Scientific editor's foreword

The development of modern civilization is increasingly founded on the use of information technology. Sweeping digitization is making it possible to place the processes of data creation, transmission, storage and processing on a new footing, guaranteeing a high degree of integrity and availability of digital information.

However, the current phase of information technology development is also characterized by a very sharp and significant increase in demands in terms of the reliability of digital data. In this context, modern error-correction coding methods have a major role to play in ensuring a high level of reliability and quality in the transmission and storage of discrete information.

The development of error-correction coding theory and techniques has a long history going back more than fifty years, during which time approaches to resolving the key problem in this discipline − namely, achieving the simplest and at the same time most effective decoding − have changed repeatedly and very significantly.

The first major direction taken by coding theory was the development of error-correction algorithms based on finite-field algebra. Then came the Viterbi algorithm era. From the theoretical point of view this algorithm was the most complex, entailing exhaustive search, yet it was so effective that for a long time all advances in coding theory as well as the development of specific error-correction systems for satellite communications centred around this method. This despite the inability to use long codes for the Viterbi algorithm because decoding complexity increases exponentially with code length.

The next stage in the continuing search for effective coding techniques led to the successful implementation of concatenated codes in communication technology. Indeed, as was predicted by theory, it actually turned out in practice that concatenated codes were able to provide much better error-correction characteristics with less decoding complexity than the original non-concatenated methods. Even with the great progress made in coding theory as well as in microelectronic technology, however, up to the end of the 1980s the characteristics of coding and subsequent decoding systems still remained very far away from the theoretically possible limits.

Only with the emergence of turbo codes in 1993 did it become apparent to specialists that something approaching full utilization of digital communication channel capacity was already a very real technical aim. A host of turbo-type codes and some others proved that there is indeed a genuine possibility of much more efficient use of the capacity of space, satellite and many other digital communication channels than hitherto possible; and these channels are very expensive. Using certain codes of these types, the power of Gaussian channels can in principle be only a few tenths of a dB higher than determined by the basic theoretical inequality for code rate R and channel capacity C: $R<C$.

Subsequent achievements and prospects in coding theory are usually associated with low-density parity check (LDPC) codes. Their decoding complexity was a bit less than for turbo codes, but still too great. Nevertheless, the industry mastered the production of this type of decoder, too, paving the way for talk of further progress in coding techniques.

That having been said, all examples of the use of turbo and LDPC codes with acceptable parameters have shown yet again that the decoder characteristics for these codes are significantly worse under real conditions, owing to unavoidable simplification of these algorithms in comparison with the ideal conditions of their simulation in research laboratories. The dramatic decline in coding gain − a parameter directly characterizing the code's usefulness for systems such as satellite communications − can be as much as ~0.5 to 1.0 dB with real decoders. This indicates that we are very far from a final solution to the key challenge in coding, i.e. correction that is easier, faster and operates at higher noise levels. It should also be pointed out that decoder processing speeds for the codes discussed above, even when implemented with state-of-the-art microelectronics, are not as high as often required, which also greatly limits their application.

The benefit derived from the use of coding, for example in binary data transmission, is usually defined by the specific parameter "code gain" (CG), which simply characterizes the extent to which the transmitter power of a communication system is increased through the use of good coding techniques and, more importantly, effective subsequent decoding of the received digital stream. As its value can represent a factor of up to 3, 5 and even 10 or more (over 10 dB), one can understand the vital importance of coding, which confers such a huge power margin on a communication system. In 1980, the famous American expert and author of seminal works on coding theory, E. Berlekamp, asserted in one of his surveys that every saved decibel of communication channel power is worth around a million dollars. On that basis, given the current scale of digital networks, the economic effectiveness of coding has grown severalfold, thanks to the significant increase in digital data transmission speeds, the smaller sizes of – very expensive – antennas, the multiple increase in communication range and the many other very important advantages for digital communication systems stemming from the use of error-correction coding. This underscores the importance for the entire telecom industry of work to develop effective decoders; indeed, error-correction coding is invariably among the most topical issues at dozens of annual international conferences.

The current situation in regard to research and application of non-binary codes and the corresponding correction algorithms is even more complex and problematic. Some 50 years after the invention of Reed-Solomon (RS) codes, it appears that RS decoders remain the best decoders implemented for non-binary codes to date. Yet even the latest decoder technology for non-binary codes still cannot support the use of RS codes longer than 255 characters. Therefore, for new Blue-Ray discs and other complex systems for the protection and safe storage of symbol (byte) data, various combinations of short RS codes are still used, including concatenated schemes. This fails to enhance the reliability and security of stored data with high levels of distortions, while significantly increasing the required redundancy for such schemes and substantially reducing processing speed.

Thus, the actual status of coding theory, with the existing set of specific coding and, most importantly, decoding methods, is still such that it is far from fulfilling its potential. Even with the implementation of high-power computing hardware, the parameters of the various types of decoders that are available are still way below what is required in terms of ease of implementation at high reliability of the decoded digital stream in the case of a strong noise level and high transmission speed.

A new, very effective solution to the problem of minimizing decoding complexity while at the same time achieving high performance characteristics of coding systems, based on *multithreshold decoding* (MTD), is clearly and accessibly described in this book, written by Russian experts in the field of error-correction coding: Y. Zubarev, Associate Member of the Russian Academy of Sciences (RAS), Professor V. Zolotarev, and Dr G. Ovechkin, Doctor of Technical Sciences. The first inventors' certificates, securing USSR priority in the development of this amazingly simple and very effective method, date back to 1972. Over the past 40 years of research and development in error-correction coding methods, the authors of this book have elaborated a comprehensive error-correction coding optimization theory, which has already helped to resolve with a good level of practical application all the basic issues involved in creating codes for MTD, to optimize the decoder parameters and to identify new types of problems that can be resolved using their new and effective theory.

MTD algorithms, like most decoders for turbo and LDPC codes, are iterative procedures. However, turbo codes appeared 20 years later than MTD decoders, which underwent very dynamic development and intensive patenting back in those days, and continue to do so to this day.

The first of the major advantages of MTD algorithms, which were originally designed for binary codes, is that the majority error-correction procedures they employ allow full parallelization of operations and the creation of hardware decoders working with theoretically maximum possible performance. This ultimately paves the way for the development of very simple high-speed decoders of this type with very high performance characteristics.

Another property of MTD algorithms, constituting a major advantage over many other error-correction procedures, is that the likelihood of their decisions strictly increases throughout the whole process of error correction in a message distorted by noise. There is as yet no evidence of other equally simple decoding procedures possessing such unique properties. Of course, when it achieves the maximum-likelihood decision an MTD decoder becomes optimum, which typically requires an exhaustive search of all possible solutions, something which is very elegantly performed, in particular, by the Viterbi algorithm. But MTD complexity is a linearly increasing function of the code length used, which should thus theoretically be the smallest possible. Therefore MTD, unlike the Viterbi algorithm, easily handles very long codes, which allows it to offer a high degree of noise immunity and power gain.

After studying this somewhat extraordinary book, readers will discover an important – some may even say dramatic – situation: why the MTD algorithm has remained unknown to experts outside of Russia. A wealth of results were published in the 1970s reporting on repeated attempts to decode messages in which a certain proportion of distorted characters had been corrected in the first step of error correction. In all of those studies, however, it transpired that, on account of the grouping of errors at the output of the decoder after the first error-correction attempt, the second decoder was useless, being able, like the first decoder, to correct only independent symbol distortions in the communication channel. The occurrence of such a severe error-grouping effect at the output of the threshold decoder (TD), called "error propagation" (EP), and the complete dearth of ideas on how to minimize this effect, brought all work on the theme of repeated decoding by majority algorithms to a complete standstill.

Why then has the topic of repeated error correction by such majority methods re-emerged, moreover displaying record efficiency with the very simplest implementation for the vast majority of encoding schemes? We can point to two basic reasons behind the high performance of MTD.

First, until the authors of this book, no-one attempted to solve the problem of modifying this simplest of known decoding methods – majority decoding – to endow it with the ability to rigorously improve its decisions with every change in decoded symbols. Indeed, endeavouring to derive, from a very simple algorithm, a method displaying equivalent characteristics to the best exhaustive search processes, while undoubtedly laudable, is also extremely risky. Yet this very difficult problem was addressed by the authors – and solved successfully! MTD algorithms, in all their modifications, possess precisely those fundamentally important properties.

Secondly, another – ideologically much more difficult – problem that the authors had to tackle – and managed to solve! – was detailed analysis of the reasons for error grouping at the output of the majority decoder. It is precisely this EP effect that is analysed in this book. Once the nature of this effect was properly understood, it became possible to find codes with a low EP level, immediately yielding a fundamental improvement in MTD characteristics. All these problems were solved by the authors of this book using special optimization procedures and mathematical methods that have never before been employed in the area of error-correction coding.

Only by solving both of these two interrelated complex problems was it possible to create special codes and iterative MTD procedures with highly simple majority functions. Even at very high noise levels, the new algorithms, in the course of many iterations, gradually improve the reliability of information received from the channel and find optimal solutions in the vast majority of cases, a result that is usually achievable only using exhaustive search algorithms. Moreover, the total number of MTD operations remains quite small.

In actual fact, however, the successful development of MTD algorithms is related to the solution of another, third, classical problem, namely optimizing functionals involving a very large number of variables. As the authors quite rightly point out, the possibility of simultaneous variation of a whole series of code and decoder parameters, in particular check weights, threshold element values and differentials in the polynomials of the codes used, creates additional conditions for improving MTD decoder performance. There may be hundreds and even thousands of such tunable elements in decoders. Incorporating automated computer search of optimized algorithm parameters based on decoder decision error probability right from the design stage significantly improves the already quite high performance of MTD without any increase in the quantity of computations in the decoder with these improved parameters. Note that for other decoding systems one cannot contemplate even posing this third problem, because it can only arise in parallel with the solution of the first two problems of effective and easy decoding for MTD algorithms described above.

Thus, success in terms of both theory and application in the area of simple and effective decoding using MTD is determined by simultaneous successful solution of the three extremely difficult problems described above, each of which is in one way or another related to meeting the challenge of optimizing functionals involving a large number of variables. It is on productive application of optimization principles in all stages of MTD research and development that the success of new coding theory for information theory and communication technology depends. Inadequate solutions in any single one of them would reduce the value of potential achievements in the area of iterative majority decoding schemes, insofar as in this case all their characteristics would steadily deteriorate.

Some other very unusual methods in relation to traditional ways of decoding are presented in this book. In practice, they all serve the ultimate goal: maximizing decoder efficiency while minimizing design complexity.

These include various concatenating methods, especially parallel concatenation, as well as codes with separate branches.

Among the algorithms proposed in this book, symbolic (non-binary) codes and their corresponding special high-speed symbolic MTD decoders occupy an important place. They were discovered in 1984, and still do not have any equivalents or competitors in other coding areas. They have already completely resolved the problem of coding and high-speed decoding of large byte arrays with any given reliability at very high speeds by simple technical means. They overlap considerably in terms of effectiveness with RS codes, remaining just as simple to implement as their counterpart – binary MTD. Surprisingly, in many decades of development of coding theory, such an obvious extension of majority methods to non-binary digital data streams has not been explored by any researcher in the world, apart from one of the authors of this book. The advantage of non-binary MTD over RS codes is so immediately great that it can be stated to all intents and purposes that these codes and algorithms, discovered thirty years ago [11], mark a new era in the processing of symbolic information. There are quite simply no other methods with reasonable implementation complexity that can compete with this type of MTD in terms of performance. As in the development of other useful approaches to error correction, the authors have achieved very significant results for non-binary codes by switching to much longer codes than the only ones available to date, RS codes. Obviously, symbolic MTD displays the extremely low decoding complexity inherent in majority methods. A non-negligible factor in achieving such uniquely high performance with the new decoder lies in the fact that there is no need to use multiplications in non-binary fields, as well as the fact that symbolic code lengths are entirely independent of the size of the symbols used. It is precisely these two properties that make symbolic MTD so extremely simple to implement. Therefore, these codes will inevitably find widespread application in the processing, storage and transmission of audio, video and other data types.

It is also interesting to note that MTD algorithms, created to combat errors in data transmission paths, could be successfully applied as they stand to improve error-correction coding, and for data compression, in other words simultaneously solving the second most important problem in information theory – coding of certain types of sources. Details of these methods can be found in other publications by the authors.

Finally, it should be emphasized that all these MTD algorithms are so simple to implement and at the same time so highly effective that their software versions have successfully passed stringent tests and have been adopted for use in special digital television systems, and the process of standardization has been started for the corresponding codes.

This book should certainly have come out some 15 years ago, in which case we would now be talking about the next revised edition. All this time, however, the authors have being striving to acquaint the scientific and technical community with their new developments in coding theory. An important event for the Russian specialists was the release of the handbook "Error-correction coding. Methods and algorithms", written by V. Zolotarev and G. Ovechkin and edited by Y. Zubarev, published by *Goryachaya liniya – Telekom* ("Hot Line - Telecom") in 2004 in Moscow, Russia. The handbook helped to expand the sphere of application of MTD algorithms and significantly boost their development for a wide range of applications.

Some additional theoretical and applied results can be found in three monographs on MTD, published in the USSR and Russia (see References) and in more than 200 other texts on these algorithms.

A major contribution to coding theory and techniques has been made by the regular reports presented by the book's authors at the traditional annual International Conference on Digital Signal Processing in Moscow over the last seven years. There are very well-stocked specialized bilingual websites on MTD methods, widely known among specialists, namely the Russian Academy of Science's Space Research Institute (SRI RAS) site at www.mtdbest.iki.rssi.ru, and the Ryazan State Radio Engineering University (RSREU) site at www.mtdbest.ru, containing around 400 theoretical, methodological, educational and demonstration items about these algorithms, even including a series of special video clips providing particularly clear and dynamic demonstrations of the properties of MTD decoders. These portals are possibly the largest of all telecommunication sites and without doubt of the sites of all telecommunication corporations in Russia devoted to error-correction coding.

The above sites are also home to a unique collection of demo programs for the best known error-correction methods. Anyone wishing to do so may download the demo programs presented, together with instructions for their use, and compare their real potential. The exceptional usefulness of this accomplishment by the authors of this book cannot be overstated.

In concluding the presentation of what is clearly a quite extraordinary book, in terms of its methods and results, about the most prominent achievements of Russian scientists in the field of error-correction coding in recent years, we wish the authors and their colleagues and students further success in the boundless domain of high-quality digital networks. The launch of such an important monograph for communications theory and engineering at a time when the world community is intensively engaged in the transition to all-digital methods of data creation, storage, processing and transmission will without doubt speed up all communication system development processes and stimulate the further growth of information service quality in our technological civilization.

The rapid development of coding technology and hardware components will further expand the MTD algorithm's sphere of application. Of course, decoder implementation will soon become an achievable task in a number of other coding methods, too, even if they currently still seem too complex. But only error-correction coding methods that expend their computation resources in a targeted and economical manner to solve the problem of error correction in digital data will really be extremely fast, most effective and highly accessible for widespread use in communication networks and large databases.

V.K. Levin
Member of the Russian Academy of Sciences
Doctor of Technical Sciences
Professor
Holder of the Lenin, USSR and Russian Federation State Awards