Review of the polar codes problems from the standpoint of noiseproof coding Optimization Theory technologies

N.A.Kuznetsov, V.V. Zolotarev, G.V. Ovechkin, R.R. Nazirov,
D.J. Satybaldina, E.D. Omirbayev
1) Institute of Radio Engineering and Electronics named after V.A. Kotelnikov RAS, Moscow, 125009, Russia
E-mail: kuznetsov@cplire.ru;
2) Space Research Institute, Russian Academy of Sciences, Moscow, 117997, Russia
E-mail: zolotasd@yandex.ru;
3) Ryazan State Radio Engineering University named after V.F. Utkina, Ryazan, 390005, Russia
E-mail: g_ovechkin@mail.ru;
4) Eurasian National University named after L.N. Gumilyov, Kazakhstan, 010008, Astana
E-mail: satybaldina_dzh@enu.kz

The general situation in applied questions of coding theory is considered. The main problems arising in the field of noiseproof codes decoding are stated. Along with the correcting ability, special attention is paid to the computational complexity of these algorithms. The latest results in the field of decoding algorithms for polar codes (PC) are presented, the main problems of their development are outlined. A comparison of the applied results of the Optimization Theory (OT) and the available extremely limited materials for the PC is carried out. Results for low density (LDPC) codes are briefly mentioned. The results of comparative analysis of PC characteristics and the block version of the Viterbi algorithm (BVA) for short codes are presented. Comparison of the capabilities of PC and MTD algorithms is also made, including cascading. The main directions of development and improvement of the OT characteristics algorithms are given. Based on the results of the comparison, it was concluded that the OT is unconditional leadership and that there is no need to use a PC and a number of other codes anywhere in general due to the inevitably weak capabilities and a large list of shortcomings of decoders of these directions and methods of their development in research for new projects of satellite and space communications, and also for remote sensing systems.

Keywords: Shannon border, channel capacity, algorithms complexity, polar codes (PC), list decoding, optimal decoding (OD), Optimization Theory (OT), block Viterbi algorithm (BVA), Reed-Solomon codes (RS), low density (LDPC) and turbo codes, multi-threshold decoders (MTD), parallel cascading, divergent principle, direct metric control decoders (DPCM).

General situation in applied questions of coding theory

As you know, turbo codes, like decoders for low-density parity-check codes (LDPC), which appeared in the last decade of 20th century, have not become the locomotives of the development of any new directions in the coding theory of binary data streams, which we have repeatedly pointed out in our main monographs and key publications [4, 5, 6, 10, 37, 39], in reviews [7, 11, 12, 20, 38], reference book [9] and on our network portals www.mtdbest.ru and www .mtdbest.iki.rssi.ru. There were many reasons for this.

When considering only binary codes of these classes, efficient algorithms for their decoding do not fit into the framework of fixed-point operations, i.e. such decoders are forced to use the arithmetic of real numbers, which immediately translated them into the category of not the simplest methods. In addition, the complexity these codes decoding is either significantly greater than linear, or requires the computation of very inconvenient functions, which additionally significantly slows down their work. Moreover, the estimates of the complexity N for these codes themselves are actually very mysterious, since a unit of such complexity is usually taken not the simplest operations such as additions and comparisons, but cycles of procedures [22], features of the hardware implementation [1], the sizes of lists solutions and other strange parameters [1, 15, 16, 26, 35].

Additional difficulties arise for these methods when applied to convolutional codes. Structural problems of codes of this class can be noted [24]. For a long time specialists in decoders of turbo and LDPC codes preferred not to notice the contradiction that, seemingly, assuming the goal of these algorithms to successfully decode near the bandwidth of the channels, for example, with additive white Gaussian noise (AWGN), any significant results that, for example, the sequence of solutions of these iterative methods near the Shannon boundary approaches the solution of the optimal decoder. These algorithms do not measure the distance of their decisions to the received message, like decoders with direct metric control (DPCM) [5, 14, 19, 37], which in many cases makes it difficult for them to achieve the best possible results at sufficiently high noise levels, even if we assume further significant complication of such decoders. But we should not expect from an algorithm that does not measure the distance of its decisions to the accepted one, that it will generally “see” the achievement of the required result.

Another unexpected problem arises in comparing these problematic algorithms, which is simply related to the fact that decoding refusals are allowed in some works in these areas. How to relate to the results of such decoding when it turns out that blocks can simply "disappear"? Allow over-demand? But this is a fundamentally different formulation of the problem. To a large extent, the presence of this and other similar problems with failures in the "old" coding theory, which manifested itself even in sequential decoding algorithms, makes the adherents of the "classics" come up with all sorts of frank "tricks" to justify the existence of their "methods", the results of which are even incomprehensible how to interpret. We have to remember that decoding by lists (due to the extremely low efficiency of the original algorithms) within the framework of both “old” and “new” methods also completely destroys the original formulation of the decoding problem, since it is not clear, what the communication system should do, having received 10 or 1000 messages to choose from, if it is configured for the uniqueness of the message used in the future, in which very rare errors are still allowed. All the situations considered above, extremely uncertain for real communication systems, clearly illustrate the system-wide crisis of the "classical" coding theory, which initially could not solve all real problems of highly reliable digital data transmission and many system problems of communication networks, especially at high noise levels.

Note, that in theoretical and applied research within the framework of the Optimization Theory (OT), the above-mentioned crisis situations in the algorithms are generally not allowed. In OT there are only correct and erroneous decisions of decoders in some symbols. Of course, in OT it is always possible to clarify the statement of the problem, and in the case of low reliability of certain symbols of the decoder solution, they can occasionally be declared erased, which will significantly increase the reliability of the remaining data bits. But for a communications system, this is a much more natural situation than using lists. And in this case, there is no loss of large fragments of transmitted information. However, when analyzing and publishing the characteristics of multi-threshold decoders (MTD) and various versions of Viterbi (VA) algorithm in OT, they practically never use even this admissibility of rare erasures of the least reliable symbols, showing the natural capabilities of our decoders in the traditional original format.

But the situation in coding theory in the second half of the last century is actually even more dramatic. The fact is that the problems of decoding non-binary codes have been in an even more critical state for 60 years, since more technological algorithms for their decoding than for the case of Reed-Solomon (RS) codes have not appeared since then. Turbo and LDPC codes here also did not show particularly significant success in decoding results (see [20] and the links given there).

We note so far only briefly that polar codes (PCs) have also suffered an absolute fiasco in the implementation of the decoding procedure for non-binary codes. This is evidenced by the well-known results, where PC decoding algorithms based on PC codes have a complexity of the order of N ~ n2, and in some cases N ~ Ln3log (n), where n is the length of the code, moreover, operations are performed in the field of real numbers, and the size of the decision list L should grow exponentially for sufficiently efficient decoding [22]. It is despite the fact, that since 1984 it has been known about the simplest, with linear complexity, i.e. with N ~ n, symbolic (non-binary!) MTD decoders [5, 9, 11, 20], which, moreover, implement the best optimal decoding (OD!), equivalent to enumeration, at very high noise levels.

It is a pity, but, as in the binary case, the authors of a few "new" methods for RS codes did not present any really complete data on the efficiency and complexity of their decoding methods. All published "theoretical" monographs on fast RS codes decoding and some other structures, then formalized as doctoral "achievements", remained, like decades ago, at the complexity levels of the N ~ n2 order or, at best, N ~ nlog (n) 2) [26, 27], which practically does not change anything at all on the essence of the issue. In other words, there have been no new results in the field of non-binary codes for 60 years, and the linear complexity of OT algorithms, even with a high channel noise level, has not been noticed at all for almost 40 years.

Here again it is useful to remind, that our scientific OT school has been insisting for many years, that a full-fledged publication on applied coding issues should always contain a complete testable program for modeling a new algorithm. But over the past decade, no scientific groups have presented a single real efficient algorithm with a detailed description and a full-fledged simulation program that could be fully verified by the standard complex criterion "reliability-noise immunity-complexity", although some authors declare the use of C ++ in modeling, not giving no information about the speed of the implemented algorithms (see, for example, [22]).

In this regard, we are forced to note again, that in OT all, including non-binary (symbolic!) codes with MTD decoding, are processed with the theoretically minimal possible complexity N ~ n. Moreover, MTD decoders of all types work only in fixed-point arithmetic (i.e. with integers!) reach OD solutions at noise levels very close to the Shannon boundary [4, 5, 6, 9, 10, 37, 39].

Let us emphasize that the stubborn ignorance of the OT results by all Russian "theoreticians" in fact only nullifies their own "scientific" fabrications over the past 30 years. Note, that this real tragic crisis has developed for the simplest methodological reason that the error probabilities of any decoding algorithms can nowhere and never be obtained analytically for areas of channel noise near their capacity. But in our country many "ministers of science" have been unable to grasp this indisputable fact for a long time. We will not consider even more complex approaches to modeling, in particular, to hardware prototyping (which we are also actively and very successfully engaged in [5, 37, 39]).

The problem of the relationship between theory and experiment arose in science about 40 years ago, and it is so acute and dramatic that some publications on this topic were even presented on the RAS website [21]. A rigorous and comprehensive assessment of this long stagnant extremely conflicting system-organizational situation in the field of coding theory over the past 40 years is also beyond the scope of this article.

Problems of polar codes as an extreme manifestation of the "old" theory crisis.

Polar codes (PC) [29], the decade of their appearance has passed almost unnoticed, are a sad illustration of the continuation of the classical coding theory deep crisis, which has been going on for more than 30 years. That is why, with an extreme shortage of positive ideas for the development of coding theory and, apparently, the absence of any knowledge at all about the simple Viterbi block algorithm (BVA) [5, 13, 14, 37], as well as with a complete lack of understanding of the OT ideas depth, according to which every change in the decoded symbol in all types of MTD decoders corresponds to the transition to a strictly more plausible solution, the world community of code specialists began to analyze and develop exactly the PC. For the umpteenth time it seemed to the "theoreticians", that these codes can be analyzed and developed analytically, without carrying out any large-scale experimental work on computer modeling. And this deepened the crisis of coding theory further.

However, it quickly became clear about PCs that they did not contribute to the progress of applied coding theory, i.e. development of simple and, most importantly (!), high-tech decoders near the Shannon border. Algorithms for PCs immediately had to work in the realm of real numbers, and the guaranteed decoding error probability decreased with the code length n extremely slowly, only as ~ n – 0.25 [29], about which the adherents of this method preferred not to dwell at all. But this, for example, means that even for a typical small error probability Pw (e) ~ 10–5 in a block for a DSC-type channel, the PC block size should be n ~ 1020! This is certainly not encouraging in any way. And it is obvious that even for this reason alone, it was already initially an extremely unsatisfactory idea. From such despair, the apologists of this direction pointed, as we quoted them in [8], to the very limited capabilities of the PC, especially at small code lengths n. We have to add to this that the accuracy of quantization of real numbers will probably remain a very problematic issue of PC forever [26, 34]. This, as usually is indicated, is associated with the problem of the inadequacy of conventional floating point arithmetic to accurately account for the dynamic range of the data used. And in general, as the authors of [31] admitted, for example, in addition to decoding algorithms, PCs themselves could be better. In particular, it was shown [32] that polar codes are significantly inferior to turbo codes in terms of energy gain for all possible code rates and channel types at the same lengths. Finite-length polar codes are inferior in efficiency and LDPC codes [28, 35].

As a result, initially weak PC decoding procedures began to be converted into list decoders [33, 35], cascading methods and other measures were applied to them, which somehow were not very connected, more precisely, it turned out to be absolutely unrelated to the original PC ideas. Although the decoding characteristics were somewhat improved in comparison with the algorithm declared in [29] [8], as it should be according to the foundations of coding theory and the possibilities of cascading.

Finally, we point out again, that in any publications we reviewed, we have never found any data on the real complexity of PC decoders. We mean that, as has been repeatedly noted in [5, 37], the only really useful way to determine the complexity of specific decoding algorithms, which is minimally prone to errors, distortions and outright deception, is to present the editorial boards of journals, reviewers and to opponents on theses of a working tested simulation program in a Gaussian or in another channel, indicating the type of processor, its clock frequency and, most importantly, the number of decoded symbols per second when implementing algorithms, for example, in C ++. We remind you, that in order to facilitate the task of correct comparison of decoding algorithms, on www.mtdbest.ru and www.mtdbest.iki.rssi.ru we have proposed means for comparing decoder speeds, highlighting for this program for calibrating decoding speeds, also in C ++, for the case using different computers. The methods are simple and widely available.

Based on the described situation, we only point out, that supporters of PC methods should not continue to test the patience of specialists in decoding algorithms for the next 10 years, since, apparently, the period of excessive enthusiasm for PC decoders is already passing. After all, there are still no specific and, most importantly, reliable data on the real volume of calculations for at least one PC decoder.

The ratio of OT and PC capabilities

As indicated in [5, 6, 9, 37, 39], nowadays on the basis of the Optimization Theory (OT) of noiseproof coding for all classical channels with independent distortions, a number of new decoders of the MTD class and new modifications of VA have been created, including BVA [5, 13, 14, 37]. Together with the best already known approaches based on convolutional AV, concatenated or divergent schemes and other methods, they ensure successful decoding with the complexity of MTD decoders linear in the code length in the immediate vicinity of these channels capacity. The results obtained allow us to assert, that OT has created algorithms and specific technologies for the development of decoders, that correspond to the solution of the great Shannon problem, formulated by him more than 70 years ago [5, 12, 14, 20, 37, 38, 39]. Thus, the new OT for all classical channels has turned into an extensive set of technologies and paradigms, within which it is now possible to create very different decoders that are extremely simple to implement, operating near the bandwidth of communication channels and providing, with the correct choice of codes, OT solutions achievement.

The PC capabilities correspond to a fundamentally different formulation of the problem of noise-proof coding. Due to the weak capabilities of PCs and their decoding algorithms, as well as the inefficiency of long codes, specialists focused their attention mostly on PCs for relatively short codes, which immediately ruled out the discussion of the proximity of their area of operation to the channel capacity. Permanent references of the PC adherents to the fact that these codes "got" to the Shannon border only testify to the lack of understanding by such authors of the absolute futility of this whole venture. We have already noted this above. In addition, as it is often the case in the scientific community, the "theorists" of the PC refused to discuss the complexity of these algorithms at all and again (erroneously again!) decided, that almost everything could be calculated for the PC, which means, as before, not to study at all full-fledged modeling of algorithms. Like attempts to solve problems with turbo and LDPC codes, this did not lead to the correct realistic attitude to these codes in 10 years of PC “promotion”. Of course, theorists have not received any really useful for technology relationships between the capabilities of OD and PC. More than 40-year OT history has long deserved attention to its methods, which the supporters of the PC did not use in vain. As far as we understand, like half a century ago, when it was not possible to create such vast spaces of discrete fields in digital spaces, that they completely cover all admissible noise configurations (which immediately limited the possibilities of algebraic codes), the same thing happened with a PC at a qualitative level. But for them it happened on the basis of the difficulties of accurate calculations with real numbers and other reasons indicated above, which in itself led to a lot of all PC problems kinds. Thus, research in the field of PC does not contain important results, at least in part close to the OT level, such as, for example, the

Basic Theorem of Multi-Threshold Decoding (BTMTD). And again, the complexity of MTD decoders is linear in n. And the complexity of the PC is still N ~ nln (n), and there are mainly operations with real numbers in large computation cycles [22]. And all this, of course, is quite difficult in terms of implementation and extremely vague for any generally reasonable estimates of complexity.

Further, taking into account the above, consider the capabilities of OT and PC for short code lengths, i.e. in the field, which OT methods have not been specially adapted to work. From this class of algorithms, we chose only ordinary schemes and codes with already known good characteristics. Let us also emphasize that OT contains as the main algorithms all modifications of MTD decoders and various versions of VA, including BAV patented by our school [5, 13, 14]. Of course, with the appearance of at least preliminary specific and understandable data on the real efficiency and complexity of PCs, we will be able to publish more accurate results of comparing OT and PC methods.

Comparative characteristics of PC and VA for short codes

Since there is no data on the real complexity of PC (alas, we will repeat this), we will consider the few available information about the reliability of PC algorithms and already known or updated results for OT methods. Let's start with comparing PC and BVA. We will consider the simplest schemes of biologically active substances, and, so far, without any connection with PC methods and without adaptation to them.

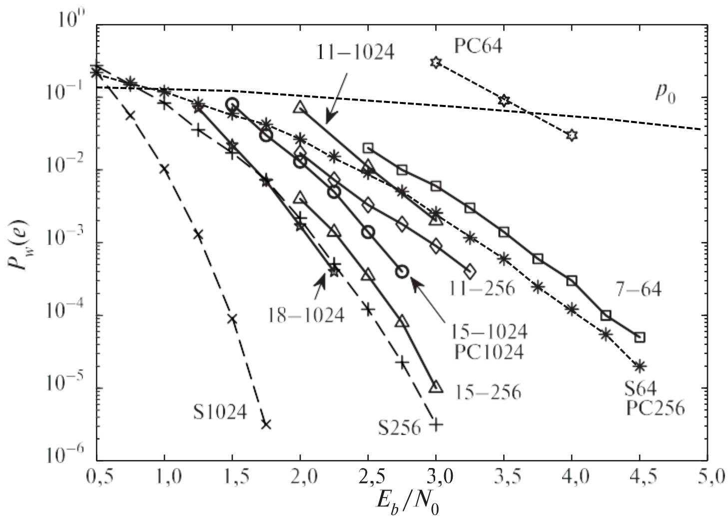

Fig. 1. Spherical packing boundaries and characteristics of BVA for codes with R = 1/2 in a channel with AWGN

In fig. 1 shows the spherical packing boundaries (dashed lines) for block codes in a Gaussian channel with 4-bit quantization at a code rate R = 1/2 in the Sn format, where n is a block length. The rest of the solid graphs of the form K-n correspond to the probabilities of the decoding error of the block Pw (e) of the quasi-cyclic block code using BVA for the generating code polynomial of the convolutional code of length K and length of the block code n. The short line PC64 shows, for polar codes of length n = 64, the lower bounds in theses [22, 26] for decoders of several PCs with conventional decoding, i.e., when the size of the decision list is L = 1. The rest of the results (about 10) in the same figures from these theses correspond to different sizes of the decision list up to L = 256. All of them are close to the graph (7-64) for biologically active substances or significantly higher than it, since the S64 boundary is known to be unattainable. We should remind that there is no data on the real complexity of such PC decoders, for example, on the number of operations or on the decoding speed.

Now let's clarify the data on graph 7-64. The program for BVA in C ++ on a 3 GHz processor runs at a speed of ~ 95 Kbit / s. Block VA has practically the same complexity as convolutional, which for K = 7 was developed 50 years ago [7, 8, 9, 13]. Even if a program for a PC is written at least for n = 64, then, taking into account all the unsolved problems of these codes, with a very successful operation of BVA, it will still not be possible to consider that this direction of PC is a new step in coding theory and technique.

Consider codes of length n = 256. The same fig. 1 shows the results for codes with BVA decoding for different K = 11 and K = 15. And the lower bound Pw (e) for a number of different PC decoders types, including concatenated ones presented in [34], coincided with the boundary S64, which is therefore marked by the signature PC256. As can be seen from the comparison of codes of length n = 256, even a BVA with a generating convolutional polynomial of length K = 11 has better characteristics than a PC. And with biologically active substances at K = 15, none of the PCs can be compared at all: extremely weak results.

Why was BVA with K = 15 taken for comparison? Here we can recall, that the elemental base of electronics over the past decades has become faster and faster, as well as the technology itself. And then it is not surprising at all, that even in that millennium NASA implemented VA with K = 15 and launched it on the Cassini spacecraft to Saturn. The project finished with triumph in 2017. And the codes did not fail! This means, that K = 15 is an absolutely real parameter. The speed of operation of biologically active substances with K = 15 under the same conditions is close to 1000 bit / s. By the way, different simple versions of VA have been developed for OT up to lengths K ~ 24. So we have something to discuss if necessary in other cases.

We would like to pay readers attention to the most important moment, that further there will be many characteristics of biologically active substances with K = 15. It is 16000 times easier to implement than OD based on VA in [18]; in the publication that is proposed to be used for teaching students. Also in this manual, a discussion of the features of the methods is highlighted, which even for not very long codes give the complexity of decoding N> 10200, which exceeds the number of atoms in the Universe. We are agree, that this is also very impressive.

Finally, let's turn to codes of length n = 1024. 2 BVA with K = 15 and K = 18 are presented here explicitly. We would add, for greater completeness of the estimates, it is useful to evaluate the capabilities of the code with K = 11. It is represented by the line (11-1024). Consider data on PC of length n = 1024. The lower bound for a group of such codes from [34], using various "improvements", including cascading, marked as PC1024, practically coincides with the curve (15-1024). And then from the graphs for codes with n = 1024 it follows that BAS at K = 11 is slightly weaker than PC, at K = 15 the reliability is close, but at K = 18 BVA is much better. Is such a biologically active substance difficult? It is only ~ 8 times slower than with K = 15 and decodes a digital stream using the same software at a bit over 100 bps. But let us recall, that VA with K = 15 was created in that millennium. So you shouldn't reject a code with K = 18. In addition, the hardware implementation of VA has long been easily fully parallelized and, in general, VA is very technological. So, even here, at least any advantages of PC are very difficult to see, especially when you consider all their problems, both listed and not mentioned here. And in the software implementation it is very convenient to use any specialized processors [36]. Here VA, that work only with integers have even greater advantages over PCs, which are hostage to unstable computations with real numbers and a host of other problems.

At the end of the discussion of the graphs in Fig. 1, we can also indicate, that the capabilities of a PC of length n = 1024 even at L = 16 reach a level not better than the curve (15-1024) shows, and the lower estimate for a PC with n = 2048 in [31] actually corresponds to curve S256. This means that PCs do not begin to acquire at all any special advantages with an increase in the length of codes n or even taking into account the possibilities of decoding by a list, which our OT school (we emphasize this main point again!) completely rejects the comparison of results with a completely different formulation of the problem as illegal decoding. And here no specific data on the real provable complexity of decoding specific PCs could be found.

Note that in Fig. 1, the characteristics and boundaries were presented, including cascade PC options. And the whole set of BVA methods were presented only by their simplest basic versions. It is obvious, that the involvement of concatenated structures for VA’s in the comparison of decoders, as it is always the case with the correct design and analysis of concatenated circuits, will certainly greatly improve the efficiency and at the same time reduce the complexity of decoders created according to technological OT paradigms.

We leave the possibility of confirming the simple and absolutely guaranteed results stated here to all our colleagues, who in this case will be provided with our most diverse technological and ideological support. The successful completion of such an important work on a deep technological update of the applied coding theory by other research teams may create a new high-tech platform for them to deploy broad research activities in all new promising areas of OT.

PC and MTD algorithms capabilities comparison

Let us compare briefly the characteristics of PC and MTD algorithms, which are basic in OT theory, subject to the choice of small lengths of the used codes. Further, the codes and MTD algorithms are also considered without any adaptation to the conditions of their comparison with the PC. Figure 2 shows the error probabilities per codeword Pw (e) in the same format for the PC codes in the same Gaussian channels. According to [23], the Arik2048 curve corresponds to the original Arikan method [29] for a PC of length n = 2048. MS256 estimates from below the capabilities of several PCs of length n = 256, some of which have a cascade structure [34], MS1024 is a lower estimate for a significant number of PCs, including the list algorithms for decoding them [22, 34], which, as we always emphasize, should not be compared with traditional methods, including OT algorithms. But we even took them into account! Finally, for a group of PC codes of length n = 2048, the graphs of the lower bounds G2048 from [31] and DS2048 [23] are presented with reference to [30].

Let's compare MTD algorithms for short codes in the same designations OT256, OT1024 and OT2048 with the PC. In all these cases, codes with R = 1/2 were applied without using cascading or other efficiency measures. The number of decoding iterations for all MTDs is not more than I = 80. Since these algorithms are extremely simple, even for MTD for a code of length n = 2048, the decoding rate on the same software exceeds 140 Kbit / s. The reliability of MTD results either coincides with PCcapabilities or is close to them. And this is very important because, for comparison, the simplest MTD schemes were taken, while PC decoders admit the use of lists, and the complexity of the decoders themselves for PCs is completely undefined, although in [22] an exponential increase in the size of lists (L >> 1000), the need for calculations with real numbers and in some cases a large number of iterations (I> 104). But the results for MTD are very close even to the PC capabilities in the listed variants. It follows, that the use of OT and its proposed decoders with direct control of the metric, among which MTD and various VA’s, generally solve all decoding problems even with restrictions on the lengths of codes, and on the basis of well-studied and long-studied methods that OT offers. It is also significant, that the results were obtained without any adaptation to the new matching conditions for small code lengths, which, of course, would add many times more efficiency to the OT methods. It is also extremely important that all decisions of MTD decoders taken for comparison, as in the case of choosing long codes, converge to optimal solutions for any small length. This explains their advantage over all other algorithms.

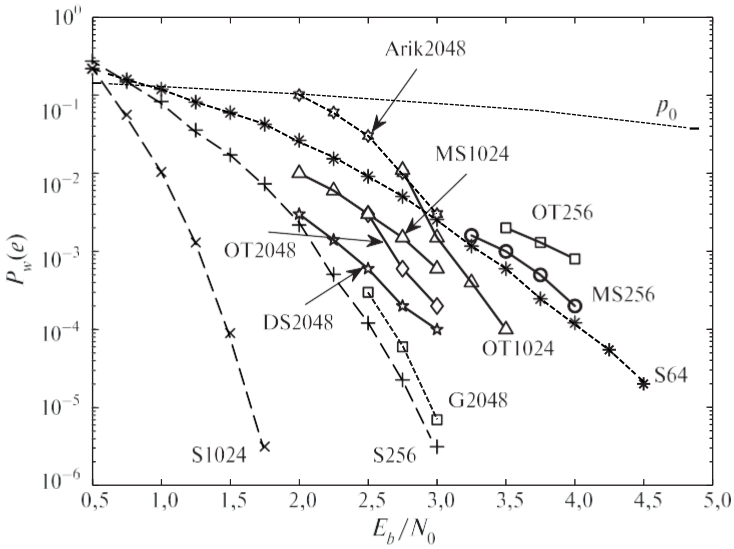

Fig. 2. MTD and PC characteristics for codes $R = 1/2$ in channels AWGN

More efficient algorithms of OT class, for example, cascade ones, can be analyzed if codes and algorithms with a specific presented and convincingly proven decoding complexity are identified among the PCs. However, this situation can not be expected, since the PC adepts have not bothered to present any such decoder over the past 10 years, and the efficiency of OT methods is constantly growing, moreover, often even with some simplification of their algorithms.

On methods of OT algorithm's characteristics improving

As shown above, PC decoding technology has very little to do with the basic ideas of error-correcting coding and, moreover, with optimal decoding (OD) or powerful OT methods, which provide fast convergence to OD, linear with the block length. That is why the idea of a PC is not rescued even by always powerful cascade circuits and list methods with a large size of such lists L that are absolutely contradicting the original formulation of the problem of noise-proof coding. Additional adaptation showed undoubtedly very decent high performance.

Nevertheless, for readers who are ready to agree that PC is also a dead-end branch in applied research of coding theory, we will indicate some options for the development of research in the OT field, which will allow, we are absolutely sure, to further significantly improve the characteristics of OT and decoding of short codes. Moreover, as follows from the data already presented above, it is really necessary to adapt the OT methods for codes of length n = 1024..2048, since the problem has already been successfully solved for shorter codes. And here it is immediately necessary to clearly indicate that in this formulation of the problem, the problem of comparing OT with a PC is not posed at all. As we saw, the PC has no results at all, which means that there is simply no reason for such a discussion. In other words, methods for increasing the efficiency of OT, VA and MTD are proposed below, regardless of any other methods, simply because at present, technologies comparable to OT simply do not exist at all.

Consider for this in Fig. 3 preliminary data for codes of length n = 1024 with R = 1/2 in error probability per bit Pb (e) for the case of BVA decoding. In addition to the spherical packing boundary S1024 in the form of Pw (e) and the dotted line Pb (e) for a convolutional code with a generator polynomial of length K = 18, all the other graphs Pb (e) correspond to generator polynomials of length K = 15, which, as we discussed above, for a very long time technologically quite affordable. Basic plots of the form Sk and Nk, where k = 0 and k = 8, correspond to systematic and non-systematic ordinary codes at k = 0 and punched codes at k = 8, i.e. with the exclusion of every 8th code symbol of the second code branch from the source code with k = 0.

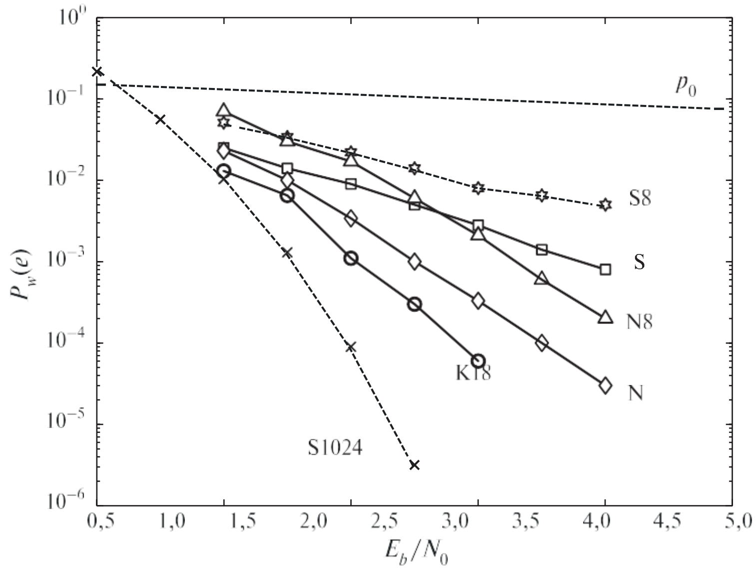

Figure: 3. Preliminary characteristics of BVA for codes with R = 1/2 in a channel with AWGN

As follows from Fig. 3, the reliability of the punched codes is very high. This guarantees the efficiency of punched code cascading, which should be successful in both traditional sequential circuits and parallel circuits where OT has many useful results. We advise specialists who are really interested in the development of decoding technologies to consider the real possibilities of the proposed approaches, which can provide a significant increase in the efficiency of error correction algorithms. School OT guarantees support for these works, consulting and software support.

It is useful to pay attention to the code with K = 18. We believe that a careful analysis of BVA will reduce the complexity of its decoding by looking at a smaller number of paths or based on other methods, for example, using the principle of divergence. Finally, it is obvious that even without the use of punched codes in the BVA, it is possible to consider concatenation of codes with k = 0 with external high-speed codes in both variants of possible concatenation. These guaranteed positive results, which can be obtained very quickly, will also be very useful, at lower final speed than R = 1/2, but at a significantly higher final reliability, which usually happens when cascading with such methods.

Once again, we emphasize the need for a clear understanding that the error probabilities of any decoding algorithms at a high noise level can never be estimated analytically. Coding theory, as OT has absolutely no alternative proved, is not a mathematical problem [5, 14, 19]. All the final characteristics of the efficiency and complexity of error correction algorithms for a high level of channel noise can always be very quickly obtained exclusively by full-scale modeling of decoding algorithms based on various optimization techniques [5, 37, 39]. OT technologies and the results of their application fully and comprehensively confirm this most important severe circumstance, so long unrecognized by some of the individuals who exist in the vicinity of code science.

Notes on the other algorithms "development"

Let us briefly note the main theoretical results of the current millennium, which are published by some supporters of LDPC codes, where the extremely modest level of the discussed methods is visible, simply because these authors also avoid any work on modeling and reasonable estimates of the decoding complexity.

In [15], a method of “simple” decoding of an LDPC code is proposed with a strangely defined decoding complexity through a link to another publication, nevertheless equal to N ~ nln (n). But simple methods have long been known to have complexity N ~ n [4, 5, 6, 9, 14, 37, 39]. The estimates of Pw (e) in [15] are made through the use of more than 50 complex completely unverifiable cross-referenced relations, which is a typical extremely problematic case described in [21]. The result of the article is a result in a style that has become popular in a certain community, but completely incomprehensible from the point of view of research on methods to improve decoding performance. The authors of [15] determined the relative proportions of successfully decoded block code symbols. In this work the fraction for the code rate R = 1/2 is ~ 0.0003. But it can be noted that even 60 years ago, very weak BCH codes, as indicated for reference in [25, p. 308, fig. 9.1], almost always corrected the share of errors of the order of 0.03, which is 2 decimal orders better than the result of the article. Moreover, MTD experimentally demonstrated as early as 1981 the possibility of correcting almost all errors with the error probability in the channel (this is here: also the equivalent of the share of corrected errors) p0 ~ 0.05, moreover, with the linear complexity of MTD algorithm, i.e. when N ~ n [9, 12]

In a recent work [16] on 27 pages, methods for estimating the share of correctable errors, which are unrealistic to be verified, are also proposed. The value of such a share at R = 1/2 is less than 0.002, which is also much weaker, 15 and 25 times, respectively, than the BCH and MTD decoders that have been known for 40 years provide.

In [1, 2, 17] analytical estimates of the fraction of corrected erasures are considered for LDPC and other codes. At the same time, the authors again interpret the complexity issues in a peculiar way and even propose not to pay attention to them. The upper bounds for the fractions of corrected erasures for these methods at R = 1/2 are everywhere less than 0.1. All these authors did not present any experimental data. At the same time, in [4, 5, 6, 37, 39], the fraction of erasures that can be corrected in the simplest way is significantly greater than 0.3, moreover, with linear complexity.

We have to remind again that in [4, 5, 6, 37, 39] and in a number of other publications, very simple estimates of OD in erasure channels are given, which determine lower bounds for the probability of non-recovery of erasure in some symbol of the code Pb, er (s) ~ per d , where per is the probability of an erased symbol appearing in the channel, d is the minimum code distance of the used code. The conclusion of the assessment took half a page of extremely simple reasoning. In [4, 5, 6, 37, 39] it is also shown that the proposed estimate (corresponding to OD!) is achieved using simplified MTD with linear complexity even near the capacity of the erasure channel. So all the possible conclusions from this comparison of codes and decoders are obvious. In the process of such decoding, at each iteration, one simplest equation is solved for three small integers with unknown X: A = B + X. It is not surprising, therefore, that OT has no competitors at all for channels with erasures. Algorithms for MTD in such channels have the minimum possible complexity, the best possible reliability and tremendous speed [4, 5, 6, 37, 39].

Finally, we note that it is very difficult to obtain reasonable estimates of the complexity for all decoding methods of non-binary codes, except MTD, which was noted in [20]. There are also a number of references to non-OT algorithms for non-binary codes. In the absolute majority of cases, where an approximate comparison of MTD complexity and other methods was generally somehow possible, the advantage of MTD in decoding speed was in thousands and more times for a simple reason of optimal decoding with a linear complexity of the code length.

Thus, the theory of OT is significantly ahead of the results for absolutely all other known classes of decoding algorithms due to the correct ratio of theoretical and experimental research methods. A number of OT results, quite possibly, will remain unattainable for other decoding algorithms for a long time.

About OT perfection

Finally, consider the resulting Figure 4 illustrating the main OT capabilities. As can be seen from this figure, OT for codes selected in accordance with some typical technical task easily constructs the probability of the initial decoder's initial error, usually called the error probability in the first symbol $P_1(e)$ [4, 5, 9] and equally simple formulas for the maximum possible probabilities of MTD decoding error at the level of the optimal solution $P_{opt}(e)$in absolutely all channels considered in the coding theory. And all the other worries of the OT specialist are simply to select the best OT codes and technologies using optimization methods that can provide the required level of $P_{opt}(e)$In this case, one should not forget that the required characteristics should always be sought at the code rate of the designed decoder R, which is between the bandwidth of the C channel and the computational rate of the R1 channel. The main variable parameter in OT is the number of decoding iterations I = 5..50 or more. That is why the whole OT theory is extremely compact and is, in fact, a systemic-philosophical treatise on solving optimization problems focused on the correcting digital data problems. And all of this is done to a large extent by means of highly intelligent software. It can be considered that OT is less in meaningful volume than the "classic", by about 3 decimal orders, which especially emphasizes OT perfection as a complete applied theory that develops decoding algorithms, comparable to which among other methods of the "classic" or "new" directions so far just not at all. And it is very difficult to expect the emergence of new and more preferable methods, since OT offers methods with OT characteristics with the lowest possible complexity linear in the length of the codes, which are workable even in the immediate vicinity of the Shannon boundary. In other words, it is difficult to come up with a completely different method that would be better than the OT algorithms, which already provide the best possible characteristics in all the main parameters "noise immunity-reliability-complexity".

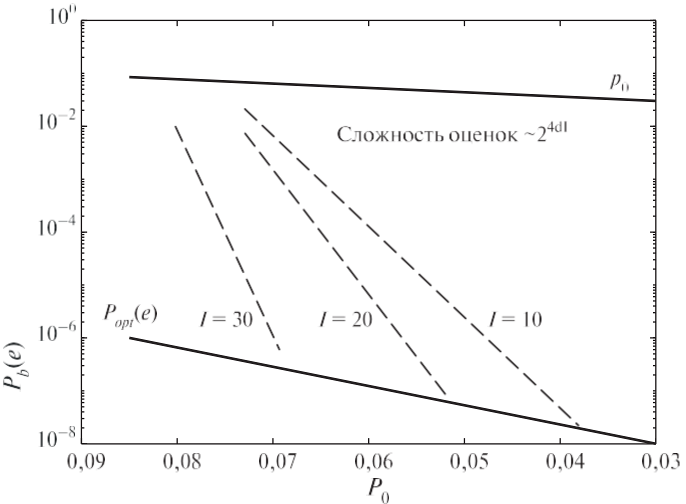

Figure: 4. Variants of MTD aspirations to the solution of the optimal decoder

Referring to Fig. 4, we emphasize that the complexity of the analytical assessment of the MTD error probability, for example, at the 9th or 38th iteration, as it seems to us so far and as shown in the line "Estimation complexity", is extremely high and exponentially depends on the decoder parameters indicated on the slide: the product of the code distance and the number of iterations. It means that it will probably never be done. At the same time, experimental results for various numbers of decoding iterations can be obtained on software or hardware models within an extremely limited period of time from several seconds to tens of hours, since the complexity of all MTD algorithms is extremely small, linear in the code length. This is how all the issues of technologies and developments in OT based on MTD are solved. It is obvious that all modifications of VA, as we have shown in this review, are modeled just as quickly and no less convincingly.

Thus, the current absolute preference for OT over other approaches to the problems of applied coding theory, which have not given any really new results over the past decades, is comprehensively and certainly justified.

Conclusion

Reasonably balanced development of OT as a new "quantum mechanics" in information theory together with large-scale, for many decades, software development of MTD code design systems and their optimization allowed solving the great Shannon problem and creating unique technologies for the development of noise-proof coding methods, which, apparently, for a long time, they will not even be able to partially repeat it anywhere precisely because of the lack of any reasonable understanding of the correct relationship and interaction between theoretical and experimental research methods [19, 21]. OT is the basis for the development of new communication satellites and remote sensing projects with performance and reliability parameters that are fundamentally inaccessible for any other coding methods.

Although this article was formally more oriented towards the discussion of the problems of polar codes, it was quite natural that the most general overview of the extremely dramatic, if not to say more, long-term absolutely critical state of error-correcting coding applied theory was obtained. It showed real and very significant advantages of OT not only in terms of reaching the Shannon bound [5, 14, 19, 37], but also in the case of considering the efficiency of short codes decoding. It is important that it did not require any preliminary adaptation of the OT methods to the formulation of the comparison problem, which was greatly changed in this version, and, moreover, does not stand up to criticism in a number of aspects of codes and decoders comparison.

The OT Scientific School invites all digital processing specialists to join our efforts to diversify OT technologies to all areas of digital informatics, increasing the reliability of discrete data to the greatest extent and in the simplest ways. Our support to all enthusiasts and specialists who would come into the new branch of coding theory is guaranteed with all our resources and knowledge.

This work was financially supported by the Russian Foundation for Basic Research (grant no. 18-47-620001), the Space Research Institute, and the Ryazan State Radio Engineering University. A large amount of additional information on multi-threshold decoders can be found on the websites of the IKI RAS www.mtdbest.iki.rssi.ru and RGRTU www.mtdbest.ru.

References

1. Afanas'ev V.B., Davydov A.A., Zigangirov D.K., Otsenka doli stiranii, ispravlyaemykh lineinymi kodami (Estimation for part of erasures recovered with linear codes), *Informatsionnye protsessy*, 2016, Vol. 16, No. 4, pp. 352–404.
2. Zigangirov D. K., Zigangirov K. Sh., Dekodirovanie nizkoplotnostnykh kodov s proverochnymi matritsami, sostavlennymi iz perestanovochnykh matrits, pri peredache po kanalu so stiraniyami (Decoding of low density codes with parity check matrixes consisting of permutation matrixes for erasure channels), *Problemy peredachi informatsii*, 2006, Vol. 42, pp. 44–52.

3. Zolotarev V. V., Mnogoporogovoe dekodirovanie v nedvoichnykh kanalakh (Multithreshold decoding over non-binary channels), *Voprosy radioelektroniki*, *Ser. EVT*, 1984, Vol. 12.
4. Zolotarev V.V., *Teoriya i algoritmy mnogoporogovogo dekodirovaniya* (Theory and algorithms for multithreshold decoding), Moscow: Goryachaya liniya — Telekom, 2006, 266 p.
5. Zolotarev V. V., *Teoriya kodirovaniya kak zadacha poiska global'nogo ekstremuma* (*Optimizatsionnaya Teoriya pomekhoustoichivogo kodirovaniya – novaya "kvantovaya mekhanika" teorii informatsii*) (Coding Theory as a Global Extremum Search Task (Optimization Theory of error-correcting coding — a new "quantum mechanics" of information theory)), Moscow: Goryachaya liniya — Telekom, 2018, 222 p.
6. Zolotarev V. V., Ovechkin G. V., *Pomekhoustoichivoe kodirovanie. Metody i algoritmy* (Error-correcting cod- ing. Methods and algorithms), Moscow: Goryachaya liniya — Telekom, 2004, 126

7. Zolotarev V. V., Ovechkin G. V., Effektivnoe mnogoporogovoe dekodirovanie nedvoichnykh kodov (Effective multithreshold decoding for non-binary codes), *Radiotekhnika i elektronika*, 2010, Vol. 55, No. 3, pp. 324–329.
8. Zolotarev V.V., Ovechkin G.V., Primenenie mnogoporogovykh metodov dekodirovaniya pomekhoustoichivykh kodov v vysokoskorostnykh sistemakh peredachi dannykh (Application of multuthreshold decoders for error-correcting codes in high-speed digital communications), *Elektrosvyaz'*, 2014, No. 12, pp. 10–14.
9. Zolotarev V. V., Ovechkin G. V., O sopostavlenii novykh metodov pomekhoustoichivogo kodirovaniya (About comparison of new error correction coding methods), *Tsifrovaya obrabotka signalov i ee primenenie, Proc. 18th Intern. Conf.*, Moscow, 2016, Vol. 1, pp. 59–65.
10. Zolotarev V.V., Ovechkin P.V., *Sposob kodirovaniya i dekodirovaniya blokovogo koda s ispol'zovaniem algoritma Viterbi* (A method for encoding and decoding of block code with using of Viterbi algorithm), Patent RU 2608872, Reg. 25.01.2017.
11. Zolotarev V. V., Zubarev Yu. B., Ovechkin G. V., *Mnogoporogovye dekodery i optimizatsionnaya teoriya kodirovaniya* (Multithreshold decoders and optimization coding theory), Moscow: Goryachaya liniya — Telekom, 2012, 238 p.
12. Zolotarev V.V., Zubarev Yu.B., Ovechkin G.V., Averin S.V., Ovechkin P.V., 25 let optimizatsionnoi teorii kodirovaniya: novye perspektivy (25 years of optimization coding theory: new perspectives), *Problemy peredachi i obrabotki informatsii v setyakh i sistemakh telekommunikatsii*, *Proc. 18th Intern. Scientific and Technological Conf.*, 2015, pp. 10–17.
13. Zolotarev V.V., Ovechkin G.V., Nazirov R.R., O peredache Optimizatsionnoi Teorii liderstva ot prikladnoi klassicheskoi teorii pomekhoustoichivogo kodirovaniya (Optimization Theory: the Reception of the Baton of Leadership from the Applied Classic Theory of Error-Correcting Coding), In: *Nekotorye aspekty sovremennykh problem mekhaniki i informatiki: sbornik nauchnykh statei*, Moscow: IKI RAN, 2018, pp. 82–90.
14. Zubarev Yu. B., Zolotarev V. V., Ovechkin G. V., Novye tekhnologii i paradigmy pomekhoustoichivogo kodirovaniya: posle resheniya problemy Shennona (New technologies and paradigms of error correction coding: after solving of Shannon problem), *Elektrosvyaz'*, 2019, No. 9, pp. 56–61, available at http://www.mtdbest.ru/articles/elsv2020.pdf.
15. Zyablov V.V., Rybin P.S., Ispravlenie stiranii kodami s maloi plotnost'yu proverok (Recovering erasures with low density codes), *Problemy peredachi informatsii*, 2009, Vol. 45, pp. 15–32.
16. Zyablov V. V., Rybin P. S., Analiz svyazi svoistv MPP-kodov i grafa Tannera (The analysis of connection between properties of LDPC codes and Tanner graph), *Problemy peredachi informatsii*, 2012, Vol. 48, pp. 3–29.
17. Zyablov V.V., Iokhannesson R., Lonchar M., Prosto dekodiruemye kody s maloi plotnost'yu proverok na osnove kodov Khemminga (Simple for decoding low density parity check codes based on Hamming codes), *Problemy peredachi informatsii*, 2009, Vol. 45, pp. 25–40.
18. Kudryashov B.D., *Osnovy teorii kodirovaniya* (Basics of error correction codes), Saint Petersburg: BKhV-Sankt-Peterburg, 2016, 393 p.
19. Kuznetsov N.A., Zolotarev V.V., Ovechkin G.V., Ovechkin P.V., Nedvoichnye mnogoporogovye dekodery i drugie metody korrektsii oshibok v simvol'noi informatsii (Non-binary multithreshold decoders and other methods for error correction in symbolical information), *Radiotekhnika*, 2010, No. 6, pp. 4–9.
20. Kuznetsov N. A., Zolotarev V. V., Zubarev Yu. B., Ovechkin G. V., Nazirov R. R., Averin S. V., *Problemy i otkrytiya Optimizatsionnoi Teorii pomekhoustoichivogo kodirovaniya* (Problems and discoveries of optimization error correction coding theory), Moscow: Goryachaya liniya — Telekom, 2020, 36 p., available at: http://www.mtdbest.ru/articles/comics.pdf.
21. Magarshak Yu., Chislo, vozvedennoe v absolyut (Elevating number to an absolute), *Nezavisimaya gazeta*, 09.09.2009, abailable at: https://www.ng.ru/science/2009-09-09/11_maths.html.
22. Miloslavskaya V. D., *Metody postroeniya i dekodirovaniya polyarnykh kodov: Diss. kand. tekhn. nauk* (Methods for construction and decoding of polar codes, Cand. techn. sci. thesis), Saint Petersburg, 2015, 206 p.
23. Morozov R. A., Dekodirovanie polyarnykh kodov s pomoshch'yu algoritma Dumera – Shabunova (Decoding polar codes using the Dumer – Shabunov algorithm), *Spisok-2013: materialy vserossiiskoi nauchnoi konferentsii po problemam informatiki*, Saint Petersburg: Izd. VVM, 2013, pp. 261–262, available at spisok.math.spbu.ru/2013/txt/papers/s7_1.odt.
24. Nazarov L.E., Sheglov M.A., Kharakteristiki polnykh i ukorochennykh pomekhoustoichivykh nizkoplotnostnykh kodov na osnove konechnykh geometrii (Characteristics of length-compatible low-density parity-check codes on finite geometries), *Uspekhi sovremennoi radioelektroniki*, 2017, No. 6, pp. 38–48.
25. Piterson U., Ueldon E., *Kody, ispravlyayushchie oshibki* (Codes correcting errors), Moscow: Mir, 1976, 594 p.
26. Trifonov P.V., *Metody postroeniya i dekodirovaniya mnogochlennykh kodov*: *Diss. dokt. tekhn. nauk* (Methods for construction and decoding of multi-partition codes, Dr. techn. sci. thesis), Saint Petersburg, 2018, 254 p.
27. Fedorenko S. V., *Metody bystrogo dekodirovaniya lineinykh blokovykh kodov* (Methods of fast decoding for linear block codes), Saint Petersburg, 2008, 198 p.

28. Ammar H., *Optimisation and analysis of polar codes in communication systems: Doctoral Thesis*, University of Manchester, 2018, 177 p.
29. Arikan E., Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels, *IEEE Trans. Information Theory*, 2009, Vol. 55, No. 7, pp. 3051–3073.
30. Dumer I., Shabunov K., Soft-Decision Decoding of Reed-Muller Codes: Recursive Lists, *IEEE Trans. Information Theory*, 2006, Vol. 52, No. 3, pp. 1260–1266.
31. Giard P., Sarkis G., Thibeault C., Gross W. J., Fast Software Polar Decoders, *Proc. IEEE Intern. Conf. Acoustics*, *Speech and Signal Processing* (*ICASSP*), 2014, 5 p.
32. Ozgur U., *A Performance Comparison of Polar Codes with Convolutional Turbo Codes: Doctoral Thesis*, 2009, 166 p.
33. Sarkis G., Giard P., Vardy A., Thibeault G., Gross W. J., Fast list decoders for polar codes, *IEEE J. Selected Areas in Communications*, 2016, Vol. 34, No. 2, pp. 318–328.
34. Seidl M., Huber J. B., An Efficient Length- and Rate-Preserving Concatenation of Polar and Repetition Codes, *Computer Science*, *Mathematics*, 2013.
35. Tal I., Vardy A., List decoding of polar codes, *IEEE Trans. Information Theory*, 2015, Vol. 61, pp. 2213–2226.
36. Zolotarev V. V., *Coding Theory as a Simple Optimal Decoding near Shannon's Bound* (*Optimization Theory of error-correcting coding – is a new "quantum mechanics" of information theory*), Moscow: Hot Line — Telecom, 2018, 333 p., available at https://mtdbest.ru/articles/mtd_book_2019.pdf.
37. Zolotarev V. V., Zubarev Yu. B., Ovechkin G. V., *Optimization Coding Theory and Multithreshold Algorithms*, Switzerland: ITU, 2015, 158 p., available at https://mtdbest.ru/articles/Zolotarev_ITU.pdf.
38. Zolotarev V., Ovechkin G., Ovechkin P., Satybaldina D., Tashatov N., Sankibayev D., High Throughput Software Multithreshold Decoder on GPU, *3rd Intern. Conf. Mathematics and Computers in Sciences and in Industry* (*MCSI*), 2016.
39. Zolotarev V. V., Ovechkin G. V., Chulkov I. V., Ovechkin P. V., Averin S. V., Satybaldina D. Zh., Kao V. T., Review of Achievements in the Optimization Coding Theory for Satellite Channels and Earth Remote Sensing Systems: 25 Years of Evolution, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kos- mosa*, 2017, Vol. 14, No. 1, pp. 9–24.