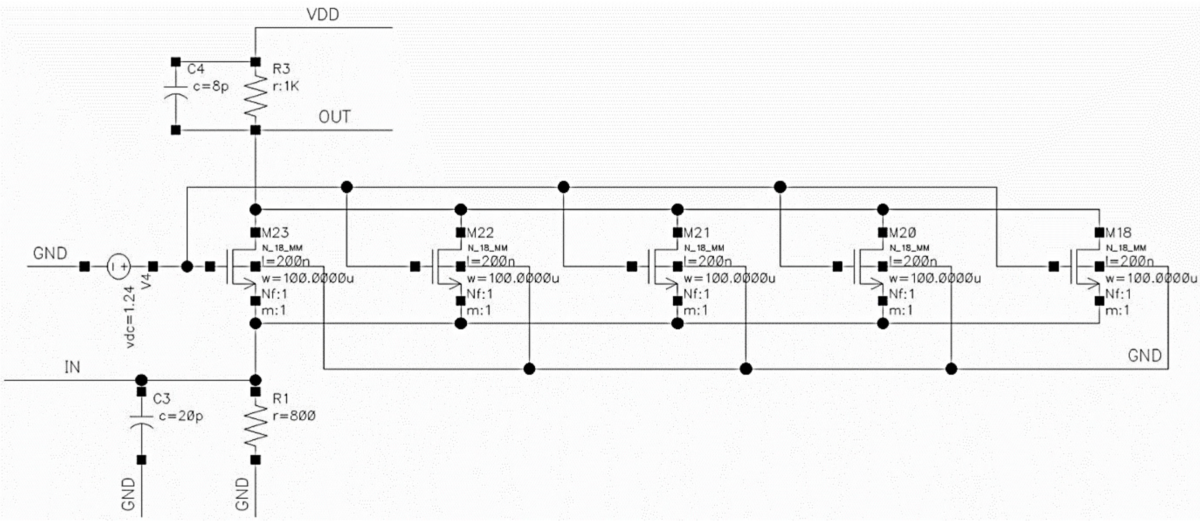


Fig. 5. Schematic of the TIA

Table 1

Dependence of the LNA performances on R_{FB1} and R_{FB2}

R_{FB1}, Ohm	0	50	100	200	200	300
R_{FB2}, Ohm	0	150	300	400	600	700
G, mA/V	32.1	52.2	58.8	60.9	63.5	64
K_S, dB	16.8	11.2	7.9	6.3	5.2	4.3
NF, dB	3.40	3.03	2.86	2.70	2.69	2.61

where K_S is the in-band interference suppression coefficient, which is determined by the ratio transconductance *G* of the LNA for the desired signal to *G* for in-band interference; *NF* is the noise figure of the LNA.

The simulation results in Table 1 show that when resistors R_{FB1} and R_{FB2} are absent, although high in-band interference suppression K_S = 16.8 dB is achieved, the transconductance *G* and noise figure *NF* of the LNA are deteriorated by 7.7 dB and 1.3 dB compared to the case without the commutated network. With an increase in R_{FB1} and R_{FB2} causing $|Z_{eq}(j\omega)|$ to increase at high frequencies, these deteriorations are reduced. For R_{FB1} = 200 Ohm and R_{FB2} = 400 Ohm, the dependence of the LNA performances on the width *W* with "the number of fingers" being one of the switches is presented in Table 2.

Table 2

Dependence of LNA performances on the width *W* of the switches when R_{FB1} = 200 Ohm and R_{FB2} = 400 Ohm

W, μm	2	5	10	15	20	30
G, mA/V	71	66.5	60.9	56.5	53	47.7
K, dB	5.2	6.3	6.3	5.9	5.5	4.8
NF, dB	2.34	2.52	2.70	2.85	2.97	3.17

The simulation results in Table 2 show that with increasing width *W* of the switches the transconductance *G* and noise figure *NF* of the LNA deteriorate due to the increase in the total parasitic capacitance of the switches. Since these parasitic capacitances reduce $|Z_{A\_in}(j\omega)|$ and $|Z_{A\_out}(j\omega)|$ (and consequently

$|Z_{eq}(j\omega)|$) at high frequencies. Based on the simulation result in Tables 1 and 2, W = 10 μm with "the number of fingers" being one, R_{FB1} = 200 Ohm and R_{FB2} = 400 Ohm are selected. In this case, the simulation results of the LNA show that although the conversion gain and the noise figure of the LNA are deteriorated by 2.1 dB and 0.6 dB, respectively, its linearity increased by 8–10 dB.

The narrowband DCR based on current-driven passive mixer was simulated with R_{SW} = 5 Ohm for two cases: with and without commutated network. Its performances are shown in Table 3.

Table 3

Performances of DCR

	DRC with commutated network	DRC without commutated network
K_G, dB	23.5	26.5
NF, dB	3.42	2.45
P_{1dB}, dBm	from –14 to –10	from –20 to –18.5
IIP3, dBm	from –1 to +3	from –9 to –4

where KG is the conversion gain of the DRC, which is calculated by the ratio of the differential output voltage of DRC to its input voltage.

The simulation results in Table 3 indicate that when using commutated network (applying Miller N-path filter), although the conversion gain K_G decreased by 3 dB and the noise figure NF increased by 1 dB, the in-band linearity (P_{1dB} and IIP3) of the DRC increased by 6–9 dB compared to the DRC without commutated network.

Conclusion

The issues when applying a Miller N-path filter to narrowband DCR based on current-driven passive mixer are considered: conversion gain reduction, noise figure increase and non-linear effects of switches. The reason for these issues lies in the small impedances of the parallel LC circuit at high frequencies. Due to these imperfections, when using the commutated network, the conversion gain and the noise figure of the LNA are deteriorated by 7.7 dB and 1.3 dB, respectively. A commutated network with the addition of resistors R_{FB1} and R_{FB2} is proposed to reduce the influence of imperfections. Based on the simulation result, W = 10 μm with "the number of fingers" being one, R_{FB1} = 200 Ohm and R_{FB2} = 400 Ohm are selected. When using commutated network, although the conversion gain decreased by 3 dB and the noise figure increased by 1 dB, the in-band linearity (P_{1dB} and IIP3) of the DRC increased by 6–9 dB compared to the DRC without commutated network.

REFERENCES

1. **Sosio M., Liscidini A., Castello R.** An intuitive current-driven passive mixer model based on switched-capacitor theory. IEEE Transactions on Circuits and Systems II: Express Briefs, 2013, Vol. 60, No. 2, Pp. 66–70. DOI: 10.1109/TCSII.2012.2234993

2. **Naseh N., Bardeh M.G., Entesari K.** A signal generator and down-conversion mixer / TIA unit for a 5.8-GHz FMCW receiver in 65nm CMOS. 2023 IEEE Radio and Wireless Symposium (RWS), 2023, Pp. 70–72. DOI: 10.1109/RWS55624.2023.10046310

3. **Korotkov A.S., Tran T.D.** A method for analysis of current-driven passive mixer with considering arbitrary input impedance and time constant of RC-load. 2022 International Conference on Electrical Engineering and Photonics (EExPolytech), 2022, Pp. 21–24. DOI: 10.1109/EExPolytech56308.2022.9950911

4. **Elsayed O., Zarate-Roldan J., Abuellil A., Hussien F.A.-L., Eladawy A., Sánchez-Sinencio E.** Highly linear low-power wireless RF receiver for WSN. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2019, Vol. 27, No. 5, Pp. 1007–1016. DOI: 10.1109/TVLSI.2018.2890093

5. **Gebhard A., Sadjina S., Tertinek S., Dufrêne K., Pretl H., Huemer M.** A harmonic rejection strategy for 25% duty-cycle IQ-mixers using digital-to-time converters. IEEE Transactions on Circuits and Systems II: Express Briefs, 2020, Vol. 67, No. 7, Pp. 1229–1233. DOI: 10.1109/TCSII.2019.2937654

6. **Han J., Kwon K.** RF receiver front-end employing IIP2-enhanced 25% duty-cycle quadrature passive mixer for advanced cellular applications. IEEE Access, 2020, Vol. 8, Pp. 8166–8177. DOI: 10.1109/ACCESS.2020.2964651

7. **Jiang J., Kim J., Karsilayan A.I., Silva-Martinez J.** A 3–6-GHz highly linear I-channel receiver with over +3.0-dBm in-band P1dB and 200-MHz baseband bandwidth suitable for 5G wireless and cognitive radio applications. IEEE Transactions on Circuits and Systems I: Regular Papers, 2019, Vol. 66, No. 8, Pp. 3134–3147. DOI: 10.1109/TCSI.2019.2909115

8. **Ghanad M.A., Dehollain C., Green M.M.** TIA linearity analysis for current mode receivers. 2018 16th IEEE International New Circuits and Systems Conference (NEWCAS), 2018, Pp. 53–56. DOI: 10.1109/NEWCAS.2018.8585587

9. **Korotkov A.S., Tran T.D.** Optimization of current-driven passive mixer conversion gain taking into account the parameters of the low noise amplifier. 2023 International Conference on Electrical Engineering and Photonics (EExPolytech), 2023, Pp. 22–25. DOI: 10.1109/EExPolytech58658.2023.10318787

10. **Mirzaei A., Darabi H., Leete J.C., Chen X., Juan K., Yazdi A.** Analysis and optimization of current-driven passive mixers in narrowband direct-conversion receivers. IEEE Journal of Solid-State Circuits, 2009, Vol. 44, No. 10, Pp. 2678–2688, DOI: 10.1109/JSSC.2009.2027937

11. **Mirzaei A., Darabi H., Leete J.C., Chang Y.** Analysis and optimization of direct-conversion receivers with 25% duty-cycle current-driven passive mixers. IEEE Transactions on Circuits and Systems I: Regular Papers, 2010, Vol. 57, No. 9, Pp. 2353–2366. DOI: 10.1109/TCSI.2010.2043014

12. **Park J.W., Razavi B.** Channel Selection at RF Using Miller Bandpass Filters. IEEE Journal of Solid-State Circuits, 2014, Vol. 49, no. 12, pp. 3063–3078. DOI: 10.1109/JSSC.2014.2362843

13. **Luo C.-k., Gudem P.S., Buckwalter J.F.** A 0.4–6-GHz 17-dBm B1dB 36-dBm IIP3 channel-selecting low-noise amplifier for SAW-less 3G/4G FDD diversity receivers. IEEE Transactions on Microwave Theory and Techniques, 2016, Vol. 64, No. 4, Pp. 1110–1121. DOI: 10.1109/TMTT.2016.2529598

14. **Park J.W., Razavi B.** 20.8 A 20mW GSM/WCDMA receiver with RF channel selection. 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2014, Pp. 356–357. DOI: 10.1109/ISSCC.2014.6757468

15. **Luo C.-k., Gudem P.S., Buckwalter J.F.** 0.4–6 GHz, 17-dBm B1dB, 36-dBm IIP3 channel-selecting, low-noise amplifier for SAW-less 3G/4G FDD receivers. 2015 IEEE Radio Frequency Integrated Circuits Symposium (RFIC), 2015, Pp. 299–302, DOI: 10.1109/RFIC.2015.7337764

16. **Tran T.D.** Enhanced frequency band wideband receiver using N-path Miller bandpass filter. Computing, Telecommunications and Control, 2024, Vol. 17, No. 3, Pp. 124–130. DOI: 10.18721/JCSTCS.17312

17. **Korotkov A.S., Chan T.D.** Analysis of a Current-Driven Passive Mixer at an Arbitrary Intermediate Frequency with Account of Input and Output Impedances. Journal of Communications Technology and Electronics, 2023, Vol. 68, pp. 77–87. DOI: 10.1134/S1064226923010072

18. **Nguyen T.-K., Kim C.-H., Ihm G.-J., Yang M.-S., Lee S.-G.** CMOS low-noise amplifier design optimization techniques. IEEE Transactions on Microwave Theory and Techniques, 2004, Vol. 52, No. 5, Pp. 1433–1442. DOI: 10.1109/TMTT.2004.827014

INFORMATION ABOUT AUTHOR / СВЕДЕНИЯ ОБ АВТОРЕ

Tran Thanh Dat
Чан Тхань Дат
E-mail: thanhdat140495@gmail.com

Submitted: 09.10.2024; Approved: 12.12.2024; Accepted: 13.12.2024.
Поступила: 09.10.2024; Одобрена: 12.12.2024; Принята: 13.12.2024.

Software of Computer, Telecommunications and Control Systems
Программное обеспечение вычислительных, телекоммуникационных и управляющих систем

Research article
DOI: https://doi.org/10.18721/JCSTCS.17405
UDC 004.021

DEVELOPMENT OF AN ALGORITHM FOR SYNCHRONIZATION OF TRAFFIC LIGHT CONTROLLER PROGRAMS

A.M. Sazanov ✉

Peter the Great St. Petersburg Polytechnic University,
St. Petersburg, Russian Federation

✉ arseny.sazanov@gmail.com

Abstract. This article develops an algorithm for synchronizing traffic light controller programs on one route in accordance with the general plan from the core. The relevance of the problem under consideration is described within the framework of improving the quality of functioning of adaptive and intelligent traffic control systems. An analysis of domestic and foreign modern solutions is carried out that ensure synchronization of traffic light modes in both local and distributed control systems of traffic light objects. Based on the necessary initial data and entered conditions that provide the required performance indicators of traffic control systems, an algorithm for synchronizing traffic controller programs is developed. For clarification, a general diagram of the synchronization algorithm is given. Conclusions are made on the results of testing the algorithm in real conditions and the prospects for further development of this study are described, including integration with existing control systems and the possibility of adaptation to various urban infrastructures.

Keywords: traffic light controller, synchronization algorithm, traffic light control, control, traffic management

Citation: Sazanov A.M. Development of an algorithm for synchronization of traffic light controller programs. Computing, Telecommunications and Control, 2024, Vol. 17, No. 4, Pp. 56–63. DOI: 10.18721/JCSTCS.17405

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.17405
УДК 004.021

РАЗРАБОТКА АЛГОРИТМА СИНХРОНИЗАЦИИ ПРОГРАММ ДОРОЖНОГО КОНТРОЛЛЕРА

А.М. Сазанов ✉

Санкт-Петербургский политехнический университет Петра Великого,
Санкт-Петербург, Российская Федерация

✉ arseny.sazanov@gmail.com

Аннотация. В данной статье осуществляется разработка алгоритма синхронизации программ дорожного контроллера на одном маршруте в соответствии с общим планом от ядра. Описывается актуальность рассматриваемой задачи в границах повышения качества функционирования адаптивных и интеллектуальных систем управления дорожным движением. Проводится анализ отечественных и зарубежных современных решений, обеспечивающих синхронизацию светофорных режимов как в локальных, так и в распределенных системах управления светофорными объектами. На основе необходимых исходных данных и введенных условий, обеспечивающих требуемые показатели эффективности систем управления дорожным движением, реализуется разработка алгоритма синхронизации программ дорожного контроллера. Для пояснения приводится общая схема алгоритма синхронизации. Делаются выводы о результатах тестирования алгоритма в реальных условиях и описываются перспективы дальнейшего развития данного исследования, включая интеграцию с существующими системами управления и возможность адаптации под различные городские инфраструктуры.

Ключевые слова: дорожный контроллер, алгоритм синхронизации, управление светофором, управление, контроль дорожного движения

Для цитирования: Sazanov A.M. Development of an algorithm for synchronization of traffic light controller programs // Computing, Telecommunications and Control. 2024. Т. 17, № 4. С. 56–63. DOI: 10.18721/JCSTCS.17405

Introduction

In the modern world of cities with increasing mobility and movement of the population, traffic flow is a vital part of the logistical, economic, and social life of the city [1, 2]. Therefore, the development of traffic management systems is important and relevant. By improving its functionality [3], a wide range of complex transport problems is solved today, such as: traffic jams and reduced speed caused by increased traffic density; accidents and road traffic incidents resulting from traffic rule violations or decreased driver attentiveness; optimization of priority passage modes for accelerated and continuous movement of special transport and others. The key role in such traffic management systems is assigned to traffic light objects, ensuring safety, efficiency, and smooth traffic flow on urban road networks.

The concepts of modern urban development within the framework of innovative technological approaches [4, 5] also indicate the improvement of traffic light objects' performance. The latter, according to [5], affects aspects of various types of management in both local and network modes. When implementing transitional modes, activating priority passage modes, changing coordination plans and other management policies, the relevance and critical importance of ensuring program synchronization with smooth transition to the appropriate phase without introducing additional service modes (e.g., a special “all red” phase after which synchronization occurs and the traffic light begins to operate in the planned mode) increase.

Thus, the development of a synchronization algorithm for traffic light objects is a serious scientific task, the solution of which can improve the efficiency of the urban transport.

Domain Analysis

To date, domestic and foreign scientists have investigated a significant number of approaches, methods and algorithms, regarding the issues of synchronization of modes and programs of traffic light objects.

In [6], a comprehensive analysis of modern approaches to optimizing traffic control systems depending on the structural and functional features of intersections was performed.

In [7], a two-level approach to local and global synchronization based on traffic flow was proposed. Local synchronization ensures autonomous adjustment of traffic light operating times based on current traffic density within the traffic lane boundaries, while global synchronization determines the green signal time of the traffic light based on the density of peer connections at various levels, as well as a set of related parameters.

In [8], modern approaches to managing traffic light objects in real time at road junctions in smart cities were analyzed. Features of traffic light signal synchronization algorithms on various busy routes to ensure uninterrupted traffic at intersections were also disclosed.

In [9], the influence of vehicle density on optimizing traffic light retiming policies on Jalan Bukit Gambir was investigated.

In [10], an IoT-based method for synchronizing traffic light signals was proposed to increase non-stop runs between intersections by adapting traffic light phases, reducing average travel time compared to other fixed-time traffic management strategies and unsynchronized control.

In [5], procedures for centralized and decentralized management of traffic light objects were developed and analyzed based on the formulation of the Hamilton-Jacobi kinematic wave model. The proposed control strategies were tested on the real road network of London.

In [11, 12], modern architectures of intelligent traffic management systems were analyzed, allowing the implementation of modern synchronization algorithms and ensuring compatibility when expanding the preferred functionality through additional services.

Problem Statement

The development of a synchronization algorithm for traffic controller (TC) programs on a single route, in accordance with the master plan from the core to the end point through precise time control, is an important step towards optimizing transport infrastructure, ensuring efficient traffic functioning on urban transport networks. In this work, we focus on the task of developing algorithms for synchronizing current and new traffic light plans, which are changed by dispatchers (or a remote computer) of the Traffic Control Center. Each program update command at a specific time plays a key role in balancing safety, efficiency and convenience for all road users.

The object of the research is the traffic light control system. The subject of the research is the methods and algorithms for synchronizing traffic light plans.

Main Part

Initial Data:

- Traffic light objects (TCs) managed by an intermediate device (ID);
- Minimum number of phases in the program: 2;
- Maximum number of phases in the program: not specified (infinite);
- TC can be:
  - Under the control of the ID-coordinated mode, in this case, the core knows exactly the current TC program and its current phase;
  - In local mode:
    - If at least one TC cycle has passed, its local program and current TC phase are known;
    - If not a single cycle has passed, we only know the number of the current phase.

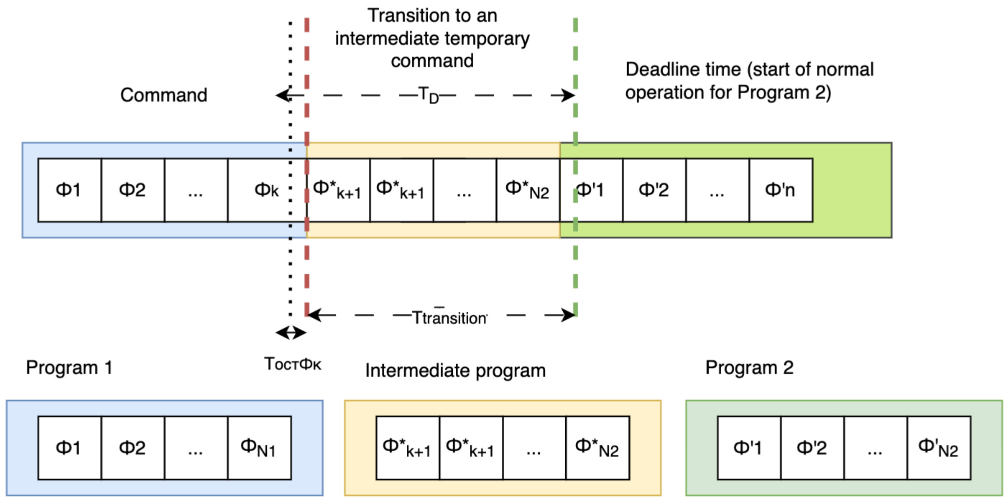


Fig. 1. Scheme of program synchronization on the traffic light controller

Conditions:

1. **System time** on the ID (at least within one route) is synchronized with an accuracy of **±1 second**.
2. The **current operating phase of the TC** is known.
3. $\mathbf{T_D}$ is the time after which (**deadline**) the new program should be running on the TC, starting from the first phase, and it is greater ($T_D > T_{\text{remains } \Phi}$) than the remaining operating time of the current phase of the old TC program;

a. If we apply a hard switch, then:

$$T_D > \min[T_{\min\Phi}; T_{\min\Phi} - T_{\text{completed part } \Phi}],$$

where $T_{\text{completed part } \Phi} = T_\Phi - T_{\text{remains } \Phi}$.

General scheme of the program synchronization on the TC is shown in Fig. 1.

Known:

- **Program No.1** consists of $\mathbf{N_1}$ phases: $\mathbf{\Phi_1}$, $\mathbf{\Phi_1}$, $\mathbf{\Phi_k}$, ..., $\mathbf{\Phi_{N1}}$;
- **Program No.2** consists of $\mathbf{N_2}$ phases: $\mathbf{\Phi'_1}$, $\mathbf{\Phi'_2}$, $\mathbf{\Phi'_n}$, $\mathbf{\Phi'_{n+1}}$, ..., $\mathbf{\Phi'_{N2}}$;

where

- $\mathbf{\Phi_1}$ is phase 1 of Program No. 1 (initial phase);
- $\mathbf{\Phi_2}$ is phase 2 of Program No. 1;
- $\mathbf{\Phi_k}$ is the k-th phase of Program No. 1 (current phase of the program at the moment the message about the program change arrives);
- $\mathbf{\Phi'_n}$, $\mathbf{\Phi'_{n+1}}$ are the transitional phases based on the phases of the new Program No. 2, considering the phase time correction matrix;
- $\mathbf{T_{transition}}$ is the transition process time (the time during which special transitional phases based on the new program phases operate on the TC);
- **n** is the program number following k → k+1 (provided: k+1 if $k+1 \le N_2$, otherwise k = 1);
- $\mathbf{\Phi_1}$ is the phase with corrected phase 1, according to the phase time correction matrix;
- $\mathbf{\Phi'_1}$ is the phase 1 of the new Program No. 2 (the initial phase of the new program; we assume that it is from this phase that the program should start after T_D occurs);
- $\mathbf{T_{\Phi i}}$ is the duration of phase No. i;

- $\mathbf{T_{min\Phi}}$ is the minimum duration of the phase, usually individually set for each phase; if not set, we assume the value $\geq T_{int}$ (duration of the intermediate phase);
- $\mathbf{T_{transition} = T_{\Phi i} + T_{corr\Phi i}}$ is the duration of each specific transitional synchronization phase until time T_D; this formula is the essence of the synchronization algorithm (which will be explained below).

Description of the Algorithm

Assumptions:

- All calculations are performed on the ID.
- The algorithm is the same for all IDs controlling TCs on the route; calculations are performed on the TC to reduce the load on the core, and because, when synchronizing times for program synchronization, one ID does not need to synchronize (knowing the phase times of another ID) with another ID.

General scheme of the program synchronization algorithm for TC on a single route according to the master plan from the core via program update command at a specific time ($T_D = T_{Deadline}$) is shown in Fig. 2.

ALGORITHM START

Upon receiving a command to switch the phase at the time when this switch should occur (i.e., upon receiving a new program and T_D):

1. Perform a check of the received T_D to ensure it is not less than:

a. The minimum time required to perform a hard program switch:

$$T_D > \min[T_{int}; T_{int} - T_{elapsed\Phi}]$$

(If this condition is not met, the program ends with an error) END

b. The minimum time required to perform a soft program switch:

$$T_D > T_{remaining\_phase}$$

(If this condition is not met and an additional option for a hard switch is not set – return an error; in case of a set option, from the next phase instead of the next phase of the old program, immediately activate Φ_1 of the new program).

2. If the ID does not have information about which program the TC is operating on (e.g., the TC was operating on a local program, the ID has recently turned on and has not accumulated statistical information on the TC's operation), then the ID must perform a hard switch to Φ'_1 of the new program at the moment TD occurs (ensuring traffic safety by observing T_{min_Φ}). END

3. If the ID has information about which program the TC is operating on, the ID calculates the remaining time of the current phase $T_{remaining\_phase}$. After the end of the current phase (Φ_k) of Program No. 1, it is further assumed that the time of the next phase (Φ_n*) is taken from the new program. Then $T_{remaining\_phase}$ is compared with TD. If $T_{remaining\_phase} > T_D$, the ID must return an error. END

4. After determining $T_{remaining\_phase}$, the ID calculates the remaining time for the transition process:

$$T_{transition} = T_D - T_{remaining\_phase}$$

5. Then, it calculates the number of sequential transitional phases that can complete within $T_{transition}$ and their total time $T_{sum\_transition\_phases}$

$$\text{(Important! } T_{sum\_transition\_phases} \leq T_{transition}\text{)}$$

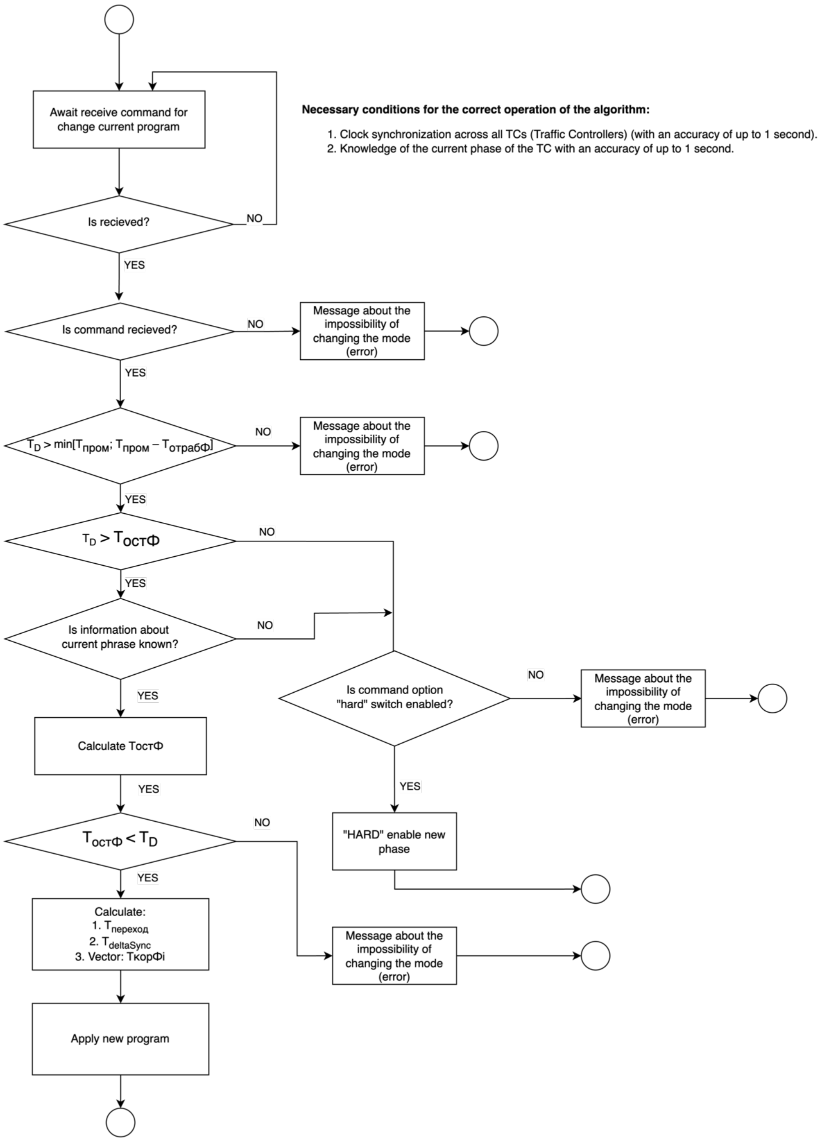


Fig. 2. Scheme of program synchronization algorithm
for a traffic light object (road controller) on a single route

$$T_{sum\_transition\_phases} = SUM(T_{\Phi n}, T_{\Phi n+1}, \ldots)$$

$$N_{transition\_phases} = COUNT(T_{\Phi n}, T_{\Phi n+1}, \ldots)$$

6. Next, it calculates the remaining time (which is insufficient to process the next phase of the new program cycle):

$$T_{delta\_sync} = T_{transition} - T_{sum\_transition\_phases}$$

7. Then, it calculates the average number of corrected seconds per transition cycle:

$$T_{avg\_corr} = DIV(T_{delta\_sync} / N_{transition\_phases}) \text{ (integer division)}$$

8. It calculates the final correction vector for all phases:

$T_{corr\_\Phi i} = T_{avg\_corr}$, where i is for all transitional phases from 2 to N_2; for i = 1, a special formula is applied:

$T_{corr\_\Phi 1} = T_{avg\_corr} + MOD(T_{delta\_sync} / N_{transition\_phases})$ (accounts for the undistributed remainder from integer division)

END OF ALGORITHM

Thus, after receiving the command to change the program with synchronization, the intermediate device, having calculated the correction vector of phases according to the algorithm described above, after the end of the current phase and until the deadline (T_D), begins to operate according to the new program, with the phase durations of this program changing according to the formula:

$T_{transition\_\Phi i} = T_{\Phi i} + T_{corr\_\Phi i}$ (this is why the phase is called transitional).

The developed algorithm ensures smooth phase synchronization and transition to the new program after its receipt, i.e., bringing the TC from the current state to the required one in the time necessary to activate the first phase of the new program, i.e., within T_D.

Considering time synchronization, if such an algorithm is successfully executed on all TCs, then all TCs are synchronized by the time T_D.

The algorithm was implemented in Go 1.18. The program text of the developed algorithm is posted on GitHub[1].

Testing and practical approbation of the algorithm were carried out in the traffic light control systems of the Moscow city transport system in 2023.

Conclusions

In this work, an algorithm for synchronizing TC programs on a single route in accordance with the master plan from the core was developed.

The developed algorithm will allow for effective implementation of synchronization modes both within a single intersection (traffic light object) and in coordination mode (on urban networks and city highways), when organizing priority passage, activating the green wave mode, or switching to a preferred traffic light plan. The algorithm's operation was successfully tested in the traffic light control systems of the Moscow Traffic Control Center. The accuracy of program synchronization for TC according to the master plan from the core was less than 1 second.

The proposed algorithm effectively aligns with the software of existing traffic light control systems and can be integrated into next-generation intelligent control systems.

REFERENCES

1. **Seliverstov S.A., Seliverstov Ya.A., Lukomskaya O.Yu., Sazanov A.M., Shkodyrev V.P.** Analysis of the functional features of intelligent traffic management systems. Transport: science, equipment, management (Scientific Information Collection), 2023, Vol. 12, Pp. 25–31. DOI: 10.36535/0236-1914-2023-12-3

2. **Seliverstov Ya.A., Gergel G.Yu., Seliverstov S.A., Nikitin K.V.** Improvement of intellectual transport systems by mobile technologies and procedures of analysis of social activity of urban population. St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunications and Control Systems, 2018, Vol. 11, No. 1, Pp. 47–64. DOI: 10.18721/JCSTCS.11105

3. **Seliverstov S.A., Seliverstov Ya.A., Shatalova N.V., Borodina O.V., Lukomskaya O.Yu., Vydrina E.O.** Transformation of digital technologies in the process of transition to cognitivity in transport. Transport: science, equipment, management (Scientific Information Collection), 2021, Vol. 8, Pp. 3–9. DOI: 10.36535/0236-1914-2021-08-1

4. **Seliverstov S., Seliverstov Ya., Gavkalyk B., Fahmi S.** Development of transport infrastructure organization model for modern cities with growing effectiveness. Transportation Research Procedia, 2020, Vol. 50, Pp. 614–625. DOI: 10.1016/j.trpro.2020.10.073

5. **Chow A.H.F., Sha R., Li S.** Centralised and decentralised signal timing optimisation approaches for network traffic control. Transportation Research Procedia, 2019, Vol. 38, Pp. 222–241. DOI: 10.1016/j.trpro.2019.05.013

6. **Seliverstov S.A., Sazanov A.M., Lukomskaya O.Y., Nikitin K.V., Shatalova N.V., Benderskaya E.N.** Analysis of modern approaches to optimizing traffic control systems. 2021 XXIV International Conference on Soft Computing and Measurements (SCM), 2021, Pp. 106–108. DOI: 10.1109/SCM52931.2021.9507147

7. **Sankara Nayaki K., Nithin Krishnan N.U., Joby V., Sreelakshmi R.** Traffic light synchronization. Advances in Computing and Information Technology. Advances in Intelligent Systems and Computing, 2013, Vol. 177, Pp. 97–106. DOI: 10.1007/978-3-642-31552-7_11

8. **Tomar I., Sreedevi I., Pandey N.** State-of-Art review of traffic light synchronization for intelligent vehicles: Current status, challenges, and emerging trends. Electronics, 2022, Vol. 11, No. 3, Article no. 465. DOI: 10.3390/electronics11030465

9. **Ahmad Rafidi M.A., Abdul Hamid A.H.** Synchronization of traffic light systems for maximum efficiency along Jalan Bukit Gambier, Penang, Malaysia. SHS Web of Conferences. Sustanable Built Environment Symposium (SBES 2014), 2014, Vol. 11, Article no. 01006. DOI: 10.1051/shsconf/20141101016

10. **Aleko D.R., Djahel S.** An IoT enabled traffic light controllers synchronization method for road traffic congestion mitigation. 2019 IEEE International Smart Cities Conference (ISC2), 2019, Pp. 709–715. DOI: 10.1109/ISC246665.2019.9071667

11. **Lee W.-H., Chiu C.-Y.** Design and implementation of a smart traffic signal control system for smart city applications. Sensors, 2020, Vol. 20, No. 2, Article no. 508. DOI: 10.3390/s20020508

12. **Seliverstov S.A., Sazanov A.M., Seliverstov Y.A., Benderskaya E.N., Nikitin K.V.** Development of the intelligent traffic management system architecture. 2021 XXIV International Conference on Soft Computing and Measurements (SCM), 2021, Pp. 200–203. DOI: 10.1109/SCM52931.2021.9507125

INFORMATION ABOUT AUTHOR / СВЕДЕНИЯ ОБ АВТОРЕ

Sazanov Arseniy M.
Сазанов Арсений Михайлович
E-mail: arseny.sazanov@gmail.com

Submitted: 11.07.2024; Approved: 10.10.2024; Accepted: 28.10.2024.
Поступила: 11.07.2024; Одобрена: 10.10.2024; Принята: 28.10.2024.

Computing, Telecommunication and Control, 2024, Vol. 17, No. 4, Pp. 64–77.
Информатика, телекоммуникации и управление. 2024. Том 17, № 4. С. 64–77.

Simulations of Computer, Telecommunications, Control and Social Systems

Моделирование вычислительных, телекоммуникационных, управляющих и социально-экономических систем

Research article
DOI: https://doi.org/10.18721/JCSTCS.17406
UDC 004.052.42

A METHOD FOR FINDING THE CORRESPONDENCE BETWEEN A RAILWAY STATION MODEL AND ITS VISUAL REPRESENTATION BASED ON GRAPHS

V.E. Ustinova ✉ , *A.S. Lutsenko*,
A.V. Shpak, *G.V. Mironenkov*, *V.A. Ivlev*

Peter the Great St. Petersburg Polytechnic University,
St. Petersburg, Russian Federation

✉ ustlera@list.ru

Abstract. The paper proposes a method for searching and comparing objects of the railway station model in the database format to the corresponding objects of the station visual representation in the SVG file format. The method is based on reducing the structure of a railway station to a directed asymmetric graph and using comparison algorithms to find correspondences between topological and actual characteristics of objects. The method includes comparing nodes and connections of the model graph with structural elements of the station visual representation. The paper also proposes a software tool that implements the proposed method. The software tool was tested in an experiment involving three employees, which revealed the average number of inconsistencies found, as well as the average time of the inconsistency search process before and after automation. The result of the research is that the method increases the accuracy of the process by two times and accelerates it by five times.

Keywords: graph, model representation, comparison of representations, railway station, visualization

Citation: Ustinova V.E., Lutsenko A.S., Shpak A.V., el al. A method for finding the correspondence between a railway station model and its visual representation based on graphs. Computing, Telecommunications and Control, 2024, Vol. 17, No. 4, Pp. 64–77. DOI: 10.18721/JCSTCS.17406

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.17406
УДК 004.052.42

МЕТОД ПОИСКА СООТВЕТСТВИЯ МОДЕЛИ ЖЕЛЕЗНОДОРОЖНОЙ СТАНЦИИ ЕЕ ВИЗУАЛЬНОМУ ПРЕДСТАВЛЕНИЮ НА ОСНОВЕ ГРАФОВ

В.Е. Устинова ✉ , *А.С. Луценко*, *А.В. Шпак*, *Г.В. Мироненков*, *В.А. Ивлев*

Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Российская Федерация

✉ ustlera@list.ru

Аннотация. В работе представлен метод поиска и сопоставления объектов модели железнодорожной станции в формате базы данных соответствующим объектам визуального представления станции в формате SVG-файла. Метод основан на приведении структуры железнодорожной станции к направленному несимметричному графу и использовании алгоритмов сравнения для поиска соответствий топологических и фактических характеристик объектов. Метод включает в себя сопоставление узлов и связей графа модели со структурными элементами визуального представления станции. Также в работе представлено программное средство, которое реализует предложенный метод. Программное средство было апробировано на эксперименте, в котором приняло участие три работника и который выявил среднее число найденных несоответствий, а также среднее время процесса поиска несоответствий до и после автоматизации. Результатом работы является то, что метод повышает точность процесса в 2 раза и ускоряет его в 5 раз.

Ключевые слова: граф, модельное представление, сравнение представлений, железнодорожная станция, визуализация

Для цитирования: Ustinova V.E., Lutsenko A.S., Shpak A.V., el al. A method for finding the correspondence between a railway station model and its visual representation based on graphs // Computing, Telecommunications and Control. 2024. Т. 17, № 4. С. 64–77. DOI: 10.18721/JCSTCS.17406

Introduction

Today, the Russian railway system is one of the most developed in the world and ranks third in terms of total length [1]. Due to such a large volume of the railway network, there are a large number of railway stations and, as a rule, several representations are created for them – a model and a visual [2, 3]. In the Russian Railways company, the model representation (electronic map) of the station exists in the form of an SQL file, while the objects are described in interconnected tables. The visual representation is demonstrated by an SVG file, while the objects are described using a set of tags [4, 5] (Fig. 1).

Discrepancies are often observed between the two representations of the same station: inconsistencies in the topology of individual sections of the electronic map and the graphical representation, names of objects, directions of direct passage and exit along the arrow, directions of traffic light adjustment. The presence of discrepancies means that it is impossible to design new stations until all discrepancies are eliminated. Manual updating of models and searching for inconsistencies is a time-consuming task, accompanied by a large number of human errors. Hence, the need to automate [6] this process and reduce the complexity of finding a match between the railway station model and the visual representation due to the algorithm implemented in the software for constructing and comparing graphs based on two representations of a railway station. This action can be presented as a goal, to achieve which it is necessary to perform a number of tasks, namely:

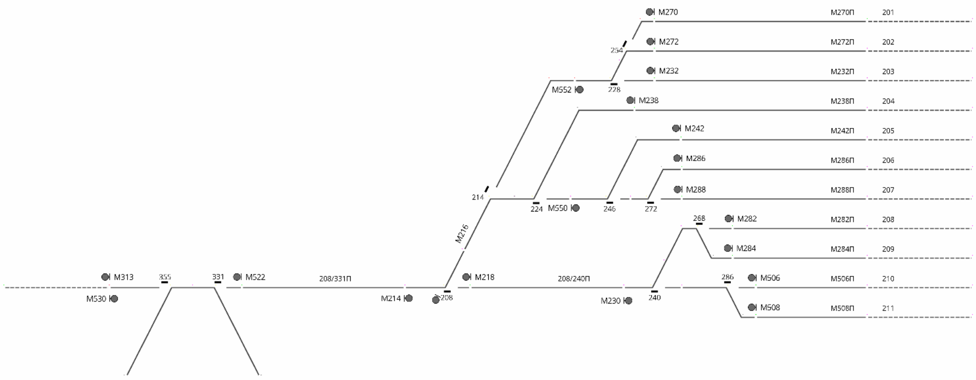


Fig. 1. Visual representation of the station

- analysis of existing solutions;
- development of the system architecture and its own method of automated comparison of the model and visual representations of the railway station plan for the Russian Railways company [7, 8];
- implementation of the method in the form of a software tool;
- conducting experiments and analysis of the obtained results to identify a reduction in the complexity of the process.

Overview of existing solutions

At the first stage, a list of criteria for the automation system was formed, namely:

- open source;
- ease of use;
- affordability (less than 50000 rubles);
- user-friendly interface;
- ability to build a railway station model and its visualization in the form of a graph.

The authors assume that this list of criteria for the automation system is necessary to achieve this goal. Thus, based on these criteria, a number of the following software systems were selected.

1. *Topomatic Robur* software system

The system [9] contains tools for designing railway crossings and stations and forming models of transport infrastructure facilities, providing an opportunity to design new and reconstruct existing railways and stations.

2. *GeoniCS* software package

The complex [10, 11] is a CAD system for railway design. It is focused on assistance in the design of new tracks, restoration and overhaul of existing railways.

3. *Graphviz* tool

This tool [12] is an open source utility package for automatic graph visualization. It offers its own templates for graph visualization and provides the opportunity to create user's own version of the representation.

4. *Higres* tool

The system [13, 14] allows you to create models, which are hierarchical graphs with a given semantics set by the user in the editor, as well as external modules for processing graphs with a certain semantics.

5. *Gephi* software

This software [15] is a multiplatform open source software focused on the visual representation and exploration of graphs.

Below is a comparative table of the results of research on the functionality of these systems.

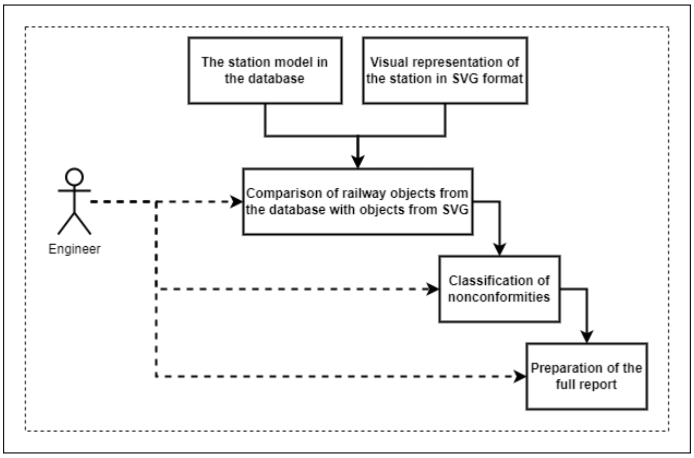


Fig. 2. Use case diagram: before automation

Table 1

Overview of existing solutions

System	Open Source	Ease of use	Price, RUB	User-friendly interface	Building of a railway station model and its visualization in the form of a graph
Topomatic Robur	–	–	139900	+	–
GeoniCS	–	–	129600	+	–
Graphviz	+	+	–	–	–
Higres	–	+	–	+	–
Gephi	+	+	–	+	–

From the comparative data presented above, it follows that no system supports simultaneously the construction of a railway station model and its visualization in the form of a graph, which entails the need to implement our own method of automated comparison of the model and visual representations of a railway station plan.

Architecture and algorithm development

To study the architecture of the system, use case diagrams were designed [16]. The traditional method of comparing railway objects requires manual work from the tester (Fig. 2).

The proposed automation method [17, 18] for object matching is presented in Fig. 3.

Due to the absence of a human factor, the proposed method has a significant advantage – the probability of error during automated verification is much lower, and the verification speed is several times higher.

Then the flowchart of the algorithm was developed, which subsequently formed the basis of the software project (Fig. 4).

Next, it is necessary to consider the algorithm in more detail based on the above scheme.

Getting railway objects from the database

To construct a graph [19] describing the relations of neighboring railway facilities and their characteristics, it is necessary to obtain appropriate data. The graph elements include arrows, traffic lights, and track circuits.

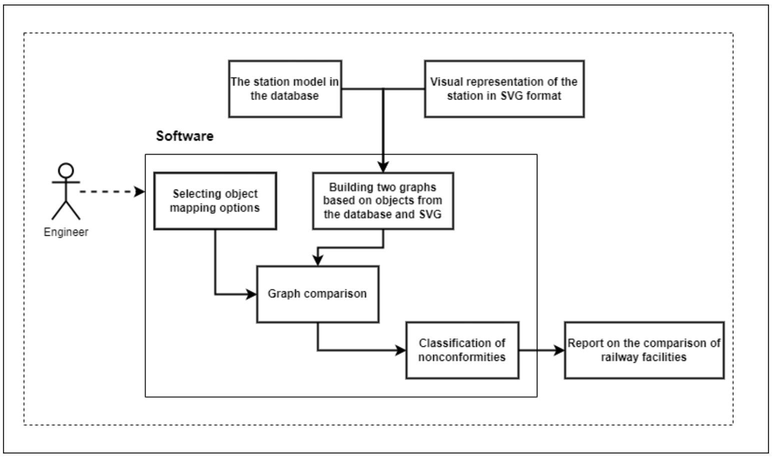


Fig. 3. Use case diagram: after automation

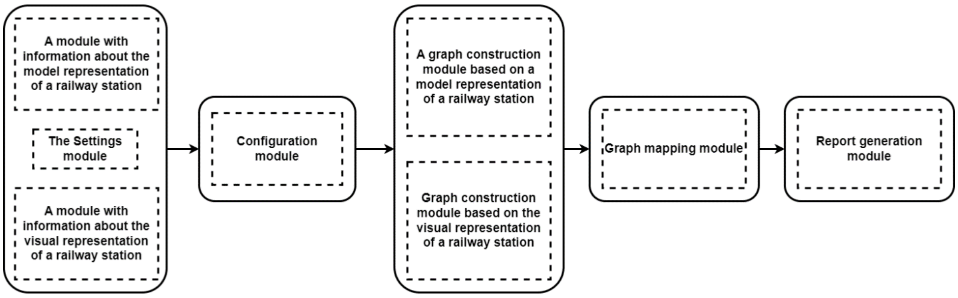


Fig. 4. The flowchart of the project

A database with railway objects is usually a model of a real railway station, part of the tracks, etc. Therefore, the database already contains all the necessary characteristics and values in order to determine which objects are connected. Thus, several SQL queries are enough to extract useful information from the database [20]. It is necessary to take into account that the graph can either coincide or diverge from the representation displayed in the database. The actual and reflected visual representation of one railway station is shown in Fig. 5.

Geolines are directional objects that make up track circuits and arrows. Each geoline has a “beginning” and “end” fields. To solve the orientation problem (related to the fact that the graph may differ from the database representation), it is necessary to add an option that reflects the order of objects from left to right. This includes reassigning the “beginning” and “end” of geolines, as well as reflecting the coordinates according to the rule $Lnew = Lmax + Lmin - L$, where L is the linear coordinate of the object, and $Lmax$ and $Lmin$ are the maximum and minimum coordinates among the list of all objects.

Getting railway objects from SVG file

The task is to read information about railway objects from SVG file and save the objects for further processing by the program. SVG file is a set of tags. Objects consist of lines and other elements and have a name, coordinates of key points, etc. in the description [21]. Operations that are performed on coordinates are also described. The task is to perform operations on all points to obtain the absolute coordinates of objects [22], and then save them along with the rest of the information about the objects.

First, the desired line is found and the coordinates of the object points are stored. Then all operations on the object are performed step by step. When all the points of the object are remembered and saved,

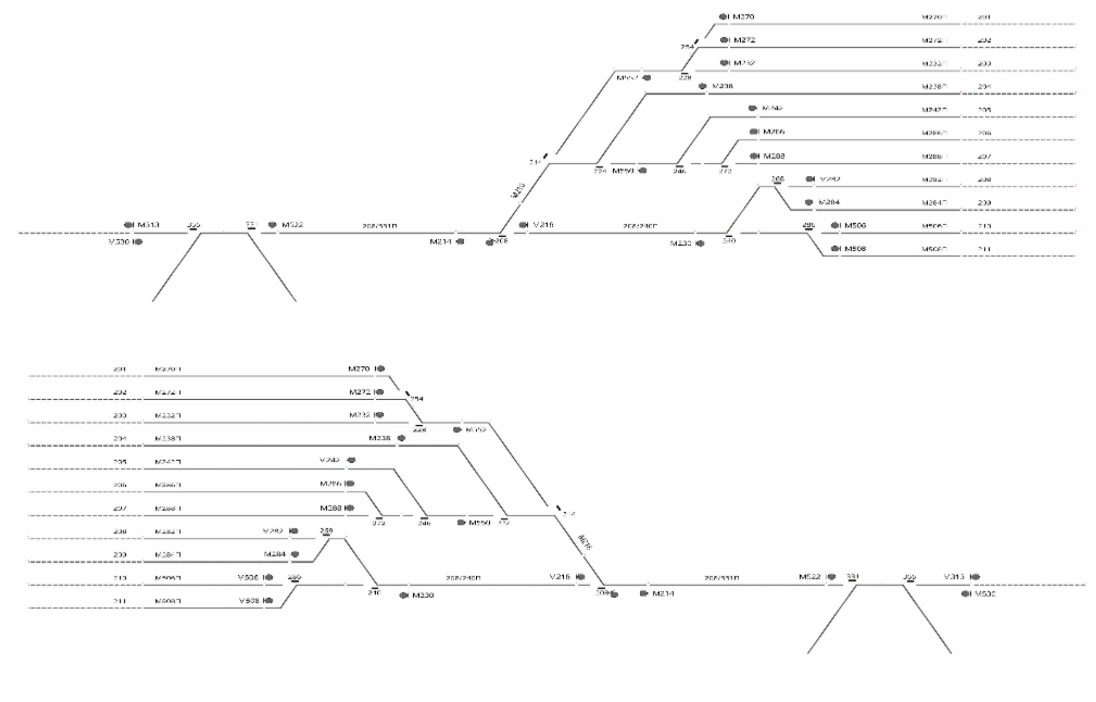


Fig. 5. Visual representations with different orientations

another one is created – *zero* (zero of the object), relative to which the rotation and reflection of the remaining points are carried out.

There are three types of operations used for SVG in this project. **Translate** moves an object in two-dimensional space; **rotate** rotates it by some angle; **scale** reflects the object along one or both axes (parameters – 1 or –1). These are affine transformations; they can be expressed in terms of affine transformation matrices [23, 24], however, frequent matrix multiplication takes longer than an algorithm based on the features of affine transformations [25]. During the writing and testing process, it was recorded:

1. When an object is rotated/reflected, and then moved along any of the axes, the axes are also rotated/reflected, and movement is carried out along the rotated/reflected axis, and not along the absolute one.
2. After rotation, the object is reflected along the rotated axes.
3. After reflection along one of the axes, the direction of rotation changes to the opposite.

All the observations obtained were incorporated into the algorithm. The operation implementation is based on the ***refindCoordWithCorner*** function. It allows you to find new coordinates of a point when the axes are rotated by a certain angle. It takes as parameters the old coordinates and the angle by which the point should be rotated. Let *xTranslate* be the old coordinate of the point on the x axis, *yTranslate* be the old coordinate of the point on the y axis, *curRotate* be the current angle between the vector drawn to the point from the origin and the vector of the positive direction of the x axis. If *xTranslate* = 0, this angle is 90 or –90. Otherwise, arctg(*yTranslate*/*xTranslate*) is taken (with an addition of 180 – in case of a negative *xTranslate*). Then, the angle to which the axes should be rotated, is added to the calculated angle. The new coordinates are calculated as follows:

1. The distance between the origin and the point is measured using the formula of the hypotenuse of a right triangle:

$$dist = \sqrt{xTranslate^2 + yTranslate^2}\,.$$

2. The coordinates are determined by multiplying the distance by the sine and cosine of the new angle: x = *dist* * cos(*curRotate*), y = –*dist* * sin(*curRotate*).

To implement the operations, we introduce a number of variables:

1. *sumRotation* is the sum of all the angles that the point was rotated by during the operations;

2. *rotationDirection* is the current direction of rotation of the object, equal to 1 or –1 (counterclockwise or clockwise, depending on previous reflection operations; if the object is reflected along one of the axes, its direction of rotation is opposite to the standard one);

3. *xDirection*, *yDirection* are equal to 1 or –1; show whether the object is currently reflected along two-dimensional axes or not;

4. *zero* is the zero of the object relative to which the reflection and rotation operations are performed.

Implementation of **rotation**: the angle to which the object should be rotated is indicated in parentheses. The angle is multiplied by the *rotationDirection* and added to the *sumRotation*. The coordinates of the current point relative to zero are determined; they are passed to the *refindCoordWithCorner* function along with the angle. The resulting values are added to zero.

Reflection implementation (**scale**): *xDirection* and *yDirection* are the values of the point reflection. The initial values of the current point's coordinate relative to zero are determined; they are multiplied by *xDirection* and *yDirection* and passed to *refindCoordWithCorner* along with the *sumRotation* angle. The resulting values are added to zero. These are the coordinates of the reflected point. Then the *rotationDirection*, *xDirection*, and *yDirection* are updated.

Translation implementation: *xTranslate*, *yTranslate* are the values of the movement. They are multiplied by *xDirection* and *yDirection* and passed to the *refindCoordWithCorner* function along with the *sumRotation*. The resulting values are added to the current point.

After calculating the absolute coordinates, the objects are filled in.

Plotting graphs based on data from a database and an SVG file

After all the necessary objects have been obtained from each representation of the railway station, you can start building the graph. It was decided to divide the construction into two stages: building a graph of track circuits and arrows and attaching traffic lights to the graph built in the previous stage.

This solution was chosen because the traffic light has a single coordinate, which is the “junction” between the track circuits and the arrows.

Thus, first, a graph is constructed, determining the neighborhood of track circuits and arrows, and then for each node of the graph it is determined which traffic lights belong to it.

Ultimately, the graph is represented as a structure with fields; the *structure* field is a dictionary in which the key is the node of the graph, and the value is a list of neighbor nodes. Below is an example in the form of five entries from this dictionary.

```
203 Circuit: none; 355 Switcher; none;
355 Switcher: 203 Circuit; 331 Switcher; none;
331 Switcher: 355 Switcher; 208/331П Circuit; none;
208/331П Circuit: 331 Switcher; 208 Switcher; none;
208 Switcher: 208/331П Circuit; 260/240П Circuit; 214 Switcher;
```

Listing 1 – *Structure* dictionary entries

Consider object 203:

1. The value “Circuit” indicates that this object is a track circuit.

2. The first element has the value “none” – the object in question is the leftmost in its group and has no neighbors on the left.

3. The second element of the list has the value “355 Switcher” – to the right of our object there is an object 355, which is an arrow.

4. The third element of the list has the value "none", which is typical for objects of the track circuit type, because the third element is designed to show which element is adjacent to an object of the arrow type along the turn line (for example, for arrow 208 along the turn line, arrow 214 will be the neighbor).

The *trafficlights* field is a dictionary in which the key is the node of the graph, and the value is a list of traffic lights located at the "ends" of this object. Below is an example in the form of five entries from this dictionary.

```
203 Circuit: M530, M313
355 Switcher: M530, M313;
331 Switcher: M522;
208/331П Circuit: M522, M214;
208 Switcher: M214, M218, M216;
```

Listing 2 – *Trafficlights* dictionary entries

It can be seen that the M530 and M313 traffic lights are located on the track circuit 203, as well as on the arrow 355, which allows us to conclude that these traffic lights are located between these two objects.

The *ends* field is an auxiliary dictionary for determining the neighbor when adding another track circuit or arrow. In this dictionary, the key is a point that is the "end" of a geoline belonging to a track circuit or an arrow. Therefore, the track circuit has only two such points, because it consists of one geoline, and the arrow has four such points, because the arrow consists of three geolines converging at one point.

The point at which the arrow's geolines converge must also be considered as the "end point", because there are situations in which two arrows have common geolines.

Because the track circuits and arrows differ from each other, a structure was introduced that transforms each object into a node of the graph.

After converting all objects into nodes, it is necessary to sort them by coordinates, after which each of the nodes joins the graph in turn.

The algorithm for attaching a node X to a graph:

1. Check if there are points in the auxiliary dictionary of the *ends* graph that coincide with the ends of X located in the end field.

2. If none of the points match, then X does not have any neighbors yet, so X is added to the *structure* dictionary as a key, and the value for this key is "none, none, none". We also need to add all the endpoints of X to the dictionary of endpoints of the *ends* graph so that the next nodes that are being added can join X.

3. If at least one point coincides with an already existing point in the dictionary of graph endpoints, then further steps are taken depending on what type of object X has.

4. If X is a track circuit, then you need to consider two situations of joining a neighbor Y:

a. Y is the neighbor on the right.

If **Y is a track circuit**, then we change the value of the key Y in the *structure* dictionary to the same list, only X will be the first element.

If **Y is an arrow**, then we look at which of the endpoints X coincides with Y, and depending on this we decide whether to put X as the neighbor on the left (the first element of the list) or the neighbor along the rotation line (the third element of the list).

After that, an entry for the key X is added to the *structure*, and the value for this key is "none, Y, none". Next, the endpoint is removed from the auxiliary dictionary, which is now the junction point of objects X and Y, and the remaining point is added, which was not used when joining X.

b. Y is the neighbor on the left.

Similar to the previous situation, only the first and second elements are interchanged for each object.

5. If X is an arrow, then we need to consider the situations when adding to 1 neighbor and to 2 neighbors, but all actions will be similar to attaching a track circuit:

a. update the data in the nodes that the new node is joining;

b. take into account the possible positions of the “turn line” for objects of the arrow type, checking which end of the arrow the new node is attached to;

c. clear the dictionary containing the ends of the graph from the already occupied ends in order not to check once again those to which it should not be physically possible to attach the object. The exception is the center points of arrow type objects.

Once all the track circuits and arrows have been added to the graph, we can proceed to the next stage.

This stage is divided into 2 parts:

1. Filling in the *tl_to_attach* dictionary, the key of which is the id of the geo point (for objects from the database) or XY coordinates (for objects from SVG), and the value is a list of traffic lights attached to this geo point/coordinate.

2. For each node of graph X (track circuit or arrow), a list of traffic lights belonging to one of the endpoints of X. This list is placed in the *trafficlights* dictionary by the key X.

At this stage, the construction of the graph of railway objects can be considered fully completed.

Graph comparison

Here we compare asymmetric graphs whose nodes are different objects [26]: track circuits, arrows, traffic lights. A node structure was invented to compare these objects, which brought them to a single view. In addition, the algorithm must identify different types of errors and mark nodes with different colours [27]: green – no errors, yellow – name error, red – location error or node is missing.

The comparison algorithm consists of the following sequential steps:

1. Checking all nodes for presence in another graph.
2. Check for duplicate nodes, which are track circuits and arrows.
3. Check for non-repeating nodes, which are track circuits and arrows.
4. Checking the nodes that are traffic lights.

Since the graphs are asymmetric, most of the steps are duplicated for both graphs. Next, let us look at each of the steps of the algorithm.

1. Checking all nodes for presence in another graph.

At this step of the algorithm, the program sequentially passes through all nodes in both graphs and searches for a matching pair in the other graph for each node. If a match is found, the node is considered paired. Nodes are only considered paired if they have the same name and type.

2. Check for duplicate nodes, which are track circuit and arrow objects.

At this stage of the algorithm, we check the neighbours of each node, looking for duplicates. A node can have at most three neighbours: the top, front, and turn neighbours. Only arrow type objects can have three neighbours; the rest always have at most two neighbours. This is because of physical factors, as this algorithm was designed for railway station layouts, not abstract graphs.

The check begins with the first paired node in the graph. Its neighbours also have a paired node in another graph. If the neighbouring nodes match, the node is marked as green; if not, it's marked red. Then the same process is repeated for all other nodes.

3. Check for non-repeating nodes, which are track circuits and arrows.

At this stage of the algorithm, the neighbours of the nodes are checked, which are not repeated. The check starts from the first unpaired node in the graph. All its neighbours that have a pair are found, as well as its position relative to the top node: the front neighbour or the neighbour on the turn. And then it is checked whether their pairs in another graph have a neighbour with the same location and type. If the same node is located from each paired neighbour, then both nodes are marked yellow, if not, they are marked red.

4. Checking the nodes that are traffic lights.

The traffic light is located at the junction of two track circuits, so it will always have two neighbours. When comparing traffic lights, their name and the nearest neighbours are compared. Since checking traffic lights is the final stage of the comparison algorithm, when comparing traffic lights, the colour of their neighbours' markings is considered.

If the traffic light has a pair in another column, then it can only be marked in red or green. If the traffic light already has a pair in another graph, it makes no sense to check the name match again, so at this step the neighbours of the traffic lights, which are marked green, play the main role. If the nearest neighbour of the traffic light turns out to be marked red, the nearest green neighbours are taken. If they match, then the traffic lights are marked green, if not, they are marked red.

If the traffic light does not have a pair in another column, then it can only be marked in red or yellow. This decision is made based on its neighbours and the neighbours of the traffic light with which the comparison takes place. The nearest neighbour with a green mark is located, based on it, a traffic light without a pair is searched in another graph, and if the other neighbours match, the traffic lights turn yellow, otherwise red.

This comparison algorithm works correctly in most cases. Errors occur only in some atypical connections. Most of the atypical connections are associated with arrows and the problem of unpredictable rail connections to them.

Software implementation

Based on the created method of searching for the correspondence of the railway station model to its visual representation, software was developed.

Requirements have been formulated for the product – the user, by uploading two files containing a model and a visual representation of the railway station, should receive an output including a log file describing the inconsistencies found, as well as SVG file on which:

1. the objects for which a complete match has been found in the electronic map are displayed in green;
2. displayed in yellow are those objects for which a topological correspondence has been found in the electronic map, but there is some discrepancy in characteristics;
3. displayed in red are those objects for which no match has been found in the electronic map.

Non-functional requirements have been formulated for the product:

1. programming language – Python 3.11;
2. the interface language is Russian;
3. platform – Windows 10;
4. minimalistic application design.

The requirements were met by implementing a simple graphical application using the PySide6 library [28]. The application contains buttons for uploading files, selecting the type of data reflection and starting file comparison – all the necessary functionality.

The Git version control system was used during development [29].

Results of the system operation

In order to demonstrate the reduction of labor intensity and increase the accuracy of the process of searching for compliance with the railway station model using a visual representation, a comparison of the time spent on work and the number of inconsistencies found with and without using the program was carried out.

Three employees of the same qualifications participated in the measurements, who independently compared two representations of the same station.

It is demonstrated in Table 2 how long each stage of the work took, from which it is possible to obtain an average value. The average number of inconsistencies found is also determined.

Then the employees compared the representations of the railway station (already different, but similar in complexity, structure and number of inconsistencies of the first station).

Table 2

Assessment of the complexity of the non-automated process

	Worker 1	**Worker 2**	**Worker 3**	**Average**
SVG and DB compliance analysis, min.	16	12	20	16
SVG adjustment, min.	8	10	15	11
Search and cataloging of DB inconsistencies, min.	27	24	24	25
DB adjustment, min.	13	13	16	14
Re-validation of SVG and DB compliance, min.	8	4	4	5
The whole process, min.	72	63	79	**71**
The number of inconsistencies found	3	7	5	**5**

In this experiment, automation of the process was used (Table 3).

The calculations are similar to the previous table.

Table 3

Assessment of the complexity of the automated process

	Worker 1	**Worker 2**	**Worker 3**	**Average**
File selection, program launch, min.	1	1	1	1
SVG correction based on the results of automated analysis, min.	4	5	3	4
Database correction based on the results of automated analysis, min.	6	8	10	8
Re-validation of SVG and DB compliance, min.	2	2	2	2
The whole process, min.	12	15	15	**14**
The number of inconsistencies found	10	10	10	**10**

The assessment of labor intensity is carried out according to a formula combining a simple summation of the resources spent on individual components and division by the number of these components [30]. The average time value in each of the sections is determined by the following formula:

$$S = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} Tij}{n},$$

where Tij is the time spent on stage j by employee i, n is the number of employees, $n = 3$; m is the number of stages; for the first section ("Before automating the process") $m = 5$, for the second section ("After automating the process") $m = 4$.

Time was measured in seconds, rounded up to minutes; rounding up occurred if the number of seconds that made up the remainder of the minutes exceeded 30. Thus, we observe that the automation of the process has significantly reduced the average time required to complete the work – from 71 minutes to 14 minutes, that is, by about 5 times. To demonstrate the results of reducing labor intensity, a histogram was constructed (Fig. 6) as one of the most common and effective ways to visualize statistical data [31].

The experiment was conducted for a small station (up to 30 graphic elements). During the further operation of the software, it was revealed that the time gain of the automated process over the non-automated one increases with the growth of the station size (average values are presented):

1. small station (up to 30 graphic elements) – 5 times;
2. medium-sized station (from 31 to 100 graphic elements) – 5.5 times;
3. large station (from 101 graphic elements) – 6.5 times.

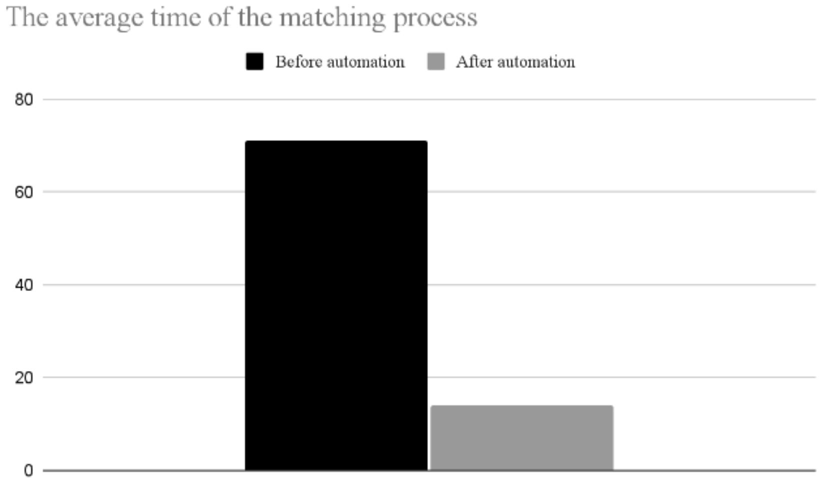


Fig. 6. The average time of the matching process

In addition, automation improves accuracy: instead of five inconsistencies found, 10 were found – and these are all the inconsistencies that were between the stations initially. Thus, automation made it possible to increase accuracy by $N = S2/S1 = 2$ times, where $S1$ is the number of inconsistencies found before automating the process, $S2$ is the number of inconsistencies found after automating the process.

Conclusion

This article presents an overview and comprehensive comparative analysis of existing approaches to comparing models to a real object. An innovative method for searching and comparing objects of the railway station model to the corresponding objects of the visual representation of the station is proposed. The proposed algorithm is implemented in the model matching software. The results of reducing labor intensity by 5 times and increasing accuracy by 2 times are demonstrated.

The work was carried out under the supervision of the Scientific Research and Design Institute of Informatization, Automation and Communications in Railway Transport.

REFERENCES

1. **Okladnikov S.M. et al.** Transport v Rossii. 2020: Statisticheskii sbornik [Transport in Russia. 2020: Statistical Digest]. Moscow: Federal'naia sluzhba gosudarstvennoi statistiki, 2020, 108 p.

2. **Kanashin N.V.** Creation of digital models of railway stations by terrestrial laser scanning. Transport of the Russian Federation, 2010, Vol. 28, No. 3, Pp. 66–67.

3. **Umansky V.I., Dolganuk S.I.** Digital models of station track development. World of Transport and Transportation, 2014, Vol. 12, No. 1(50), Pp. 126–133.

4. **Telegin A.I., Timofeev D.N., Chitalov D.I., Pudovkina S.G.** SVG-marking of two-dimensional graphics: The experience of SVG-using in creating of two-dimensional graphics. Miass: South Ural State University (National Research University), 2015, 73 p.

5. **Mikheev P.N.** SVG – novyi standart vektornoi grafiki v Web [SVG – the new standard for vector graphics on the Web]. Journal of Radio Electronics, 2001, Vol. 9, Pp. 7.

6. **Ivlev V.A., Nikiforov I.V., Yusupova O.A.** Automation method for configuring IT infrastructure for IT projects. Proceedings SPIE 12637, International Conference on Digital Transformation: Informatics, Economics, and Education (DTIEE2023), 2023. DOI: 10.1117/12.2680779

7. **Lutsenko A.S., Ustinova V.E., Ivlev V.A., Mironenkov G.V.** Realizatsiia programmnogo obespechenie dlia verifikatsii sootvetstviia model'nogo i vizual'nogo predstavleniia plana zheleznodorozhnoi stantsii [Implementation of software for verification of conformity between model and visual representation of a railway station plan]. Sovremennye tekhnologii v teorii i praktike programmirovaniia [Modern technologies in the theory and practice of programming], 2024, Pp. 59–61.

8. **Shalin A.P., Batrakov V.N.** Verification and validation in conformity assessment. Production Quality Control, 2022, Vol. 4, Pp. 47–50.

9. **Smogunov V.V., Mitrokhina N.Yu.** Sistemnyi analiz metodov proektirovaniia avtomobil'nykh dorog [System analysis of road design methods]. University proceedings. Volga region. Technical sciences, 2011, Vol. 20, No. 4, Pp. 116–127.

10. **Buchkin V.A., Ryzhik E.A., Lenchenkova E.P.** Comparative Analysis of Software Packages. World of Transport and Transportation, 2013, Vol. 11, No. 2(46), Pp. 112–121.

11. **Soboleva E.L., Archipova O.B.** Research of geoengineering systems for computer-aided design. Geo-Siberia, 2010, Vol. 1, No. 2, Pp. 42–45.

12. **Kozonogova E.V., Kurushin D.S., Dubrovskaya J.V.** Computer visualization of the identify industrial clusters task using GVMap. Scientific Visualization, 2019, Vol. 11, No. 5, Pp. 126–141. DOI: 10.26583/sv.11.5.11

13. **Gordeev D.S.** A survey of visualization techniques of algorithms on graphs. Scientific Visualization, 2018, Vol. 10, No. 1, Pp. 18–48. DOI: 10.26583/sv.10.1.02

14. **Kasyanov V.N., Zolotuhin T.A.** Visual graph – a system for visualization of big size complex structural information on the base of graph models. Scientific Visualization, 2015, Vol. 7, No. 4, Pp. 44–59.

15. **Ulizko M.S., Antonov E.V., Artamonov A.A., Tukumbetova R.R.** Visualization of Graph-based representations for analyzing related multidimensional objects. Scientific Visualization, 2020, Vol. 12, No. 4, Pp. 133–142. DOI: 10.26583/sv.12.4.12

16. **Janushko V.V., Erkin S.N.** Construction of the process automation product design based on UML (use case) diagrams. Izvestiya SFedU. Engineering Sciences, 2009, Vol. 101, No. 12, Pp. 64–71.

17. **Petrov A.A., Nikiforov I.V., Ustinov S.M.** Algorithm of ESXi cluster migration between different vCenter servers with the ability to rollback. Informatsionno-upravliaiushchie sistemy [Information and Control Systems], 2022, Vol. 2, Pp. 20–31. DOI: 10.31799/1684-8853-2022-2-20-31

18. **Kovalev A.D., Nikiforov I.V., Drobintsev P.D.** Automated approach to semantic search through software documentation based on Doc2Vec algorithm. Informatsionno-upravliaiushchie sistemy [Information and Control Systems], 2021, Vol. 1, Pp. 17–27. DOI: 10.31799/1684-8853-2021-1-17-27

19. **Magomedov A.M.** Postroenie i vizual'noe redaktirovanie grafa s sokrashchennymi spiskami smezhnosti vershin [Construction and visual editing of a graph with reduced adjacency lists of vertices]. State registration of a computer program RU 2017617670, 2017.

20. **Berezutskaia L.A., Troshkina G.N., Iudintsev A.Iu.** SQLite main commands and their comparison to SQL for an Android app. LinguaNet, 2019, Pp. 289–293.

21. **Sedykh D.V., Zuyev D.V., Gordon M.A.** Branch format of technical documentation on devices of railway automation and remote control. Part 4. Presentation of elements. Transport Automation Research, 2017, Vol. 3, No. 4, Pp. 563–577.

22. **Ivlev V.A., Nikiforov I.V., Leont'eva T.V.** Obrabotka dannykh v geoinformatsionnykh sistemakh dlia vybora mestopolozheniia reklamy [Processing data in geographic information systems to select advertising locations]. Sovremennye tekhnologii v teorii i praktike programmirovaniia [Modern technologies in the theory and practice of programming], 2019, Pp. 27–30.

23. **Lisyak V.V., Lisyak M.V.** The resulting matrix of interactive geometrical transformations composition search method in CAD. Izvestiya SFedU. Engineering Sciences, 2008, Vol. 81, No. 4, Pp. 73–78.

24. **Kudrina M.A., Murzin A.V.** Affinnye preobrazovaniia ob"ektov v komp'iuternoi grafike [Affine transformations of objects in computer graphics]. Trudy mezhdunarodnogo simpoziuma "Nadezhnost' i kachestvo" [Proceedings of the International Symposium "Reliability and Quality"], 2014, Vol. 1, Pp. 307–310.

25. **Ustinova V.E., SHpak A.V., Ivlev V.A., Mironenkov G.V.** Sistema preobrazovaniia koordinat ob"ektov v svg-faile dlia poiska sootvetstviia modeli zheleznodorozhnoi stantsii ee vizual'nomu predstavleniiu [System of transformation of coordinates of objects in svg-file for search of correspondence of model of railway station to its visual representation]. Sovremennye tekhnologii v teorii i praktike programmirovaniia [Modern technologies in the theory and practice of programming], 2024, Pp. 132–134.

26. **Sysoev V.V.** A Framework Similarity Estimation of Unweighted Undirected Graphs. Modern Information Technologies and IT-Education, 2022, Vol. 18, No. 3, Pp. 655–665. DOI: 10.25559/SITITO.18.202203.655-665

27. **Khonina O.I., Zabrodin A.V.** Graph Coloring Problem in the Context of Schedule Optimization: Software Solution. Intellectual Technologies on Transport, 2023, Vol. 35, No. 3, Pp. 32–37. DOI: 10.24412/2413-2527-2023-335-32-37

28. **Fataliyev A.** Most important TPPs for performing graphical user interface and scientific calculations in Python programming environment. The Scientific Heritage, 2023, Vol. 124, Pp. 37–41. DOI: 10.5281/zenodo.10077616

29. **Voinov N., Rodriguez Garzon K., Nikiforov I., Drobintsev P.** Big Data Processing System for Analysis of GitHub Events. 2019 XXII International Conference on Soft Computing and Measurements (SCM), 2019, Pp. 187–190. DOI: 10.1109/SCM.2019.8903782

30. **Zharinov I., Zharinov O., Shek-Iovsepyantz R., Suslov V.** Robustness drop estimation of design documentation preparation by CALS-technologies in instrument making. Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2012, Vol. 80, No. 4, Pp. 151–153.

31. **Egorova N.E., Arbuzova A.A.** On the formation of the skill of processing and visualization of statistical data. Pozharnaia i avariinaia bezopasnost' [Fire and emergency safety], 2019, Vol. 13, No. 2, Pp. 38–45.

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Ustinova Valeria E.
Устинова Валерия Евгеньевна
E-mail: ustlera@list.ru

Lutsenko Anton S.
Луценко Антон Степанович
E-mail: anton.lutsenko.03@mail.ru

Shpak Adelina V.
Шпак Аделина Владимировна
E-mail: adelina.shpak@yandex.ru

Mironenkov Grigorii V.
Мироненков Григорий Васильевич
E-mail: mironenkov97@gmail.com

Ivlev Vladislav A.
Ивлев Владислав Александрович
E-mail: nevidd@yandex.ru

Submitted: 24.06.2024; Approved: 05.11.2024; Accepted: 18.11.2024.
Поступила: 24.06.2024; Одобрена: 05.11.2024; Принята: 18.11.2024.

Computing, Telecommunication and Control, 2024, Vol. 17, No. 4, Pp. 78–89.
Информатика, телекоммуникации и управление. 2024. Том 17, № 4. С. 78–89.

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

Research article
DOI: https://doi.org/10.18721/JCSTCS.17407
UDC 621.3.049.774.2

DESIGN PROCESS OF THE MASH 2-2 SIGMA-DELTA MODULATOR

A.S. Kozlov[1] ✉ , *M.M. Pilipko*[2] ,
A.T. Tulaev[1], *Ya.V. Belyaev*[1]

[1] Concern CSRI Elektropribor, JSC,
St. Petersburg, Russian Federation;

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

✉ kas573@yandex.ru

Abstract. This paper presents the design process of the cascaded (MASH) 2-2 sigma-delta modulator. For this purpose, several sigma-delta modulator topologies were studied at the system and schematic levels. Simulation at the system level was used to define the requirements for the operational transconductance amplifier (OTA), which is a part of the integrator. Sigma-delta modulators were designed using a 0.18 µm CMOS technology library. The influence of temperature swing, process variations and noise on the characteristics of second-order sigma-delta modulators was taken into account during simulation. As a result of the simulation, the optimal second-order topology was selected. This topology was used to design the cascaded (MASH) 2-2 sigma-delta modulator. The developed modulator has a resolution of 15.97 effective bits, a working frequency band of 20 kHz, consumes 12 mW of power at a supply voltage of 3.3 V. The occupied area of the circuit on the chip is 0.22 mm^2. A device with such characteristics can be used in interfaces of micromechanical sensors.

Keywords: ADC, sigma-delta, OTA, comparator, topology

Acknowledgements: M.M. Pilipko expresses gratitude for the support of the basic research provided within the framework of the state assignment “Functional nanocomposite and nanostructured materials for advanced micro- and nanoelectronic devices” (FSEG-2023-0016).

Citation: Kozlov A.S., Pilipko M.M., Tulaev A.T., Belyaev Ya.V. Design process of the MASH 2-2 sigma-delta modulator. Computing, Telecommunications and Control, 2024, Vol. 17, No. 4, Pp. 78–89. DOI: 10.18721/JCSTCS.17407

Научная статья
DOI: https://doi.org/10.18721/JCSTCS.17407
УДК 621.3.049.774.2

РАЗРАБОТКА MASH 2-2 СИГМА-ДЕЛЬТА МОДУЛЯТОРА

А.С. Козлов[1] ✉ , *М.М. Пилипко*[2] , *А.Т. Тулаев*[1], *Я.В. Беляев*[1]

[1] АО «Концерн «ЦНИИ «Электроприбор», Санкт-Петербург, Российская Федерация;

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

✉ kas573@yandex.ru

Аннотация. В данной статье представлена разработка каскадного (MASH) 2-2 сигма-дельта модулятора. В ходе работы были рассмотрены различные архитектуры сигма-дельта модуляторов на системном и схемотехническом уровнях. Моделирование на системном уровне позволило определить характеристики операционного транскондуктивного усилителя (ОТУ), входящего в состав интегратора. Для разработки схемотехнической реализации сигма-дельта модуляторов использовалась технологическая библиотека КМОП-технологии с нормами 0,18 мкм. На схемотехническом уровне было проведено моделирование структур второго порядка с учетом шума, температуры, а также разброса параметров технологии. В результате моделирования была выбрана оптимальная структура второго порядка. На основе этой структуры была построена схема каскадного сигма-дельта модулятора. Разработанный модулятор имеет разрешающую способность 15,97 эффективных бит, полосу рабочих частот 20 кГц, занимает на кристалле площадь в 0,22 мм^2 и потребляет 12 мВт мощности при напряжении питания 3,3 В. Устройство с такими характеристиками может применяться в интерфейсах микромеханических датчиков и сенсоров.

Ключевые слова: АЦП, сигма-дельта, ОТУ, компаратор, топология

Благодарности: М.М. Пилипко выражает благодарность за поддержку базового исследования, оказанную в рамках государственного задания «Функциональные нанокомпозитные и наноструктурированные материалы для перспективных устройств микро- и наноэлектроники» (FSEG-2023-0016).

Для цитирования: Kozlov A.S., Pilipko M.M., Tulaev A.T., Belyaev Ya.V. Design process of the MASH 2-2 sigma-delta modulator // Computing, Telecommunications and Control. 2024. Т. 17, № 4. С. 78–89. DOI: 10.18721/JCSTCS.17407

Introduction

The rapidly growing market for microelectromechanical systems (MEMS) is driving the demand for precision interface chips. Analog-to-digital converters (ADCs) are an essential part of such circuits, since without them, further digital processing of the signal from the MEMS device is impossible. The output signals of MEMS sensors usually do not exceed tens of kilohertz, so ADCs based on sigma-delta (ΣΔ) modulation are most often used to convert them into digital form. A sigma-delta ADC consists of an analog (sigma-delta modulator) and a digital (filter-decimator) part. This type of ADC is capable of providing high resolution (up to 24 bits). Typically, sigma-delta ADCs consume much less power than faster ADC architectures. Sigma-delta ADCs achieve high resolution by using noise shaping and high sampling rates above the Nyquist frequency (oversampling).

Sigma-delta modulators can be roughly divided into two general categories: continuous-time (CT) and discrete-time (DT) circuits. In CT circuits, the only clocked unit is the comparator, while the integrator is implemented as a Gm-C, active RC or passive RC filter. Sigma-delta modulators of this type consume less power and are capable of operating at higher clock rates than DT modulators. According to [1], the highest operating frequency band for the CT modulator achieved to date is 465 MHz, while for the DT modulator this value is only 40 MHz. However, CT modulators also have their drawbacks. First, they have increased sensitivity to jitter, since errors in the digital-to-analog converter (DAC) occur due to an unstable clock signal. In addition, delays in the DAC and quantizer can introduce additional poles in the signal transfer function (SiTF) and noise transfer function (NTF), which leads to degraded performance and possible instability of the modulator. This effect is called excess loop delay (ELD) [2].

In contrast, DT sigma-delta modulators have lower bandwidth and higher performance requirements for an operational amplifier, which is a part of an integrator. Nevertheless, they have become widespread due to their immunity to jitter and ELD. It is also worth noting that the characteristics of DT modulators are less sensitive to process variations.

Simulation of second-order topologies at system level

First-order sigma-delta modulators are easy to design and inherently immune to instability. Nevertheless, the relatively low level of the suppression of quantization noise at moderate values of the oversampling ratio significantly limits their scope of application. This is why second-order sigma-delta modulators have become the most widely used.

Interest in second-order sigma-delta modulators has led to the emergence of a number of different structures. The best known is Boser–Wooley modulator – cascade of integrators with feedback (CIFB) [3]. Reducing the gains of the integrators improves the stability of the modulator for signals with high amplitude, and the use of delaying integrators makes it possible to reduce the speed requirements for the operational transconductance amplifier (OTA). Integrators' gains are chosen such that $STF(z) = z^{-2}$ and $NTF(z) = (1 - z^{-1})^2$.

The so-called feedforward topology was first introduced in [4]. Some papers also referred to it as the ideal topology and the broadband topology. The main advantages are the reduced requirements for the signal swing of the OTA (since the integrator processes only the quantization noise), as well as an increased dynamic range. This modulator implements $STF(z) = 1$ and $NTF(z) = (1 - z^{-1})^2$.

Another version of the sigma-delta modulator, called the cascade of integrators with distributed feedback topology with capacitive input feedforward (CIFB-CIF), was presented in [5]. The proposed modifications made it possible to reduce the speed requirements for the modulator components (it is especially important in the case of a modulator with a multi-bit quantizer, since in this case a multi-bit DAC with digital element matching is used), and also to omit the adder before the quantizer. Like the feedforward topology, this modulator implements $STF(z) = 1$ and $NTF(z) = (1 - z^{-1})^2$.

The following structure was first considered in [6]. A similar structure was published in [7].

It can be noted that this architecture is also characterized by the absence of an adder in front of the comparator. In addition, the modulator has only one feedback branch, therefore, to implement this structure, a single DAC is required, which makes it possible to reduce the area of the circuit on the chip, as well as power consumption. In this structure, $STF(z) = 1$ and $NTF(z) = (1 - z^{-1})^2$.

It is also possible to design a topology with $STF(z) = z^{-1}$ and $NTF(z) = (1 - z^{-1})^2$. This topology was considered in [8]. A hybrid topology is a kind of CIFB architecture. The hybrid topology, like the feedforward one, is characterized by reduced requirements for integrator nonlinearities. In addition, in this case, the number of signal branches is smaller and there is no adder in front of the comparator, therefore the power consumed by the modulator is reduced.

The noise performance of second-order topologies can be a limiting factor for their use in high-precision applications. Higher order sigma-delta modulators present an effective way to reduce quantization noise in signal band. The main drawback of high-order single-loop architectures is their vulnerability to instability. Multi-stAge noise SHaping (MASH) architecture can be immune to instability, while demonstrating noise suppression comparable to high-order single-loop sigma-delta modulator. This is why MASH topologies are of interest.

Theoretically, any of the second-order topologies mentioned above can be used to design a MASH 2-2 sigma-delta modulator. However, different non-idealities should be taken into account. This is why simulation at a system level is an essential part of the design process. In [9, 10], an analysis of the influence of various factors on the performance of the first-order sigma-delta modulator was carried out. Nevertheless, the authors note that such an analysis needs to be performed for various architectures. In this regard, the effect of the finite gain, the finite gain-bandwidth and the finite slew rate will be discussed.

According to [11, 12], the finite value of the gain leads to an effect called "integrator leakage". The parameter p, which characterizes this effect, is defined as:

$$p = \frac{k-1}{k},$$

where k is the amplifier gain. Due to leakage, the gain of the integrator turns out to be different from the nominal value. It should also be noted that leakage changes the position of zeros in the noise transfer function, thereby increasing the noise power in the useful frequency band. These effects can greatly affect the operation of sigma-delta modulators based on the MASH architecture.

The influence of the gain bandwidth on the transfer function of the integrator was considered in [13]. Taking into account the gain bandwidth modifies the transfer functions as follows:

Non-inverting integrator:

$$H(z) = \frac{C_1}{C_2}\frac{z^{-1}}{1-z^{-1}}\left[1-e^{-k_1}\left(\frac{C_1}{C_1+C_2}\right)\right].$$

Inverting integrator:

$$H(z) = -\frac{C_1}{C_2}\frac{1}{1-z^{-1}}\left[1-e^{-k_1}+e^{-k_1}\left(\frac{C_2}{C_1+C_2}\right)\right],$$

$$k_1 = \pi\left(\frac{C_2}{C_1+C_2}\right)\left(\frac{f_t}{f_s}\right),$$

where f_t is the unity gain frequency, f_s is the clock frequency, C_1 is the input capacitor, C_2 is the feedback capacitor.

In practice, in switched-capacitor circuits, it is recommended to use amplifiers with gain bandwidth of at least 5 times the clock signal frequency [14, 15].

The slew rate (SR) of the op-amp output is also an important parameter that determines the overall performance of the sigma-delta modulator. The finite SR causes the gain to be nonlinear [11]. To account for this parameter, changes in the output signal are described using the equations from [16].

To implement a high SR, it is necessary to increase the current of the amplifier output stage. This leads to an increase in power consumption and an increase in the occupied area on the chip (due to wider devices). It is also possible to increase the SR by using class AB output stage amplifiers; however, these are difficult to design and generally require frequency compensation.

According to [17], the SR requirements are highly dependent on the gain bandwidth. To achieve high signal-to-noise-and-distortion ratio (SNDR) values, it is necessary to provide one of two operating modes of the integrator: slow or fast.

The normalized SR is determined by the expression:

$$SR_N = SR\frac{T_s}{2V_{pp}},$$

where T_s is the clock period, V_{pp} is the value of the feedback signal of the sigma-delta modulator (determined by the supply voltage).

In the slow mode, it is assumed to use an amplifier with a relatively small gain bandwidth (few times of f_s). The normalized SR in this mode must exceed the value of n_τ, which is defined as follows:

$$n_\tau = \frac{T_s}{2\tau},$$

where τ is the amplifier's time constant

$$\tau = \frac{1}{2\pi f_t}.$$

When working in this mode, two difficulties arise. First, a high SR is required for a broadband amplifier to operate correctly. Satisfying this criterion is especially problematic if the amplifier is loaded with a large capacitance. Second, if the gain bandwidth of the amplifier is higher than expected after the circuit is fabricated, the SR_N may occur less than n_τ, which will lead to a decrease in the SNDR of the modulator.

To implement a circuit operating in a fast mode, an amplifier with a bandwidth such that $n_\tau > 12$ is required. In this case, the SR requirements are significantly reduced.

To summarize all of the above, it should be noted that for each sigma-delta modulator topology, system-level models with OTA's non-idealities were designed. The model for the feedback topology is shown in Fig. 1.

The simulation results show that for $f_t = 20f_s$ the dependence of the SNDR on the amplitude of the input signal becomes close to ideal (with an infinitely high gain) when the normalized SR is higher than 5.

Similar dependencies were obtained for a finite gain and finite gain-bandwidth. All considered sigma-delta modulator topologies have shown approximately the same results.

System-level simulations show that for the MASH 2-2 modulator with SNDR > 90dB, input signal bandwidth of 20 kHz and oversampling ratio of 128, non-idealities of an OTA become negligible (do not degrade overall modulator performance) if its parameters meet the following requirements:

- DC gain of more than 40 dB;
- Gain bandwidth of more than 100 MHz;
- Slew rate of more than 82.5 V/us.

Simulation of the second-order topologies at the schematic level

In order to design second-order sigma-delta modulators at the schematic level, several main blocks are required: a CMOS switch, a comparator, and an OTA. The switch has a finite resistance, which, in the case of using a single transistor with a certain type of conductivity, turns out to be highly dependent on the input signal level. Using a switch based on a complementary pair of transistors allows one to effectively cope with this problem. The clock feedthrough effect, however, remains. One of the options for reducing the clock feedthrough is to use a CMOS switch with additional dummy transistors [18].

A comparator is required to convert the signal from the integrator output to a pulse train. Clocked comparators are of interest because they do not consume static power. An example of such a comparator

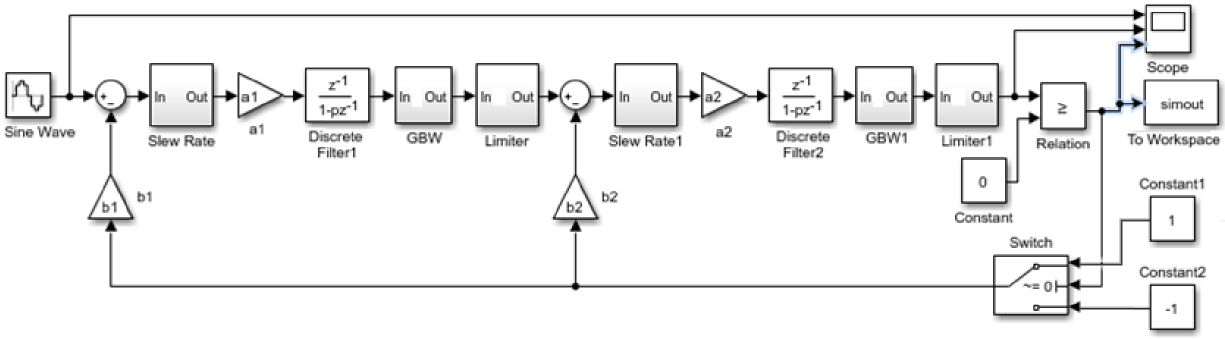


Fig. 1. System-level model of the feedback topology sigma-delta modulator

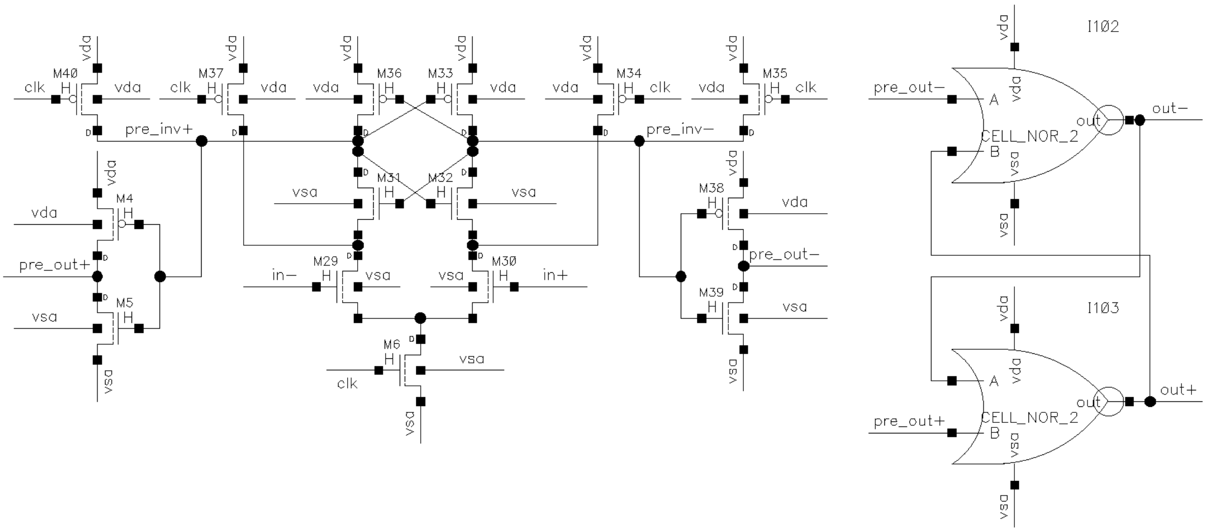


Fig. 2. StrongARM latch

is a structure called StrongARM latch [19], shown in Fig. 2. When a low-level signal is applied to the clock inputs of the comparator, the p-transistors connected in parallel with the bistable cell in the upper part of the circuit are turned on, and the current source in the form of an n-transistor in the lower part of the circuit is locked. As a result of resetting the comparator value, a logical zero value is set at both its outputs, regardless of the signals at the comparator inputs. In order to store the information signal while the comparator is reset, an RS-latch is used.

In the previous paragraph, the following OTA's requirements were obtained: DC gain of more than 40 dB, gain bandwidth of more than 100 MHz, slew rate of more than 82.5 V/us. In addition, in order to ensure high linearity, the OTA should have a rail-to-rail output signal swing. A folded-cascode amplifier meets these requirements. The amplifier with feedforward compensation from [20] was taken as a basis. For better stabilization of the operating point, a common mode feedback circuit (CMFB) in the form of a differential amplifier with diode-connected transistors as a load was used. Fig. 3 shows the final OTA schematic. Simulation results showed that the designed OTA has the following characteristics: DC gain of 75 dB, gain bandwidth of 200 MHz and slew rate of 102 V/us with a capacitive load of 4 pF. Performance margin is added in order to ensure the reliability of the design across all process corners.

After designing the main blocks, the design of the sigma-delta modulators themselves follows. Based on the analysis carried out earlier, an implementation with switched capacitors was chosen. All structures are fully differential.

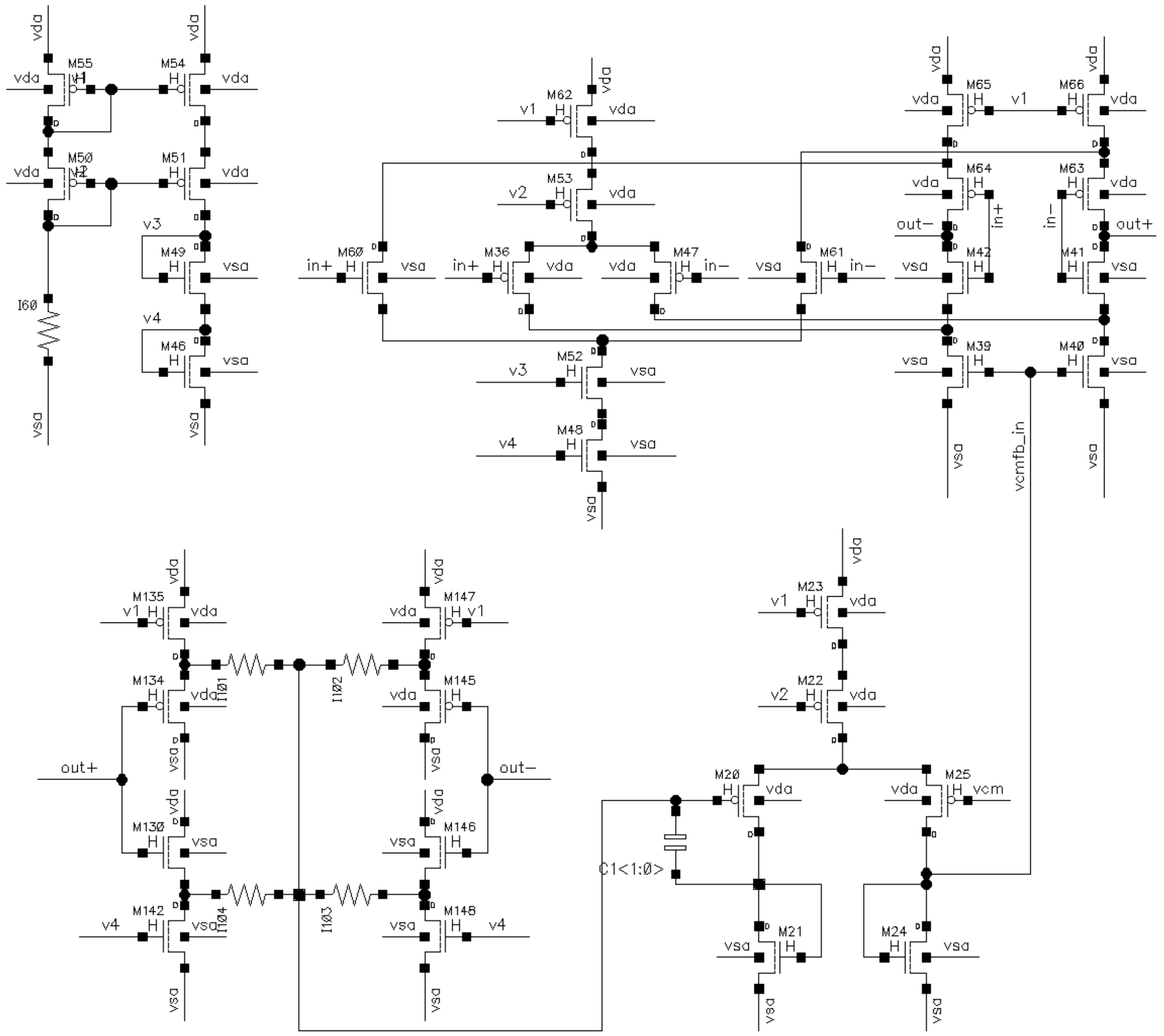


Fig. 3. Folded-cascode OTA

As a result of the simulation, the dependence of the SNDR of the output signal versus the amplitude of the input signal was obtained for different second-order topologies (Fig. 4). The graphs for feedback (FB), feedforward (FF [4]), hybrid (HY [21]), Gharbiya (GH [5]) and Murayama (MU [6]) topologies are shown.

Next, the simulation was carried out taking into account the influence of temperature swing, process variation and the intrinsic electrical noise of the circuit. Further simulation was carried out at a single signal amplitude of 1.2 V.

To study the effect of temperature on the characteristics of modulators, three temperature values were chosen: –40°C, 27°C, 85°C. Based on the data obtained, it can be concluded that the Murayama structure [6] is the most sensitive to the effect of temperature: the decrease in SNDR at the temperature of –40°C was 24 dB compared to the nominal value.

The technology libraries used in the project allow for simulation with process variations. The largest deviations of the parameters from the nominal values are called “corners”. The simulation showed that for all five structures, the change in characteristics turns out to be insignificant.

Simulation with intrinsic noise was performed using the transient noise option. The SNDR value for the Murayama structure decreased by 24 dB, for the Gharbiya structure – by 6.4 dB, and for the

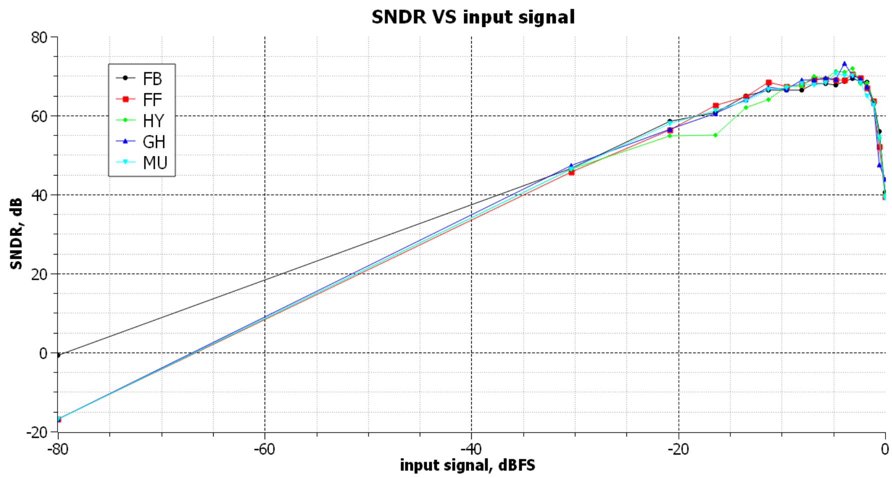


Fig. 4. SNDR vs input signal amplitude for different second-order topologies

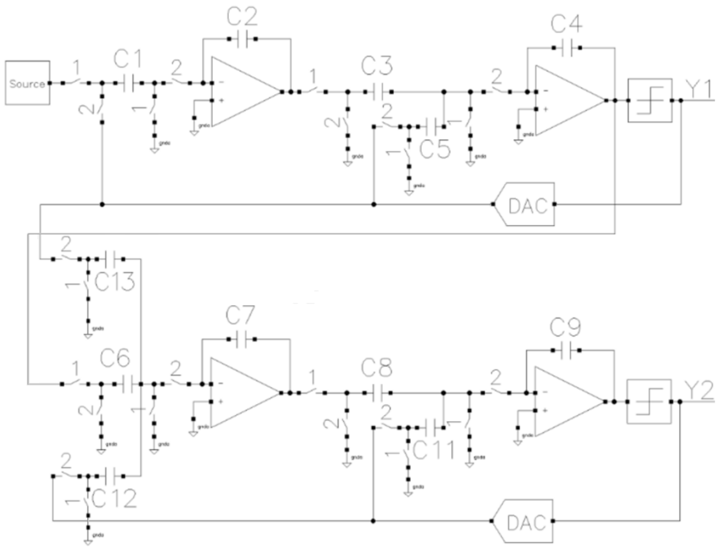


Fig. 5. Analog part of the MASH 2-2 sigma-delta modulator

feedforward structure – by 2.5 dB. The SNDR value for the feedback and hybrid structures even improved slightly after adding noise.

According to the simulation results, it is clear that the most reliable structures are the feedback, feedforward, and hybrid. In addition, as the feedback structure requires the least number of capacitors, it was decided to build the MASH sigma-delta modulator based on it.

The MASH 2-2 sigma-delta modulator

Next, the schematic of MASH 2-2 sigma-delta modulator was designed. The analog part is shown in Fig. 5. The first stage is presented by the feedback topology shown earlier. In the second stage, changes were made to this circuit. The modulator feedback is implemented using a separate capacitor C12. Capacitors C6 and C13 are used to subtract the code of the first stage from its residue analog signal. Signals Y1 and Y2 are then fed to the digital processing circuit. The resulting multibit signal has the properties of fourth-order noise shaping.

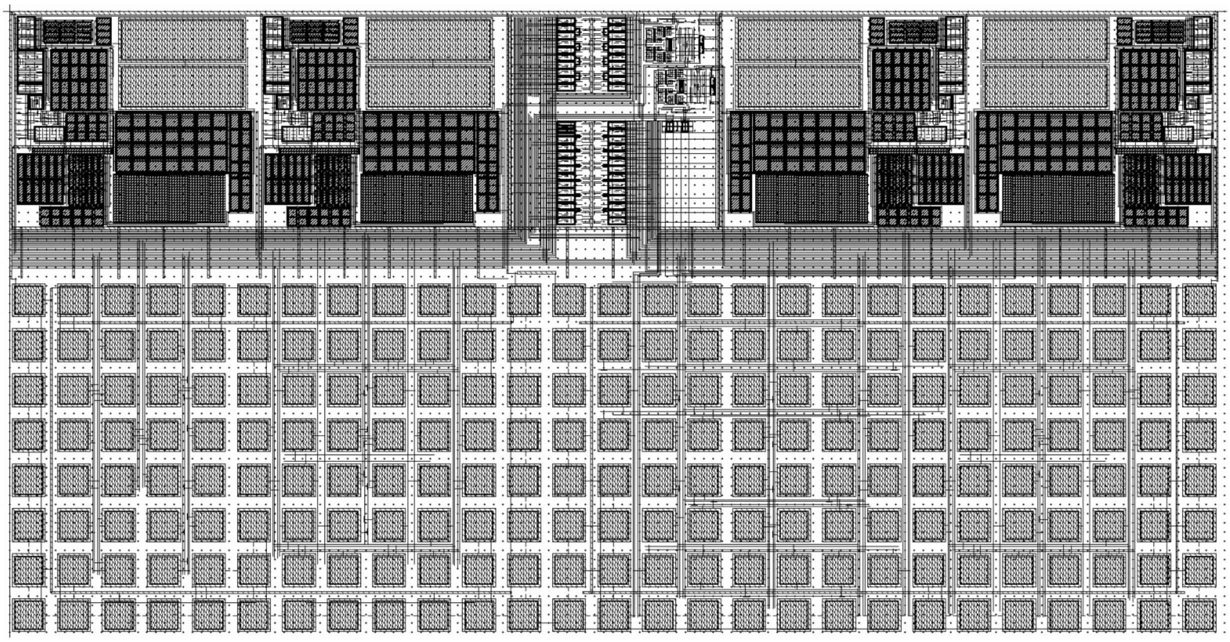


Fig. 6. The layout of the MASH 2-2 modulator core

The digital processing circuit is built based on adders and D-flip-flops. The output of the circuit is a parallel 5-bit digital code. The least significant bit is taken from the output of one of the D-flip-flops, the other bits are signals at the outputs of the adders. One of the inputs of the last adder is set to logic one in order to ensure that there are no negative values in the code.

The simulation of the MASH 2-2 sigma-delta modulator at the schematic level showed that the designed circuit can achieve an SNDR value of 97 dB.

The assembled layout of the MASH 2-2 sigma-delta modulator is presented in Fig. 6. The capacitor arrays and switches are seen in the bottom part while four OTAs, two comparators and the digital part are placed in the upper part. Overall characteristics of the designed modulator are as follows:

- Effective number of bits: 15.97;
- Power consumption: 12 mW (at a supply voltage of 3.3 V);
- Signal band: 20 kHz;
- Chip area: 0.22 mm^2.

Table 1 shows the comparison of the designed circuit with circuits from previous publications.

Table 1

Main parameters summary and comparisons

	This work	[22]	[23]	[24]	[25]
DR (bit)	15.97	17.9/14.8	12.29	14.49	14.32
BW, kHz	20	8000/20000	20000	20	20
OSR	128	20/8	12.5	128	100
Architecture	2-2	2-2-1	2-2	2-2	2-2
Supply voltage, V	3.3	3.3	1.2	0.9	1
Power, mW	12	68	20.4	0.065	0.66

Conclusion

The paper presents the design process of the MASH 2-2 sigma-delta modulator. Simulation at the system level was used to define the requirements for the requirements for the operational amplifier, which is the key building block of the sigma-delta modulators.

The circuits of the main blocks of a sigma-delta modulator based on a 0.18 μm CMOS technology library have been designed, and simulation has been carried out. Based on the developed blocks, five structures of second-order sigma-delta modulators were assembled, simulated and compared. The influence of temperature swing, process variations and noise on the characteristics of second-order sigma-delta modulators was taken into account during simulation. Simulation results showed that the feedback topology is the most suitable for implementing a cascaded sigma-delta modulator.

A schematic for the MASH 2-2 sigma-delta modulator has been designed, and the characteristics of the developed device have been obtained. With an effective number of bits equal to 15.97 in a signal band of 20 kHz, the circuit is suitable for audio systems and other high-resolution low-power applications.

REFERENCES

1. **Liu Q., Edward A., Briseno-Vidrios C., Silva-Martinez J.** Design Techniques for Mash Continuous-Time Delta-Sigma Modulators. Switzerland: Springer Cham, 2018. DOI: 10.1007/978-3-319-77225-7

2. **Nowacki B., Paulino N., Goes J.** A third-order MASH ΣΔ modulator using passive integrators. IEEE Transactions on Circuits and Systems I: Regular Papers, 2017, Vol. 64, No. 11, Pp. 2871–2883. DOI: 10.1109/TCSI.2017.2704164

3. **Boser B.E., Wooley B.A.** The design of sigma-delta modulation analog-to-digital converters. IEEE Journal of Solid-State Circuits, 1988, Vol. 23, No. 6, Pp. 1298–1308.

4. **Silva J., Moon U., Steensgaard J., Temes G.C.** Wideband low-distortion delta-sigma ADC topology. Electronics Letters, 2001, Vol. 37, No. 12, Pp. 737–738. DOI: 10.1049/el:20010542

5. **Gharbiya A., Johns D.A.** On the implementation of input-feedforward delta-sigma modulators. IEEE Transactions on Circuits and Systems II: Express Briefs, 2006, Vol. 53, No. 6, Pp. 453–457. DOI: 10.1109/TCSII.2006.873829

6. **San H., Konagaya H., Xu F., Motozawa A., Kobayashi H., Ando K., Yoshida H., Murayama C.** Second-order ΔΣAD modulator with novel feedforward architecture. 2007 50th Midwest Symposium on Circuits and Systems, 2007, Pp. 148–151.

7. **Guinea J., Sentieri E., Baschirotto A.** An instrumentation 128dB-SNR 750μA SDM. 2012 Proceedings of the ESSCIRC (ESSCIRC), 2012, Pp. 205–208. DOI: 10.1109/ESSCIRC.2012.6341294

8. **Morgado A., del Río R., Rosa J.M.** Nanometer CMOS Sigma-Delta Modulators for Software Defined Radio. NY: Springer New York, 2011. DOI: 10.1007/978-1-4614-0037-0

9. **Lorenz M., Bruckner T., Ritter R., Ortmanns M.** Concurrent estimation of amplifier nonidealities and excess loop delay in continuous-time sigma-delta modulators. 2013 IEEE International Symposium on Circuits and Systems (ISCAS), 2013, Pp. 1031–1034.

10. **Meyer A., Wunderlich R., Heinen S.** Analysis of opamp non-idealities in quadrature bandpass sigma-delta modulators. 2020 27th IEEE International Conference on Electronics, Circuits and Systems (ICECS), 2020, Pp. 1–4. DOI: 10.1109/ICECS49266.2020.9294849

11. **Wagner J., Reich S., Ritter R., Anders J., Ortmanns M.** Finite GBW in single OpAmp CT ΣΔ modulators. 2016 IEEE International Conference on Electronics, Circuits and Systems (ICECS), 2016, Pp. 468–471. DOI: 10.1109/ICECS.2016.7841240

12. **Feely O., Chua L.O.** The effect of integrator leak in Sigma-Delta modulation. IEEE Transactions on Circuits and Systems, 1991, Vol. 38, No. 11, Pp. 1293–1305.

13. **Martin K., Sedra A.** Effects of the op amp finite gain and bandwidth on the performance of switched-capacitor filters. IEEE Transactions on Circuits and Systems, 1981, Vol. 28, No. 8, Pp. 822–829. DOI: 10.1109/TCS.1981.1085052

14. **Bafandeh A., Yavari M.** Digital calibration of amplifier finite DC gain and gain bandwidth in MASH ΣΔ modulators. IEEE Transactions on Circuits and Systems II: Express Briefs, 2015, Vol. 63, No. 4, Pp. 321–325. DOI: 10.1109/TCSII.2015.2504027

15. **Hussain A., Hanif M.** Active-pasisve delta-sigma modulator ADC for MEMS accelecrometers. 2017 International Symposium on Recent Advances in Electrical Engineering (RAEE), 2017, pp. 1–5.

16. **Williams L.A., Wooley B.A.** A third-order sigma-delta modulator with extended dynamic range. IEEE Journal of Solid-State Circuits, 1994, Vol. 29, No. 3, Pp. 193–202. DOI: 10.1109/4.278340

17. **Dumitru F.-S., Mihalache S., Enachescu M.** OPAMP's finite gain and slew rate impact on a 16-bit ΣΔ ADC performance: A case study. 2017 International Semiconductor Conference (CAS), 2017, Pp. 161–164. DOI: 10.1109/SMICND.2017.8101187

18. **Eichenberger C., Guggenbuhl W.** On charge injection in analog MOS switches and dummy switch compensation techniques. IEEE Transactions on Circuits and Systems, 1990, Vol. 37, No. 2, Pp. 256–264. DOI: 10.1109/31.45719

19. **Razavi B.** The StrongARM latch. IEEE Solid-State Circuits Magazine, 2015, Vol. 7, No. 2, Pp. 12–17. DOI: 10.1109/MSSC.2015.2418155

20. **Lampinen H., Vainio O.** An optimization approach to designing OTAs for low-voltage sigma-delta modulators. IEEE Transactions on Instrumentation and Measurement, 2001, Vol. 50, No. 6, Pp. 1665–1671. DOI: 10.1109/19.982965

21. **Kozlov A.S., Pilipko M.M.** A second-order sigma-delta modulator with a hybrid topology in 180nm CMOS. 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), 2020, Pp. 144–146. DOI: 10.1109/EIConRus49466.2020.9039246

22. **Li D., Qian X., Li R., Fei C., Jiang L., Chen X.** High resolution ADC for ultrasound color Doppler imaging based on MASH sigma-delta modulator. IEEE Transactions on Biomedical Engineering, 2019, Vol. 67, No. 5, Pp. 1438–1449. DOI: 10.1109/TBME.2019.2938275

23. **Kwak Y.-S., Cho K.-I., Kim H.-J., Lee S.-H., Ahn G.-C.** A 72.9-dB SNDR 20-MHz BW 2-2 discrete-time resolution-enhanced sturdy MASH Delta-Sigma modulator using source-follower-based integrators. IEEE Journal of Solid-State Circuits, 2018, Vol. 53, No. 10, Pp. 2772–2782. DOI: 10.1109/JSSC.2018.2859401

24. **Honarparvar M., Safi-Harb M., Sawan M.** An amplifier-shared inverter-based MASH structure ΔΣ modulator for smart sensor interfaces. 2016 IEEE International Symposium on Circuits and Systems (ISCAS), 2016, Pp. 2250–2253. DOI: 10.1109/ISCAS.2016.7539031

25. **Kuo C.-H., Shi D.-Y., Chang K.-S.** A low-voltage fourth-order cascade delta–sigma modulator in 0.18-μm CMOS. IEEE Transactions on Circuits and Systems I: Regular Papers, 2010, Vol. 57, No. 9, Pp. 2450–2461. DOI: 10.1109/TCSI.2010.2046231

INFORMATION ABOUT AUTHORS / СВЕДЕНИЯ ОБ АВТОРАХ

Kozlov Alexey S.
Козлов Алексей Сергеевич
E-mail: kas573@yandex.ru

Pilipko Mikhail M.
Пилипко Михаил Михайлович
E-mail: m_m_pilipko@rambler.ru
ORCID: https://orcid.org/0000-0003-3813-6846

Tulaev Artyom T.
Тулаев Артем Толибович
E-mail: artulaev@gmail.com

Belyaev Yakov V.
Беляев Яков Валерьевич
E-mail: jak0b@mail.ru

Submitted: 26.10.2024; Approved: 19.12.2024; Accepted: 23.12.2024.
Поступила: 26.10.2024; Одобрена: 19.12.2024; Принята: 23.12.2024.