Program realization of simulation tools of low-density codecs: effective architecture of a decoder for massive parallel computations on graphic processors

A V Bashkirov¹, A A Pirogov¹, O Y Makarov¹, A V Muratov¹, S Y Beletskaya¹, V V Glotov¹

¹Voronezh State Technical University, 14, Moscow ave., Voronezh, 394026, Russian

E-mail: vadik-livny@mail.ru

Abstract. The article deals with the peculiarities of modeling the characteristics of low-density parity-check codes (LDPC) in terms of their decoding. When decoding low-density parity-check codes in various communication standards, the iterative belief propagation algorithm (BP) and its various modifications - Pearl’s algorithm, sum-of-products algorithm - are widely used. In the article, special attention is given to this algorithm. The BP algorithm is a message-passing algorithm on a graphical model and consists of two basic messaging procedures between the nodes in the Tanner graph for LDPC code. The issues of constructing an effective architecture for an iterative decoder operating on the belief propagation algorithm in a heterogeneous system with massively parallel computing on a graphics processing unit are considered. It is shown that the use of the GPU allows to achieve a significant gain in decoding time with a sufficient size of the test matrix (a sufficient number of code and test nodes), obtained due to the parallelism of the architecture. Expressions for calculation of the size of the “working group” are given. The benefit of using a modified decoder architecture is estimated.

1. Introduction
Information-telecommunication system functioning in the context of diverse interference effects forces designers to take measures against errors appearing in the channels of communication. Application of correcting coding, namely, low density (LDPS – Low density parity-check code), is considered to be one of the most effective ways to increase interference resistance of means of communications. Due to their energy efficiency LDPC application is recommended by a vast amount of modern standards in the sphere of telecommunications (802.11 (WiFi), DVB-x2 family and others [1, 2, 4].

A very important contribution to the solution of this problem is made by the Anti-Noise Coding Theory. On the basis of this theory there are worked-out methods of error protection, based on the application of noise-combating (anti-noise) codes. Utilization of such codes provides for obtaining the Energy Gain of Coding (EGC), which characterizes the degree of possible diminishing of the transfer energy due to the coding in comparison to the absence of the coding, if the transfer credibility is the same in both cases [1 – 3]. This gain can be used for the improvement of parameters and characteristics of many important qualities of the data transfer systems, for example: for downsizing very expensive aerials, extension of communication range, increase of data transfer rate, reduction of the required capacity of the transmitter, etc.
Binary LDPC Codes, after their re-discovery at the end of the 1990-ies, attracted the researchers greatly due to their excellent efficiency in error correction and high parallel iterative scheme of decoding. They became one of the branch standards for correction of coding errors in the wireless communications. Nevertheless, when the length of a codeword is small enough, or when a modulation of higher order is in use [5], then the non-binary LDPC codes, determined in the Galois field GF (q) [6] can outdo their binary counterparts.

The basic disadvantage of the codes in the fields of higher order is the complexity of the decoding algorithm. Both binary and non-binary LDPC codes are decoded by means of the Belief Propagation algorithm (BP algorithm) on their factor-graphs; however, the decoding complexity of GF (q) codes is scaled exponentially with the number of bits in the symbol.

2. Fundamentals and Exploring the importance of the problem

Issues of simulation characteristics of interference resistance codecs at the preliminary stages of the information-telecommunication system design have been discussed in the studies [2-7], whereby, task of code decoding will be further considered in the context of simulation modeling on ECM.

Iterative “belief propagation” algorithm and its modifications are most widely used for decoding of low density codes. Belief propagation algorithm (BP, Perl algorithm, “sum-product” algorithm) is an iterative soft-decision probability algorithm. An algorithm estimates precise a posteriori probabilities after a certain number of iterations if a Tanner graph (a bipartite graph incident to check matrix H) for the given code does not contain cycles. A complete algorithm description is given, for example, in [1-6]. It is instructive to recall that BP algorithm is a message passing algorithm on a graph model and is composed of two basic procedures of message exchange between Tanner graph nodes for LDPC. For the ease of convenience, we consider them in the context of BP algorithm logarithmic version (Figure 1).

![Figure 1. Basic stages of message exchange in the belief propagation algorithm: a) a procedure of message passing from code peaks to check peaks and estimation in them; b) a procedure of message passing from code peaks to check peaks and estimation in them.](image)

It is a fact that the BP algorithm complexity can be lessened with its duel form in the frequency domain. Another algorithm uses the logarithmic domain with logarithmic density or logarithmic ratio of
densities of messages’ notation, which require fewer quantizing levels due to low sensitivity of quantization. Besides, there is applied the Extended Min-Sum algorithm (EMS algorithm).

Figure 1 demonstrates: a stage a) corresponds to the first step of BP algorithm message processing. As a result, each $i$ check peak (type 1 node on the graph) independently estimates $K_i$ of $\lambda_{i,k}^{0,1}$ values. Due to independence of these estimations they may be performed either successively for each peak, one after another, or directly parallelly for all $M$ peaks. A similar situation is also typical for the second step of BP algorithm message processing schematically showed in Figure 1 b). Each $i$ code peak (type 2 node on the graph) independently estimates $J_i$ of $\lambda_{i,j}^{1,1}$ values. These calculations may also be performed either successively or parallelly, directly for $N$ peaks. Depending on successive or parallel calculations at these stages a decoder may be designed on several variants of its architecture. We are mainly interested in parallel architecture. General architecture of a parallel decoder is represented in Figure 2.

**Figure 2.** Architecture of the parallel log-BP decoder.
3. Methods
Parallel architecture makes significant demands to a certain number of necessary computational elements, at the same time, possessing evident advantages regarding rate of calculations over successive realization.

This parallelism opens a way to use graphics processing units in calculations; they, on the one side, are absolutely suitable for solving tasks of parallel calculations; on the other side, they are massively widespread and always appear to be components of ECM, namely, its video-accelerators. Application of these resources allows considering ECM as the heterogeneous calculating system with the basic calculating device – a central processing unit (CPU) and non-basic – a video-accelerator graph processor [8-9].

The architectural decision described earlier is possible to optimally realize by means of an open standard Open CL (Open Computing Language). If check matrix size is sufficient enough (a sufficient number of code and check nodes), time decoding improvement achieved by means of architecture parallelism is sufficient enough to exceed simulating successive realization with concentration of calculations on the central ECM processor.

Better results can be obtained using the following modification of the given architecture. Considering blocks $F$ and $F'$ of $\chi_{m1}^0$ and $\pi_{m1}^0$ calculations we can conclude that calculations inside them may also be parallelized; this will demand $K$ processing elements in blocks type $F$ and $J$ processing elements in blocks type $F'$ (Figure 3).

![Figure 3. Model of BP decoder parallelism: a) classic; b) modified](image)

The architecture corrected in accordance with the scheme of Figure 3b) is presented in Figure 4.

A number of performed $Th_{s}^{0,1}$ streams (global Threads) and a two-dimensional work-group size $Th_{l}^{1,1}$ (local Threads) for implementation of $N$ code exemplars on the graph processor are determined by the expression:

$$Th_{s}^{0} = Th_{l}^{0} = J; mxN = floor\left(\frac{Th_{\max}}{J}\right);$$

$$\begin{cases} N > mxN, \\ Th_{l}^{1} = N + mxN - \left[N \mod (mxN)\right], \\ Th_{l}^{1} = mxN; \\ N \leq mxN, Th_{s}^{1} = Th_{l}^{1} = N. \end{cases}$$

where $Th_{\max}$ - a maximally permissible size of a local one-dimensional work-group (CL KERNEL WORK GROUP SIZE), and $floor(x)$ function returns $x$ value rounded to the closest whole number downward.
4. Results
Application of this model allowed obtaining reduction of calculations performance time. Figure 5 presents a graph of time improvements achieved due to application of the modified parallelism model depending on the code length ($N$, 3, 6).

Figure 5 represents $T_2$ – decoding time with classic architecture, and $T_1$ – decoding time with modified architecture in 10 iterations of BP algorithm and 3408 simulations per one SNR value (totally 10 SNR values were modified: from 0.1 to 2.5 with resolution equal 0.25 dB). Calculations were performed on the graph card AMD Radeon HD 5770 (800 streaming processors with 850 MHz frequency).

Calculation time is the time in which the BER (Bit Error Rate) of low-density codes with a speed of 0.5, a length of 96-9972 bits, a noise level between 0.1 and 2.5 dB with an increment of 0.25 dB was calculated.

![Diagram](image)

**Figure 4.** Modified architecture of the parallel log-BP decoder suggested by the authors of the article.
5. Conclusion

Thus, it is possible to significantly increase calculation performance carried out when simulating interference resistant low-density codec characteristics by additionally parallelized iterative decoding scheme. This fact is considered to be of great importance for applications crucially demandable to calculating resources, and, eventually, allows reducing time expenditures on the simulation process.

This method of parallelization of calculations should be used as a means of solving the problem of reducing the time spent on resource-intensive modeling, produced in the preliminary stages of the development of noise-resistant radio systems. The application of the proposed means, which provides a significant increase in the speed of calculations even on computers with budget graphics cards, is relevant for communication system developers, especially considering its hardware independence and the prospects of the layout as an accelerated modelling of noise-resistant low-density codecs produced to optimize the noise-resistant coding system. Such environment could constitute an alternative to the frequently used low-performance modeling packages (MatLab, Simulink, Mathcad, SigLib), which due to their versatility are expensive and do not provide enough opportunities when modeling for highly specialized tasks.

References

[1] Hale U K 1980 Frequency assignment. Theory and applications. 68 112
[2] John R B 1990 Electronic System Design: Interference and Noise Control Techniques (Moscow: Mir)
[3] Sliykin V G 2004 Electromagnetic compatibility of electrical equipment of information technologies under the influence of pulsed electromagnetic fields (Samara, SGTU)
[4] Tryapitsyn A V Development of the methodology for the design of wireless local area networks with EMC (Moscow: Mir)
[5] Williams T E 2004 EMC for Product Designers (Moscow: Publishing house "Technologies")
[6] Habiger E H 1995 Electromagnetic Compatibility. Fundamentals of its provision in technology. (Moscow: Energoatomizdat)
[7] Kechiev L N 2007 Proektirovanie pechatnykh plat dlya tsifrovoy bystrodeystvuyushchey apparatury (Moscow: Gruppa IDT)
[8] Surikov V V 2010 Elektromagnitnaya sovmestimost' priemno-peredayushchikh ustroystv, raspolozhennykh na elementakh konstruktssi slozhnoy formy (St. Petersburg:Eva)
[9] Knyazev A D 1984 Elementy teorii i praktiki obespecheniya elektromagnitnnoy sovmestimosti radioelektronnykh sredstv (Moscow: Radio i syvaz)