

Computation Optimization for Digital Signal Processing Operations in Homogeneous Computing Environments and Structural Transformations of Their Parallel Algorithms

O. V. Klimova^{1, 2}

¹*Institute of Engineering Science, Ural Branch of the Russian Academy of Sciences,
34 Komsomolskaya St., Ekaterinburg, 620049, Russia*

²*B. N. Yeltsin Ural Federal University (UrFU), 19 Mira St., Ekaterinburg, 620002, Russia*

Corresponding author: ovk31@mail.ru

Abstract. The issues of computation optimization for digital signal processing operations are considered when choosing the variant of their organization in the space-time environment. The implementation of such research is based on the obtained model description of computation organization, which allows us to synthesize a variety of their variants and perform the concurrent exploration of algorithms and architectures. Methods of computation optimization obtained as a result of analyzing parameters characterizing the above model description are discussed. Using these methods in the concurrent exploration of algorithms and architectures allows you to find the best variants for computation organization.

INTRODUCTION

Improving the efficiency of modern parallel computing is associated with the implementation of the stage of the concurrent exploration of algorithms and architectures [1–3], which is built into the process of designing advanced computation systems. Such studies require a formal tool characterized by both algorithmic and architectural parameters, which allows synthesizing a variety of variants of computation organization and making a well-founded choice of a rational variant. The possibility of making such a choice implies the presence of a formal parametrized description of computation organization, which allows one to perform optimization actions in the process of synthesis and analysis of their variants within the framework of the given restrictions on hardware and time costs. Such an instrument was created [3, 4] for the class of digital signal processing (DSP) operations on the basis of the development of a group-theoretic decomposition approach [3], which made it possible to study the intrinsic space-time structure of the above operations and to reveal composition forms of their representation. As a result, a transition was made from the original analytical expressions representing the algorithms of these operations to their equivalent composition forms and then to a single parameterized model description of computation organization in a space-time environment. The resulting formal tool, a model description, allows one to synthesize a variety of computation organization variants for a whole class of operations and serves as a basis for the implementation of the concurrent exploration of algorithms and architectures. However, for such an instrument to be used effectively, it is necessary to study and describe its capabilities aimed at optimizing the structures of synthesized parallel algorithms. The paper discusses ways of optimizing computation, representing the above capabilities. These methods are the basis for optimizing the search of the best variants of computation organization when conducting the concurrent exploration of algorithms and architectures. The study of the optimization properties of the obtained model description is based on the consideration of its structural components – parameters, the symbiosis of which allows us to present a whole variety of computation organization variants.

MODEL DESCRIPTION: THE BASIS FOR COMPUTATION OPTIMIZATION

The basis for creating a model description of the computation organization for DSP operations was the transition from the original analytical forms of their presentation to compositional forms. The equivalent transformations of the original analytical expressions of DSP operations performed on the basis of the proposed decomposition approach made it possible to obtain their composition forms and thereby transfer the computations from the time domain to the space-time domain. In the process of implementing this transition, the computational structures characteristic of the operations of the above class were changed. These changes consist in the restructuring of the computing environment. Indeed, if a time computing environment generated by the operation of multiplying a matrix of size $N \times N$ by the vector $x(t)$ of length N could be used to describe structures corresponding to the original forms of representing these operations, then, to describe the structures corresponding to the obtained composition forms, a parameterized space-time computing environment has become necessary. The computation organization in such an environment is described by the obtained compositional forms of operations (CFO), presented using their single model description $CFO(A_i^{jp}(t_1), CCE_i(j, p))$. Its components are the parameterized coordination-computing environment (CCE) – $CCE_i(j, p)$ and algorithms $A_i^{jp}(t_1)$ obtained as a result of decomposition, immersion into the space environment, and time compression of the sequential algorithms $A_i(t)$. The presented model description forms a parameterized space-time environment consisting of L^2 matrices of size $h_1 \times h_1$ and characterized by the corresponding coordination of the computation results obtained by multiplying these matrices by the vectors $x_j(t_1)$ of length h_1 formed as a result of the group-theoretic decomposition of the input data $x(t)$. The parameters $N=h_1L$, $t=j+t_1L$, $t=0, \dots, N-1$, $t_1=0, \dots, h_1-1$, $j=0, \dots, L-1$ characterize the decomposition performed. The type of computation coordination is determined by the operation under study. Figure 1 shows changes in the internal structure of operations of the class in question when moving from the original form of their presentation (see Fig. 1a) to the composition form (see Fig. 1b). The resulting internal structure of operations can be characterized as a parameterized homogeneous computing environment represented using a two-dimensional lattice. Its L^2 nodes contain computing units that implement the operation under study. This operation is compressed to the size h_1 and represented using the obtained parameterized environment described by the parameters j, p . Using these parameters, you can work with the established internal structure of operations in order to identify the most effective variants of parallel computation organization. To perform such actions, it is necessary to study the issues of computation optimization in the resulting homogeneous computing environments. A brief presentation of the results of such a study will be presented in the next section. The basis for considering optimization issues was a model description of variants of parallel computation implementation, namely $CFO(A_i^{jp}(t_1), CCE_i(j, p))$.

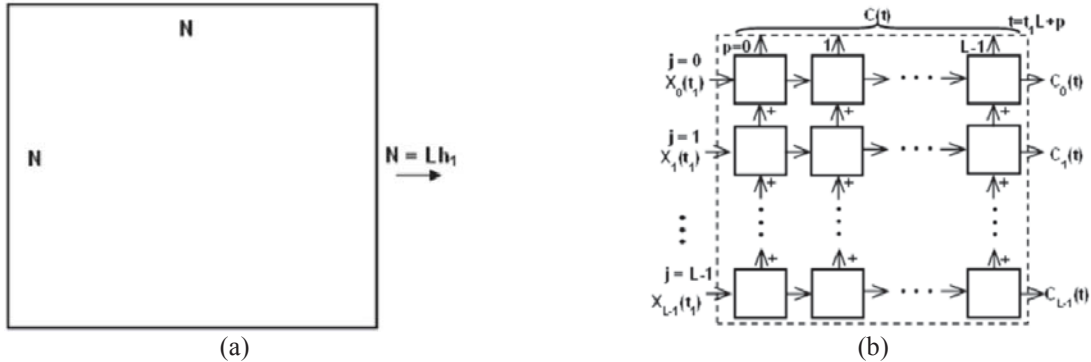


FIGURE 1. Illustration of the internal structures of computational operations for two forms of their representation: the original form (a) and the parameterized composition form (b)

THE PARAMETERS AND METHODS OF OPTIMIZATION

The transition performed to a model description of parallel algorithms for DSP operations made it possible to control the change in their structures by parameters and thereby synthesize their variety. The obtained opportunity to form a variety of algorithm structures opens the way for computation optimization in a space-time environment. Such optimization can be performed on a formal basis in the process of finding the best variant of computation organization for the given conditions of its implementation. This process consists of the following stages. The stage

of parametrized synthesis of parallel algorithms, the stage of analysis by forming estimates of the complexity of their implementation, and the stage of reasonable choice within the framework of the given restrictions on the parameters of hardware and time costs. The model of computation organization for various DSP operations: *Model* ($O_m, N, L, h_1, A_{mi}^{jp}(t_1), CCE_{mi}(j, p)$) built on the basis of the above-mentioned CFO is the platform for implementing this process. The specific form of the CFO is determined by the index i . The variety of CFO types obtained in the framework of the equivalent transformations performed over the initial forms of operations O_m is determined by the possibilities of changing the decomposition parameters $P_D(N, L, h_1)$ used in this case, decomposition methods, domains of implementation of the computations (time or frequency), as well as types of orthogonal transformations. As the latter, both discrete Fourier transforms (DFTs), characterized by calculations in the field of complex numbers, and number-theoretic transformations (NTTs) [5, 6] with calculations in finite fields can be used. The above methods of changing the CFO and, therefore, the structures of parallel algorithms can be set using the following parameters: the parameter P_{MD} determining the decomposition method, the parameter P_{CD} determining the computation domain, and the parameter P_{OT} characterizing the type of orthogonal transformation. The dependence on the P_{OT} parameter is characteristic of convolution and correlation operations. The transformation of the internal structure of convolution computation when changing the P_{CD} and P_{OT} parameters is illustrated in Fig. 2.

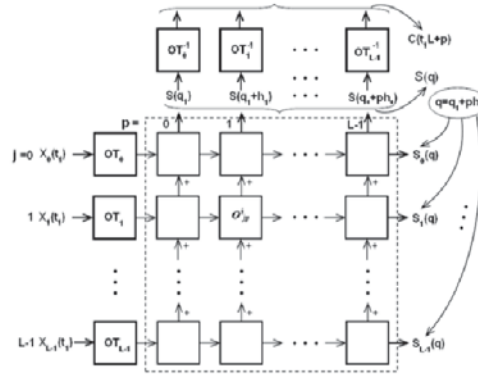


FIGURE 2. The internal structure of the convolution computations in the frequency domain using a two-dimensional parameterized environment

Thus, changing the value of the P_{CD} parameter and transferring the computations to the frequency domain, we can go from the space-time structure shown in Fig. 1 to the structure defined using Fig. 2. This transition is based on the synthesis of a compositional form that provides a parameterized description of the space-time organization of computations in the frequency domain. In the process of creating such a form, the coordination-computing environment is transformed. In each cell of such an environment, the multiplication of the corresponding length spectra h_1 is performed. Convolution computations in the frequency domain are performed (see Fig. 2) using direct and inverse orthogonal transformations (OT, OT^{-1}) of length h_1 . And since the obtained compositional structure (see Fig. 1) of DSP operations allows us to solve the problem [5, 6] of using effective, but short number-theoretic transformations (NTTs) for calculating long convolutions, the structure shown in Fig. 2 can be changed due to the transition from DFT to NTT. Such structural changes can be controlled by changing the values of the P_{OT} parameter. The analysis of the above possibilities of changing the algorithm structures shows ways of optimizing the computations that open up in the process of forming various CFOs for DSP operations.

When working with a specific CFO, the possibilities of changing the algorithm structures and optimizing computations are opened up by adjusting their coordination-computing environments $CCE_{mi}(j, p)$ by the decomposition parameters $P_D, j, p=0, \dots, L-1$. Indeed, changing the values of the parameter L within the given restrictions $L=N/h_1$, it is possible to synthesize various parallel structures for DSP operations and the corresponding compositional forms obtained as a result of the above transformations of the original forms of operations. This ability to control the structure changes of the algorithms by the parameters allows you to choose the space characteristics of the parametrized homogeneous computing environment that represents the internal structure of the operations in question. This possibility allows you to coordinate the relationship of algorithmic parameters with architectural ones in the process of their concurrent explorations, enabling you to optimize the computations and to choose the best variant of the computation organization.

The next opportunity for changing the algorithm structures and optimizing computations is associated with the introduction of restrictions on the space size of homogeneous environments that provide the computations generated by the CFO. The introduction of such restrictions will make it possible to change the algorithm structures and to perform the computation optimization by synthesizing their various variants, characterized by a parallel-pipelined organization of the computing process. Such restrictions can be introduced by using the matrix computing environment $MCE(k_1, k_2)$ defined by the parameters k_1 and k_2 , $k_1, k_2 = 1, \dots, L$, structurally similar to $CCE_{mi}(j, p)$ and setting the space size of homogeneous environments that implement computations according to the CFO in the parallel-pipelined processing conditions. To form an augmented model that can describe such variants of the computation organization along with the above ones, we complement the model presented above with the component $MCE(k_1, k_2)$. The resulting augmented model $Model(O_m, N, L, h_1, A_{mi}^{jp}(t_1), CCE_{mi}(j, p), MCE(k_1, k_2))$ allows us to synthesize a variety of variants of the computation organization and to optimize them due to its inherent ability to control changes in the algorithm structures by parameters.

The results can be summarized as the following methods for optimizing computations obtained on the basis of the model description of computation organization, which is characterized by the possibilities of changing the algorithm structures:

- optimization by changing the spatial structure of the coordination-computing environment;
- optimization based on the introduction of space restrictions on the implementation of parameterized compositional forms and the use of parallel-pipelined processing;
- optimization due to the use of various orthogonal transformations in the CFO structures;
- optimization based on the change in the computation implementation domain (frequency/time);
- optimization by developing other compositional forms for the same operation and their use in the synthesis of parallel structures.

CONCLUSION

A comprehensive study of the optimization of DSP operation computations in a space-time environment has become possible thanks to the development of a model for their organization in a homogeneous computing environment. The model provides a platform for optimization and allows you to form methods for its implementation. The basic characteristics of the model, which open up the possibility of optimization, are its abilities to generate a variety of parallel algorithms and manage the changes of their structures. It is the changes of the algorithm structures within the framework of established formal rules due to changes in the values of its parameters of the model description that open the way to computation optimization. The analysis of the developed model of computation organization has made it possible to identify the parameters by which changes of the algorithm structures can be controlled. The changes of the structures open the way to optimization; therefore, the parameters by which you can change the structure of the computations form the basis for the development of optimization methods. An analysis of these parameters has been performed, and a characteristic of these methods has been presented.

REFERENCES

1. G. G. Lee, Y. K. Chen, M. Mattavelli, and E.S. Jang, [IEEE Trans. Circuits and Systems for Video Technology](#) **19**(11), 1576–1587 (2009).
2. G. G. (Ch.) Lee, H.-Y. Lin, Ch.-F. Chen, and T.-Y. Huang, [IEEE Trans. on Parallel and Distributed Systems](#) **23**(5), 944–957 (2012).
3. O. V. Klimova, Russ. Toms State University Journal of Control and Computer Science **2**(23), 112–120 (2013).
4. O. V. Klimova, Russ. Journal of Information Technologies and Computing Systems **2**, 11–22 (2016).
5. J. H. McClellan and C. M. Rader, *Number Theory in Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1979).
6. O. V. Klimova, Russ. Izvestiya RAS. Tech. Cybernetics **2**, 183–191 (1994).