Sergey I. Kumkov
Sergey Shabunin
Stavros Syngellakis *Editors*

Advances in Information Technologies, Telecommunication, and Radioelectronics

Springer

Innovation and Discovery in Russian Science and Engineering

This Series provides rapid dissemination of the most recent and advanced work in engineering, science, and technology originating within the foremost Russian Institutions, including the new Federal District Universities. It publishes outstanding, high-level pure and applied fields of science and all disciplines of engineering. All volumes in the Series are published in English and available to the international community. Whereas research into scientific problems and engineering challenges within Russia has, historically, developed along different lines than in Europe and North America. It has yielded similarly remarkable achievements utilizing different tools and methodologies than those used in the West. Availability of these contributions in English opens new research perspectives to members of the scientific and engineering community across the world and promotes dialogue at an international level around the important work of the Russian colleagues. The broad range of topics examined in the Series represent highly original research contributions and important technologic best practices developed in Russia and rigorously reviewed by peers across the international scientific community.

More information about this series at http://www.springer.com/series/15790

Sergey I. Kumkov · Sergey Shabunin ·
Stavros Syngellakis
Editors

Advances in Information Technologies, Telecommunication, and Radioelectronics

Editors
Sergey I. Kumkov
Ural Federal University
Ekaterinburg, Russia

Sergey Shabunin
Ural Federal University
Ekaterinburg, Russia

Stavros Syngellakis
Wessex Institute of Technology
Southampton, Hampshire, UK

Preface

This book is mainly a collection of ideas, techniques, and results in the field of video information technologies and various related applications of numerical methods. It comprises 18 chapters grouped under four main topics, namely, image processing and computer vision, signal processing and navigation, simulation of some practical processes and computations for antennas and applications of microwaves. The research described in this volume is addressed to a wide audience of scientists, engineers, and mathematicians involved in the above mentioned four scientific topics.

Part I of the book (Image Processing and Computer Vision) comprises several inter-connected topics. Chapter 1 presents an overview of a new theoretical framework for multichannel image processing using the hypercomplex algebras. The latter constitute a generalization of the algebras of complex numbers. The main hypotheses are: the brain of primates operates with hypercomplex numbers during retinal image processing while the visual systems of animals with different evolutionary history use different hypercomplex algebras for color and multicolor image processing. In the Chap. 2, procedures for formation of satellite images and satellite data are considered with their application to practical problems arising from Earth surface observation. Chapter 3 is devoted to problems of constructing special fast computers for image processing in embedded computer systems. A new promising approach to image processing and fusion is described that is based on wavelet transformations. Theory and effective algorithms are suggested in the Chap. 5, which is devoted to the development of many-factor MIMO-filters for processing various image data. Part I finishes with a Chap. 6 on a new approach to the examination of the clarification effect under colorimetric object illumination. This problem and its solution are very important for providing high quality multicolor printing.

Part II comprises chapters on two interesting topics. Chapter 7 is devoted to the multi-fractal analysis of bio-signals recorded simultaneously from many sensors; the situation is typical in investigations of the human brain and neural system. Chapter 8 describes the main aspects of elaboration algorithms and software for the

3-D navigation in a closed environment; for example, in navigation of transportation means and robots inside buildings and pack-houses.

Part III is devoted to problems of simulating important practical processes. Chapter 9 delivers simulation of a money turnover process in a computer agent-based system; the investigation allows one to look at the process from a new point of view and to understand more details in the organization of money turnover. Chapter 10 presents results of simulating real-time processes in industrial systems. Here, a detailed analysis is performed for determining the components of manufacturing processes and for the organization of optimal sequences in their implementation. Chapter 11 considers simulation of elaboration of artificial networks for sophisticated estimation of chemical element contents in the soil. It is noted that building such instruments is very important for practical investigations of current situations associated with contamination of the environment. This part of the book ends with the Chap. 12 devoted to the simulation of urban passenger transportation systems and organization of their management. A new type of management building is suggested on the basis of its five-module concept.

The final part, Part IV, presents investigations on the antennas and on the scope of microwaves. Chapter 13 describes a special application of Green functions for the description of multi-layered cylindrical structures. Here, effective algorithms were developed for computation of radiation propagation and scattering waves in such structures. Chapter 14 is devoted to detailed investigation on the elaboration of topologies and practical constructing micro-strip units, particularly, the micro-strip ring coupler. The theoretical aspects and computational results were implemented in the experimental units. In the Chap. 15, numerical analysis is performed for spherical and geodesic shelters of antennas. Here, computational algorithms are based on the sophisticated application of Green functions. The results of the investigation allow one to provide the necessary radiation pattern of antennas with shelters. Chapter 16 is devoted to a comparative analysis of methods for excitations of TE_{01} mode of waves in circular waveguides. The investigation provides a new insight on problems appearing in excitation of waves of the mentioned type and corresponding aspects of energy transferring. Chapter 17 presents very interesting details of the application of antenna-sensors in investigations of the state and dynamics of processes in the human brain. To provide high precision and jam-proved information, a special sensor in the form of the antenna-applicator and its properties are considered. Providing such signals, without additional noises and corruptions, is very important for implementation of necessary diagnosis of the human brain. The concluding Chap. 18 is devoted to detailed numerical investigation of the multi-fractal nature of the human brain radiation signals. An approach based on the multi-fractal concept allows one to analyze sophisticated details of the brain state and its connection with various anomalies in its internal processes.

Our thanks go to all of them. We express special gratitude to Colleagues from the Wessex Institute who helped us with this book from its first version to the final one. Lastly, we are grateful to Elena Magaril, whose endless energy brought us through the most critical stages of the book preparation.

Ekaterinburg, Russia

Sergey I. Kumkov
Sergey Shabunin

Contents

Part I
Image Processing and Computer Vision

Chapter 1
Hypercomplex Algebras as Unified Language for Image Processing and Pattern Recognition Part 1. Cliffordean Models of Multichannel Images

Valeriy G. Labunets, Juriy G. Smetanin, Victor P. Chasovskikh and Ekaterina Ostheimer

1.1 Introduction

We develop a conceptual framework and design methodologies for multi-channel image processing systems with assessment capability. The term multi-channel image is used for an image with more than one component. They are composed of a series of images $f_{\lambda_1}(\mathbf{x})$, $f_{\lambda_2}(\mathbf{x})$, ..., $f_{\lambda_K}(\mathbf{x})$ in different optical bands at wavelengths $\lambda_1, \lambda_2, \ldots, \lambda_K$, called spectral channels, where K is the number of different optical channels. A multi-channel image (MCI) can be considered as a n-D K-component (vector-valued) function

$$\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_K(\mathbf{x})) : \mathbf{R}^n \to \mathbf{V}^K \tag{1.1}$$

with values into K-D perceptual space $\mathbf{V}^K$ (dichromatic $\mathbf{V}^2$, color $\mathbf{V}^3_{rgb}$, multi-channel $\mathbf{V}^K$), where $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbf{R}^n$, $n = 2, 3, \ldots$ For processing and recognition of 2-D, 3-D, and n-D images, we turn the perceptual spaces into corresponding hyper-complex algebras (and call them perceptual algebras). In this work, our approach to multi-channel processing is based on noncommutative Clifford algebras. In the

V. G. Labunets (✉) · V. P. Chasovskikh
Ural State Forest Engineering University, Yekaterinburg, Russia
e-mail: vlabunets05@yahoo.com

Ural Federal University, Yekaterinburg, Russia

J. G. Smetanin
Federal Research Center "Information and Control" of the RAS, Moscow, Russia

E. Ostheimer
Capricat LLC, Pompano Beach, FL, USA

algebraic-geometrical approach, each multi-channel pixel is considered not as a K-D vector, but as a K-D hypercomplex number. We will interpret multi-channel images as Cliffordean-valued signals

$$\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_K(\mathbf{x})) = f_1(\mathbf{x})J_1 + f_2(\mathbf{x})J_2 + \cdots + f_K(\mathbf{x})J_K, \quad (1.2)$$

which take values in so-called Clifford algebra $Alg_{2^K}^{Vis(u,v,w)}(\mathbf{R}|1, J_1, J_2, \ldots, J_{K-1}) = Alg_{2^K}^{Vis(u,v,w)}$ where $J_1, J_2, \ldots, J_K$ are hyper-imaginary units with the following noncommutative multiplication rule $J_r J_s = -J_s J_r, J_s^2 = -1$ for $s, r = 1, 2, \ldots, K$.

In this context, the full machinery of ordinary gray-level signal processing theory can be transposed in multi-channel image processing theory. Our hypotheses are

1. The brain (visual cortex) of primates operates with Cliffordean numbers during image recognition. In the algebraic approach, each pixel is considered not as a multidimensional vector, but as a multidimensional Cliffordean number. For this reason, we assume that the human retina and human visual cortex use Commutative hypercomplex and noncommutative Clifford numbers to process and recognition of color (RGB) images, respectively.
2. The brain uses different algebras for retina and for visual cortex (VC) levels. Multi-channel images appear on the retina as functions with values in a multiplet K-D algebra (see [1–3]), where K is the number of spectral channels. For example, RGB color images as they appear on the human retina are represented as triplet-valued functions. But multi-channel images in a human visual cortex are functions (1.2) with values in a 8-D Clifford algebra.
3. Visual systems of animals with different evolutionary history use different hypercomplex algebras for color and multi-channel image processing. Thus, visual cortex might have the ability to operate as a Clifford algebra computing device.

1.2 Clifford Algebras and Cayley-Klein Geometries of Visual Spaces

As we see in (1.1), n-D K-component images

$$\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_K(\mathbf{x})) : \mathbf{R}^n \to \mathbf{V}^K$$

have two attributes: n-D physical $\mathbf{R}^n$ and K-D perceptual space $\mathbf{V}^K$. We suppose that a brain operates with hypercomplex numbers when processing an image and calculates hypercomplex-valued invariants of an image recognizing it. In order to operate with n-D vectors $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbf{R}^n$ and K-D vectors $\mathbf{f} = (f_1, f_2, \ldots, f_K) \in \mathbf{V}^K$ as with numbers, we embed $\mathbf{R}^n$ and $\mathbf{V}^K$ into spatial $Alg_{2^n}^{Sp}(\mathbf{R}|1, I_1, \ldots, I_K)$ and visual $Alg_{2^K}^{Vis}(\mathbf{R}|1, J_1, \ldots, J_K)$ hypercomplex Clifford algebras:

$$\mathbf{R}^n \to Alg_{2^n}^{Sp}(\mathbf{R}|1, I_1, \ldots, I_n), \quad \mathbf{V}^K \to Alg_{2^K}^{Vis}(\mathbf{R}|1, J_1, \ldots, J_K).$$

When one speaks about both algebras simultaneously, then they are denoted by $Alg_{2^t}(\mathbf{R}|1, B_1, \ldots, B_t)$ or Alg_{2^t}. Obviously,

$$Alg_{2^t}(\mathbf{R}|1, B_1, \ldots, B_t) = \begin{cases} Alg_{2^n}^{Sp}(\mathbf{R}|1, I_1, \ldots, I_n), & \text{if } t = n, B_1 = I_1, \ldots, B_n = I_n, \\ Alg_{2^K}^{Vis}(\mathbf{R}|1, J_1, \ldots, J_K), & \text{if } t = K, B_1 = J_1, \ldots, B_K = J_K. \end{cases}$$

Let "small" t-D space $\mathbf{R}^t$ be spanned on the orthonormal basis of t *hyper-imaginary units* B_i, $i = 1, 2, \ldots, t$ (basic set of imaginaries—BSoI). We assume

$$B_i^2 = \begin{cases} +1, & \text{for } i = 1, 2, \ldots, u, \\ -1, & \text{for } i = u+1, u+2, \ldots, u+v, \\ 0, & \text{for } i = u+v+1, \ldots, u+v+w = t, \end{cases}$$

and $B_i B_j = -B_j B_i$. Now, we construct the "big" multicolor 2^t-D hypercomplex space $\mathbf{R}^{2^t}$ as a direct sum of subspaces of dimensions $C_t^0, C_t^1, C_t^2, \ldots, C_t^t$:

$$\mathbf{R}^{2^t} = \mathbf{R}^{C_t^0} \oplus \mathbf{R}^{C_t^1} \oplus \cdots \oplus \mathbf{R}^{C_t^s} \oplus \cdots \oplus \mathbf{R}^{C_t^{t-1}} \oplus \mathbf{R}^{C_t^t},$$

where subspaces $\mathbf{R}^{C_t^s}$, $s = 0, 1, 2, \ldots, t$ are spanned on the s-products of units $B_{m_1} \cdot B_{m_2} \cdot \cdots \cdot B_{m_s}(m_1 < m_2 < \cdots < m_s)$. By definition, we suppose that $B_0 \equiv 1$ is the classical real units 1. So

$$\mathbf{R}^{C_t^0} = \{a_0 B_0 | a_0 \in \mathbf{R}\},$$
$$\mathbf{R}^{C_t^1} = \{a_1 B_1 + a_2 B_2 + \cdots + a_t B_t | a_1, a_2, \ldots, a_t \in \mathbf{R}\},$$
$$\mathbf{R}^{C_t^2} = \{a_{1,2} B_1 B_2 + a_{1,3} B_1 B_3 + \cdots + a_{t-1,t} B_{t-1} B_t | a_{1,2}, \ldots, a_{t-1,t} \in \mathbf{R}\},$$
$$\ldots,$$
$$\mathbf{R}^{C_t^t} = \{a_{1,2,3,\ldots,t} B_1 B_2 \cdots B_t | a_{1,2,3,\ldots,t} \in \mathbf{R}\}.$$

Example 1 Let us consider 1-, 2-, 3-D small spaces $\mathbf{R}^t = \mathbf{R}^1, \mathbf{R}^2, \mathbf{R}^3$. The corresponding "big" spaces are

$$\mathbf{R}^{2^1} = \mathbf{R}^2 = \mathbf{R}^1 \oplus \mathbf{R}^1 = \mathbf{R} \cdot B_0 + \mathbf{R} \cdot B_1,$$
$$\mathbf{R}^{2^2} = \mathbf{R}^4 = \mathbf{R}^1 \oplus \mathbf{R}^2 \oplus \mathbf{R}^1 = \mathbf{R} \cdot B_0 + \underbrace{[\mathbf{R} \cdot B_1 + \mathbf{R} \cdot B_2]}_{\mathbf{R}^2} + \mathbf{R} \cdot B_1 B_2,$$
$$\mathbf{R}^{2^3} = \mathbf{R}^8 = \mathbf{R}^1 \oplus \mathbf{R}^3 \oplus \mathbf{R}^3 \oplus \mathbf{R}^1 = \mathbf{R} \cdot B_0 + \underbrace{[\mathbf{R} \cdot B_1 + \mathbf{R} \cdot B_2 + \mathbf{R} \cdot B_3]}_{\mathbf{R}^3}$$
$$\oplus \underbrace{[\mathbf{R} \cdot B_1 B_2 + \mathbf{R} \cdot B_1 B_3 + \mathbf{R} \cdot B_2 B_3]}_{\mathbf{R}^3} \oplus \mathbf{R} \cdot B_1 B_2 B_3.$$

Every element of $\mathbf{R}^{2^t}$ has the following representation. Let $\mathbf{b} = (b_1, b_2, \ldots, b_t) \in \mathbf{B}_2^t$ be an arbitrary t-bit binary vector, where $b_i \in \mathbf{B}_2 = \{0, 1\}$ and $\mathbf{B}_2^t$ is the t-D Boolean. Let us introduce 2^t elements $\mathbf{B}^{\mathbf{b}} := B_1^{b_1} B_2^{b_2} \ldots B_t^{b_t}$, where $\mathbf{b} = (b_1, b_2, \ldots, b_t) \in \mathbf{B}_2^t$. Let $w(\mathbf{b}) = b_1 + b_2 + \cdots + b_t$ be the weight of $\mathbf{b} = (b_1, b_2, \ldots, b_K) \in \mathbf{B}_2^K$. Elements $B^{\mathbf{b}} = B_1^{b_1} B_2^{b_2} \ldots B_t^{b_t}$ form a basis (i.e., FSoI) of 2^t-D space. For all $\mathcal{C} \in \mathbf{R}^{2^t}$, we have the following hypercomplex representations

$$\mathcal{C} := \sum_{\mathbf{b} \in \mathbf{B}_2^k} a_{\mathbf{b}} \mathbf{B}^{\mathbf{b}} = \sum_{s=0}^{K} \sum_{w(\mathbf{b})=s} a_{\mathbf{b}} \mathbf{B}^{\mathbf{b}}$$

$$= \sum_{w(\mathbf{b})=0} a_{\mathbf{b}} \mathbf{B}^{\mathbf{b}} + \left(\sum_{w(\mathbf{b})=1} a_{\mathbf{b}} \mathbf{B}^{\mathbf{b}} + \sum_{w(\mathbf{b})=2} a_{\mathbf{b}} \mathbf{B}^{\mathbf{b}} + \cdots + \sum_{w(\mathbf{b})=s} a_{\mathbf{b}} \mathbf{B}^{\mathbf{b}} + \cdots + \sum_{w(\mathbf{b})=t} a_{\mathbf{b}} \mathbf{B}^{\mathbf{b}} \right)$$

$$= \mathbf{Sc}(\mathcal{C}) + \left(\mathbf{Vec}^1(\mathcal{C}) + \mathbf{Vec}^2(\mathcal{C}) + \cdots + \mathbf{Vec}^s(\mathcal{C}) + \cdots + \mathbf{Vec}^t(\mathcal{C}) \right), \quad (1.3)$$

where

- $\mathbf{Sc}(\mathcal{C}) = \sum\limits_{w(\mathbf{b})=0} a_{\mathbf{b}} \mathbf{B}^{\mathbf{b}} = a_0 B_0 \in \mathbf{R}^{C_t^0}$ is the *scalar part* of the Clifford number c,
- $\mathbf{Vec}^1(\mathcal{C}) = \sum\limits_{w(\mathbf{b})=1} a_{\mathbf{b}} \mathbf{B}^{\mathbf{b}} \in \mathbf{R}^{C_t^1}$ is the pure vector part,
- $\mathbf{Vec}^2(\mathcal{C}) = \sum\limits_{w(\mathbf{b})=2} a_{\mathbf{b}} \mathbf{B}^{\mathbf{b}} \in \mathbf{R}^{C_t^2}$ is the bivector part,
- ...,
- $\mathbf{Vec}^s(\mathcal{C}) = \sum\limits_{w(\mathbf{b})=s} a_{\mathbf{b}} \mathbf{B}^{\mathbf{b}} \in \mathbf{R}^{C_t^s}$ is the S-vector part,
- ..., and
- $\mathbf{Vec}^t(\mathcal{C}) = \sum\limits_{w(\mathbf{b})=t} a_{\mathbf{b}} \mathbf{B}^{\mathbf{b}} \in \mathbf{R}^{C_t^t}$ is the t-vector part.

Imaginary units $\mathbf{B}^{\mathbf{b}} := B_1^{b_1} B_2^{b_2} \cdots B_t^{b_t}$ with $w(\mathbf{b}) = 1$ (i.e., $B_1, \ldots, B_t$) are called the *basic imaginaries* and units $\mathbf{B}^{\mathbf{b}} := B_1^{b_1} B_2^{b_2} \cdots B_t^{b_t}$ with $w(\mathbf{b}) > 1$ are called the *derivative imaginaries*.

Binary vector $\mathbf{b} = (b_1, b_2, \ldots, b_t) \in \mathbf{B}_2^t$ can be considered as binary code of an integer $\mathbf{b} = (b_1, b_2, \ldots, b_t) \in [0, 2^t - 1]$, hence $\mathbf{B}^{\mathbf{b}} \in \{\mathbf{B}^0, \mathbf{B}^1, \ldots, \mathbf{B}^{2^t-1}\}$ and

$$B_0 = \mathbf{B}^0 = \mathbf{B}^{(0,0,0,\ldots,0)}, \ B_1 = \mathbf{B}^1 = \mathbf{B}^{(1,0,0,\ldots,0)}, \ B_2 = \mathbf{B}^2 = \mathbf{B}^{(0,1,0,0\ldots,0)},$$

$$B_3 = B_1 B_2 = \mathbf{B}^{(1,1,0,\ldots,0)}, \ldots, \mathbf{B}^{2^t-1} = B_1 B_2 \ldots B_t = \mathbf{B}^{(1,1,1,\ldots,1)}.$$

Definition 1 Elements of the form (1.3) are called the physical (*if $t = n$*) or hyperspectral (*if $t = K$*) Clifford numbers.

If $\mathcal{A} = \sum\limits_{\mathbf{b} \in \mathbf{B}_2^t} a_{\mathbf{b}} \mathbf{B}^{\mathbf{b}}$, $\mathcal{B} = \sum\limits_{\mathbf{c} \in \mathbf{B}_2^t} b_{\mathbf{c}} \mathbf{B}^{\mathbf{c}} \in \mathbf{R}^{2^t}$ are two Clifford numbers, then their product is

$$\mathcal{C} = \mathcal{A}\mathcal{B} := \left(\sum_{\mathbf{b} \in \mathbf{B}_2^t} a_{\mathbf{b}} \mathbf{B}^{\mathbf{b}} \right) \cdot \left(\sum_{\mathbf{c} \in \mathbf{B}_2^t} b_{\mathbf{c}} \mathbf{B}^{\mathbf{c}} \right) = \left(\sum_{\mathbf{b} \in \mathbf{B}_2^t} \sum_{\mathbf{c} \in \mathbf{B}_2^t} a_{\mathbf{b}} b_{\mathbf{c}} \mathbf{B}^{\mathbf{b}} \mathbf{B}^{\mathbf{c}} \right)$$

$$= \left(\sum_{\mathbf{b} \in \mathbf{B}_2^t} \sum_{\mathbf{c} \in \mathbf{B}_2^t} (-1)^{\langle \mathbf{b} | R | \mathbf{c} \rangle} a_{\mathbf{b}} b_{\mathbf{c}} \mathbf{B}^{\mathbf{b} \oplus \mathbf{c}} \right)$$

$$= \left(\sum_{\mathbf{d} \in \mathbf{B}_2^t} \sum_{\mathbf{b} \in \mathbf{B}_2^t} (-1)^{\langle \mathbf{b} | R | \mathbf{d} \oplus \mathbf{b} \rangle} a_{\mathbf{b}} b_{\mathbf{d} \oplus \mathbf{b}} \mathbf{B}^{\mathbf{d}} \right) = \left(\sum_{\mathbf{d} \in \mathbf{B}_2^t} \sum_{\mathbf{b} \in \mathbf{B}_2^t} (-1)^{\langle \mathbf{b} | R | \mathbf{d} \oplus \mathbf{b} \rangle} a_{\mathbf{b}} b_{\mathbf{d} \oplus \mathbf{b}} I^{\mathbf{d}} \right)$$

$$= \sum_{\mathbf{d} \in \mathbf{B}_2^K} c_{\mathbf{d}} \mathbf{B}^{\mathbf{d}}, \tag{1.4}$$

where

$$c_{\mathbf{d}} = \sum_{\mathbf{b} \in \mathbf{B}_2^t} (-1)^{\langle \mathbf{b} | R | \mathbf{d} \oplus \mathbf{b} \rangle} a_{\mathbf{b}} b_{\mathbf{d} \oplus \mathbf{b}}, \ R := \begin{bmatrix} 1 & & & & \\ 1 & 1 & & & \\ 1 & 1 & 1 & & \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}. \tag{1.5}$$

There are 3^t possibilities for $B_s^2 = +1, 0, -1, \forall s = 1, 2, \ldots, t$. Every possibility generates an algebra. Therefore, the space $\mathbf{R}^{2^t}$ with 3^t rules of the multiplication forms 3^t different 2^t-D algebras, which are called the *Clifford algebras*. We denote these algebras by $Alg_{2^t}^{(u,v,w)}(\mathbf{R}|1, B_1, \ldots, B_t)$, or Alg_{2^t} if $B_1, \ldots, B_K$ and u, v, w are fixed.

If $t = n$ and $B_1 = I_1, B_2 = I_2, \ldots, B_n = I_n$, then we have 3^n different 2^n-D spatial Clifford algebras $Alg_{2^n}^{Sp(p,q,r)}(\mathbf{R}|1, I_1, \ldots, I_n)$. For $t = K$ and $B_1 = J_1, B_2 = J_2, \ldots, B_K = J_K$, we have 3^K different 2^K-D visual Clifford algebras $Alg_{2^K}^{Vis(u,v,w)}(\mathbf{R}|1, J_1, \ldots, J_K)$. Spatial and visual Clifford algebras can have different signatures: $(u, v, w) \neq (p, q, r)$.

In Alg_{2^t}, we introduce the conjugation operation which maps every Clifford number $\mathcal{C} := c_0 J_0 + \sum_{\mathbf{b} \neq 0} c_{\mathbf{b}} \mathbf{B}^{\mathbf{b}}$ to the number $\overline{\mathcal{C}} := c_0 J_0 - \sum_{\mathbf{b} \neq 0} c_{\mathbf{b}} \mathbf{B}^{\mathbf{b}}$. The algebras Alg_{2^t} are transformed into 2^t-D pseudo-metric spaces designed as $\mathbf{Geo}_{2^t}$, if the pseudo-distance between two Clifford numbers $\mathcal{A}$ and $\mathcal{B}$ is defined by

$$\rho(\mathcal{A}, \mathcal{B}) = |\mathcal{A} - \mathcal{B}| = \sqrt{(\mathcal{A} - \mathcal{B})\overline{(\mathcal{A} - \mathcal{B})}} = |\mathcal{U}| = \mathcal{U}\overline{\mathcal{U}} = \sqrt{\sum_{\mathbf{b} \in \mathbf{B}_2^t} (-1)^{w(\mathbf{b})} u_{\mathbf{b}} (B^{\mathbf{b}})^2}, \tag{1.6}$$

where $\mathcal{A} - \mathcal{B} = \mathcal{U} = \sum_{\mathbf{b} \in \mathbf{B}_2^t} u_{\mathbf{b}} B^{\mathbf{b}}$.

Subspaces of pure vector Clifford numbers $\mathbf{R}^t = \{\mathbf{x} | \mathbf{x} = \sum_{w(\mathbf{b})=1} x_{\mathbf{b}} \mathbf{B}^{\mathbf{b}} = x_1 B_1 + \cdots + x_t B_t\}$ in this case is transformed into different t-D pseudo-metric spaces $\mathbf{R}^t \rightarrow$

$\mathbf{Geo}_t = \langle\langle \mathbf{R}^t, \rho(\mathcal{A}, \mathcal{B})\rangle\rangle$, since pseudo-metrics $\rho(\mathcal{A}, \mathcal{B})$ constructed in $\mathbf{R}^{2^t}$ induce corresponding pseudo-metrics in $\mathbf{R}^t$. The pseudo-metric spaces $\mathbf{Geo}_t$ are *the Cayley-Klein geometries* [4–9]. Obviously,

$$\mathbf{R}^n \to \mathbf{Geo}_n^{Sp(p,q,r)} = \langle\langle \mathbf{R}^n, \rho^{Sp(p,q,r)}\rangle\rangle, \mathbf{V}^K \to \mathbf{Geo}_K^{Vis(u,v,w)} = \langle\langle \mathbf{V}^K, \rho^{Vis(u,v,w)}\rangle\rangle,$$

where $\mathbf{Geo}_n^{Sp(p,q,r)}$ is a spatial geometry for the physical space $\mathbf{R}^n$ with a metric $\rho^{Sp(p,q,r)}$ and $\mathbf{Geo}_K^{Vis(u,v,w)}$ is a geometry for the visual space $\mathbf{V}^K$ (in the Visual Cortex) with a metric $\rho^{Vis(u,v,w)}$.

Every algebra $Alg_{2^K}^{Vis(u,v,w)}$ can be decomposed as

$$Alg_{2^t} = {}^0Alg_{2^t} + {}^1Alg_{2^t} = \sum_{s=0}^{]t/2[} \mathbf{Vec}^{2s} + \sum_{s=0}^{]t/2[} \mathbf{Vec}^{2s+1}, \tag{1.7}$$

where ${}^0Alg_{2^t} = \sum_{s=0}^{]t/2[} \mathbf{Vec}^{2s}$ and ${}^1Alg_{2^t} = \sum_{s=0}^{]t/2[} \mathbf{Vec}^{2s+1}$ are even and odd parts of Alg_{2^t}. We will see that all orthogonal transforms of "small" perceptual t-D space $\mathbf{R}^t$ live in ${}^0Alg_{2^t}$. Clifford numbers $\mathcal{E} \in {}^0Alg_{2^t}$ of unit modulus represent the rotation group for the corresponding space $\mathbf{R}^t$ which is called *spinor group* and is denoted by $\mathbf{Spin}(Alg_{2^t})$. We know that complex numbers and quaternions of unit modulus have the following forms:

$$\mathbf{e}_0 = e^{i\varphi} = \cos\varphi + i \cdot \sin\varphi, \quad Q_0 = e^{\mathbf{u}_0\varphi} = \cos\varphi + \mathbf{u}_0 \sin\varphi, \tag{1.8}$$

where $\cos\varphi$ and $\sin\varphi$ are trigonometric functions in the corresponding $2-D$ geometries, respectively, φ is a rotation angle around vector-valued quaternion $\mathbf{u}_0$ of unit modulus ($|\mathbf{u}_0| = 1$, $\mathbf{u}_0 = -\bar{\mathbf{u}}_0$). In general case, Clifford spinors $\mathcal{E}_0 \in \mathbf{Spin}(Alg_{2^t})$ with unit modulus have the same form

$$\mathcal{E}_0 = e^{\mathcal{U}\varphi} = \cos\varphi + \mathcal{U} \cdot \sin\varphi \in \mathbf{Spin}(Alg_{2^t}), \tag{1.9}$$

where $\mathcal{U} = \sum_{k_1=1}^t \sum_{k_2=1}^t u_{k_1 k_2} B_{k_1} B_{k_2} = \sum_{w(\mathbf{b})=2} u_{\mathbf{b}} \mathbf{B}^{\mathbf{b}} \in \mathbf{Vec}^2$ is a unit bivector ($\mathcal{U}^2 = -1$), and φ is a rotation angle.

Theorem 1 [5–10]. *The transforms*

$$Q' = e^{\mathcal{U}_1\varphi_1/2}Q, \; Q'' = Qe^{-\mathcal{U}_2\varphi_2/2}, \; Q''' = e^{\mathcal{U}_1\varphi_1/2}Qe^{-\mathcal{U}_2\varphi_2/2} \tag{1.10}$$

are the rotations of the "big" space $\mathbf{Geo}_{2^t}$, *where* $Q, Q', Q'', Q''' \in \mathbf{Geo}_{2^t}$ *and* $e^{\mathcal{U}_1\varphi_1/2}, e^{-\mathcal{U}_2\varphi_2/2} \in \mathbf{Spin}(Alg_{2^t})$. *They form groups* $\mathbf{Rot}_L(\mathbf{Geo}_{2^t})$, $\mathbf{Rot}_R(\mathbf{Geo}_{2^t})$, $\mathbf{Rot}_{LR}(\mathbf{Geo}_{2^t})$ *of left, right, and double-side rotations of "big" space* $\mathbf{Geo}_{2^t}$ *and transforms*

$$\mathbf{x}' = e^{\mathcal{U}_1\varphi_1/2}\mathbf{x}e^{-\mathcal{U}_2\varphi_2/2}, \tag{1.11}$$

where $\mathbf{x}$, $\mathbf{x}' \in \mathbf{Geo}_t$, and $e^{\mathcal{U}_1\varphi_1/2}$, $e^{-\mathcal{U}_2\varphi_2/2} \in \mathbf{Spin}(Alg_{2^t})$ are rotations of "small" space $\mathbf{Geo}_t$. They form group of rotations $\mathbf{Rot}_{LR}(\mathbf{Geo}_t)$.

Theorem 2 [5–10]. *The transforms*

$$\mathcal{Q}' = e^{\mathcal{U}_1\varphi_1/2}\mathcal{Q} + \mathcal{P}, \quad \mathcal{Q}'' = \mathcal{Q}e^{-\mathcal{U}_2\varphi_2/2} + \mathcal{P}, \quad \mathcal{Q}''' = e^{\mathcal{U}_1\varphi_1/2}\mathcal{Q}e^{-\mathcal{U}_2\varphi_2/2} + \mathcal{P}, \quad (1.12)$$

form three groups of left, right, and double-side multicolor motions $\mathbf{Mov}_L(\mathbf{Geo}_{2^t})$, $\mathbf{Mov}_R(\mathbf{Geo}_{2^t})$, $\mathbf{Mov}_{LR}(\mathbf{Geo}_{2^t})$ *of "big" space* $\mathbf{Geo}_{2^t}$.

Theorem 3 *Every motion of "small"* $\mathbf{Geo}_t$-space *is represented in the following form*

$$\mathbf{z}' = e^{\mathcal{U}\varphi/2} \cdot \mathbf{z} \cdot e^{-\mathcal{U}\varphi/2} + \mathbf{w}, \quad \mathbf{z}, \mathbf{z}', \mathbf{w} \in \mathbf{Geo}_t. \qquad (1.13)$$

1.3 Cliffordean Models of Multi-channel Images

In classical approach multi-channel images $\mathbf{f}(\mathbf{x})$ are considered as n-D K-component (vector-valued) functions

$$\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_K(\mathbf{x})) : \mathbf{R}^n \to \mathbf{V}^K \qquad (1.14)$$

with values into K-D perceptual spaces $\mathbf{V}^K$, where $\mathbf{x} \in \mathbf{R}^n$, $n = 2, 3, \ldots$. Now we can interpret multi-channel images $\mathbf{f}(\mathbf{x})$ as $Alg_{2^K}^{Vis(u,v,w)}(\mathbf{R}|1, J_1, \ldots, J_K)$-valued signal of hypercomplex variables $\mathbf{x} \in Alg_{2^n}^{Sp(p,q,r)}(\mathbf{R}|1, I_1, \ldots, I_n)$:

$$\mathbf{f}(\mathbf{x}) : Alg_{2^n}^{Sp(p,q,r)}(\mathbf{R}|1, I_1, \ldots, I_n) \to Alg_{2^K}^{Vis(u,v,w)}(\mathbf{R}|1, J_1, \ldots, J_K). \quad (1.15)$$

From mathematical point of view, there are four algebraic models for *multi-channel* images, since we can identify $\mathbf{R}^n$ with $\mathbf{Vec}^1\left(Alg_{2^n}^{Sp(p,q,r)}\right)$, or with $Alg_{2^n}^{Sp(p,q,r)}$, a $\mathbf{V}^K$ with $\mathbf{Vec}^1\left(Alg_{2^K}^{Vis(u,v,w)}\right)$ or with $Alg_{2^K}^{Vis(u,v,w)}$.
The first model is

$$\mathbf{f}(\mathbf{x}) : \mathbf{Vec}^1\left(Alg_{2^n}^{Sp(p,q,r)}\right) \to \mathbf{Vec}^1\left(Alg_{2^K}^{Vis(u,v,w)}(\mathbf{R}|1, J_1, \ldots, J_K)\right),$$

where we used only basic spatial and multi-channel imaginaries,

$$\mathbf{f}\left(\sum_{s=1}^{n} I_s x_s\right) = \sum_{b=1}^{K} J_b \cdot f_b\left(\sum_{s=1}^{n} I_s x_s\right), \qquad (1.16)$$

i.e.,

$$\mathbf{f}(I_1 x_1 + I_2 x_2 + \cdots + I_n x_n)$$
$$= J_1 \cdot f_1(I_1 x_1 + I_2 x_2 + \cdots + I_n x_n) + J_2 \cdot f_2(I_1 x_1 + I_2 x_2 + \cdots + I_n x_n)$$
$$+ \cdots + J_K \cdot f_K(I_1 x_1 + I_2 x_2 + \cdots + I_n x_n).$$

The second model is

$$\mathbf{f}(\mathbf{x}) : \mathbf{Vec}^1\left(Alg_{2^n}^{Sp(p,q,r)}\right) \to Alg_{2^m}^{Vis(u,v,w)}(\mathbf{R}|1, J_1, \ldots, J_m),$$

where we used basic spatial and the full set of multi-channel imaginaries

$$\mathbf{f}\left(\sum_{s=1}^{n} I_s x_s\right) = \sum_{\mathbf{b} \in \mathbf{B}_2^m} \mathbf{J}^{\mathbf{b}} \cdot f_{\mathbf{b}}\left(\sum_{s=1}^{n} I_s x_s\right)$$

$$= \sum_{(b_1,b_2,\ldots,b_m) \in \mathbf{B}_2^m} J_1^{b_1} J_2^{b_2} \cdot \cdots \cdot J_m^{b_m} \cdot f_{(b_1,b_2,\ldots,b_m)}\left(\sum_{s=1}^{n} I_s x_s\right), \quad (1.17)$$

Here we suppose that $K = 2^m$ and $\mathbf{f}(\mathbf{x})$ in (1.17) is a 2^m-channel image.
The third model

$$\mathbf{f}(\mathbf{x}) : Alg_{2^n}^{Sp(p,q,r)}(\mathbf{R}|1, I_1, \ldots, I_n) \to \mathbf{Vec}^1\left(Alg_{2^K}^{Vis(u,v,w)}(\mathbf{R}|1, J_1, \ldots, J_K)\right),$$

where we used the full set of spatial and basic set of multi-channel imaginaries

$$\mathbf{f}\left(\sum_{\mathbf{s} \in \mathbf{B}_2^l} \mathbf{I}^{\mathbf{s}} \cdot x_{\mathbf{s}}\right) = \sum_{k=1}^{K} J_k \cdot f_k\left(\sum_{\mathbf{s} \in \mathbf{B}_2^l} \mathbf{I}^{\mathbf{s}} \cdot x_{\mathbf{s}}\right), \quad (1.18)$$

i.e.,

$$\mathbf{f}\left(\sum_{(s_1,s_2,\ldots,s_l) \in \mathbf{B}_2^l} I_1^{s_1} I_2^{s_2} \cdot \cdots \cdot I_l^{s_l} x_{(s_1,s_2,\ldots s_l)}\right)$$

$$= \sum_{k=1}^{K} J_k \cdot f_k\left(\sum_{(s_1,s_2,\ldots,s_l) \in \mathbf{B}_2^l} I_1^{s_1} I_2^{s_2} \cdot \cdots \cdot I_l^{s_l} x_{(s_1,s_2,\ldots,s_l)}\right)$$

Here we suppose that $n = 2^l$ and $\mathbf{f}(\mathbf{x})$ in (1.18) is a K-channel image.
The fourth model

$$\mathbf{f}(\mathbf{x}) : Alg_{2^n}^{Sp(p,q,r)}(\mathbf{R}|1, I_1, \ldots, I_n) \to Alg_{2^K}^{Vis(u,v,w)}(\mathbf{R}|1, J_1, \ldots, J_K).$$

where we used the full set of spatial and multi-channel imaginaries, i.e.,

$$\mathbf{f}\left(\sum_{\mathbf{s}\in\mathbf{B}_2^l}\mathbf{I}^{\mathbf{s}}\cdot x_{\mathbf{s}}\right)=\sum_{\mathbf{b}\in\mathbf{B}_2^m}\mathbf{J}^{\mathbf{b}}\cdot f_{\mathbf{b}}\left(\sum_{\mathbf{s}\in\mathbf{B}_2^l}\mathbf{I}^{\mathbf{s}}\cdot x_{\mathbf{s}}\right),\ \text{or}$$

$$\mathbf{f}\left(\sum_{(s_1,s_2,\ldots,s_l)\in\mathbf{B}_2^l}I_1^{s_1}I_2^{s_2}\cdot\cdots\cdot I_l^{s_l}x_{(s_1,s_2,\ldots,s_l)}\right)$$

$$=\sum_{(b_1,b_2,\ldots,b_m)\in\mathbf{B}_2^m}J_1^{b_1}J_2^{b_2}\cdot\cdots\cdot J_m^{b_m}\cdot f_{(b_1,b_2,\ldots,b_m)}\left(\sum_{(s_1,s_2,\ldots,s_l)\in\mathbf{B}_2^l}I_1^{s_1}I_2^{s_2}\cdot\cdots\cdot I_l^{s_l}x_{(s_1,s_2,\ldots,s_l)}\right).$$

$$(1.19)$$

Here we suppose that $n=2^l$, $K=2^m$, and $\mathbf{f}(\mathbf{x})$ in (1.19) is a 2^l-D and 2^m-channel image.

Example 2 Let

$$\mathbf{f}(\mathbf{x})=[f_Y(\mathbf{x}),f_R(\mathbf{x}),f_G(\mathbf{x}),f_B(\mathbf{x})]\qquad(1.20)$$

be a 4-channel image retinal images (gray-level channel plus three color RGB channels). We can consider four models for this image. For this purpose, we introduce the following physical and multi-channel Clifford algebras:

$$Alg_{2^1}^{Sp(p,q,r)}(\mathbf{R}|1,I_1),\ Alg_{2^2}^{Sp(p,q,r)}(\mathbf{R}|1,I_1,I_2),$$
$$Alg_{2^2}^{Vis(u,v,w)}(\mathbf{R}|1,J_1,J_2),\ Alg_{2^4}^{Vis(u,v,w)}(\mathbf{R}|1,J_1,J_2,J_3,J_4).$$

We are going to consider two algebraic representations for space variables $\mathbf{x}=(x,y)$:

$$\mathbf{x}=(x,y)=xI_0+yI_1\in Alg_{2^1}^{Sp(p,q,r)}(\mathbf{R}|1,I_1)=\mathbf{R}I_0+\mathbf{R}I_1,$$
$$\mathbf{x}=(x,y)=xI_1+yI_2\in Vec^1\left(Alg_{2^2}^{Sp(p,q,r)}(\mathbf{R}|1,I_1,I_2)\right)$$
$$=\mathbf{R}I_1+\mathbf{R}I_2\subset\mathbf{R}I_0+\underbrace{(\mathbf{R}I_1+\mathbf{R}I_2)}_{Vec^1}+\mathbf{R}I_1I_2$$

and two algebraic representations for perceptual variables $\mathbf{f}=[f_Y,f_R,f_G,f_B]$:

$$\mathbf{f}=(f_Y,f_R,f_G,f_B)\in Alg_{2^2}^{Vis(u,v,w)}(\mathbf{R}|J_0=1,J_1,J_1)$$
$$=\mathbf{R}J_0+\mathbf{R}J_1+\mathbf{R}J_2+\mathbf{R}J_1J_2=\mathbf{R}J_Y+\mathbf{R}J_R+\mathbf{R}J_G+\mathbf{R}J_B,$$

where $J_0\equiv1$, $J_R=J_1$, $J_G=J_2$, $J_B=J_1J_2$ are gray, red, green, and blue hyper-imaginary units, respectively. Hence,

$$f=(f_Y,f_R,f_G,f_B)\in Vec^1\left\{Alg_{2^4}^{Vis(u,v,w)}(\mathbf{R}|J_0=1,J_1,J_2,J_3,J_4)\right\}$$

$$= \underbrace{\mathbf{R}J_1 + \mathbf{R}J_2 + \mathbf{R}J_3 + \mathbf{R}J_4}_{4D-Vec^1}$$

where $J_Y = J_1$, $J_R = J_2$, $J_G = J_3$, $J_B = J_4$. Here

$$Alg_{2^4}^{Vis(u,v,w)}(\mathbf{R}|J_0 = 1, J_1, J_2, J_3, J_4)$$

$$= {}^1 \underbrace{\mathbf{R}J_0}_{1D-Sc} + \underbrace{\mathbf{R}J_1 + \mathbf{R}J_2 + \mathbf{R}J_3 + \mathbf{R}J_4}_{4D-Vec^1}$$

$$+ \underbrace{\mathbf{R}J_1J_2 + \mathbf{R}J_1J_3 + \mathbf{R}J_1J_4 + \mathbf{R}J_2J_3 + \mathbf{R}J_2J_4 + \mathbf{R}J_3J_4}_{6D-Vec^2}$$

$$+ \underbrace{\mathbf{R}J_1J_2J_3 + \mathbf{R}J_1J_2J_4 + \mathbf{R}J_1J_3J_4 + \mathbf{R}J_2J_3J_4}_{4D-Vec^3} + \underbrace{\mathbf{R}J_1J_2J_3J_4}_{1D-Vec^4}$$

Now we can consider the following four algebraic models of image (1.20).

The first model:

$$\mathbf{f}(\mathbf{x}) : Vec^1\left(Alg_{2^2}^{Sp}(\mathbf{R}|I_0, I_1, I_2)\right) \rightarrow Vec^1\left(Alg_{2^4}^{Vis}(\mathbf{R}|J_0, J_1, J_2, J_3, J_4)\right)$$

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(xI_1 + yI_2)$$

$$= f_Y(xI_1 + yI_2) \cdot J_1 + f_R(xI_1 + yI_2) \cdot J_2$$

$$+ f_G(xI_1 + yI_2) \cdot J_3 + f_B(xI_1 + yI_2) \cdot J_4.$$

The second model:

$$\mathbf{f}(\mathbf{x}) : Vec^1\left\{Alg_{2^1}^{Sp}(\mathbf{R}|I_0, I_1, I_2)\right\} \rightarrow Alg_{2^2}^{Vis}(\mathbf{R}|J_0, J_1, J_2),$$

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(xI_1 + yI_2)$$

$$= f_Y(xI_1 + yI_2)J_0 + f_R(xI_1 + yI_2)J_1$$

$$+ f_G(xI_1 + yI_2)J_2 + f_B(xI_1 + yI_2)J_1J_2.$$

The third model:

$$\mathbf{f}(\mathbf{x}) : Alg_{2^1}^{Sp}(\mathbf{R}|I_0, I_1) \rightarrow Vec^1\left(Alg_{2^2}^{Vis}(\mathbf{R}|J_0, J_1, J_2, J_3, J_4)\right),$$

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(xI_0 + yI_1)$$

$$= f_Y(xI_0 + yI_1) \cdot J_1 + f_R(xI_0 + yI_1) \cdot J_2$$

$$+ f_G(xI_0 + yI_1) \cdot J_3 + f_B(xI_0 + yI_1) \cdot J_4.$$

The fourth model:

$$\mathbf{f}(\mathbf{x}) : Alg_{2^1}^{Sp}(\mathbf{R}|I_0, I_1) \rightarrow Alg_{2^2}^{Vis}(\mathbf{R}|J_0, J_1, J_2),$$

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(xI_0 + yI_1)$$

$$= f_Y(xI_0 + yI_1)J_0 + f_R(xI_0 + yI_1)J_1$$
$$+ f_G(xI_0 + yI_1)J_2 + f_B(xI_0 + yI_1)J_1J_2.$$

Geometric properties of these images depend on signatures $Sp(p, q, r)$ and $Vis(u, v, w)$.

Example 3 Binocular vision is a vision in which both eyes are used together. Each eye (right and left) views the visual world from slightly different horizontal positions. Each eye's image differs from the other (see Fig. 1.1), i.e.,

$$\mathbf{f}_r(\mathbf{x}) = \left(f_1^r(\mathbf{x}), f_2^r(\mathbf{x}), \dots, f_K^r(\mathbf{x})\right), \quad \mathbf{f}^l(\mathbf{x}) = \left(f_1^l(\mathbf{x}), f_2^l(\mathbf{x}), \dots, f_K^l(\mathbf{x})\right). \quad (1.21)$$

Objects at different distances from the eyes project images in the two eyes that differ in their horizontal positions, giving the depth cue of horizontal disparity. Depth perception is commonly referred to as stereopsis. Stereopsis appears to be processed in the visual cortex in binocular cells having receptive fields in different horizontal positions in the two eyes. Such a cell is active only when its preferred stimulus is in the correct position in the left eye and in the correct position in the right eye, making it a disparity detector.

The two eyes can influence each other. We take into account this influence using a new hyper-imaginary binocular unit B and construct binocular image as signal

$$\mathbf{f}^{Bin}(\mathbf{x}) = \mathbf{f}^r(\mathbf{x}) + \mathbf{f}^l(\mathbf{x}) \cdot B \quad (1.22)$$

Fig. 1.1 When two eyes (or cameras) view a 3-D scene from two distinct positions, there are a number of geometric relations between the 3-D points and their projections onto the 2-D images that lead to constraints between the image points

Left view   Right view

$\mathbf{f}^l(\mathbf{x})$ $\mathbf{f}^r(\mathbf{x})$
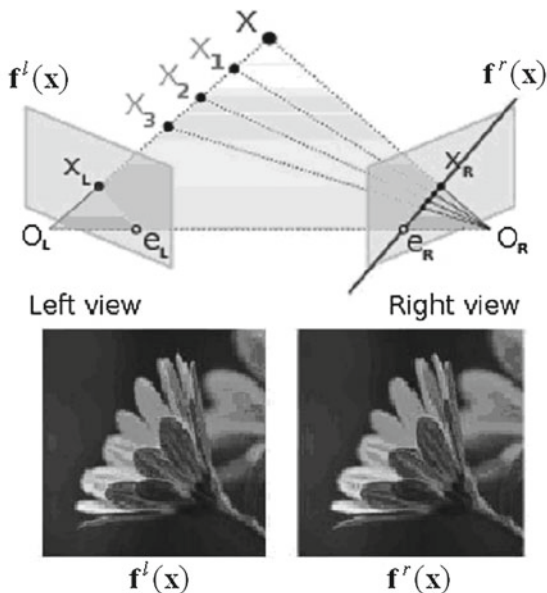
with $B^2 = \delta_3 = +1, 0, -1$. We obtain the following four models of binocular images.

The first model:

$$\mathbf{f}^{Bin}(\mathbf{x}) : \mathbf{Vec}^1\left(Alg_{2^n}^{Sp(p,q,r)}\right) \to \mathbf{Vec}^1\left(Alg_{2^{K+1}}^{Vis(u,v,w)}(\mathbf{R}|1, J_1, \ldots, J_K, B)\right),$$

$$\mathbf{f}^{Bin}\left(\sum_{s=1}^{n} I_s x_s\right) = \sum_{b=1}^{K} J_b \cdot f_b^l\left(\sum_{s=1}^{n} I_s x_s\right) + \sum_{b=1}^{K} J_b \cdot B \cdot f_b^r\left(\sum_{s=1}^{n} I_s x_s\right) \quad (1.23)$$

$$= \sum_{b=1}^{K} J_b \cdot \left[f_b^l\left(\sum_{s=1}^{n} I_s x_s\right) + B \cdot f_b^r\left(\sum_{s=1}^{n} I_s x_s\right)\right]. \quad (1.24)$$

The second model:

$$\mathbf{f}^{Bin}(\mathbf{x}) : \mathbf{Vec}^1\left(Alg_{2^n}^{Sp(p,q,r)}\right) \to Alg_{2^m}^{Vis(u,v,w)}(\mathbf{R}|1, J_1, \ldots, J_m, B),$$

$$\mathbf{f}^{Bin}\left(\sum_{s=1}^{n} I_s x_s\right) = \sum_{\mathbf{b} \in \mathbf{B}_2^m} \mathbf{J}^{\mathbf{b}} \cdot f_{\mathbf{b}}^l\left(\sum_{s=1}^{n} I_s x_s\right) + \sum_{\mathbf{b} \in \mathbf{B}_2^m} \mathbf{J}^{\mathbf{b}} \cdot B \cdot f_{\mathbf{b}}^r\left(\sum_{s=1}^{n} I_s x_s\right) \quad (1.25)$$

$$= \sum_{\mathbf{b} \in \mathbf{B}_2^m} \mathbf{J}^{\mathbf{b}} \cdot \left[f_{\mathbf{b}}^l\left(\sum_{s=1}^{n} I_s x_s\right) + B \cdot f_{\mathbf{b}}^r\left(\sum_{s=1}^{n} I_s x_s\right)\right]. \quad (1.26)$$

The third model:

$$\mathbf{f}^{Bin}(\mathbf{x}) : Alg_{2^n}^{Sp(p,q,r)}(\mathbf{R}|1, I_1, \ldots, I_n) \to \mathbf{Vec}^1\left(Alg_{2^K}^{Vis(u,v,w)}(\mathbf{R}|1, J_1, \ldots, J_K, B)\right),$$

$$\mathbf{f}^{Bin}\left(\sum_{\mathbf{s} \in \mathbf{B}_2^l} \mathbf{I}^{\mathbf{s}} \cdot x_{\mathbf{s}}\right) = \sum_{k=1}^{K} J_k \cdot f_k^l\left(\sum_{\mathbf{s} \in \mathbf{B}_2^l} \mathbf{I}^{\mathbf{s}} \cdot x_{\mathbf{s}}\right) + \sum_{k=1}^{K} J_k \cdot B \cdot f_k^r\left(\sum_{\mathbf{s} \in \mathbf{B}_2^l} \mathbf{I}^{\mathbf{s}} \cdot x_{\mathbf{s}}\right)$$

$$(1.27)$$

$$= \sum_{k=1}^{K} J_k \cdot \left[f_k^l\left(\sum_{\mathbf{s} \in \mathbf{B}_2^l} \mathbf{I}^{\mathbf{s}} \cdot x_{\mathbf{s}}\right) + B \cdot f_k^r\left(\sum_{\mathbf{s} \in \mathbf{B}_2^l} \mathbf{I}^{\mathbf{s}} \cdot x_{\mathbf{s}}\right)\right]. \quad (1.28)$$

The fourth model:

$$\mathbf{f}^{Bin}(\mathbf{x}) : Alg_{2^n}^{Sp(p,q,r)}(\mathbf{R}|1, I_1, \ldots, I_n) \to Alg_{2^K}^{Vis(u,v,w)}(\mathbf{R}|1, J_1, \ldots, J_K, B),$$

$$\mathbf{f}^{Bin}\left(\sum_{\mathbf{s}\in\mathbf{B}_2^l}\mathbf{I^s}\cdot x_{\mathbf{s}}\right)=\sum_{\mathbf{b}\in\mathbf{B}_2^m}\mathbf{J^b}\cdot f_{\mathbf{b}}^l f_{\mathbf{b}}\left(\sum_{\mathbf{s}\in\mathbf{B}_2^l}\mathbf{I^s}\cdot x_{\mathbf{s}}\right)+\sum_{\mathbf{b}\in\mathbf{B}_2^m}\mathbf{J^b}\cdot B\cdot f_{\mathbf{b}}^r\left(\sum_{\mathbf{s}\in\mathbf{B}_2^l}\mathbf{I^s}\cdot x_{\mathbf{s}}\right)$$

$$\tag{1.29}$$

$$=\sum_{\mathbf{b}\in\mathbf{B}_2^m}\mathbf{J^b}\cdot\left[f_{\mathbf{b}}^l\left(\sum_{\mathbf{s}\in\mathbf{B}_2^l}\mathbf{I^s}\cdot x_{\mathbf{s}}\right)+B\cdot f_{\mathbf{b}}^r\left(\sum_{\mathbf{s}\in\mathbf{B}_2^l}\mathbf{I^s}\cdot x_{\mathbf{s}}\right)\right].\tag{1.30}$$

Here $f_k^{Bin}(\mathbf{x})=f_k^r(\mathbf{x})+Bf_k^l(\mathbf{x})$ and $f_{\mathbf{b}}^{Bin}(\mathbf{x})=f_{\mathbf{b}}^l(\mathbf{x})+Bf_{\mathbf{b}}^r(\mathbf{x})$ are channel binocular images. In this form, these binocular images are complex-valued images, where B binocular complex unit.

For pattern recognition theory, it is very important to have mathematical models of real distortions of n-D K-channel images $\mathbf{f}(\mathbf{x})$. Changes in the perceptual spaces can be treated in the language of Clifford as the actions of some perceptual transformation groups. Geometrical distortions are caused by: (1) n-D translations $\mathbf{x}\to\mathbf{x}+\mathbf{w}$; (2) n-D rotations $\mathbf{x}\to\mathcal{E}_0(\mathbf{x}+\mathbf{w})\mathcal{E}_0^{-1}$, where $\mathbf{x},\mathbf{w}\in\mathbf{Vec}^1\left(Alg_{2^n}^{Sp(p,q,r)}\right)$, $\mathcal{E}_0\in\mathbf{Spin}\left(Alg_{2^n}^{Sp(p,q,r)}\right)$; (3) dilatation: $\mathbf{x}\to\lambda\mathbf{x}$, where $\lambda\in\mathbf{R}^+$ is a scale factor. If $\mathbf{f}(\mathbf{x})$ is an initial image and $_{\lambda\mathcal{E}_0 w}\mathbf{f}(\mathbf{x})$ is its distorted version, then $_{\lambda\mathcal{E}_0 w}\mathbf{f}(\mathbf{x}):=\mathbf{f}\left(\lambda\mathcal{E}_0(\mathbf{x}+\mathbf{w})\mathcal{E}_0^{-1}\right)$. Transforms

$$g^{Sp}(\lambda,\mathcal{E}_0,\mathbf{w}):\mathbf{f}(\mathbf{x})\to{}_{\lambda\mathcal{E}_0 w}\mathbf{f}(\mathbf{x}):=\mathbf{f}\left(\lambda\mathcal{E}_0(\mathbf{x}+\mathbf{w})\mathcal{E}_0^{-1}\right)$$

form the space affine group $\mathbf{Aff}_n^{Sp(p,q,r)}=\mathbf{Aff}\left[\mathbf{Vec}^1\left(Alg_{2^n}^{Sp(p,q,r)}\right)\right]$. Obviously, all transforms

$$g^{Sp}(\mathcal{E}_0,\mathbf{w}):\mathbf{f}(\mathbf{x})\to{}_{\lambda\mathcal{E}_0 w}\mathbf{f}(\mathbf{x}):=\mathbf{f}\left(\mathcal{E}_0(\mathbf{x}+\mathbf{w})\mathcal{E}_0^{-1}\right),$$
$$g^{Sp}(\mathcal{E}_0):\mathbf{f}(\mathbf{x})\to{}_{\lambda\mathcal{E}_0 w}\mathbf{f}(\mathbf{x}):=\mathbf{f}\left(\mathcal{E}_0(\mathbf{x})\mathcal{E}_0^{-1}\right)$$

form the space motion $\mathbf{Mot}_n^{Sp(p,q,r)}$ and rotation $\mathbf{Rot}_n^{Sp(p,q,r)}$ groups, respectively.

We are going to consider the following multicolor distortions caused by: (1) $K-$D multicolor translations $\mathbf{f}\to\mathbf{f}+\mathbf{g}$, where $\mathbf{f},\mathbf{g}\in\mathbf{Vec}^1\left(Alg_{2^K}^{Vis(u,v,w)}\right)$ and $\mathbf{g}$ is a constant; (2) $K-$D multicolor rotations $\mathbf{f}\to\mathcal{Q}_0\mathbf{f}\mathcal{Q}_0^{-1}$, where $\mathcal{Q}_0\in\mathbf{Spin}\left(Alg_{2^K}^{Vis(u,v,w)}\right)$; (3) dilatation: $\mathbf{f}\to\bar{\lambda}\mathbf{f}$, where $\bar{\lambda}\in\mathbf{R}^+$ is a scale factor. If $\mathbf{f}(\mathbf{x})$ is an initial image and $\mathbf{f}^{\bar{\lambda}\mathcal{Q}_0\mathbf{g}}(\mathbf{x})$ is its multicolor distorted version, then transforms

$$g^{Vis}(\bar{\lambda},\mathcal{Q}_0,\mathbf{g}):\mathbf{f}(\mathbf{x})\to\mathbf{f}^{\bar{\lambda}\mathcal{Q}_0\mathbf{g}}(\mathbf{x})=\bar{\lambda}\mathcal{Q}_0\mathbf{f}(\mathbf{x})\mathcal{Q}_0^{-1}+\mathbf{g}$$

form multicolor affine group $\mathbf{Aff}_K^{Vis(u,v,w)}=\mathbf{Aff}\left[\mathbf{Vec}^1\left(Alg_{2^K}^{Vis(u,v,w)}\right)\right]$ and transforms

$$g^{Vis}(\mathcal{Q}_0, \mathbf{g}) : \mathbf{f}(\mathbf{x}) \rightarrow \mathbf{f}^{\mathcal{Q}_0 \mathbf{g}}(\mathbf{x}) = \mathcal{Q}_0 \mathbf{f}(\mathbf{x}) \mathcal{Q}_0^{-1} + \mathbf{g},$$

$$g^{Vis}(\mathcal{Q}_0) : \mathbf{f}(\mathbf{x}) \rightarrow \mathbf{f}^{\mathcal{Q}_0}(\mathbf{x}) = \mathcal{Q}_0 \mathbf{f}(\mathbf{x}) \mathcal{Q}_0^{-1}$$

form the multicolor motion $\mathbf{Mot}_K^{Vis(u,v,w)}$ and the rotation $\mathbf{Rot}_K^{Vis(u,v,w)}$ groups, respectively.

For geometrical and multicolor distortions, we have

$$\left(g^{Sp}(\lambda, \mathcal{E}_0, \mathbf{w}), g^{Vis}(\bar{\lambda}, \mathcal{Q}_0, \mathbf{g}) \right) : \mathbf{f}(\mathbf{x}) \rightarrow {}_{\lambda \mathcal{E}_0 w} \mathbf{f}^{\bar{\lambda} \mathcal{Q}_0 \mathbf{g}}(\mathbf{x})$$
$$= \bar{\lambda} \mathcal{Q}_0 \mathbf{f} \left(\lambda \mathcal{E}_0 (\mathbf{x} + \mathbf{w}) \mathcal{E}_0^{-1} \right) \mathcal{Q}_0^{-1} + \mathbf{g}.$$

Transforms $\left(g^{Sp}(\lambda, \mathcal{E}_0, \mathbf{w}), g^{Vis}(\bar{\lambda}, \mathcal{Q}_0, \mathbf{g}) \right)$ form the space-multicolor affine group

$$\mathbf{Aff}_n^{Sp(p,q,r)} \times \mathbf{Aff}_K^{Vis(u,v,w)} = \mathbf{Aff} \left[\mathbf{Vec}^1 \left(Alg_{2^n}^{Sp(p,q,r)} \right) \right] \times \mathbf{Aff} \left[\mathbf{Vec}^1 \left(Alg_{2^K}^{Vis(u,v,w)} \right) \right].$$

1.4 Clifford-Valued Invariants

Let us assume that $\mathbf{f}(\mathbf{x}) : \mathbf{Vec}^1 \left(Alg_{2^n}^{Sp(p,q,r)} \right) \rightarrow \mathbf{Vec}^1 \left(Alg_{2^K}^{Vis(u,v,w)} \right)$ is an image of a multicolor $n - D$ object. Changes in the surrounding world can be treated in the language of the spatial-multicolor algebra as an action of two groups: the space affine group $\mathbf{Aff}_n^{Sp(p,q,r)}$ acting on the physical space $\mathbf{Vec}^1 \left(Alg_{2^n}^{Sp(p,q,r)} \right)$ and the multicolor affine group $\mathbf{Aff}_K^{Vis(u,v,w)}$ acting on $\mathbf{Vec}^1 \left(Alg_{2^K}^{Vis(u,v,w)} \right)$.

Let $\mathbf{G}^{Sp,Vis} = \mathbf{Aff}_n^{Sp(p,q,r)} \times \mathbf{Aff}_K^{Vis(u,v,w)}$ be the space-multicolor group, and $(g^{Sp}, g^{Vis}) \in \mathbf{G}^{Sp,Vis}$, where $g^{Sp} \in \mathbf{Aff}_n^{Sp(p,q,r)}$, $g^{Vis} \in \mathbf{Aff}_K^{Vis(u,v,w)}$. If $\mathcal{C} \in Alg_{2^K}^{Vis(u,v,w)}$ and $\mathbf{x} \in \mathbf{Vec}^1 \left(Alg_{2^n}^{Sp(p,q,r)} \right)$, then product of the form $\mathbf{x}\mathcal{C}$ is called the space-color number. They form a space-color algebra $Alg_{2^{n+K}}^{SpVis} = Alg_{2^n}^{Sp(p,q,r)} \otimes Alg_{2^K}^{Vis(u,v,w)}$, where $\otimes$ is the tensor product. Here, we assume that all spatial hyper-imaginary units commute with all color units.

Definition 2 The $Alg_{2^{n+K}}^{SpVis}$-valued functional $\Im[\mathbf{f}(\mathbf{x})]$ of the image $\mathbf{f}(\mathbf{x})$ is called the relative invariant with respect to $\mathbf{G}^{Sp,Vis}$ if

$$\Im \left[{}_{\lambda \mathcal{E}_0 w} \mathbf{f}^{\bar{\lambda} \mathcal{Q}_0 \mathbf{g}}(\mathbf{x}) \right] = \Phi_l(\bar{\lambda}, \lambda, \mathcal{C}, \mathcal{Q}) \cdot \Im[\mathbf{f}(\mathbf{x})] \cdot \Phi_r(\bar{\lambda}^{-1}, \lambda^{-1}, \mathcal{C}^{-1}, \mathcal{Q}^{-1}),$$

$\forall \left(g^{Sp}(\lambda, \mathcal{E}_0, \mathbf{w}), g^{Vis}(\bar{\lambda}, \mathcal{Q}_0, \mathbf{g}) \right) \in \mathbf{G}^{Sp,Vis}$, where $\Phi_l(\bar{\lambda}, \lambda, \mathcal{C}, \mathcal{Q})$ and $\Phi_r(\bar{\lambda}^{-1}, \lambda^{-1}, \mathcal{C}^{-1}, \mathcal{Q}^{-1})$ are left and right multiplicators, respectively, depending only on transform parameters. If $\Phi_l(\bar{\lambda}, \lambda, \mathcal{C}, \mathcal{Q}) = 1$ and $\Phi_r(\bar{\lambda}^{-1}, \lambda^{-1}, \mathcal{C}^{-1}, \mathcal{Q}^{-1}) = 1$, then $\Im[\mathbf{f}(\mathbf{x})]$ is called the absolute invariant and denoted by $\bar{\Im}[\mathbf{f}(\mathbf{x})]$.

Definition 3 Let $\mathbf{c_f}$ be the centroid of the image $\mathbf{f(x)}$, then functionals

$$\mathbb{M}_p[\mathbf{f}] := \int\limits_{\mathbf{x} \in \mathbf{Vec}^1\left(Alg_{2^n}^{Sp(p,q,r)}\right)} (\mathbf{x} - \mathbf{c_f})^p \mathbf{f(x)} dx, \quad p \in \mathbf{Q}$$

are called the central $Alg_{2^{n+K}}^{SpVis}$-valued moments of the n-D image $\mathbf{f(x)}$, where $p \in \mathbf{Q}$ are rational numbers.

Let us clarify the rules of moment transformations with respect to distortions of the following form

$$_{\lambda\mathcal{E}_{0w}}\mathbf{f}^{\bar{\lambda}\mathcal{Q}_0}(\mathbf{x}) = \bar{\lambda}\mathcal{Q}_0\mathbf{f}\left(\lambda\mathcal{E}_0(\mathbf{x} + \mathbf{w})\mathcal{E}_0^{-1}\right)\mathcal{Q}_0^{-1},$$

where $\mathbf{f(x)}$ is centered image (i.e., $\mathbf{c_f} = 0$).

Theorem 4 *The central moments* $\mathbb{M}_p\{\mathbf{f(x)}\}$ *of the multi-channel image* $\mathbf{f(x)}$ *are relative* $Alg_{2^{n+K}}^{SpVis}$-*valued invariants*

$$\mathbb{M}_p\left\{_{\lambda\mathcal{E}_{0w}}\mathbf{f}^{\bar{\lambda}\mathcal{Q}_0}(\mathbf{x})\right\} = \left(\bar{\lambda}\lambda^{-n(p+1)}\mathcal{E}_0^p\mathcal{Q}_0\right) \cdot \mathbb{M}_p\{\mathbf{f(x)}\} \cdot \left(\mathcal{Q}_0^{-1}\mathcal{E}_0^{-p}\right) \qquad (1.31)$$

with respect to the group $\mathbf{G}^{Sp,Vis}$ *with both* $Alg_{2^{n+K}}^{SpVis}$-*valued left* $\Phi_l(\bar{\lambda}, \lambda, \mathcal{C}, \mathcal{Q}) = \bar{\lambda}\lambda^{-n(p+1)}\mathcal{E}_0^p\mathcal{Q}_0$ *and right* $\Phi_r(\bar{\lambda}^{-1}, \lambda^{-1}, \mathcal{C}^{-1}, \mathcal{Q}^{-1}) = \mathcal{Q}_0^{-1}\mathcal{E}_0^{-p}$ *multiplicators, respectively, i.e.,* $\Im_p\{\mathbf{f(x)}\} = \mathbb{M}_p\{\mathbf{f(x)}\}$.
If $Alg_{2^{n+K}}^{SpVis}$-*valued invariants are denoted as* $\Im_p\{\mathbf{f(x)}\}$, *then scalar-valued (real-valued) invariants we will denote by* $\Re_p\{\mathbf{f(x)}\}$.

Theorem 5 *Norms of central moments* $\left\|\mathbb{M}_p\{\mathbf{f(x)}\}\right\|$ *of the multi-channel image* $\mathbf{f(x)}$ *are relative real-valued invari*ants

$$\left\|\mathbb{M}_p\left\{_{\lambda\mathcal{E}_{0w}}\mathbf{f}^{\bar{\lambda}\mathcal{Q}_0}(\mathbf{x})\right\}\right\| = \bar{\lambda}\lambda^{-n(p+1)} \cdot \left\|\mathbb{M}_p\{\mathbf{f(x)}\}\right\| \qquad (1.32)$$

with respect to the group $\mathbf{G}^{Sp,Vis}$ *with scalar-valued multiplicators* $\bar{\lambda}\lambda^{-n(p+1)}$, *i.e.,* $\Re_p\{\mathbf{f(x)}\} = \left\|\mathbb{M}_p\{\mathbf{f(x)}\}\right\|$.

Theorem 6 *The normalized central moments*

$$\mathbb{N}_p\{\mathbf{f(x)}\} = \frac{\mathbb{M}_0^{p-1}\{\mathbf{f(x)}\}\mathbb{M}_p\{\mathbf{f(x)}\}}{\mathbb{M}_1^p\{\mathbf{f(x)}\}}$$

are relative $Alg_{2^{n+K}}^{SpVis}$-*valued invariants, with respect to the same group* $\mathbf{G}^{Sp,Vis}$ *and with both* $Alg_{2^{n+K}}^{SpVis}$-*valued left* $\Phi_l(\bar{\lambda}, \lambda, \mathcal{C}, \mathcal{Q}) = \mathcal{E}_0^p\mathcal{Q}_0^p$ *and right* $\Phi_r(\bar{\lambda}^{-1}, \lambda^{-1}, \mathcal{C}^{-1}, \mathcal{Q}^{-1}) = \mathcal{Q}_0^{-p}\mathcal{E}_0^{-p}$ *multiplicators.*

Proof Indeed,

$$
\mathbb{N}_p\left\{{}_{\lambda\mathcal{E}_0 w}\mathbf{f}^{\bar{\lambda}\mathcal{Q}_0}(\mathbf{x})\right\} = \frac{\mathbb{M}_0^{p-1}\left\{{}_{\lambda\mathcal{E}_0 w}\mathbf{f}^{\bar{\lambda}\mathcal{Q}_0}(\mathbf{x})\right\}\mathbb{M}_p\left\{{}_{\lambda\mathcal{E}_0 w}\mathbf{f}^{\bar{\lambda}\mathcal{Q}_0}(\mathbf{x})\right\}}{\mathbb{M}_1^p\left\{{}_{\lambda\mathcal{E}_0 w}\mathbf{f}^{\bar{\lambda}\mathcal{Q}_0}(\mathbf{x})\right\}}
$$

$$
= \left(\mathcal{E}_0^p \mathcal{Q}_0^p\right)\frac{\mathbb{M}_0^{p-1}\{\mathbf{f}(\mathbf{x})\}\mathbb{M}_p\{\mathbf{f}(\mathbf{x})\}}{\mathbb{M}_1^p\{\mathbf{f}(\mathbf{x})\}}\left(\mathcal{Q}_0^{-p}\mathcal{E}_0^{-p}\right)
$$

$$
= \left(\mathcal{E}_0^p \mathcal{Q}_0^p\right)\mathbb{N}_p\{\mathbf{f}(\mathbf{x})\}\left(\mathcal{Q}_0^{-p}\mathcal{E}_0^{-p}\right).
$$

Theorem 7 *Norms of normalized central moments*

$$
\left\|\mathbb{N}_p\{\mathbf{f}(\mathbf{x})\}\right\| = \left\|\frac{\mathbb{M}_0^{p-1}\{\mathbf{f}(\mathbf{x})\}\mathbb{M}_p\{\mathbf{f}(\mathbf{x})\}}{\mathbb{M}_1^p\{\mathbf{f}(\mathbf{x})\}}\right\| = \frac{\left\|\mathbb{M}_0^{p-1}\{\mathbf{f}(\mathbf{x})\}\right\| \cdot \left\|\mathbb{M}_p\{\mathbf{f}(\mathbf{x})\}\right\|}{\left\|\mathbb{M}_1^p\{\mathbf{f}(\mathbf{x})\}\right\|}
$$

$$
= \frac{\left\|\mathfrak{R}_0^{p-1}\{\mathbf{f}(\mathbf{x})\}\right\| \cdot \left\|\mathfrak{R}_p\{\mathbf{f}(\mathbf{x})\}\right\|}{\left\|\mathfrak{R}_1^p\{\mathbf{f}(\mathbf{x})\}\right\|}
$$

are absolute scalar-valued invariants, with respect to the same group.

Proof Indeed,

$$
\left\|\mathbb{N}_p\left\{{}_{\lambda\mathcal{E}_0 w}\mathbf{f}^{\bar{\lambda}\mathcal{Q}_0}(\mathbf{x})\right\}\right\| = \frac{\left\|\mathfrak{R}_0^{p-1}\left\{{}_{\lambda\mathcal{E}_0 w}\mathbf{f}^{\bar{\lambda}\mathcal{Q}_0}(\mathbf{x})\right\}\right\| \cdot \left\|\mathfrak{R}_p\left\{{}_{\lambda\mathcal{E}_0 w}\mathbf{f}^{\bar{\lambda}\mathcal{Q}_0}(\mathbf{x})\right\}\right\|}{\left\|\mathfrak{R}_1^p\left\{{}_{\lambda\mathcal{E}_0 w}\mathbf{f}^{\bar{\lambda}\mathcal{Q}_0}(\mathbf{x})\right\}\right\|}
$$

$$
= \frac{\left(\bar{\lambda}^{(p-1)}\lambda^{-n(p-1)}\left\|\mathfrak{R}_0^{p-1}\{\mathbf{f}(\mathbf{x})\}\right\|\right) \cdot \left(\bar{\lambda}\lambda^{-n(p+1)} \cdot \left\|\mathfrak{R}_p\{\mathbf{f}(\mathbf{x})\}\right\|\right)}{\bar{\lambda}^p\lambda^{-2np}\left\|\mathfrak{R}_1^p\{\mathbf{f}(\mathbf{x})\}\right\|}
$$

$$
= \frac{\left\|\mathfrak{R}_0^{p-1}\{\mathbf{f}(\mathbf{x})\}\right\| \cdot \left\|\mathfrak{R}_p\{\mathbf{f}(\mathbf{x})\}\right\|}{\left\|\mathfrak{R}_1^p\{\mathbf{f}(\mathbf{x})\}\right\|}
$$

$$
= \left\|\mathbb{N}_p\{\mathbf{f}(\mathbf{x})\}\right\|.
$$

Using this theorem, we are going to calculate hypercomplex-valued invariants of gray-level, color, multicolor, and hyperspectral 2-D, 3-D and *n*-D images for Euclidean and non-Euclidean geometries in the next part of our work.

1.5 Conclusion

We developed a novel algebraic approach based on hypercomplex algebras to algebraic modeling of color, multicolor, and hyperspectral images. It is our aim to show that the use of hypercomplex algebras fits more naturally to the tasks of recognition

of multicolor patterns than does the use of color vector spaces. One can argue that Nature has, through evolution, also learned to utilize properties of hypercomplex numbers. Thus, a brain might have the ability to operate as a Clifford algebra computing device. We do not agree with L. Kronecker in that that "*the LORD created the integers, the rest is the work of man*." We assume that the LORD was the first engineer who knew hypercomplex algebras and used them for designing the visual systems of animals.

References

1. V.G. Labunets, Clifford algebra as unified language for image processing and pattern recognition, in *Topics in Computational Noncommutative Algebra and Applications*, ed. by J. Byrnes, G. Ostheimer (Kluwer Academic Publishers, Dordrect, Boston, London, 2003), pp. 197–225
2. E.V. Labunets-Rundblad, V.G. Labunets, J.Astola, Algebra and geometry of color images, in *Proceedings of the First International Workshop on Spectral Techniques and Logic Design for Future Digital Systems* (2000), pp. 231–261
3. V.G. Labunets, E.V. Kokh, E. Ostheimer, Algebraic models and methods of image computer processing. Part 1. Multiplet models of multichannel images. Comput. Opt. **42**(1), 84–95 (2018)
4. V.G. Labunets, E.V. Rundblad-Ostheimer, Spatial-color Clifford algebras for invariant image recognition, in *Topics in Geometric Computing with Clifford Algebras*, ed. by G. Sommer (Springer, Berlin, New York, 2001), pp. 155–185
5. E. Labunets-Rundblad, I. Nikitin, V.G. Labunets, Unified approach to Fourier-Clifford-Prometheus sequences, transforms and filter banks, in *Topics in Computational Noncommutative Algebra and Applications*, ed. by J. Byrnes, G. Ostheimer (Kluwer Academic Publishers, Dordrect, Boston, London, 2003), pp. 389–400
6. V.G. Labunets, E.V. Rundblad, J. Astola, Is the brain a "Clifford algebra quantum computer?", in *Topics in Applied Geometrical Algebras in Computer Science and Engineering*, ed. by L. Dorst, C. Doran, J. Lasenby (Birkhauser, 2003), pp. 486–495 (2002)
7. E.V. Labunets-Rundblad, V.G. Labunets, J. Astola, Is the visual cortex a "fast Clifford algebra quantum computer"?, in *Topics in Clifford Analysis and Its Applications, II*. Mathematics, Physics and Chemistry, NATO Science Series, vol. 25 (2001), pp. 173–183
8. V.G. Labunets, A. Maidan, E.V. Rundblad-Labunets, J. Astola, Color triplet-valued wavelets and splines, in *Proceedings of the International Conference on Image and Signal Processing and Analysis (ISPA'01)* (2001), pp. 535–541
9. V.G. Labunets, A. Maidan, E.V. Rundblad-Labunets, J. Astola, Color triplet-valued wavelets, splines and median filters, in *Proceedings of the International Conference on Spectral Methods and Multirate Signal Processing (SMMSP'2001)* (2001), pp. 61–70
10. E.V. Labunets-Rundblad, *Fast Fourier-Clifford Transforms Design and Application in Invariant Recognition*. Ph.D. thesis (Tampere University Technology, Tampere, Finland), 262 pp.

Chapter 2
Influence of Reflections from the Clouds and Artificial Structures on Fire Detection from Space

Sergey M. Zraenko

2.1 Introduction

Earth's remote sensing (ERS) data is widely used for the detection of forest fires. Currently the images of MODIS spectroradiometer [1], which is part of the equipment of American satellites Terra and Aqua, are mainly used. The operational data reception from MODIS is performed by using the personal receiving stations, one of which (UniScan-24 [2]) is mounted at Ural Federal University.

Now the MOD14 algorithm for thermal anomalies definition [3] is widely used for fire detection according to MODIS data. This algorithm uses the pixel temperature data in two infrared spectral channels: 4 microns (21 channels) and 11 microns (31 channels). However, the brightness temperature of "hot" pixel in the 21 channel should be above 312 K during the day and 305 K in the night. In addition, the higher the temperature difference of the pixel in the 21 and 31 channels and its temperature difference in the 21 channel and the environment, the greater the probability that the pixel belongs to the fire. However, as we have shown [4], the Internet services using such a data processing technology often give the different information on the actual forest fire situation. This is primarily due to the use of different approaches to eliminate interference caused by the thermal emissions of industrial plants that are close to the mirror reflections at the artificial structures, water surface, and cloud edges.

To reduce the dependence of the "hot" pixels detection on the observation conditions, we have developed an adaptive algorithm for thermal anomalies detection based on the use of sub-satellite information about the fire in the analyzed territory [5, 6]. The algorithm idea is that the thermal anomaly detection threshold is determined by the conditions of definition of the known fire. In these circumstances, the threshold is determined as an average of pixels brightness in the fire seat and in the

S. M. Zraenko (✉)
Ural Federal University, Yekaterinburg, Russia
e-mail: z_sm@mail.ru

area surrounding it. Then, the found threshold is used to detect the thermal points in the whole image, provided that the image pixels that exceed the threshold are related to fire seats.

It should be noted that in the adaptive algorithm, as in the MOD14 algorithm, the temporary instability of cloud cover is a serious problem associated with the appearance of false thermal points caused by flecks of sunlight—reflections at the cloud edges. In this regard, we have conducted research on the analysis of the impact of the cloud reflections on the performance of algorithms and a way of its decrease is proposed. To reduce the influence of reflections from artificial structures and thermal emissions of industrial enterprises during the analysis of forest fire situation, it is proposed to exclude counties. The results of these studies are presented in this article.

2.2 Cloud Impact on MOD14 Algorithm Performance

Investigation of the cloud impact on fire detection algorithms performance was carried out on a MODIS image A201208010820, which according to data of the Ministry of Emergency Situations dated August 1, 2012 should detect a fire in the 5-hectare Krasnouralsky county. When using the MOD14 algorithm with a threshold of detection in 21 spectral channels equal to 320 K (slightly higher than the recommended 312 K for detecting thermal anomalies in daily snapshots), we still find an existing fire, but we continue to detect multiple false points (Fig. 2.1).
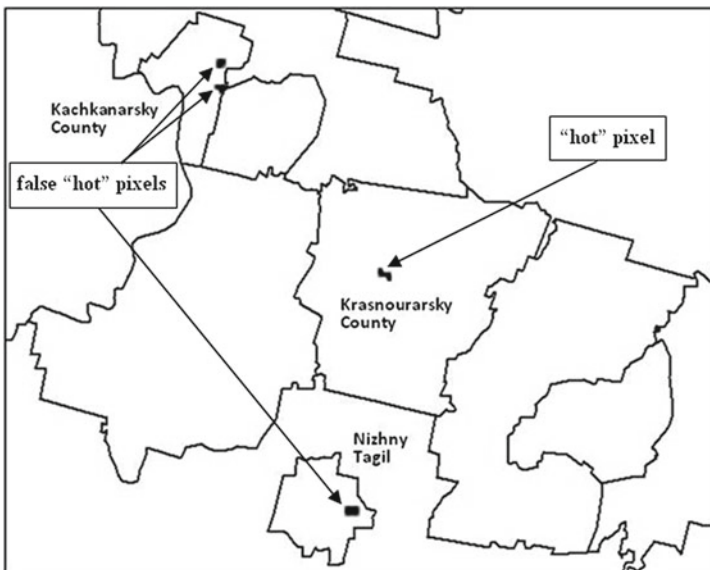
Fig. 2.1 "Hot" pixels allocated by the MOD14 algorithm (while maintaining all the clouds)
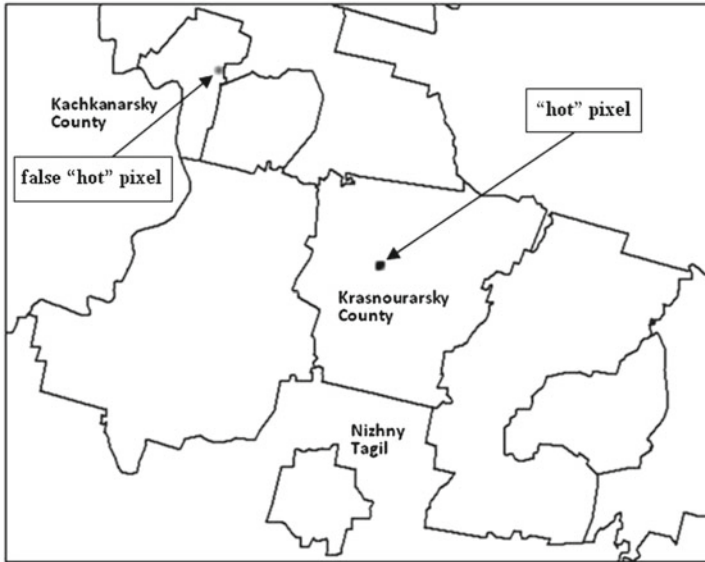
Fig. 2.2 "Hot" pixels allocated by the adaptive algorithm (while maintaining all the clouds)

To reduce the number of false thermal points, we have formed three masks consisting of zeros and ones using the information provided by the MOD35 algorithm on the cloud density in each image pixel [3]. As a result of combination of the image with these cloud masks, we have got three images of the controlled territory. The first image displays only the areas with a light cloud cover (with cloud density less than 20%), while the rest are set to zero. The second image displays only the areas with light and medium clouds (with cloud density less than 50%). The third one represents the original image without zeroing the cloud areas (the results of the work of the investigated algorithms with this image are presented in Figs. 2.1, 2.2, and 2.3).

Representation of an image with clouds of different densities allows analyzing the performance of algorithms in cloudy conditions. For example, when working with an image with a weak cloud cover (column 4 of Table 2.1), the MOD14 algorithm does not accept reflections from the edges of the cloud as thermal anomalies. However, in this case, as a result of zeroing of the areas with the density of the cloud cover more than 20%, we also zero an area with the known fire seat near Krasnouralsky. Keeping the image areas with cloud density less than 50%, we find the fire using the MOD14. This is because the thermal emission of the fire overcomes the light cloud cover with density from 20 to 50% at this point (columns 2 and 3 of Table 2.1). However, in this case, we keep the image areas with medium and large density clouds, from the edges of which the flecks of sunlight are formed. This results in false detection of thermal anomalies by MOD14 algorithm (columns 2 and 3 of Table 2.1) in Kachkanarsky, Nizhnetagilsky, Karpinsky, and Krasnoturinsky counties. If the cloud cover is not
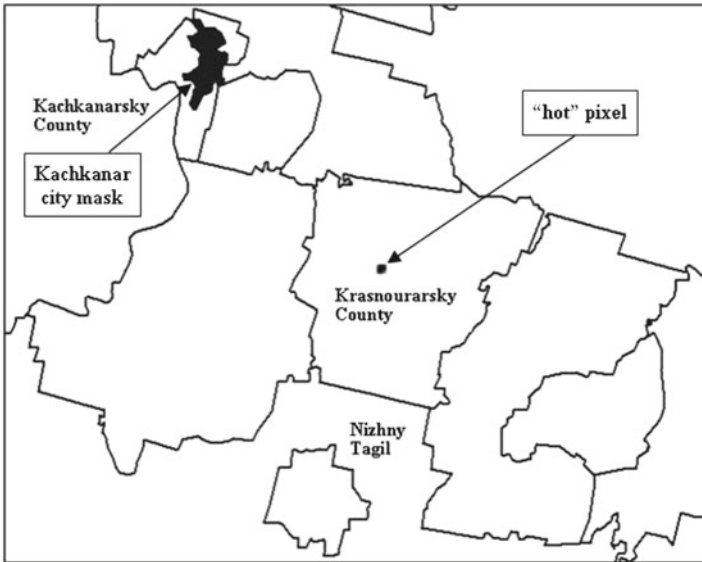
Fig. 2.3 "Hot" pixels allocated by the adaptive algorithm when masking the city of Kachkanar (while maintaining all the clouds)

Table 2.1 The number of "hot" pixels allocated by the MOD14 algorithm

County	Keeping all the clouds	Keeping the clouds with density less than 50%	Keeping the clouds with density less than 20%
Krasnouralsky	1	1	–
Kachkanarsky	2	1	–
Nizhnetagilsky	1	1	–
Karpinsky	2	2	–
Krasnoturinsky	1	1	–
Ivdelsky	7	–	–
Severouralsky	1	–	–
Sosvinsky	1	–	–

excluded at all (column 2 of Table 2.1), the false thermal anomalies appear also in Ivdelsky, Severouralsky, and Sosvinsky counties, the number of anomalies in Kachkanarsky county is increased up to two.

Thus, zeroing of the image areas with light cloud cover leads to loss of the "hot pixels" in it. Keeping the areas with more dense clouds leads to thermal anomalies due to reflections at the cloud edges. Making recommendations for cloud cover elimination is difficult, because the fire thermal emission depends on its area and temperature and therefore can lead to its omission after cloud areas zeroing.

2.3 Cloud Cover Impact on the Adaptive Algorithm Performance

Using the method of excluding the cloud cover for the adaptive algorithm, it was shown that keeping all the clouds (Fig. 2.2) correctly identifies the existing thermal anomaly (column 2 of Table 2.2—Krasnouralsky county). Nevertheless, the false detection of thermal anomalies (column 2 of Table 2.2—Kachkanarsky county) is possible due to the solar radiation reflection at the edge of a very dense cloud. By zeroing of the image areas with the cloud cover density more than 50%, there are no false "hot pixels" and the thermal emissions of the existing fire is detected, as in the case of MOD14 even with light cloud cover over them.

For a quantitative comparison of the algorithms performance with the flecks of sunlight of the cloud edges, such a parameter as false alarm probability was calculated. For the purpose of equivalence, the situation without cloud cover zeroing was selected (when the adaptive algorithm also detects the false thermal anomalies)—column 2 in Tables 2.1 and 2.2. The number of the false "hot pixels" detected by the MOD14 algorithm in this case is 15, and by the adaptive algorithm is 1 for the analyzed image. Given the fact that the image pixel area used for the MODIS 21 channel is 1 km^2, and the area of the analyzed image coincides with the area of the Sverdlovsk region (S = 194,307 km^2 [7]), we get:

$$P_{\text{FA}_{\text{MOD14}}} = \frac{15}{S} = 7.7 * 10^{-5};$$
$$P_{\text{FA}_{\text{Adapt.Alg}}} = \frac{1}{S} = 5.1 * 10^{-6}. \tag{2.1}$$

These results are explained by the higher value of the adaptive threshold of our algorithm in comparison with its fixed value in the MOD14 algorithm. Thus, for the analyzed image, the adaptive algorithm has a 15 times greater noise immunity compared with the MOD14 algorithm against the flecks of sunlight of the cloud edges.

Table 2.2 The number of "hot" pixels allocated by the adaptive algorithm

County	Keeping all the clouds	Keeping the clouds with density less than 50%	Keeping the clouds with density less than 20%
Krasnouralsky	1	1	–
Kachkanarsky	1	–	–
Nizhnetagilsky	–	–	–
Karpinsky	–	–	–
Krasnoturinsky	–	–	–
Ivdelsky	–	–	–
Severouralsky	–	–	–
Sosvinsky	1	–	–

2.4 Artificial Structures Interference Impact on the Algorithm Performance

Solar specular reflections from artificial structures and their thermal radiation are perceived in the used spectral channels as "hot" pixels. To eliminate the associated errors of the analyzed algorithms, it is proposed to exclude the counties on satellite images when fires are detected. The reason for this decision is that the monitoring of the fire situation in residential areas is carried out by other means. To do this, using the Landsat satellite images the county mask of several industrial cities of the Sverdlovsk region was created. Combination of this mask with the MODIS images allows elimination of the reflections at the artificial structures located there and thermal emissions of the industrial enterprises. As an example of this technique, Fig. 2.3 shows the masking of Kachkanar city. This mask allowed excluding a false fire in the city, which was detected by the adaptive algorithm.

We consider such an approach as reasonable, because we solve the problem of remote monitoring of fire situation in the forests far from counties. At the same time for the counties, we assumed that standard means of fire detection are used.

2.5 Conclusion

The research on the impact of cloud cover and artificial structures on the effectiveness of monitoring the forestry fire situation by MODIS spectroradiometer satellite images is carried out in this article. It is shown that the regular fire detection algorithm MOD14 often assumes reflections at the dense clouds edges as thermal anomalies. The false detection of the "hot" pixels by the adaptive algorithm proposed by us occurs at the edges of only a very dense cloud cover. It is shown that the probability of false alarm from the adaptive algorithm is 15 times less than the MOD14 algorithm. To improve the efficiency of fire detection, zeroing of cloud areas of the image is proposed. To this end, the MOD35 algorithm to build "cloud masks" is used. Thus, the use of the adaptive algorithm allows one to exclude the false "hot" pixels. As for the interference of artificial structures, it is proposed to exclude them by means of "county masks" based on Landsat images. It is shown that the use of the proposed masks significantly reduces the false thermal anomalies.

References

1. Moderate Resolution Imaging Spectroradiometer. Wikipedia. https://en.wikipedia.org/wiki/Moderate_Resolution_Imaging_Spectroradiometer. Accessed 22 Sept 2018
2. UniScan™ Ground Receiving Station. SCANEX. http://scanex.ru/en/station/unis%D1%81an/. Accessed 22 Sept 2018
3. Algorithm Theoretical Basis Documents – ATBD. ESA https://earth.esa.int/web/sppa/mission-performance/esa-missions/envisat/meris/products-and-algorithms/atbd. Accessed 22 Sept 2018
4. S.M. Zraenko, M.A. Mymrina, V.V. Ganzha, Obuchaemyi algoritm obnaruzheniya pozharov po dannym distantsionnogo zondirovaniya [Learning algorithm of fire detection from remote sensing data]. Zhurnal radioelektroniki: elektronnyi zhurnal **11**, 1–13 (2014). http://jre.cplire.ru/jre/nov14/8/text.pdf. Accessed 22 Sept 2018
5. S.M. Zraenko, M.A. Mymrina, V.V. Ganzha, Application of remote sensing data for forest fires detection with the use of a trained thresholding algorithm, in *Proceedings of the Twenty Fourth Microwave and Telecommunication Technology Conference* (2014), pp. 1151–1152
6. S.M. Zraenko, M.A. Mymrina, V.V. Ganzha, Development of trained algorithm detection of fires for multispectral systems remote monitoring, in *Supplementary Proceedings of the Fourth Analysis of Images, Social Networks and Texts Conference* (2015), pp. 187–195
7. The official website of the Sverdlovsk region government. http://www.midural.ru. Accessed 22 Sept 2018

Chapter 3
Reconfigurable Systolic 2D-Arrays of Bit-Level Processor Elements for High Speed Data Processing in Embedded Computer Systems

Nick A. Lookin

3.1 Introduction

High-speed data processing based upon special-purpose processors is one of the basic trends in the development of embedded real-time computing. The main purpose of these processors is substantial increase of computational performance for complicated algorithms. These algorithms can be implemented in real time only by means of parallel data processing on operation level, word level, or even bit level.

As an example, let us assume that we need image processing of some objects. Images are dynamically changed in space and time dimensions and have 1024 × 1024 pixels. The frame rate is 100 Hz and each frame period contains both intra-frame data processing (noise filtering, staining, etc.) and inter-frame data processing (objects recognition, motion estimation in inertial frame, etc.). If the numbers of objects are an order of thousands, then only processor architectures with parallel and independent processing of small set of pixels (may be separate pixels) can provide full-size image processing (intra- and inter-frame) during one frame period.

There are three main directions of performance increase for real-time computer systems. They are:

- Growth of the gates switching frequency. This requires essential enhancement of design and technology of chips, micro assemblies, and printed circuit boards. In addition, increasing of frequency results in growth of power consumption density and amplifies the impact of electro physical parameters on circuits operation [1]. This way of performance increasing has physical limitations already today.
- Parallel data processing. There are two kinds of parallelism in computer systems:

N. A. Lookin (✉)
Ural Federal University, Yekaterinburg, Russia
e-mail: nicklookin@mail.ru

Institute of Engineering Science, Urals Division of Russian Academy of Sciences, Yekaterinburg, Russia

- Universal parallelism that leads to general-purpose parallel architectures (multiprocessor or multicomputer systems) having numerous restrictions for performance increasing connectivity of the algorithms graphs, decision time losses because of processors interconnections and memory access conflicts [2]. This kind of parallelism is not connected with real algorithms features and widely used in supercomputers for scientific computations and modeling, where performance-cost efficiency is not as important as for real-time computer systems (RTCS). On the other hand, numerous software tools for these general-purpose parallel computer architectures are quite well developed. As a result, projects of supercomputers are performed for a short time, but efficiency of data processing is low. General-purpose parallelism is not promising for data processing systems with strong restrictions both to hardware cost and processing time, for example, for RTCS.
- Special-purpose parallelism connected with features of applied algorithms. Use of this kind of parallel data processing leads to special-purpose parallel architectures that provide efficient computations of certain functions, procedures, or fragments of algorithms. They are more effective for real-time algorithms with high complexity of computations than general-purpose architectures. So-called homogenous computer environments (HCE) are certain kinds of special-purpose parallel architectures. They can provide extremal performance but not for all algorithms, only for multi-streaming data processing with large-scale input data.

- Hardware implementation of functions. Modern computer architectures based upon hardware support for such complicated procedures as multiplication and division. This is an effective means of achieving high performance of processor and computer as a whole. Further increase of computer performance is closely connected with hardware supporting for elementary mathematical functions, vector-matrix procedures, numerical methods of solving equations, etc. Thus besides general-purpose CPU, the architecture of high-performance computer may consist of numerous special hardware units that can perform fixed set of transformations (up to one). All of these blocks (usually called as coprocessors or FPU) work in parallel, so we have special-purpose parallelism in this case. Furthermore not only computer architecture but hardware blocks architecture individually supporting complicated computations may realize this kind of parallelism. Computer performance increasing by means of additional special-purpose blocks do not require frequency increase because of these blocks based on same technology as CPU, for example, CMOS. This direction of performance increase is not connected with direct power consumption, so it may be efficient for future RTCS.

Thus special-purpose parallelism based on supporting the complicated computational procedures by special-purpose hardware blocks is efficient and promising for providing high-performance computations in RTCS. These blocks may be represented as a certain class of processors. Their instruction sets and architectures are oriented for fast computations of specific functions, procedures, or algorithms, so

they are called functional-oriented processors (FOP) [3]. There are many problems of the FOP development, which are not investigated enough, the main among them are:

– Mathematical models that describe the interconnection between cost and performance of computation (that is hardware and time complexity) are developed only for some computational procedures of computer arithmetic (multiplication and division at first), elementary functions, and some operations for linear algebra. The theory of optimization of FOP architectures on cost-performance criteria based on complexity estimation is not sufficiently developed.
– Parallel computation technologies for procedures mentioned above are poorly designed especially for VLSI implementation.

Modern CAD software does not contain different procedure for the design of rational FOP architectures on the logical gate base. Successful resolution of the different problems of theory and development of FOP architectures is important.

The paper is devoted to a certain class of FOP—reconfigurable HCE consisting of bit-level processor element (PE). These 2D-arrays permit one to achieve very high performance for digital signal and image processing and may be useful for the implementation of some complicated numerical algorithms.

3.2 Reconfigurable Processor Arrays as Base for FOP

Reconfigurable processor arrays (RPA) containing bit-level processor elements (PE) belong to the type of special-purpose processor architectures [4]. According to the opinion of many scientists and designers, RPA are preferable for real-time computations in hard time, space, and power restrictions. In the paper, we consider the special case of locally connected RPA, representing a special kind of uniform computing environments [5]. These RPA may be represented by both array of bit-level PE (fine-grained parallelism) and set of clusters consisting of RPA (coarse-grained parallelism).

High efficiency of RPA applications in RTS is enabled due to:

– mass parallelism for data processing (total number of data flows has an order of ten thousands), which provides maximum of performance and relatively low switching frequency (and hence low power consumption);
– architecture homogeneity (repeatability of PE and regularity of interconnections) which provides maximum of VLSI integration level;
– locality of information lines in RPA (interconnections between adjacent PE only) which provides minimal length of inter-circuits connections in VLSI chip and therefore minimal level of noise;
– great number of identical PE in one chip (up to several thousand PE in one chip) which provides high fault tolerance of computations.
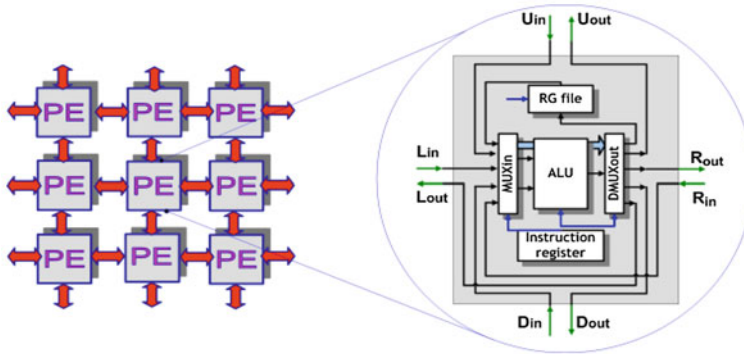
Fig. 3.1 Example of RPA architecture

The main RPA feature is distributed data processing based on a set of locally connected PE. Each PE is an elementary arithmetic and logic unit. It performs basic operations such as addition, multiplication, shift, etc. From this point of view, PE is similar to the first ALUs which were realized as a LSI in 60th. In addition, each PE uses a set of interfaces for their mutual interconnection inside RPA. Various RPAs differ in a data word width, a number of instructions, and an interconnection topology.

One example of RPA architecture is shown in Fig. 3.1.

There are two main categories which are important for the development of RPA that must be used for image and signal processing. They are data structures and informational structure of algorithms. In most cases RPAs realize data flow processing. The number of data flows depends on the number of PE and is limited only by hardware expenditures; therefore, peculiarities of homogenous architectures do not present a strong restriction for data processing. RPA is a multi-pipeline machine so if input data flows are continuous (i.e., have not any gaps), then output data flows are continuous as well. Therefore RPA architectures are scalable. Data flow algorithms commonly used in RTS are preferable for RPA. In this case, we can obtain the optimal combination between RTS algorithms and RPA architectures. Evolution of RPA counts about 40 years. During this period, many architectures, a lot of hardware and software were developed. The main areas of RPA applications in real-time systems (RTS) are as follows:

– optical systems for missiles and aircrafts;
– data processing in radar guidance;
– data processing in hydro acoustic system of underwater vehicles;
– pre-processing in inertial sensors of air- and space-born navigation systems;
– adaptive formation of antenna directional diagrams for radio navigation systems.

This paper is devoted to development of one type of RPA and implementation of some RTS algorithms by using that RPA.

3.3 RPA MiniTera-2

The architecture of RPA MiniTera-2 that was developed in 2002 [4] represents a rectangular array of identical cells connected to each other by means of only local data channels. Each cell contains:

– one PE performing bitwise data processing of data flows;
– shadow memory module for fast rewriting of the PE's operation codes that is needed for dynamically reconfiguration PE array during data processing;
– control unit for the execution of instructions written in control register before data processing.

The architecture of one cell is shown in Fig. 3.2.

The main features of RPA MiniTera-2 cell are synchronous data flows processing in "on-fly" mode and programmable number of digits for addition/subtraction. Sixteen programmable interfaces permit switching of any four channels concurrently with data flow processing. Instruction set consists of 49 instructions: 22 ALU instructions, 14 instructions for comparison operations, remaining instructions are used for data flow commutation, code conversion such as P → S and S → P. All cells are combined in rectangular arrays; each cell is connected with four adjacent cells. Data channels of each cell are intended for concurrent interchanges between (i, j)-th cell and (i-1, j), (i + 1, j), (i, j-1), (i, j + 1)-th cells. Special part of op code of each PE is needed for programming these interchanges. All cells operate synchronously with one bit clock from synchronizing generator.
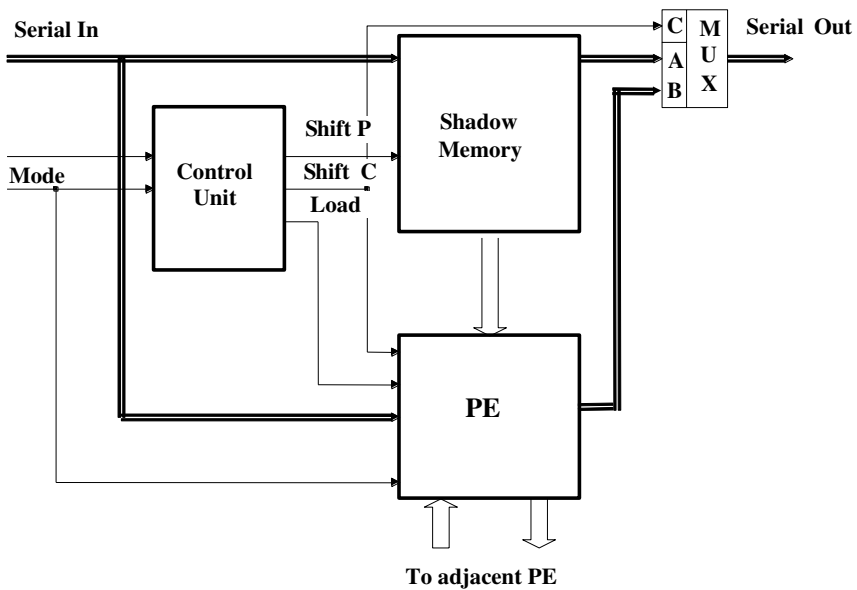

Fig. 3.2 RPA MiniTera-2 cell architecture
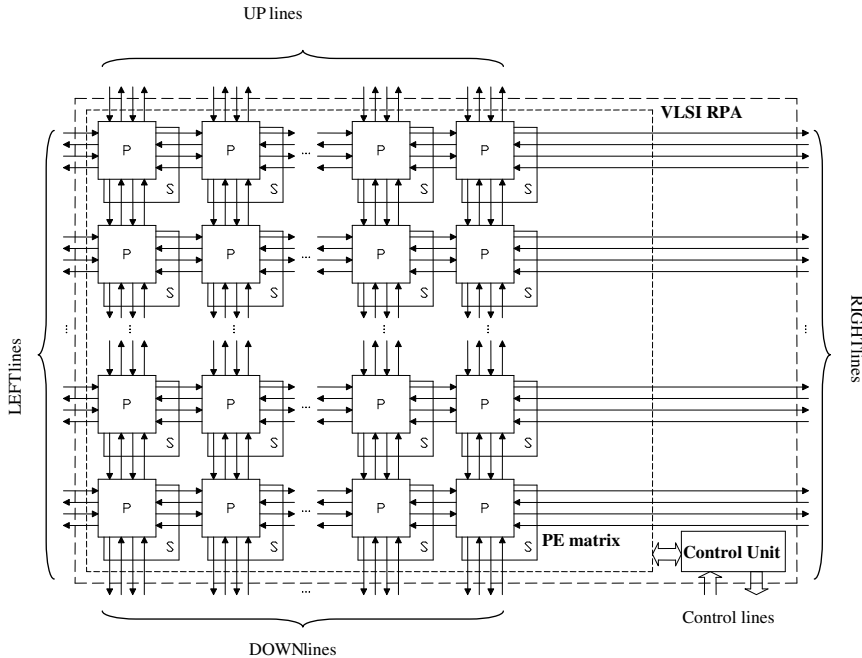
Fig. 3.3 RPA MiniTera-2 architecture

Before data processing one instruction and one mode of interchange must be written to control register of each RPA cells. This enables after initial setup all PE's univocal correspondence between algorithm graph and cell array topology. The structure of RPA MiniTera-2 is shown in Fig. 3.3.

One VLSI chip contains 2D-RPA. Its size or dimension depends on technology. In 2002, a pre-production model of VLSI MiniTera-2 was developed and fabricated. This experimental chip was made on $0.6\,\mu$m CMOS technology and contained cell matrix 5×5 PE, frequency of operation $f = 30$ MHz. Simulation by SPICE revealed for 90 nm CMOS technology 10,000 cells per one chip and 6 W power consumption for $f = 100$ MHz.

In 2003, a model of functional-oriented processor (FOP) based upon MiniTera-2 was developed. In this year, a computer workstation for real-time image processing was realized. Some application-specific image processing algorithms supported by workstation operated successfully in real time. FOP consisting of 288 VLSI (7200 PE) permitted one to get a performance about 20 GOPS for image processing. It allowed realizing HDTV algorithms with frame rate 100 Hz. Workstation and FOP based upon RPA MiniTera-2 are shown in Fig. 3.4.

In addition in 2003, the software IDE MiniTera-2 for programming and debugging application tasks was developed. The main principle of software operation is graphical programming and debugging of user programs on the base of simulation

Fig. 3.4 FOP based upon RPA MiniTera-2 and computer workstation

model of each PE and then 2D-array. The basic components of graphical part of IDE are:

- Algorithm Graph Editor;
- PE Array Editor;
- Inspector;
- ASM Tool;
- Librarian;
- Debugger;
- Project Panel;
- Console.

User interface IDE MiniTera-2 is based upon standard MDI GUI, thus any project may be created and debugged very fast. The main window of IDE is drawn in Fig. 3.5.

The main features of RPA MiniTera-2 architecture are

- data flow processing based on the homogenous computing environment;
- RPA performance may be increased by simple and unlimited expansion of array dimension (by simple splicing of additional chips);
- any topologies of PE network for specific algorithms are realized by preliminary setup every PE.

These features enable the maximal concurrency for data processing but require the adaptation of arbitrary algorithms (so-called "systolization of algorithms") [6] to homogenous RPA architecture. It is one of the most difficult problems in modern

Fig. 3.5 IDE main window

computer science, and successful solution of this task determines the efficiency of RPA application in RTS.

3.4 Realization of RTS Algorithms by Means of RPA MiniTera-2

From all possible RTS applications, let us choose three: image processing, arithmetic operations on long numbers, and real-time sorting. Implementation of algorithms of these tasks by means of MiniTera-2 is discussed below, in more details for image processing and briefly for two other tasks.

3.4.1 Image Processing

Intra-frame processing of noise dot images is one of the most widely used tasks for mobile RTS such us video-channels for missiles and aircrafts. Firstly, they belong to the most complicated algorithms requiring maximal computational performance. Secondly, these algorithms have considerable internal parallelism therefore their
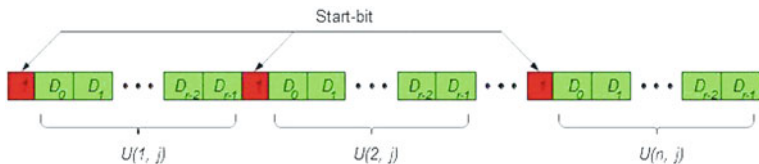
Fig. 3.6 Format for jth line of pixel matrix

systolic processing is preferable. Dot image filtering is basic algorithm of intra-frame image processing; determination and selection of local extremes (most bright pixels) in frame is most complicated procedure of filtering. Usually image matrix in real sensors contains both pixel patterns and pixel noise that is why the main task is a separation of false pixels from real ones. It is impossible to know a number of false and real pixels before image processing; therefore, we need to check every local area around every pixel. For real conditions, total number of pixels to be checked has an order of ten thousands.

Local extreme extraction algorithm is based on the comparison of the signal amplitude U in coordinate (i, j) with signal amplitudes in eight neighbor coordinates:

$$if\, U(i, j) \geq U \begin{pmatrix} i \pm 1, & j \pm 1 \\ i, & j \pm 1 \\ i \pm 1, & j \end{pmatrix}, \text{then } B(i, j) = 1, else\, B(i, j) = 0 \qquad (3.1)$$

Variable B is Boolean and presents the sign of comparison result, and $0 \leq i \leq (n - 1); 0 \leq j \leq (n - 1)$.

One of the principal conditions for correct implementation of this algorithm by RPA MiniTera-2 is a strong synchronization of pixel flows during their processing. It is the requirement for a structure of input pixel flows. All lines of pixels pass into PE array on horizontal rows. Every input data flow contains continuous sequence of pixels; each pixel moves along its line from might digit (Fig. 3.6).

Local extreme extraction by means of MiniTera-2 is based upon parallel streaming comparison each matrix element $U(i, j)$ with every adjacent element in window of size 3×3. Parallel systolic comparison of central pixel with adjacent ones in window of size 3×3 may be realized by rectangular macros that contains 15 PE. A structure of such macros is represented in Fig. 3.7.

PE macros handle three lines in parallel and every pixel in one line in serial. The number of pixels in each line may be unlimited. Continuous flows of pixels with signs of local extremes are output data of macros.

Macro described above is "building brick" for extracting local extremes for any size of image. It is important that increasing of a number of lines (number of matrix columns) is provided by simple expansion of macros. Increasing of number of pixels in line (number of matrix rows) does not require any increasing at all. So upper bound of hardware complexity estimation is represented as a linear function:
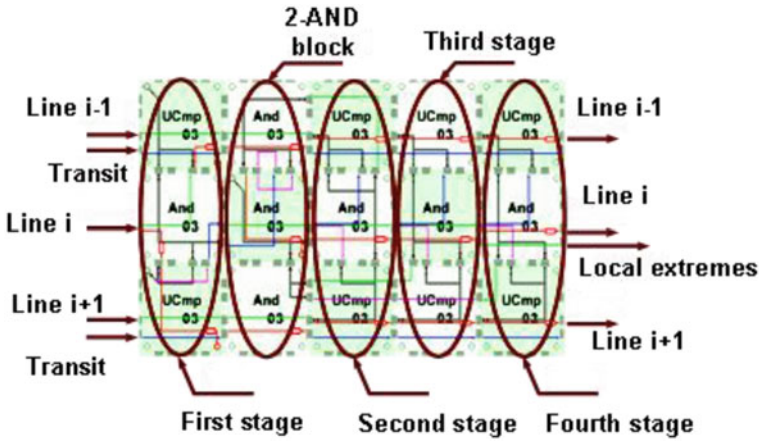
Fig. 3.7 Systolic macros for local extremes extraction into window size 3×3

$$L_{\text{h}} = O(N) \qquad\qquad (3.2)$$

where L_h is a number of PE, N is a number of pixels in row of image matrix. Comparison of a central pixel with an adjacent may be performed in four stages as follows:

$$1^0 Comp\,(U_{i-1,\,j} \leftrightarrow U_{i,\,j} \leftrightarrow U_{i-1,\,j})$$

$$2^0 Comp\,(U_{i-1,\,j-1} \leftrightarrow U_{i,\,j} \leftrightarrow U_{i-1,\,j+1})$$

$$3^0 Comp\,(U_{i,\,j-1} \leftrightarrow U_{i,\,j} \leftrightarrow U_{i,\,j+1})$$

$$4^0 Comp\,(U_{i+1,\,j-1} \leftrightarrow U_{i,\,j} \leftrightarrow U_{i+1,\,j+1})$$

Extraction of local extremes for full-size image matrix may be provided by splicing macros between each other. Using this methodology one may realize local extremes extraction algorithm for any size of the images.

The upper bound of time complexity for local extreme extraction in window of size 3×3 by means of MiniTera-2 may be represented as a linear function (similar to L_h) as follows:

$$L_t = 2r + 19 \qquad\qquad (3.3)$$

where r is a number of bits per one pixel.

RPA MiniTera-2 extracts all local extremes in image with size 512×512 16-bit pixels during 8700 cycles. For example, video processor based upon such DSP as TMS320C67XXX needs near 30000 cycles.

Extraction of local extremes is widely used in the various numerical algorithms in mathematical physics. Therefore, macros described above may be applied not only to image processing.

3.4.2 Multiplication of Very Long Numbers

Systolic RPA MiniTera-2 provides high efficiency for arithmetic algorithms. Pipeline multiplication of very long numbers is more effective than well-known parallel schemes. For example, for 8192-bit factors MiniTera-2 requires 1.5 million CMOS gates and 66.5 K cycles and matrix multiplier requires 400 million CMOS gates and 32.7 K cycles. The main goal of our today researches is decreasing multiplication time for very long numbers by parametrical decomposition of cofactors and partial products.

3.4.3 Sorting of Large Massive

Using the MiniTera-2 and Shearsort sorting algorithm, it is possible to sort a large massive with upper bounds of time complexity as follows:

$$L_t = O\left(\sqrt{M} \log M\right) \tag{3.4}$$

where M is number of massive elements [7]. Many parallel sorting algorithms on clusters and meshes provide an average upper bound of time complexity as $L_t = O(M \log M)$ [8]. This effect of systolic sorting is based on two steps of following algorithm:

- Transformation of a linear massive to square matrix.
- Parallel 2D-sorting concurrently on row of matrix ("serpent" sorting) and on column of matrix (usual sorting, for example, ascending sort). The main goal of our today researches is fast 2D-sorting on hybrid architectures (CPU + RPA).

3.5 Conclusion

Reconfigurable processor arrays are effective for high performance data processing in real-time systems. Results of the research and developments presented in the paper show that for three important tasks: image processing, high-precision arithmetic, and parallel sorting, RPA MiniTera-2 provides advantages on system criteria "cost $\times$ performance".

References

1. Bennett P (2004) The why, where and what of low-power SoC design//EEdesign.com. http://www.eetimes.com/document.asp?doc_id=1276973
2. Manke JW, Kerlick GD, Levine D, Banerjee S, Dillon E (2001) Parallel performance of two applications in the Boeing high performance computing benchmark suite. Parallel Comput 27:457–475
3. Lookin NA, oth (2007) VLSI based on two-dimensional reconfigurable array of processor elements and their implementation for numerical algorithms in real-time systems. Proceedings of IP based electronics system, 2007. Espace Congrés du World Trade Center 5 place Robert Schuman, 38500 Grenoble, France, pp 541–546
4. Lookin NA (2004) Reconfigurable processor arrays for real-time systems: architectures, efficiency, application. Proceedings of international science school high-performance computer systems, Taganrog, Russia, pp 44–59
5. Prangishvili IV, oth (1967) Microelectronics and homogenous structures for logical and computational devices, [in Russian]—science. Moscow
6. S.Y. Kung, *VLSI array processors* (Prentice Hall, New Jersey, 1988)
7. Lookin NA, Souvorova PG (2004) Realization of fast 2D-sorting by means of uniform computer environment [in Russian]. In: Proceedings of international science school high-performance computer systems, Taganrog, Russia, pp 59–63
8. Miller R, Boxer L (2000) Algorithms sequential and parallel: a unified approach. Pearson Education, Prentice Hall, New Jersey

Chapter 4
Image Fusion Based on Wavelet Transformation

Vladimir A. Trenikhin and Victor G. Kobernichenko

4.1 Introduction

One of the most serious problems of remotely sensed data complex use, received from different satellite systems, is the need to process information of different spectral bands with different spatial resolution. This problem occurs when data of optical multispectral systems and their combination with radiometric and radar images are being processed [1].

The main part of modern satellite systems provided for remote sensing (such as Landsat, SPOT, IRS, IKONOS, QuickBird, WorldView-2, Result-P, and so on) has ability for obtaining multispectral and radar images with different spatial resolution (Table 4.1).

Nowadays, there exist a lot of image fusion methods for effective use of these data. These methods presuppose the acquisition of the multispectral color image with high spatial resolution from fusing low-resolution multispectral image with high-resolution panchromatic image [2–5]. In general, the image fusion procedure consists of the following stages:

1. The source images transformation to the usability form for high-resolution details injection from the high-resolution panchromatic image to the other.
2. The sampling frequency alignment of source images and the following interpolation (bicubic, linear, nearest neighbor, and so on).
3. The high-resolution details injection to intensity components of transformed images.
4. The inverse transformation to the RGB-base.

The numerous procedures of image fusion differ by methods applied on the first stage, injection ways on the third stage, and second stage interpolation algorithms.

V. A. Trenikhin · V. G. Kobernichenko (✉)
Ural Federal University, Yekaterinburg, Russia
e-mail: vgkobern@mail.ru

41

Table 4.1 The spatial resolution of optical remote sensing space systems

Space systems (SS)		Landsat, ETM+	SPOT, HRGI	QuickBird, BHRC-60	WorldView-2, WV-110	Resurs-P, Geoton-L1
Spatial resolution (m)	Panchrom. mode	15	5 (2.5)	0.61	0.46	0.9
	Multisp. mode	30, 60	10	2.5	1.84	3–4
Panchrom. spectral band (μm)		0.52–0.90	0.48–0.71	0.45–0.90	0.45–0.80	0.58–0.86

The detailed overview is given in the article [6]. After carrying out the image fusion procedure, resulted images allow to significantly improve capability of thematic decoding. Meanwhile this procedure leads the color failure effect to the resulted image. In order to reduce the impact of the color failure effect, there are a lot of strategies, each of them is designed for the particular variant of image fusion methods and different source data. The significant color errors in the resulted image become more serious after the emergence of ultrahigh-resolution remote sensing data. In front of SPOT and IRS SS panchromatic images, wavelength band of QuickBird, OrbView, and Resurs-P are extended from visible to the near-infrared spectral region (see Table 4.1). It leads to significant changes of gray-scale level of panchromatic images. Therefore, the image processing methods, which have proved to be useful for fusion of panchromatic SPOT images with other multispectral data, cannot provide qualitative results when fusing satellite ultrahigh-resolution images [7].

4.2 Image Fusion Based on Wavelet Transformation

This paper is dedicated to one of the implementations of image fusion based on wavelet transformation. The algorithm based on wavelet transformation is presented on Fig. 4.1.

Using the fast wavelet transformation, algorithm permits one to represent an original panchromatic image as a set of an approximation component **A** and a detail component **D**.

$$\mathbf{S}(\tau) = \mathbf{A}_k(\tau) + \sum_{l=1}^{m} \mathbf{D}_{lk}(\tau) \tag{4.1}$$

where $\mathbf{S}(\tau) = \{s_1; \ldots, s_m\}$ is the original signal vector, $\mathbf{A}_k(\tau) = \{a_{1;k;\ldots}; a_{mk};_k\}$ is the approximation vector, $\mathbf{D}_{l,k}(\_) = \{d_{1,k};\ldots; d_{m\text{-}mk},_k\}$ is a detail vector, $m_k = m/2^k$, k is the level of wavelet transformation. Components of approximation and detail vectors can be written as following recursion formulae:
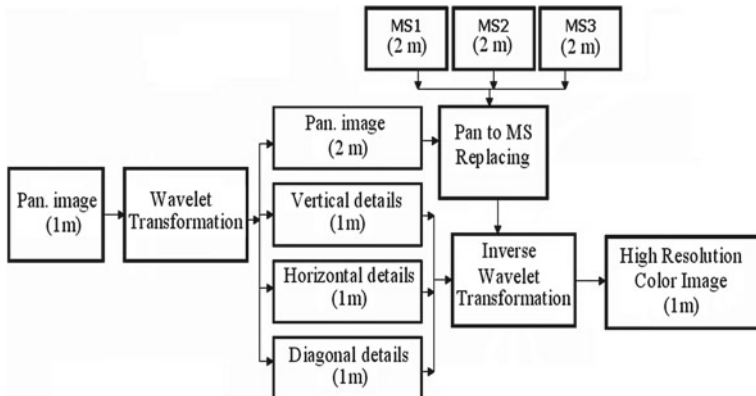
Fig. 4.1 The image fusion based on wavelet transformation scheme

$$a_{j+1,k} = \sum_m h_m a_{j,2k+m}$$

$$d_{j+1,k} = \sum_m g_m a_{j,2k+m} \qquad (4.2)$$

$$a_{0,k} = S_k$$

Approximation coefficients $\{h_1, \ldots, h_m\}$ and $\{g_1, \ldots, g_m\}$ or approximation **H** and detail **G** filters are determined by the type of wavelet being used. From the equations it is seen that at each step of wavelet transformation, one approximation and detail coefficient of n-th step image is defined by the linear combination of m samples and includes aggregated information about all incoming samples. When using wavelet transformation, the image fusion method is implemented as follows. The high-resolution panchromatic image is transformed to the low-resolution panchromatic image and wavelet coefficients for each level of transformation (spatial details). The low-resolution panchromatic image is replaced by low-resolution multispectral images with the same resolution level. After that inverse wavelet transformation is performed for aggregation spatial details and color information into high-resolution color image. The replace and the inverse transformation are performed three times—one per each color channel of multispectral image.

Therefore, high-resolution spatial details are injected into multispectral image [4]. The research of the image fusion method based on wavelet transformation (see Fig. 4.1) has demonstrated availability of this method only for using with *Haar*-wavelet. This is because an approximation **H** and detail **G** filters of this type of wavelet has the support, which length is 2, while approximation filter coefficients are equal. Therefore, at each step of wavelet transformation, one approximation coefficient of n-th step is formed by using of two samples of $n-1$-th step of transformation. As a result of n-steps transformation, we obtain a 2^n times reduced copy of source image and, that is why, it can be replaced by the low-resolution multispectral image. When

using other types of wavelets, as a result of n-th wavelet transformations, we obtain an image 2^n times less than an original one, although this image does not correspond to the reduced copy of the original image, because for all types of wavelets excluding *Haar*-wavelet, the support length of approximation and detail filter is more than 2.

4.3   The Modified Image Fusion Algorithm

To solve the problem described above, a new modified algorithm is proposed. In the first step of this algorithm, the sampling frequency of source multispectral images should be increased to the sampling frequency of panchromatic source image. After that, a forward wavelet transformation is carried out for each panchromatic and MS channel using the necessary type of wavelet. Next step involves replacing MS details with panchromatic details, and lastly, an inverse wavelet transformation should be carried out for each multispectral channel. The modified algorithm scheme is given on Fig. 4.2.

In the paper, the analysis of suggested algorithm efficiency has been performed using the image fragments of industrial territory obtained by the QuickBird satellite system in panchromatic and three multispectral bands (see Fig. 4.3). The image was provided by Ural Regional Information Analysis Center "Uralgeoinform". Special attention was paid to the analysis of resolution and color failure characteristics. In order to estimate image resolution before and after fusion, the slice analysis was used. The slice was performed along the line crossing images of buildings,
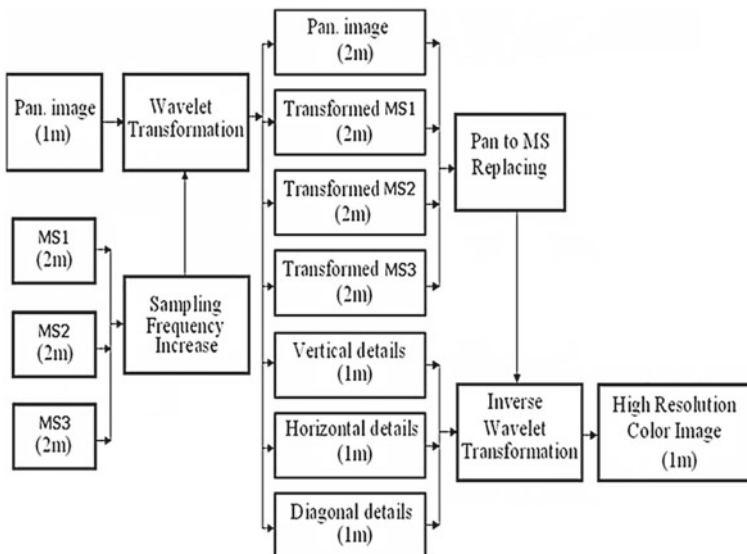

Fig. 4.2 The modified image fusion algorithm

Fig. 4.3 The fragment of
QuickBird space image.
Panchromatic channel

roads, and technological constructions with different dimension. The image slice
is the graph of brightness as a function of distance along a section line. Source
images and results of Db4-wavelet using slices of images are given on Figs. 4.4, 4.5,
4.6, 4.7, 4.8.

The color failure effect can be estimated by looking at differences between top
points of the slice image graph on the Fig. 4.8.

Fig. 4.4 The fragment of
space image. The color
image

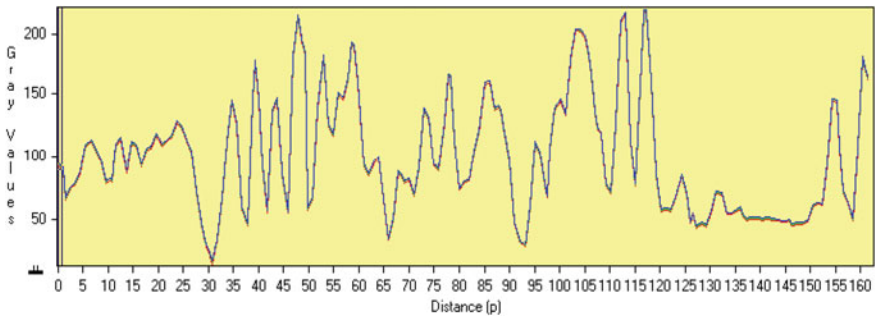**Fig. 4.5** Resulted image using Db4-wavelet

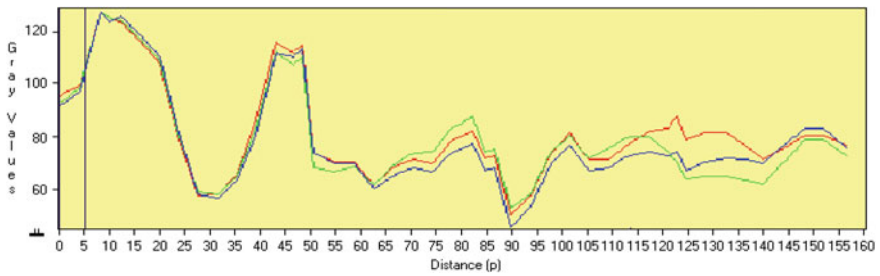**Fig. 4.6** The panchromatic image slice (resolution 0.61 m)

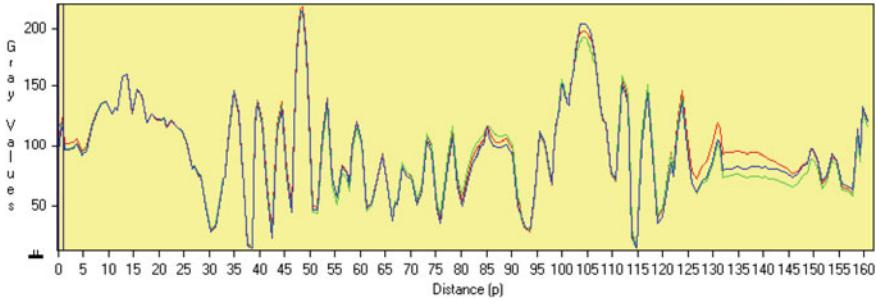**Fig. 4.7** The color image slice (resolution 2.4 m)

Fig. 4.8 The color images slice after image fusion procedure

The color failure effect analysis revealed that all types of wavelets introduce color errors. However, on-type landscape areas were clearly resolved by color. It means that final high-resolution images obtained after image fusion based on wavelet transformation processing are suitable for thematic decoding.

In the same levels of sampling frequency case, it is convenient to estimate the color failure effect using correlation coefficients between processed channels. The principle of this method is the calculation of correlation coefficients between multispectral channels and panchromatic channel before and after image fusion processing.

The calculation is produced using the well-known formula:

$$\rho_{x_{b1},x_{b2}} = \frac{\mathrm{cov}(x_{b1}(n), x_{b2}(n))}{\sqrt{D[x_{b1}(n)] \times D[x_{b2}(n)]}}, \tag{4.3}$$

where $x_{b1}(n)$, $x_{b2}(n)$—the brashness value of the pixel in the n-point of images obtained in b_1 and b_2 frequency bands; cov(..,..)—the covariance of two random variables; $D[..]$—the dispersion of the random variable.

The example of analysis for Db4-wavelets is shown in Table 4.2.

Table 4.2 Correlation coefficients obtained with Db4-wavelets using

ρ		Source resolution/resulted resolution			
b1	b2	1/1 (Source image)	1/4	1/8	1/16
RED	GREEN	0.957	0.968	0.975	0.985
RED	BLUE	0.974	0.977	0.981	0.991
GREEN	BLUE	0.983	0.984	0.974	0.981
PAN	RED	0.372	0.5	0.767	0.988
PAN	GREEN	0.396	0.527	0.717	0.995
PAN	BLUE	0.328	0.46	0.689	0.975

Analysis results demonstrate the increase of the correlation coefficient between final multispectral and panchromatic images when carrying out the image fusion procedure. The increase of the correlation coefficient is due to injection of the panchromatic information to multispectral channels. This leads to the loss of part of color information and to the reduction of the chromaticity of resulted image. Meanwhile, the image fusion processing does not lead to the conversion of analyzed objects from one class to the other. It allows one to resolve these objects totally clear by color. Therefore, the main reason of the emergence of the color failure effect is the injection of the pancrhome power to the resulted image. The comparison of processing results with different wavelets families showed the best performance with using *Coif* and *Bior* wavelets families. *Coif* and *Bior* wavelet families also allow one to resolved bright points much better. This is because these wavelets families have much more null moments.

4.4   Conclusions

When assessing the correlation coefficient of the color failure effect analysis, it appeared that this coefficient tends to one asymptotically with increasing difference between scales of source panchromatic and multispectral images. The increment of the correlation coefficient directly depends on the scale of source multispectral image. The analysis of image fusion, based on wavelet transformation, is also able to detect implementations of method, which are less susceptible to the color failure effect. However, it is worth mentioning, that no implementations of the image fusion method work for multispectral images whose scale is 16 times less than the scale of panchromatic image.

References

1. Weydahl DJ, Becquey X, Tollefsen T, Combining ERS-1 SAR with optical satellite data over urban areas, in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, vol. 3, pp. 2161–2163 (1995)
2. C. Pohl, J.L. Van Genderen, Multisensor image fusion in remote sensing: concepts, methods and applications. Int. J. Remote Sens. **19**, 823–854 (1998)
3. T. Ranchin, B. Aiazzi, L. Alparone, S. Baronti, L. Wald, Image fusion—the ARSIS concept and some successful implementation schemes. ISPRS J. Photogramm. Remote Sens. **58**, 4–18 (2003)
4. P.S. Chavez Jr., S.C. Sides, J.A. Anderson, Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic. Photogramm. Eng. Remote Sens. **57**(3), 295–303 (1991)
5. Vald Lucien, *Definitions and Architectures* (Les Presses de l Ecole des Mines. Paris, Fusion of Images of Different Spatial Resolutions, 2002)

6. V.G. Kobernichenko, V.A. Trenikhin, Methods for fusing images based on different resolution remote sensed data. Achiev. Mod. Radioelectron. **4**, 22–31 (2007). (in Russian)
7. Y. Zhang, A new merging method and its spectral and spatial effects. Int. J. Remote Sens. **20**, 2003–2014 (1999)

Chapter 5
Many-Factor MIMO-Filters

**Valeriy G. Labunets, Denis E. Komarov, Victor P. Chasovskikh
and Ekaterina Ostheimer**

5.1 Introduction

We develop a conceptual framework and design methodologies for multichannel image many-factor (bilateral, 3-, and 4-lateral) aggregation filters with assessment capability. The term "multichannel" (multicomponent, multispectral, hyperspectral) image is used for an image with more than one component. They are composed of a series of images in different optical bands at wavelengths $\lambda_1, \lambda_2, \ldots, \lambda_k$ called spectral channels:

$$\mathbf{f}(x, y) = \left(f_{\lambda_1}(x, y), f_{\lambda_2}(x, y), \ldots, f_{\lambda_K}(x, y) \right) = (f_1(x, y), f_2(x, y), \ldots, f_K(x, y)),$$

where K is the number of different optical channels, *i.e.*, $\mathbf{f} \colon \mathbf{R}^2 \to \mathbf{R}^K$, where $\mathbf{R}^K$ is multicolor space. The bold font for $\mathbf{f}(x, y)$ emphasizes the fact that images may be multichannel. Each pixel in $\mathbf{f}(x, y)$, therefore, represents the spectrum at the wavelengths $\lambda_1, \lambda_2, \ldots, \lambda_K$ of the observed scene at point $\mathbf{x} = (x, y)$.

Let us introduce the observation model and notions used throughout the paper. We consider noised image in the form $\mathbf{f}(\mathbf{x}) = \mathbf{s}(\mathbf{x}) + \mathbf{n}(\mathbf{x})$, where $\mathbf{s}(\mathbf{x})$ is the original gray-level image and $\mathbf{n}(\mathbf{x})$ denotes the noise introduced into $\mathbf{s}(\mathbf{x})$ to produce the corrupted image $\mathbf{f}(\mathbf{x})$. The aim of image enhancement is to reduce the noise as much as possible or to find a method which, given $\mathbf{s}(\mathbf{x})$, derives an image $\hat{s}(\mathbf{x})$ as close as possible to the original $\mathbf{s}(\mathbf{x})$, subjected to a suitable optimality criterion.

V. G. Labunets (✉) · D. E. Komarov · V. P. Chasovskikh
Ural State Forest Engineering University, Yekaterinburg, Russia
e-mail: vlabunets05@yahoo.com

Ural Federal University, Yekaterinburg, Russia

E. Ostheimer
Capricat LLC, Florida, USA

The standard bilateral filter (BF) [1–9] with a square N-cellular window $M(\mathbf{x})$ is located at $\mathbf{x}$, the weighted average of pixels in the moving window replaces the central pixel

$$\hat{\mathbf{s}}(\mathbf{x}) = \mathbf{BilMean}_{\mathbf{p} \in M(\mathbf{x})}\big[w(\mathbf{x}, \mathbf{p}) \cdot \mathbf{f}(\mathbf{p})\big] = \frac{1}{k(\mathbf{x})} \sum_{\mathbf{p} \in M(\mathbf{x})} w(\mathbf{x}, \mathbf{p}) \cdot \mathbf{f}(\mathbf{p}), \qquad (5.1)$$

where $\hat{s}(\mathbf{x})$ is the filtered image and $k(\mathbf{p})$ is the normalization factor

$$k(\mathbf{x}) = \sum_{\mathbf{p} \in M(\mathbf{x})} w(\mathbf{x}, \mathbf{p}). \qquad (5.2)$$

Equation (5.1) is simply a normalized weighted average of a neighborhood of a N-cellular window $M(\mathbf{x})$ (*i.e.*, the mask around pixel $\mathbf{x}$, consisting of N pixels). The scalar-valued weights $w(\mathbf{x}, \mathbf{p})$ are computed based on the content of the neighborhood. For pixels $\{\mathbf{f}(\mathbf{p})\}_{\mathbf{p} \in M(\mathbf{x})}$ around the centroid $\mathbf{f}(\mathbf{x})$, the weights $\{w(\mathbf{x}, \mathbf{p})\}_{\mathbf{p} \in M(\mathbf{x})}$ are computed by multiplying the following two factors:

$$w(\mathbf{x}, \mathbf{p}) = w_{Sp}(\mathbf{p}) \cdot w_{Rn}(\mathbf{x}, \mathbf{p}) = w_{Sp}(||\mathbf{p}||) \cdot w_{Rn}(||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{p})||_2).$$

The weight includes two factors – spatial $w_{Sp}(||\mathbf{p}||)$ and radiometric $w_{Rn}(\mathbf{x}, \mathbf{p}) = w_{Rn}(||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{p})||_2)$ weights. The first weight measures the geometric distance $||\mathbf{p}||$ between the center pixel $f(\mathbf{x})$ and the pixel $\mathbf{f}(\mathbf{p})$ (note, the centroid $\mathbf{x}$ has the position $\mathbf{0} \in M(\mathbf{x})$ inside of the mask $M(\mathbf{x})$). Here the Euclidean metric $||\mathbf{p}|| = ||\mathbf{p}||_2$ is applied. This way, close-by pixels influence the final result more than distant ones. The second weight measures the radiometric distance between the values of the center sample $f(\mathbf{x})$ and the pixel $\mathbf{f}(\mathbf{p})$, and again, the Euclidean metric $||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{p})||_2$ is chosen, too. Therefore, pixels with close-by values tend to influence the final result more than those having distant value.

This paper considers two natural extensions to the bilateral filter. Firstly, instead of the center pixel $\mathbf{f}(\mathbf{x})$ in $w_{Rn}(||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{p})||)$, we use a certain mean or median $\bar{\mathbf{f}}(\mathbf{x})$ (for example, the *Fréchet median* $\bar{\mathbf{f}}_{opt}(\mathbf{x})$ [11]) of a neighborhood of a N-cellular window $M(\mathbf{x})$ for calculating weights $w_{Rn}(\mathbf{x}, \mathbf{p}) = w_{Rn}(||\bar{\mathbf{f}}(\mathbf{x}) - \mathbf{f}(\mathbf{p})||_2)$. Secondly, instead of a scale-valued weigh, we use a matrix-valued one

$$\hat{\mathbf{s}}(\mathbf{x}) = \mathbf{BilMean}_{\mathbf{p} \in M(\mathbf{x})}\big[\mathbf{W}(\mathbf{x}, \mathbf{p}) \cdot \mathbf{f}(\mathbf{p})\big], \qquad (5.3)$$

where $\mathbf{W}(\mathbf{x}, \mathbf{p})$ are the matrix-valued weights.

5.2 The First Modification of Bilateral Filters

In this modification, we use the suboptimal *Fréchet median* $\bar{\mathbf{f}}_{opt}(\mathbf{x})$ for calculating weights $w_{Rn}(\mathbf{x}, \mathbf{p})$ instead of the center pixel $\mathbf{f}(\mathbf{x})$ in w_{Rn}: $w_{Rn}(\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{p})\|) \rightarrow w_{Rn}(\|\bar{\mathbf{f}}_{opt}(\mathbf{x}) - \mathbf{f}(\mathbf{p})\|)$.

Let $\langle \mathbf{R}^K, \rho \rangle$ be a metric space, where ρ is a distance function (*i.e.*, $\rho : \mathbf{R}^K \times \mathbf{R}^K \rightarrow \mathbf{R}^+$). Let $w_1, w_2, \ldots, w_N$ be N weights summing to 1 and let $\{\mathbf{f}^1, \mathbf{f}^2, \ldots, \mathbf{f}^N\} = \mathbf{D} \subset \mathbf{R}^K$ be N pixels in the N-cellular window $M(\mathbf{x})$.

Definition 1 [10, 11]. The *optimal Fréchet point* associated with the metric ρ, is the point, $\mathbf{f}_{opt} \in \mathbf{R}^K$, that minimizes the Fréchet cost function (FCF) $\sum_{i=1}^{N} w_i \rho(\mathbf{f}, \mathbf{f}^i)$ (the weighted sum distances from an arbitrary point $\mathbf{f}$ to each point $\mathbf{f}^1, \mathbf{f}^2, \ldots, \mathbf{f}^{N^2} \in \mathbf{R}^K$). It is formally defined as:

$$\mathbf{f}_{opt} = \mathbf{FrechMed}\big(\rho | \mathbf{f}^1, \mathbf{f}^2, \ldots, \mathbf{f}^N\big) = \arg\min_{\mathbf{f} \in \mathbf{R}^K} \sum_{i=1}^{N} w_i \rho\big(\mathbf{f}, \mathbf{f}^i\big). \qquad (5.4)$$

Note that **argmin** means the argument, for which the sum is minimized. This generalizes the ordinary median, which has the property of minimizing the sum of distances for one-dimensional data, and provides a central tendency for higher dimensions.

In computation point of view, it is better to restrict the search domain from $\mathbf{R}^K$ until the set $\mathbf{D} = \{\mathbf{f}^1, \mathbf{f}^2, \ldots, \mathbf{f}^N\} \subset \mathbf{R}^K$. In this case, we obtain definition of the *suboptimal Fréchet point or the optimal vector Fréchet median*.

Definition 2 [10, 11]. The *suboptimal weighted Fréchet point* or *optimal Fréchet median* associated with the metric ρ is the point, $\bar{\mathbf{f}}_{opt} \in \{\mathbf{f}^1, \mathbf{f}^2, \ldots, \mathbf{f}^N\} \subset \mathbf{R}^K$, that minimizes the FCF over the restrict search domain $\mathbf{D} \subset \mathbf{R}^K$:

$$\bar{\mathbf{f}}_{opt} = \mathbf{FrechMed}\big(\rho | \mathbf{f}^1, \mathbf{f}^2, \ldots, \mathbf{f}^N\big) = \underset{\mathbf{f} \in \mathbf{D}}{\mathbf{argmin}} \sum_{i=1}^{N^2} w_i \rho\big(\mathbf{f}, \mathbf{f}^i\big). \qquad (5.5)$$

Example 1 If observation data are real numbers, *i.e.*, $f^1, f^2, \ldots, f^N \in \mathbf{R}$ and distance function is the city distance $\rho(f, g) = \rho_1(f, g) = |f - g|$, then the optimal Fréchet point (5.4) and optimal Fréchet median (5.5) for gray-level pixels $f^1, f^2, \ldots, f^N \in \mathbf{R}$ to be the classical *Fréchet point* and *median,* respectively, *i.e.*,

$$\mathbf{f}_{opt} = \mathbf{FrechPt}\big(\rho_1 | f^1, f^2, \ldots, f^{N^2}\big) = \arg\min_{f \in \mathbf{R}} \sum_{i=1}^{N} \big| f - f^i \big|, \qquad (5.6)$$

$$\bar{\mathbf{f}}_{opt} = \mathbf{FrechMed}\big(\rho_1 | f^1, f^2, \ldots, f^N\big) = \arg\min_{f \in \mathbf{D}} \sum_{i=1}^{N} \big| f - f^i \big| =$$

$$= \mathbf{Med}\big(f^1, f^2, \dots, f^N\big). \tag{5.7}$$

Example 2 If observation data are vectors, *i.e.,* $\mathbf{f}^1, \mathbf{f}^2, \dots, \mathbf{f}^N \in \mathbf{R}^K$, and distance function is the city distance $\rho(\mathbf{f}, \mathbf{g}) = \rho_1(\mathbf{f}, \mathbf{g})$, then the optimal Fréchet point (5.4) and optimal Fréchet median (5.5) for vectors $\mathbf{f}^1, \mathbf{f}^2, \dots, \mathbf{f}^N \in \mathbf{R}^K$ to be the *Fréchet point* and *the Fréchet vector median*, associated with the same metric $\rho_1(\mathbf{f}, \mathbf{g})$,

$$\mathbf{f}_{opt} = \mathbf{FrechPt}\big(\rho_1 | \mathbf{f}^1, \mathbf{f}^2, \dots, \mathbf{f}^N\big) = \underset{\mathbf{f} \in \mathbf{R}^K}{\mathbf{argmin}}\left(\sum_{i=1}^{N} ||\mathbf{f} - \mathbf{f}^i||_1\right), \tag{5.8}$$

$$\bar{\mathbf{f}}_{opt} = \mathbf{FrechMed}\big(\rho_1 | \mathbf{f}^1, \mathbf{f}^2, \dots, \mathbf{f}^N\big) = \underset{f \in \mathbf{D}}{\arg\min}\left(\sum_{i=1}^{N} ||\mathbf{f} - \mathbf{f}^i||_1\right) =$$

$$= \mathbf{VecMed}\big(\rho_1 | \mathbf{f}^1, \mathbf{f}^2, \dots, \mathbf{f}^N\big). \tag{5.9}$$

Now we use Fréchet median $\bar{\mathbf{f}}_{opt}$ for calculating radiometric weights: $w_{Rn}(\mathbf{x}, \mathbf{p}) = w_{Rn}\big(||\bar{\mathbf{f}}_{opt}(\mathbf{x}) - \mathbf{f}(\mathbf{p})||_2\big)$. The modified bilateral filter is given as

$$\hat{\mathbf{s}}(\mathbf{x}) = \underset{\mathbf{p} \in M(\mathbf{x})}{\mathbf{BilMean}}\big[w(\mathbf{x}, \mathbf{p}) \cdot \mathbf{f}(\mathbf{p})\big]$$

$$= \frac{1}{k(\mathbf{x})} \sum_{\mathbf{p} \in M(\mathbf{x})} w_{Sp}(||\mathbf{p}||) \cdot w_{Rn}\big(||\bar{\mathbf{f}}_{opt}(\mathbf{x}) - \mathbf{f}(\mathbf{p})||_2\big) \cdot \mathbf{f}(\mathbf{x}), \tag{5.10}$$

where $\hat{\mathbf{s}}(\mathbf{x})$ is the filtered image.

5.3 The Second Modification: Four-Factor MIMO-Filters

In the case of multichannel images, processed data are vector-valued $\mathbf{f}(\mathbf{x}) : \mathbf{R}^2 \to \mathbf{R}^K : \mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_K(\mathbf{x})) = [f_c(\mathbf{x})]_{c=1}^K$. By this reason, we must use matrix-valued weights $\{\mathbf{W}(\mathbf{x}, \mathbf{p})\}_{\mathbf{p} \in M(\mathbf{x})}$, where $\mathbf{W}(\mathbf{x}, \mathbf{p})$ is a $(K \times K)$-matrix, and K is the number of different channels in $\mathbf{f}(\mathbf{x}) : \mathbf{R}^2 \to \mathbf{R}^K$. The 4-factor MIMO-filter suggests a weighted average of pixels in the given image

$$\hat{\mathbf{s}}(\mathbf{x}) = \underset{\mathbf{p} \in M(\mathbf{x})}{\mathbf{MIMO}^4\mathbf{FactMean}}\big[\bar{\mathbf{W}}(\mathbf{x}, \mathbf{p}) \cdot \mathbf{f}(\mathbf{p})\big] =$$

$$= \frac{1}{\mathbf{diag}\{k_1(\mathbf{x}), k_2(\mathbf{x}), \dots, k_K(\mathbf{x})\}} \sum_{\mathbf{p} \in M(\mathbf{x})} \mathbf{W}(\mathbf{x}, \mathbf{p}) \cdot \mathbf{f}(\mathbf{p}), \tag{5.11}$$

or in component-wise form

$$\hat{s}_a(\mathbf{x}) = \frac{1}{k_a(\mathbf{x})} \sum_{\mathbf{p} \in M(\mathbf{x})} \sum_{b=1}^{K} w^{ab}(\mathbf{x}, \mathbf{p}) \cdot f_b(\mathbf{p}) = \sum_{\mathbf{p} \in M(\mathbf{x})} \sum_{b=1}^{K} \bar{w}^{ab}(\mathbf{x}, \mathbf{p}) \cdot f_b(\mathbf{p}), \quad (5.12)$$

where $\hat{\mathbf{s}}(\mathbf{x})$ is the filtered multichannel image, $\hat{s}_a(\mathbf{x})$ is its ath channel, $\bar{w}^{ab} = w^{ab}/k_a$, $\bar{\mathbf{W}}(\mathbf{x}, \mathbf{p}) = \mathbf{diag}\{k_1^{-1}(\mathbf{x}), k_2^{-1}(\mathbf{x}), \ldots, k_K^{-1}(\mathbf{x})\} \cdot \bar{\mathbf{W}}(\mathbf{x}, \mathbf{p})$ and $k_a(\mathbf{p})$ is the normalization factor in the ath channel:

$$k_a(\mathbf{x}) = \sum_{\mathbf{p} \in M(\mathbf{x})} \sum_{b=1}^{K} w^{ab}(\mathbf{x}, \mathbf{p}) \quad (5.13)$$

and $\mathbf{diag}\{k_1(\mathbf{x}), k_2(\mathbf{x}), \ldots, k_K(\mathbf{x})\}$ is diagonal matrix with channel normalization factors. Note, that

$$\frac{1}{\mathbf{diag}\{k_1(\mathbf{x}), k_2(\mathbf{x}), \ldots, k_K(\mathbf{x})\}} \mathbf{W}(\mathbf{x}, \mathbf{p}) \cdot \mathbf{f}(\mathbf{p}) = \bar{\mathbf{W}}(\mathbf{x}, \mathbf{p}) \cdot \mathbf{f}(\mathbf{p}) =$$

$$= \begin{bmatrix} k_1^{-1}(\mathbf{x}) & & \vdots \\ & k_2^{-1}(\mathbf{x}) & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ & & \vdots & k_K^{-1}(\mathbf{x}) \end{bmatrix} \begin{bmatrix} w^{11}(\mathbf{x}, \mathbf{p}) & w^{12}(\mathbf{x}, \mathbf{p}) & \cdots & w^{1K}(\mathbf{x}, \mathbf{p}) \\ w^{21}(\mathbf{x}, \mathbf{p}) & w^{22}(\mathbf{x}, \mathbf{p}) & \cdots & w^{2K}(\mathbf{x}, \mathbf{p}) \\ \vdots & \vdots & \vdots & \vdots \\ w^{K1}(\mathbf{x}, \mathbf{p}) & w^{K2}(\mathbf{x}, \mathbf{p}) & \cdots & w^{KK}(\mathbf{x}, \mathbf{p}) \end{bmatrix} \begin{bmatrix} f_1(\mathbf{p}) \\ f_2(\mathbf{p}) \\ \vdots \\ f_K(\mathbf{p}) \end{bmatrix}.$$

The normalized matrix-valued weights $\bar{\mathbf{W}}(\mathbf{x}, \mathbf{p})$ are computed based on the content of the neighborhood. For pixels $\mathbf{f}(\mathbf{p})$, $\mathbf{p} \in M(\mathbf{x})$ around the Fréchet centroid $\bar{\mathbf{f}}_{opt}(\mathbf{x})$, the scalar-valued weights $\bar{w}^{cd}(\mathbf{x}, \mathbf{p})$ of the matrices $\bar{\mathbf{W}}(\mathbf{x}, \mathbf{p})$, $\mathbf{p} \in M(\mathbf{x})$ are computed by multiplying the following four factors:

$$\bar{w}^{cd}(\mathbf{x}, \mathbf{p}) =$$
$$= \bar{w}_{Sp}(||\mathbf{p}||) \cdot \bar{w}_{Ch}(|c - d|) \cdot \bar{w}_{Rn}\big(||\bar{\mathbf{f}}_{opt}(\mathbf{x})) - \mathbf{f}(\mathbf{p})||_2\big) \cdot \bar{w}_{Rn}\big(|\bar{f}_{c,opt}(\mathbf{x}) - f_d(\mathbf{p})|\big).$$

The weight includes four factors: spatial $\bar{w}_{Sp}(||\mathbf{p}||)$, interchannels $\bar{w}_{Ch}(|c - d|)$, global radiometric $\bar{w}_{Rn}\big(||\bar{\mathbf{f}}_{opt}(\mathbf{x})) - \mathbf{f}(\mathbf{p})||_2\big)$, and radiometric interchannel weights $\bar{w}_{Rn}\big(|\bar{f}_{c,opt}(\mathbf{x}) - f_d(\mathbf{p})|\big)$. The first factor $\bar{w}_{Sp}(||\mathbf{p}||)$ measures the geometric distance between the center pixel $\bar{\mathbf{f}}_{opt}(\mathbf{x})$ and the neighborhood pixel $\mathbf{f}(\mathbf{p})$, $\mathbf{p} \in M(\mathbf{x})$. The second factor $\bar{w}_{Ch}(|c - d|)$ measures the spectral (interchannel) distance. The third factor $w_{Rn}\big(||\bar{\mathbf{f}}_{opt}(\mathbf{x}) - \mathbf{f}(\mathbf{p})||_2\big)$ measures the global radiometric distance between the values of the Fréchet center $\bar{\mathbf{f}}_{opt}(\mathbf{x})$ and the pixels $\mathbf{f}(\mathbf{p})$, $\mathbf{p} \in M(\mathbf{x})$. The fourth factor $\bar{w}_{Rn}\big(|\bar{f}_{c,opt}(\mathbf{x}) - f_d(\mathbf{p})|\big)$ measures the radiometric distance between the values of the center sample $\bar{f}_{c,opt}(\mathbf{x})$ of the c-channel and the pixel $f_d(\mathbf{p})$, $\mathbf{p} \in M(\mathbf{x})$ of the d-channel. All weights $\bar{w}_{Rn}^{cd}(\mathbf{x}, \mathbf{p}) = \bar{w}_{Rn}\big(|\bar{f}_{c,opt}(\mathbf{x}) - f_d(\mathbf{p})|\big)$ form N radiometric interchannel $(K \times K)$-matrices

$$\{\bar{\mathbf{W}}_{Rn}(\mathbf{x}, \mathbf{p})\}_{\mathbf{p} \in M(\mathbf{x})} = \left\{ \big[\bar{w}_{Rn}^{cd}(\mathbf{x}, \mathbf{p})\big]_{c,d=1}^{K} \right\}_{\mathbf{p} \in M(\mathbf{x})} = \left\{ \big[\bar{w}_{Rn}\big(|\bar{f}_{c,opt}(\mathbf{x}) - f_d(\mathbf{p})|\big)\big]_{c,d=1}^{K} \right\}_{\mathbf{p} \in M(\mathbf{x})}.$$

if N-cellular window is used. We obtain 3-factor MIMO-filters if we are going to use only three ingredients, for example,

$$\bar{w}^{cd}(\mathbf{x}, \mathbf{p}) = \bar{w}_{Sp}(||\mathbf{p}||) \cdot \bar{w}_{Rn}\big(||\bar{\mathbf{f}}_{opt}(\mathbf{x}) - \mathbf{f}(\mathbf{p})||_2\big) \cdot \bar{w}_{Rn}\big(\big|\bar{f}_{c,opt}(\mathbf{x}) - f_d(\mathbf{p})\big|\big)$$

or $\bar{w}^{cd}(\mathbf{x}, \mathbf{p}) = \bar{w}_{Sp}(||\mathbf{p}||) \cdot \bar{w}_{Ch}(|c - d|) \cdot \bar{w}_{Rn}\big(\big|\bar{f}_{c,opt}(\mathbf{x}) - f_d(\mathbf{p})\big|\big).$

5.4 Simulation Experiment

Some variants of the proposed filters are tested. They are compared on real image "LENA". Noise is added (see Fig. 5.1) with different Peak Signal to Noise Rations (PSNRs). The noised image has 1% noised pixels (PSNR = 25.36 dB), 5% noised pixels (PSNR = 18.34 dB), 10% noised pixels (PSNR = 15.41 dB), 20% noised pixels (PSNR = 12.64 dB), 50% noised pixels (PSNR = 9.22 dB). Figures 5.2, 5.3 and 5.4 summarize the results for "Salt and Pepper" noise and bilateral filters with Laplacian weights $w(f, g) = \exp(-\alpha|f - g|)$ for different $\alpha = 0.035$ (Fig. 5.2), $\alpha = 0.07$ (Fig. 5.3), $\alpha = 0.1$ (Fig. 5.4). Figures 5.2, 5.3 and 5.4 show the results obtained by the following bilateral filters with $\boxplus_3$-mask

- the classical bilateral filter (5.1) (**BF3 × 3**),
- modified bilateral filter (5.10) (**BF3 × 3Med**), where $\bar{\mathbf{f}}(\mathbf{x})$ is calculated as classical median in each channel,
- modified bilateral filter (5.10) (**BF3 × 3Fr1**), where $\bar{\mathbf{f}}(\mathbf{x})$ is calculated as Fréchet median $\bar{\mathbf{f}}(\mathbf{x}) = \bar{\mathbf{f}}_{opt}(\mathbf{x})$ with distance $\rho(\mathbf{f}, \mathbf{g}) = \rho_1(\mathbf{f}, \mathbf{g})$,
- modified bilateral filter (5.10) (**BF3 × 3Fr2**), where $\bar{\mathbf{f}}(\mathbf{x})$ is calculated as Fréchet median $\bar{\mathbf{f}}(\mathbf{x}) = \bar{\mathbf{f}}_{opt}(\mathbf{x})$ with distance $\rho(\mathbf{f}, \mathbf{g}) = \rho_2(\mathbf{f}, \mathbf{g})$,
- modified bilateral filter (5.10) (**BF3 × 3Fr∞**), where $\bar{\mathbf{f}}(\mathbf{x})$ is calculated as Fréchet median $\bar{\mathbf{f}}(\mathbf{x}) = \bar{\mathbf{f}}_{opt}(\mathbf{x})$ with distance $\rho(\mathbf{f}, \mathbf{g}) = \rho_\infty(\mathbf{f}, \mathbf{g})$.

It is easy to see that results for all modified bilateral filters are better, compared to the classical bilateral filter **BF3 × 3**.

5.5 Conclusions and Future Work

A new class of nonlinear generalized 2-, 3-, and 4-factor MIMO-filters for multichannel image processing is introduced in this paper. Weights in 4-factor MIMO-filters include four components: spatial, radiometric, interchannel, and interchannel radiometric weights. The fourth weight measures the radiometric distance (for gray-level images) between the interchannel values of the center scalar-valued channel pixels and local neighborhood channel pixels. Here, the 1D Euclidean metric is used, too. We are going to use in (5.10) and (5.11) a generalized average (aggregation) [11–13] instead of ordinary mean.

Fig. 5.1 Original and noised images. Noise: Salt-Pepper. **a** Original image; **b** noised images, PSNR = 25.36; **c** noised images, PSNR = 18.34; **d** noised images, PSNR = 15.41; **e** noised images, PSNR = 12.64; **f** noised images, PSNR = 9.22
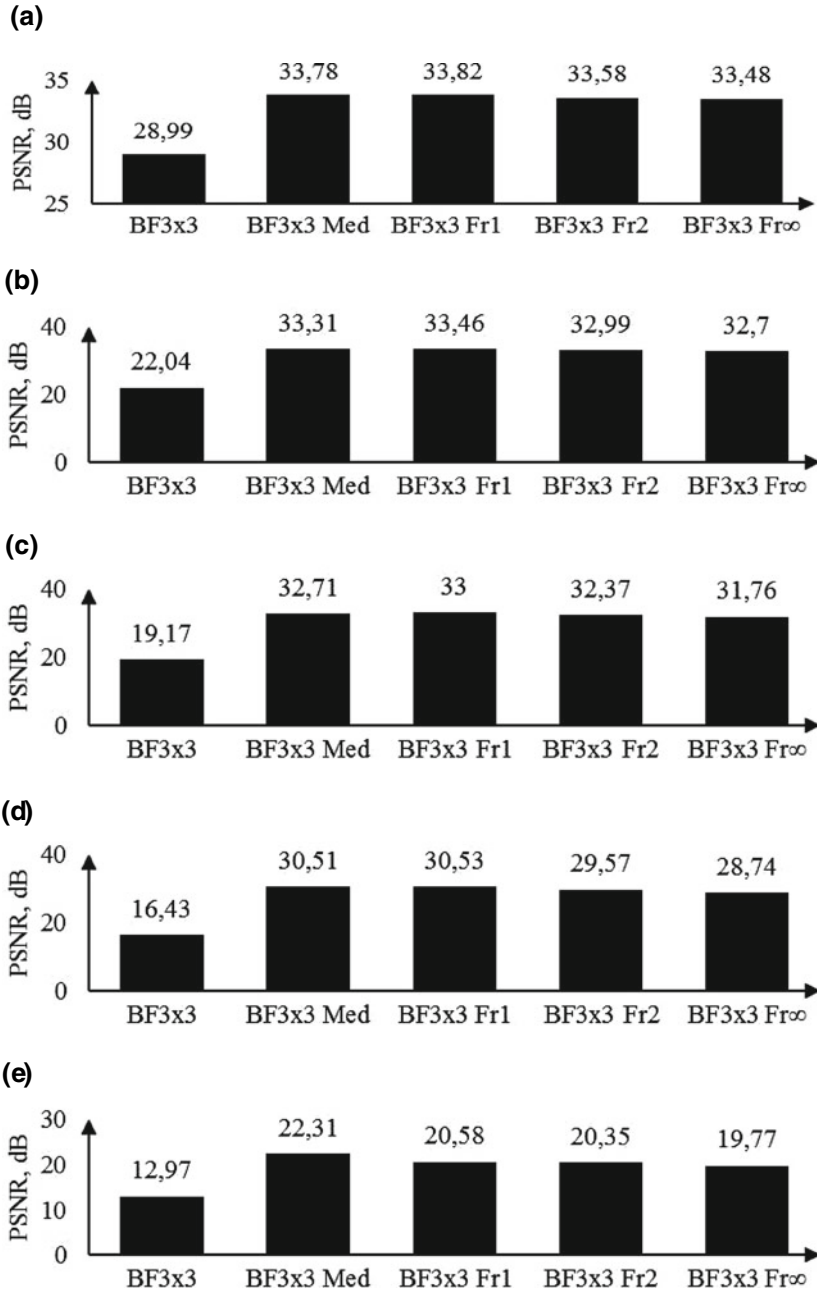
(a)

(b)

(c)

(d)

(e)

Fig. 5.2 The results for "Salt and Pepper" noise and bilateral filters with Laplacian weights for $\alpha = 0.035$. **a** Filtration of "Salt and Pepper" noise with 1% noised pixels; **b** filtration of "Salt and Pepper" noise with 5% noised pixels; **c** filtration of "Salt and Pepper" noise with 10% noised pixels; **d** filtration of "Salt and Pepper" noise with 20% noised pixels; **e** Filtration of "Salt and Pepper" noise with 50% noised pixels
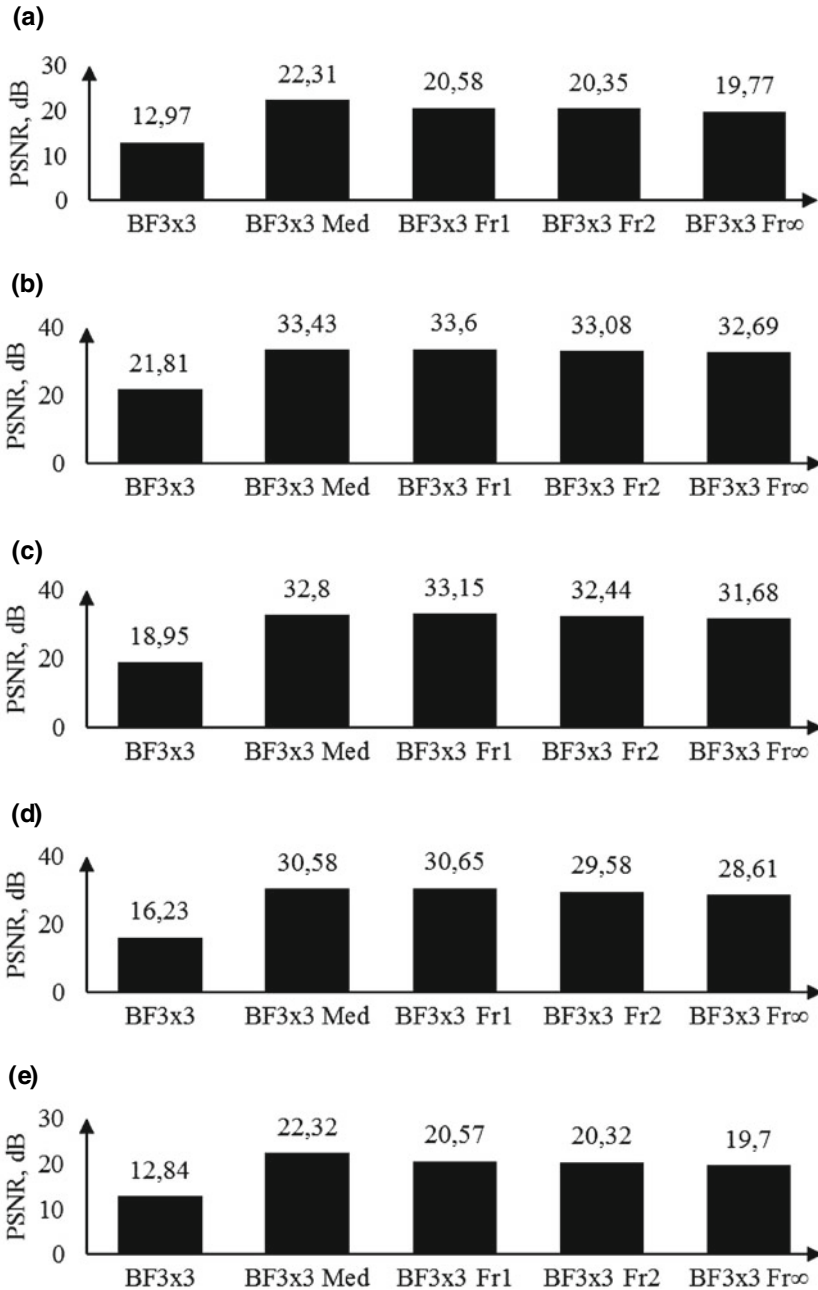
Fig. 5.3 The results for "Salt and Pepper" noise and bilateral filters with Laplacian weights for $\alpha = 0.07$. **a** Filtration of "Salt and Pepper" noise with 1% noised pixels; **b** filtration of "Salt and Pepper" noise with 5% noised pixels; **c** filtration of "Salt and Pepper" noise with 10% noised pixels; **d** filtration of "Salt and Pepper" noise with 20% noised pixels; **e** Filtration of "Salt and Pepper" noise with 50% noised pixels
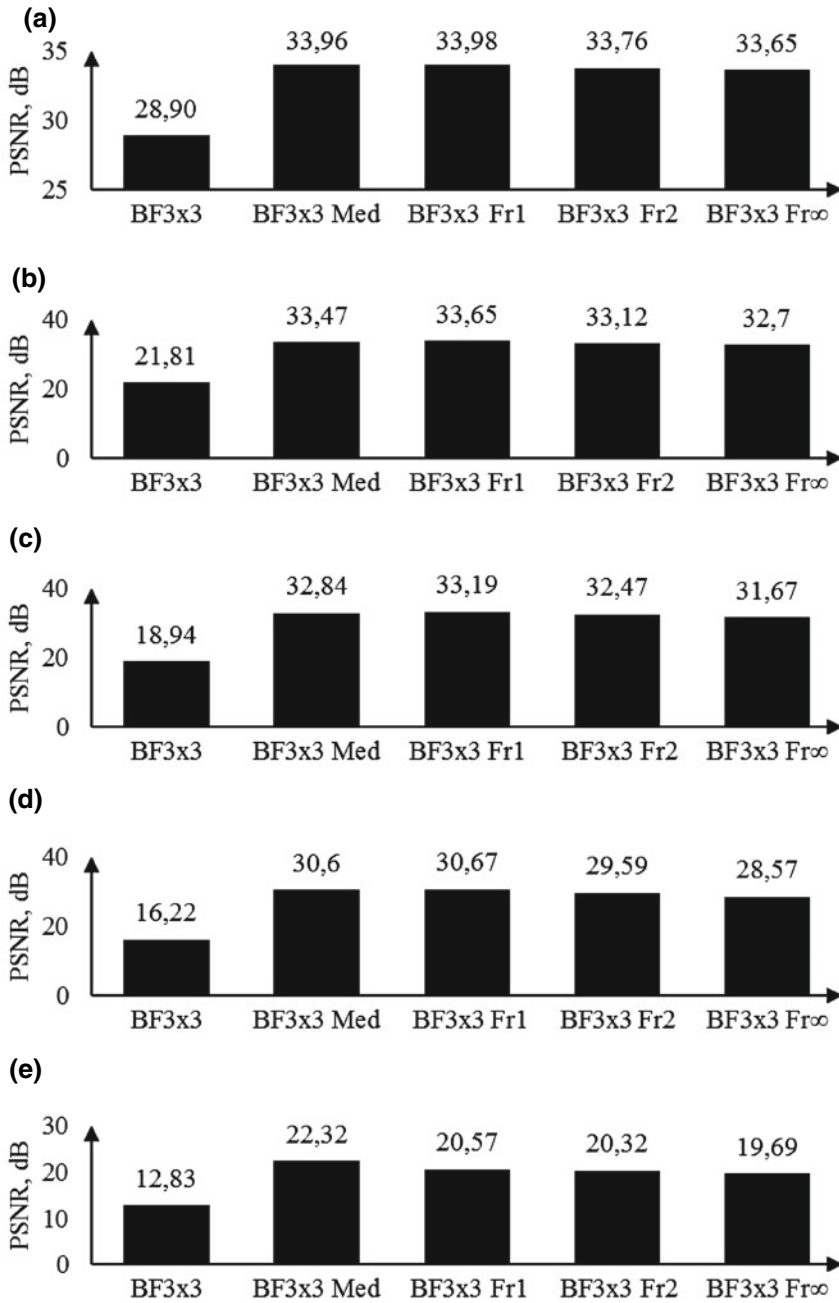
(a)

(b)

(c)

(d)

(e)

Fig. 5.4 The results for "Salt and Pepper" noise and bilateral filters with Laplacian weights for $\alpha = 0.1$. **a** Filtration of "Salt and Pepper" noise with 1% noised pixels; **b** filtration of "Salt and Pepper" noise with 5% noised pixels; **c** filtration of "Salt and Pepper" noise with 10% noised pixels; **d** filtration of "Salt and Pepper" noise with 20% noised pixels; **e** Filtration of "Salt and Pepper" noise with 50% noised pixels

The aggregation problem [11–13] consists in aggregating n-tuples of objects all belonging to a given set S, into a single object of the same set S, i.e., **Agg** $: S^n \to S$. In the case of mathematical aggregation operator (AO) the set S is an interval of the real $S = [0, 1] \subset \mathbf{R}$ or integer numbers $S = [0, 255] \subset \mathbf{Z}$. In this setting, an AO is simply a function, which assigns a number y to any N-tuple $(x_1, x_2, \ldots, x_N)$ of numbers: $y = \mathbf{Aggreg}(x_1, x_2, \ldots, x_N)$ that satisfies:

- **Agg** $(x) = x$,

- **Agg** $\underbrace{a, a, \ldots, a}_{N} = a$. In particular, **Agg** $(0, 0, \ldots, 0) = 0$ and **Agg** $(1, 1, \ldots, 1) = 1$, or

  Agg $(255, 255, \ldots, 255) = 255$.

- $\mathbf{min}(x_1, x_2, \ldots x_N) \leq \mathbf{Agg}(x_1, x_2, \ldots x_N) \leq \mathbf{max}(x_1, x_2, \ldots x_N)$.

Here $\mathbf{min}(x_1, x_2, \ldots, x_N)$ and $\mathbf{max}(x_1, x_2, \ldots, x_N)$ are respectively the *minimum* and the *maximum* values among the elements of $(x_1, x_2, \ldots, x_N)$.

All other properties may come in addition to this fundamental group. For example, if for every permutation $\forall \sigma \in \mathbf{S_N}$ of $\{1, 2, \ldots, N\}$ the AO satisfies:

$$\mathbf{Agg}\big(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(N)}\big) = \mathbf{Agg}(x_1, x_2, \ldots, x_N),$$

then it is invariant (symmetric) with respect to the permutations of the elements of $(x_1, x_2, \ldots, x_N)$. In other words, as far as means are concerned, the *order* of the elements of $(x_1, x_2, \ldots, x_N)$ is—and must be—completely irrelevant.

We list below a few particular cases of means:

(1) Arithmetic mean: $\mathbf{Mean}(x_1, x_2, \ldots, x_N) = \frac{1}{N} \sum_{i=1}^{N} x_i$.
(2) Geometric mean: $\mathbf{Geo}(x_1, x_2, \ldots, x_N) = \exp\left(\frac{1}{N} \sum_{i=1}^{N} \ln x_i\right)$.
(3) Harmonic mean: $\mathbf{Harm}(x_1, x_2, \ldots, x_N) = \left(\frac{1}{N} \sum_{i=1}^{N} \frac{1}{x_i}\right)^{-1}$.
(4) Quasi-arithmetic (power or Hölder) mean of the form: $\mathbf{Power}_p(x_1, x_2, \ldots, x_N) = \left(\frac{1}{N} \sum_{i=1}^{N} x_i^p\right)^{\frac{1}{p}}$. This family is particularly interesting, because it generalizes a group of common means, only by changing the value of p.

The very notable particular cases correspond to the logic functions (min; max; median): $y = \mathbf{Min}(x_1, x_2, \ldots, x_N)$, $y = \mathbf{Max}(x_1, x_2, \ldots, x_N)$, $y = \mathbf{Med}(x_1, x_2, \ldots, x_N)$.

In a 2D standard linear and median scalar filters with a square N-cellular window $M(\mathbf{x})$ and located at $\mathbf{x}$, the mean and median replace the central pixel

$$\hat{s}(\mathbf{x}) = \underset{\mathbf{p} \in M(\mathbf{x})}{\mathbf{Mean}}\big[f(\mathbf{p})\big], \ \hat{s}(\mathbf{x}) = \underset{\mathbf{p} \in M(\mathbf{x})}{\mathbf{Med}}\big[f(\mathbf{p})\big], \tag{5.14}$$

where $\hat{s}(\mathbf{x})$ is the filtered gray-level image, $\{f(\mathbf{p})\}_{\mathbf{p} \in M(\mathbf{x})}$ is an image block of the fixed size N extracted from f by moving N-cellular window $M(\mathbf{x})$ at the position $\mathbf{x}$, **Mean** and **Med** are the mean (average) and median operators. When filters (5.14) are modified as follows

$$\hat{s}(\mathbf{x}) = \underset{\mathbf{p} \in M(\mathbf{x})}{^M\mathbf{Agg}}\big[f(\mathbf{p})\big], \tag{5.15}$$

we the unique class of nonlinear *aggregation SISO-filters* proposed in [14–16], where $^M\mathbf{Agg}$ is an aggregation operator on the mask $M(\mathbf{x})$.

For MIMO-filters, we have to introduce a vector-valued aggregation. Note that for ordinary vector-matrix product $\mathbf{g} = \bar{\mathbf{W}}\mathbf{f}$, we have in component-wise form

$$g_c = \bar{w}_{c1} f_1 + \bar{w}_{c2} f_2 + \ldots + \bar{w}_{cK} f_K = \sum_{d=1}^{K} \bar{w}_{cd} f_d. \tag{5.16}$$

Let us introduce vector-matrix aggregation product $\mathbf{g} = \bar{\mathbf{W}}_{\mathrm{Agg}} \odot \mathbf{f}$ in component-wise form by the following way

$$g_c = \mathrm{Agg}\{\bar{w}_{c1} f_1, \bar{w}_{c2} f_2, \ldots, \bar{w}_{cK} f_K,\} = \mathrm{Agg}_{d=1}^{K}\{\bar{w}_{cd} f_d\}, \tag{5.17}$$

where Agg is an aggregation operator. Obviously, we can use different aggregation operators in different channels

$$g_c = {}^c\mathrm{Agg}\{\bar{w}_{c1} f_1, \bar{w}_{c2} f_2, \ldots, \bar{w}_{cK} f_K,\} = {}^c\mathrm{Agg}_{d=1}^{K}\{\bar{w}_{cd} f_d\}, \tag{5.18}$$

for $c = 1, 2, \ldots, K$. Let $\mathbf{Agg} = \{{}^1\mathrm{Agg}, {}^2\mathrm{Agg}, \ldots, {}^K\mathrm{Agg}\}$ is the K-element set of aggregation operators. In this case, we write $\mathbf{g} = \bar{\mathbf{W}}_{\mathbf{Agg}} \odot \mathbf{f}$

When 4-factor MIMO-filter (5.11) is modified as follows

$$\hat{\mathbf{s}}(\mathbf{x}) = \mathbf{MIMO^4Fact}\underset{\mathbf{p} \in M(\mathbf{x})}{^M\mathbf{Agg}}\big[\bar{\mathbf{W}}_{\mathrm{Agg}}(\mathbf{x}, \mathbf{p}) \odot \mathbf{f}(\mathbf{p})\big], \tag{5.19}$$

or in component-wise form

$$\hat{s}_c(\mathbf{x}) = \underset{\mathbf{p} \in M(\mathbf{x})}{^M\mathbf{Agg}}\big\{{}^c\mathbf{Agg}_{d=1}^{K}\big\{\bar{w}^{cd}(\mathbf{x}, \mathbf{p}) \cdot f_d(\mathbf{p})\big\}\big\} \tag{5.20}$$

we get the unique class of nonlinear *aggregation MIMO-filters* that we are going to research in future works. They are based on $K + 1$ of aggregation operators: 1) $^M\mathrm{Agg}$ (aggregation on the mask $M(\mathbf{x})$) and 2) $\mathbf{Agg} = \{{}^1\mathrm{Agg}, {}^2\mathrm{Agg}, \ldots, {}^K\mathrm{Agg}\}$

(interchannel aggregation), which could be changed independently of one another. For each set of aggregation operators, we get the unique class of new nonlinear filters.

References

1. C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images, in *Proceedings of the 6th International Conference on Computer Vision*, (1998), pp. 839–846
2. J. Astola, P. Haavisto, Y. Neuvo, Vector median filters. IEEE Trans. Image Process. **78**, 678–689 (1990)
3. K. Tang, J. Astola, Y. Neuvo, Nonlinear multivariate image filtering techniques. IEEE Trans. Image Process. **4**, 788–798 (1996)
4. D. Barash, Bilateral filtering and anisotropic diffusion: towards a unified view point. in *Third International Conference on Scale- Space and Morphology* (2001), pp. 273–280
5. F. Durand, J. Dorsey, Fast bilateral filtering for the display of high-dynamic-range images, in *Proceedings of ACM SIGGRAPH* (2002), pp. 257–266
6. M. Elad, Analysis of the bilateral filter. in *The 36th Asilomar on Signals, Systems and Computers*, (Pacific Grove, CA, 2002)
7. M. Elad, On the origin of the bilateral filter and ways to improve it. IEEE Trans. Image Process. **11**(10), 1141–1151 (2002)
8. S. Fleishman, I. Drori, D. Cohen, Bilateral mesh filtering, in *Proceedings of ACM SIGGRAPH*, (San Diego, TX, 2003), pp. 950—953
9. D. Barash, A fundamental relationship between bilateral filtering, adaptive smoothing and the non-linear diffusion equation, PAMI **24**(6) (2002)
10. M. Fréchet, Les elements aleatoires de nature quelconque dans un espace distancie. Ann. Inst. H. Poincare **10**(3), 215–310 (1948)
11. G. Mayor, E. Trillas, On the representation of some aggregation functions, in *Proceeding of ISMVL*, vol. 20, pp. 111–114 (1986)
12. S. Ovchinnikov, On Robust aggregation procedures. in *Aggregation and Fusion of Imperfect Information*, ed. by B. Bouchon-Meunier, (Springer, Berlin, Heidelberg, 1998), pp. 3–10
13. A. Kolmogorov, Sur la notion de la moyenne. Atti Accad. Naz. Lincei **12**, 388–391 (1930)
14. V.G. Labunets, Filters based on aggregation operators. Part 1. aggregation operators, in *Proceedings of the 24th Int. Crimean Conference Microwave& Telecommunication Technology*, vol. 24 (2014), pp. 1239–1240
15. V.G. Labunets, D.N. Gainanov, E. Ostheimer, Filters based on aggregation operators. Part 2. The Kolmogorov filters, in *Proceedings of the 24th International Crimean Conference Microwave & Telecommunication Technology*, vol. 24 (2014), pp. 1241–1242
16. V.G. Labunets, D.N. Gainanov, A.D. Tarasov, E. Ostheimer, Filters based on aggregation operators. Part 3. The Heron filters, in *Proceedings of the 24th International Crimean Conference Microwave& Telecommunication Technology*, vol. 24 (2014), pp. 1243–1244

Chapter 6
Examination of the Clarification Effect Under Colorimetric Object Illumination

S. P. Arapova, S. Yu. Arapov, D. A. Tarasov and A. P. Sergeev

6.1 Introduction

The clarification phenomenon attracted the attention of the authors during the laboratory experiments with combined light source [1], which was originally designed to study deviations from the classical colorimetry described in [2–5]. The effect is to increase the amount of color reflecting surfaces under a light source with a highly pronounced uneven spectral power distribution (SPD). As shown in [6–8], the laboratory light source [1] allows to create in neighboring working areas colorimetrically close lighting with very different SPD (Fig. 6.1). This enables direct comparative observations of clarification. The notion of "colorimetric illuminant" is introduced in [1].

Hunt refers to the clarification in [9] (Chapter 10.13). Manifestation of the clarification effect occurs in the chromaticity diagrams in the works devoted to criticism of the calculation method of CIE color rendering index (CRI) Ra [7]. The point of criticism is the discrepancy between the value of Ra and visually observable color rendering. The different approach to the calculation of CRI is offered instead.

The apparent causal relationship between the clarification effect and low color rendering index forces the clarification to be considered as undesirable and even harmful. This arises from the fact that "the illuminant with a low Ra" is habitually associated with the notion of "bad illuminant", although one should not speak of a direct link. For example, in [10] the approach is based on the calculation of gamut volume as a number of distinguishable colors. It is not ruled out that in this case, the opposite colorimetric lighting could outperform others.

S. P. Arapova · S. Yu. Arapov · D. A. Tarasov (✉) · A. P. Sergeev
Ural Federal University, Ekaterinburg, Russia
e-mail: datarasov@yandex.ru

D. A. Tarasov · A. P. Sergeev
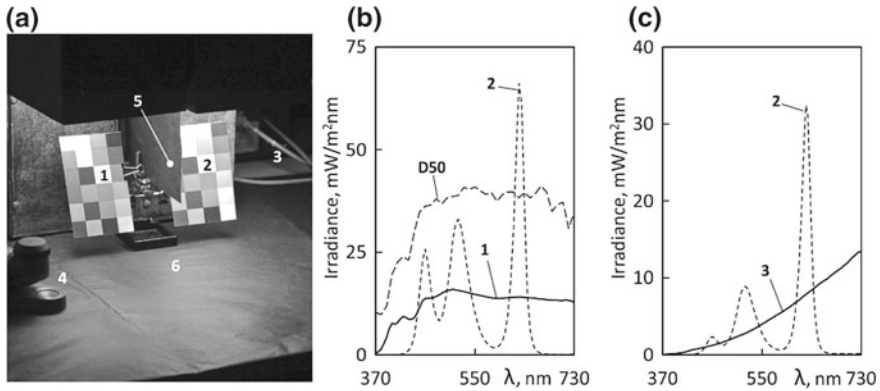Institute of Industrial Ecology, UB RAS, Ekaterinburg, Russia

Fig. 6.1 Schematic of the experiment. **a** The light source combined with 2 work areas: 1 (close to D50) and 2 (equivalent to RGB). Areas 1 and 2 are separated by an opaque partition 5. Colorimetric equalization area 2 and an external source located in the zone 3 is possible. 4—spectrophotometer, 6—mixing illumination zone. **b**, **c**—spectral irradiance distribution. Signatures 1, 2, 3 correspond to working areas

Thus, the aim of this paper is to explain the effect of clarification and to evaluate the possibility of its controlled use in reproduction technology.

6.2 Causes of the Clarification Effect

The essence of the effect is quite simple. As is known, the most visually rich colors are created by monochromatic spectral stimuli. The locus of pure spectral colors in different chromaticity diagrams defines the boundary of perception of color saturation by the observer. With the expansion of the stimulus from the narrowband SPD to the distribution over the entire visible range, there is a reduction of its color saturation, and an approximation to the achromatic, neutral colors (Fig. 6.2), and in the end to the equal-energy stimulus (Illuminant E).

The reflection spectra of most reflective surfaces in the natural environment have quite smooth curves, which together with uniform standard illuminants such as A or D-series create stimuli with a wide range of apparently distributed incentives. Thus, the maximum saturated color is essentially unavailable for a large number of observed objects, including printed and prints obtained using a standard colorants triad. The most frequently cited estimate gamut of incentives is found in the real world [11], but later and other options are put forward, for example in [12].

The situation changes significantly when using light sources with a very uneven spectrum. When lighting a flood greatly reduced capacity in some areas of the spectrum, there may be an increase in the perceived color saturation of the reflecting surface due to narrowing of the spectral distribution of the stimulus.
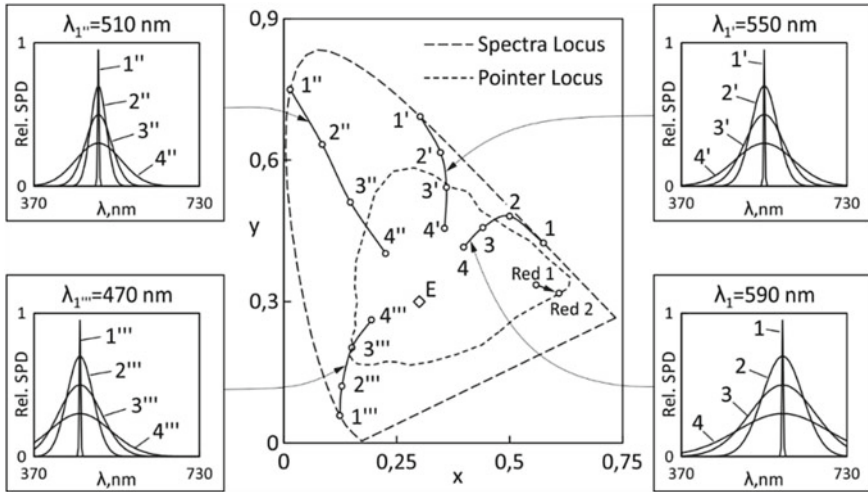
Fig. 6.2 Changing of chromaticity xy (2-deg CIE 1931), depending on the "spectral width" of stimulus. Pointer's locus is given to the lightness L* = 50 E—energy-equal stimuli, SPD of Red 1 and Red 2 given in Fig. 6.3
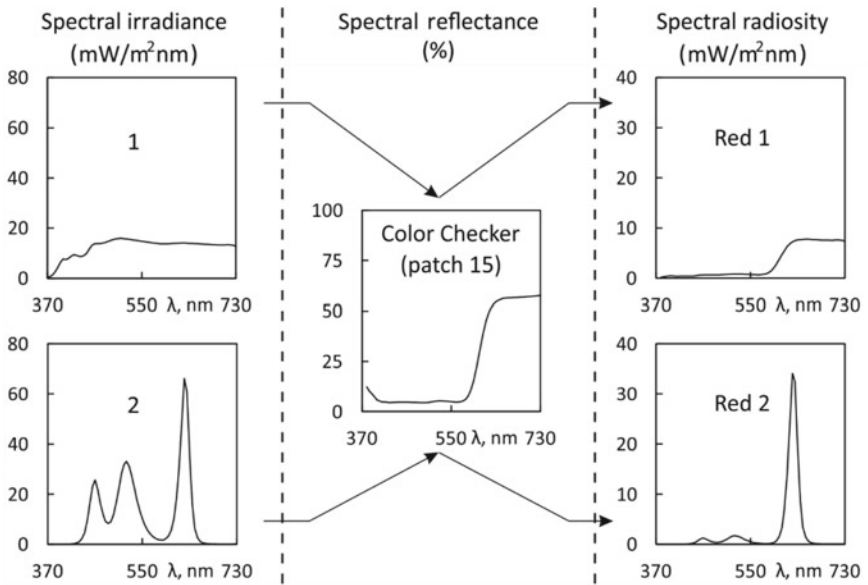

Fig. 6.3 The emergence of clarification effect on the example of the 15th patch of Color Checker scale

The obvious disadvantage of the colorimetric light is the loss of some shades of orange stimuli (skin-colors) heat-green color. At the same time, modern color management systems have the feature of accounting SPD of illumination for prints viewing conditions. Thus, the task of testing of profiling in a practical application of colorimetric light sources is of undoubted interest. In the case of correct operation implemented in software solutions, this currently offers the prospect of using the phenomenon of extraction, a clarification aimed to expand the color gamut of printed products and to increase the effect of its visual perception.

6.3   Description of the Experiment

The aim of the study was to obtain two prints from one original, which, when placed in different areas of the light source (Fig. 6.1a), would be visually perceived the same, thus, the performed experimental study differed from [6–8]. The solution was to create the colorimetric equal to D50 illumination of prints (Fig. 6.1b), analysis of the color rendering of the printer in these conditions and building an icc-profile, taking into account the colorimetric lighting.

The most accurate option for visual evaluation of the results obtained with the use of print profiles is to directly observe the differences in the simultaneous observing a pair of prints. It is essential to comply with the colorimetric equal illumination in both zones to exclude the need to consider the chromatic adaptation. The key issue in this case—an error in the classical colorimetry is described in [2, 3]. The best adjustment results in our experiments were obtained by automatic adjustment with the addition of 10-deg XYZ CMFs transformed from the CIE (2006) 2-deg LMS cone fundamentals. For spectral measurements of light and printer profiling, a widespread spectrophotometer X-Rite i1Pro in conjunction with the open source software Argyll CMS were used.

The images were printed on dye-sublimation printer Canon Selphy CP800. The printer has a wide color gamut. It is identified in the system as RGB device. The profiling was performed according to standard procedures in three steps. In the beginning, the calibration was performed to improve the transfer of gradations (calibration scale of 80 patches), then followed the preliminary profiling (scale of 480 patches) and the final profiling (scale of 1120 patches).

SPD in the area of colorimetric lighting was measured and taken into account in the construction of the profile by means of Argyll CMS. The calibration scale was the common source for both zones, but two different calibration files were created with the different lighting. Scales for preliminary and final profiling were created separately for each area based on individual calibration. Thus, two icc-profiles for areas with a standard D50 and colorimetric light were obtained.

Test photo-realistic images and synthesized colorimetric scale replica of Color Checker based on the reflectance spectra were printed with application of the profiles. The prints were placed in corresponding zones and compared visually.
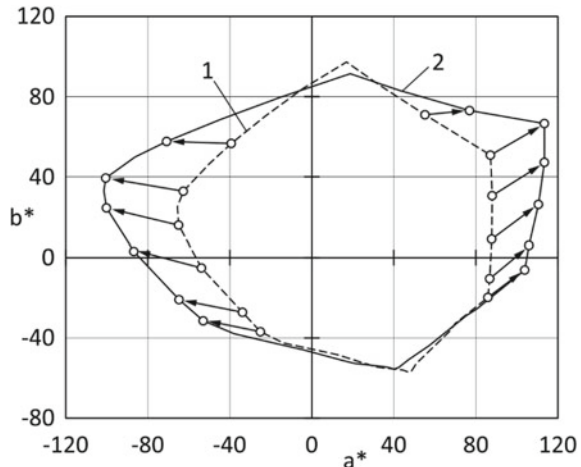
6.4 Results

Color gamut of the printer for the areas of standard and colorimetric light on the basis of the color values of the scales fields of the final profiling obtained by calculation in Argyll CMS were built (Fig. 6.4).

Obviously, the effect of clarification is very noticeable. The main saturation increase falls on the red-green direction in the chromaticity diagram.

The performed visual assessment of printed material is preliminary, as carried out within a group of authors of this paper. For more reliable and accurate results, large-scale experiments using different printing devices and attracting a large contingent of observers are required. The methodical side of the research needs further analysis, as in this case, color management of prints is a complex problem that requires taking into account the circumstances of [2, 3, 5], which themselves are still insufficiently explored.

The preliminary visual comparison of the samples gives good results with the standard method of profiling using widespread equipment. A satisfactory gradation and color match of prints placed in the respective areas of illumination, for photorealistic images, and for the Color Checker scale, are observed. It was found that the prints prepared for the D50 conditions (Fig. 6.1b) and perceptually equal to its colorimetric illumination demonstrate satisfactory agreement, as for the similar conditions corresponding to Illuminant A (Fig. 6.1c).

Fig. 6.4 Comparison of gamut built according to profiled printer Canon Selphy CP800. 1—for the illuminant D50, 2—for the equivalent colorimetric lighting

6.5 Conclusions

There is an expansion of the body of color gamut in the colorimetric illumination. Some colors, inaccessible with normal viewing conditions of print results become achievable.

Such a significant change in colorimetric values of the prints can be properly taken into account by existing color management systems. When used at this stage of studies, no significant degradation of the printed images has been detected.

The practical application of the clarification phenomenon is to improve the showiness of perception of prints with using an additional outer colorimetric illumination. The same effect in light-boxes is to be expected.

The effect of clarification deserves serious study because of the widespread introduction of LED lighting systems, including adjustable chroma ones based on RGB diodes. Inadequate accounting of color saturation change and gradation of transmission can not only distort the content of illustrations in printed materials, but also lead to undesirable hygiene consequences, such as fatigue, or vice versa, and irritability of the perception of the environment. Perhaps, RGB-LED sources should only be used for specialized tasks such as those considered in this paper.

References

1. S.P. Arapova, S.Y. Arapov, M.S. Solodova, Automated laboratory light sources for research color rendering under the colorimetric lighting, in *Proceedings of International Scientific Conference on Transmission, Processing and Perception of Textual and Graphical Information* (Ekaterinburg, 19–20 March 2015). Ekaterinburg, UrFU, pp. 34–46
2. Y. Shanda, P. Shuty, The improved description of metameric LED light clusters. Svetotechnica **6**, 8–12 (2009)
3. P. Csuti et al., Getting color right: improved visual matching with LED light sources, in *PLDC 3rd Global Lighting Design Convention* (2011), pp. 19–22
4. B. Oicherman et al., The uncertainty of colour-matching data, in *Color and Imaging Conference. Society for Imaging Science and Technology*, vol. 2005 (2005), pp. 326–332
5. B. Oicherman, R.M. Luo, A.R. Robertson, Observer metamerism and colorimetric additivity failures in soft-proofing, in *Society for Imaging Science and Technology Color and Imaging Conference*, vol. 2006 (2006), pp. 24–30
6. J. Schanda, The concept of colour rendering revisited, in *Society for Imaging Science and Technology Conference on Colour in Graphics, Imaging, and Vision*, vol. 2002 (2002), pp. 37–41
7. C. Li, M.R. Luo, C. Li, Evaluation of light sources based on visual colour rendering, in *Society for Imaging Science and Technology Color and Imaging Conference*, vol. 2008 (2008), pp. 201–204
8. Y. Park, Approach to CRI (Colour Rendering Index) for full colour RGB LED source lighting, in *Society for Imaging Science and Technology Color and Imaging Conference*, vol. 2010 (2010), pp. 371–374

9. R.W.G. Hunt, *The Reproduction of Colour*, 6 edn. Chichester, West Sussex, England; Hoboken, NJ: Wiley (2004), p. 724

10. E. Perales et al., Computing the number of distinguishable colors under several illuminants and light sources, in *Society for Imaging Science and Technology Conference on Colour in Graphics, Imaging, and Vision*, vol. 2006 (2006), pp. 414–419

11. M.R. Pointer, The Gamut of real surface colours. Color Res. Appl. **5**(3), 145–155 (1980)

12. M. Inui et al., Color Gamut of SOCS and its Comparison to Pointer's Gamut, NIP & Digital Fabrication Conference. Society for Imaging Science and Technology **2004**, 410–415 (2004)

Part II
Signal Processing and Navigation

Chapter 7
Peculiarities of Application of Multifractal Analysis to Simultaneously Recorded Biomedical Signals

Vladimir S. Kublanov, Vasilii I. Borisov and Anton Yu. Dolganov

7.1 Introduction

Human brain functioning is a complex interconnected system which is controlled by neurogenic, humoral, metabolic, and myogenic regulatory loops. The loops of regulation that control brain blood supply are in dynamic interaction. Their activity is directed to physiological homeostats establishment, which is defined by the balance of the process of water filtration from blood to tissue forced by the hydrostatic pressure in the arterial part of capillary and by its absorption in the venous part of capillary under the influence of the oncotic pressure of blood plasma. The activity is also directed to establishment of the chemical homeostats of inner environment of the brain [1].

The executive elements of blood supply regulation in the brain are vascular muscles. The muscles can perceive several types of influence: irritation of the specialized receptors like mediators allocated in nervous endings (in accordance with the myogenic theory), chemical compounds that fall into environment and blood (in accordance with the metabolic theory), mechanical stress of the blood pressure into the vessels, and sensitivity to a particular product of cellular metabolism [2].

The increase in neuron activity determined either by motional, sensorial, or mental processes results in the vasodilation and bloodstream growth caused by activation of the metabolic processes in brain tissues and secretion of the metabolites in neurohumoral environment [3].

The modernized Radiophysical complex MRTHR allows one to record the microwave radiation signals of human brain in frequency bands from 650 to 850 MHz and signals of the heart rate variability time series [4] simultaneously. The simultaneous record of these signals gives new possibilities that allow one to form the informational base for investigations of human brain functional changes both in the

---

V. S. Kublanov (✉) · V. I. Borisov · A. Yu. Dolganov
Ural Federal University, Yekaterinburg, Russia

prenosological state and for later stages of the changes in development [5]. It is suitable to apply the principles of statistical self-similarity to study these processes since dynamic control systems that form them are complex.

One has to evaluate the applicability of multifractal methods to justify its inclusion in the software of the modernized Radiophysical complex. For that, it is necessary to use modeled time series for:

- accuracy assessment of the multifractal approaches;
- investigation of the assessments dependency on the different features of the modeled time series.

7.2 Methods

The first step in the multifractal study of model signals is to estimate the accuracy of the multifractal detrended fluctuation analysis (MFDFA) approach. The main steps of obtaining the MFDFA are as follows [5, 6].

The values of interpolated time series (TS) $y(i)$, $i = \overline{1, N}$ are divided into integer number $N_s = (N/s)$ of non-overlapping segments of equal length s. Argument s is related to the selected time windows for detecting biomedical signals' fluctuations and defied as uniform logarithmic apportionment between time windows boundary points.

To implement detrending method, we find the function with most suitable polynomial trend $y_v(i) = \sum_{k=0}^{m} C_k i^{m-k}$ for each segment v from TS.

In Fig. 7.1, fluctuations of the examined signal around quadratic trend for eight-second segments are shown.

Then the variation is calculated according to the equation

$$F_y^2(s, v) = \frac{1}{s} \sum_{k=1}^{s} [y(k) - y_v(k)]^2 \tag{7.1}$$

The variability of all segments can be expressed by a function [3] with index q in the range q = [−5, 5] and is defined for nonzero values by the equation

$$F_y(q, s) = \left\{ \frac{1}{N_s} \sum_{v=1}^{N_s} [F_y^2(s, v)]^{q/2} \right\}^{1/q} \tag{7.2}$$

For q = 0, the following equation is obtained.

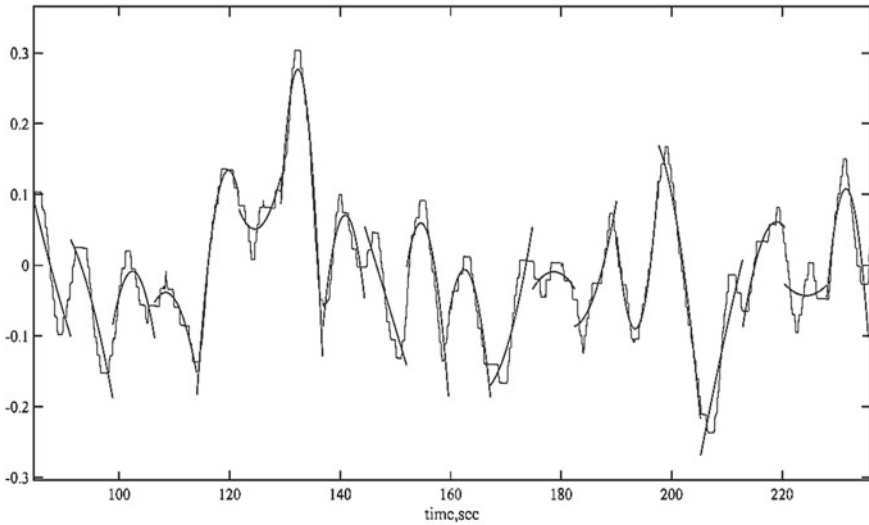$$F_y(0, s) = \frac{1}{2N_s} \sum_{v=1}^{N_s} \ln[F_y^2(s, v)]. \tag{7.3}$$

Fig. 7.1 Detrending of the brain microwave radiation signal

The fluctuation function for self-similar TS depends on the window width s as degree [6]

$$F_y(q, s) \approx s^{h_y(q)}. \tag{7.4}$$

The index $h(q)$ is called the generalized Hurst exponent (GHE) and calculated from the slope of fluctuation function against s in logarithmic coordinates. For monofractal signals, GHE does not depend on q. For multifractal signals, positive q denotes behavior of large fluctuations; negative q denotes behavior of small fluctuations.

In the general case, a multifractal set is characterized by the scaling exponent $s(q)$. Function $\tau(q)$ shows how heterogeneous the selected set is. GHE is connected with $\tau(q)$ function as follows

$$\tau(q) = q \cdot h_y(q) - 1. \tag{7.5}$$

The multifractal spectrum width is determined by the spectral distribution function D using Legendre transformation as probability distribution of q

$$D(\alpha) = q \cdot \alpha - \tau, \tag{7.6}$$

where α is the Hölder exponent.

The multifractal cross-correlation analysis method (MFCCA) is used to evaluate correlations between selected time series [7]. The correlation function is constructed with the following substitutions to Eqs. (7.1)–(7.3)
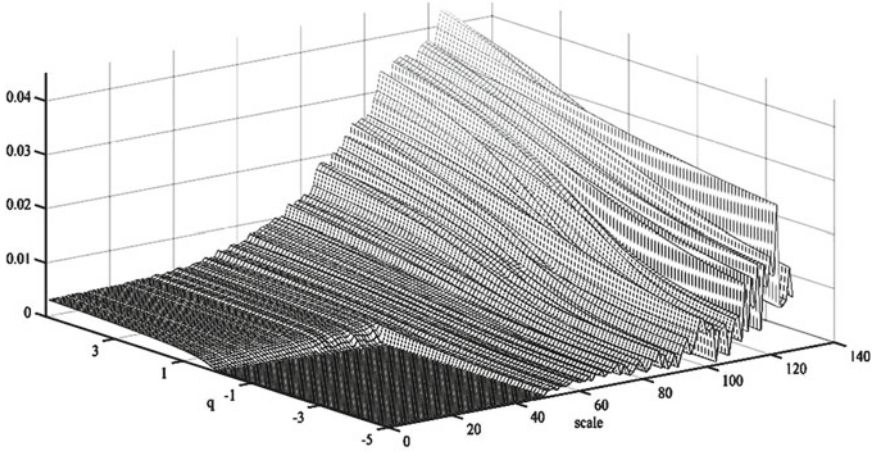
Fig. 7.2 Correlation function of two signals

$$F_{xy}^2(s, v) = \frac{1}{s} \sum_{k=1}^{s} [x(k) - x_v(k)][y(k) - y_v(k)], \qquad (7.7)$$

$$F_{xy}(q, s) = \{\frac{1}{N_s} \sum_{v=1}^{N_s} [F_{xy}^2(s, v)]^{q/2}\}^{1/q} \approx s^{h_{xy}(q)}, \qquad (7.8)$$

$$F_{xy}(0, s) = \frac{1}{N_s} \sum_{v=1}^{N_s} \ln[F_y^2(s, v)] \approx s^{h_{xy}(q=0)}. \qquad (7.9)$$

Figure 7.2 shows the correlation function of two signals.

Figure 7.2 represents dependence of correlation function F_{xy} on degree q (X-axis) and scale s (Y-axis).

Similar to the MFDFA method, the Legendre transformation is used to calculate the cross-correlated Hurst exponents (7.5), (7.6). This characteristic can be used to detect long-term cross-correlation between the signals [8].

Nowadays, the cross-correlation multifractal method is used for analysis in various spheres: financial market data [9], DNA structure investigation [10], electroencephalography signal processing [11], etc.

7.3 Results and Discussion

In order to evaluate the accuracy of the multifractal approaches, the wavelet-based synthesis for the fractional Brownian motion (WFBM) model was used. The WFBM model implies application of the wavelet transform and Abry-Sellan algorithm, which

Table 7.1 Accuracy evaluation of MFDFA approach, %

h	$|H-h|/h$	$|H_{int}-h-1|/h$
0.1	38	10
0.2	6	7
0.3	1	6
0.4	2	5
0.5	1	6
0.6	1	5
0.7	2	2
0.8	3	4
0.9	2	4

is based on the ARIMA fractal process for evaluation of the approximation coefficients [12]. The model is expected to plot time series with particular Hurst exponent h. In addition, this model was used to evaluate integrated model signals. As is known, Hurst exponent of the integrated time series with Hurst exponent H is $H_{int} = H+1$. Overall, model signals were generated 100 times for each $h = 0.1 \times n$, where $n \in N$, $n = 1,..,9$. The length of the model time series corresponded to the length of the biomedical signals recorded during functional loads by the Radiophysical complex MRTHR.

Table 7.1 presents results of the MFDFA accuracy evaluation averaged over 100 model signals: h is the expected value of the Hurst exponent, H is the value obtained by the MFDFA approach, H_{int} is the MFDFA evaluation for integrated time series.

Data in Table 7.1 shows that MFDFA approach can evaluate Hurst exponent of WFBM-modeled signals with relatively low error (< 10%) for both original and integrated time series for each h > 0.1. The maximal error of the Hurst exponent evaluation is less or equal to 3% for h > 0.3.

The second step of the model studies is stability assessment of the multifractal estimates in case of noise addition. Real time series can be corrupted by noise. Therefore, it is advisable to model the impact of the noise on the multifractal estimates [13]. For that, we have tested previously generated WFBM-modeled signals with addition of the white noise time series. The amplitude of the noise depended on the signal/noise ratio and varied from 26 to –14 dB. Overall, 12 different scales of the signal/noise ratio was used. By the white noise in this study, we have considered fractal Brownian motion with h = 0.5.

Table 7.2 and Fig. 7.3 present estimations of MFDFA approach stability in case of the white noise addition. In Table 7.2, evaluations of errors are presented averaged over 100 generations of the model signals.

According to the Fig. 7.3 and Table 7.2, the addition of the white noise is less noticeable for values of the Hurst exponent higher than 0.5 (error is less than 10%). For Hurst exponent $0.2 < h < 0.5$ reliable evaluations are possible in case of signal/noise ratio higher than 6 dB.

Table 7.2 Relative values of error for the Hurst exponent evaluation by MFDFA approach for different signal/noise ratio, %

| h	Signal/Noise ratio, dB											
	26	20	16	14	12	6	0	−4	−6	−10	−12	−14
0.1	34.0	39.9	47.6	52.4	63.2	102.5	202.7	257.3	305.6	334.1	354.6	356.5
0.2	11.7	12.5	11.7	10.6	13.1	26.8	62.4	92.5	104.9	119.4	127.3	130.3
0.3	0.7	0.9	1.3	0.7	1.5	6.8	24.4	31.4	37.6	47.4	51.2	54.6
0.4	0.4	4.9	1.4	4.2	3.7	2.5	6.0	9.2	10.1	12.1	15.7	13.4
0.5	1.7	0.4	0.8	3.4	6.1	2.9	3.1	4.1	3.8	7.1	4.2	6.0
0.6	1.4	0.7	3.4	2.9	0.1	2.6	4.6	3.8	8.8	9.8	9.3	15.8
0.7	3.5	5.2	4.4	4.1	0.4	3.1	5.3	4.3	4.4	11.7	15.1	15.1
0.8	4.1	0.4	2.9	2.1	3.2	2.6	1.7	4.3	5.5	6.2	12.6	13.3
0.9	2.0	3.3	1.7	0.6	1.5	1.2	2.8	3.5	6.0	1.6	9.0	7.8
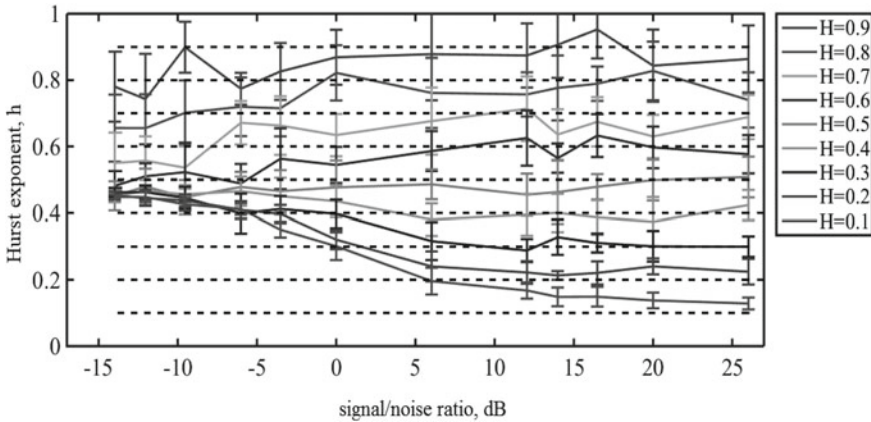
Fig. 7.3 Hurst exponent evaluation by MFDFA in case of white noise addition

In order to evaluate the MFCCA approach stability, we have used the afore-mentioned technique. The cross-correlation Hurst exponent were obtained for two signals:

- time series with predefined Hurst exponent in range from 0.1 to 0.9;
- same time series with addition of white noise with different signal/noise ratio in range from 26 to –14 dB.

In Table 7.3, evaluations of errors for MFCCA approach are presented averaged over 100 generations of the model signals for different scales of signal/noise ratio.

According to the data presented on Fig. 7.4 and Table 7.3 for the Hurst exponent $h > 0.2$, there is relative low level of relative error (less than 10%) for all scales of signal/noise ratio. With that in mind, one can conclude that cross-correlation Hurst exponent of two signals with same Hurst exponent, one of which is noised by the addition of white noise, tends to evaluate original Hurst exponent. Such results of the noise addition evaluation do not contradict conclusions in state-of-art works [12, 13].

All things considered, both MFDFA and MFCCA approaches can be used for evaluation of the biomedical signals of the Radiophysical complex MRTHR with certain degree of accuracy.

7.4 Conclusions

Among the known methods of multi- and monofractal formalism for the evaluation of short-term biomedical signals, the multifractal fluctuation (MFDFA) and cross-correlation (MFCCA) analysis approaches are of the greatest interest. These

Table 7.3 Absolute values of error for the Hurst exponent evaluation by MFDFA approach for different signal/noise ratio, %

h	Signal/Noise ratio, dB												
	26	20	16	14	12	6	0	−4	−6	−10	−12	−14
0.1	33.2	36.4	41.5	37.7	40.3	34.8	40.1	49.4	74.8	70.0	105.5	95.0
0.2	11.3	10.6	8.9	6.7	6.2	4.7	0.7	9.7	7.9	19.2	27.5	35.5
0.3	0.8	0.4	0.3	0.7	1.4	1.0	3.0	1.4	4.0	8.2	9.5	16.6
0.4	0.4	5.1	1.7	4.7	3.8	5.0	0.2	1.0	0.7	4.0	1.0	4.7
0.5	1.7	0.2	0.8	3.6	6.1	2.7	2.7	4.4	4.0	10.9	3.2	8.9
0.6	1.4	0.7	3.5	2.9	0.0	2.4	4.2	0.3	6.5	5.3	3.0	8.0
0.7	3.5	5.2	4.4	4.0	0.3	2.8	3.9	2.8	1.9	6.4	6.7	6.4
0.8	4.1	0.4	2.9	2.1	3.2	2.3	2.2	2.7	3.6	2.3	5.8	5.2
0.9	2.0	3.3	1.7	0.6	1.5	1.1	2.5	2.8	4.7	0.9	5.3	4.2
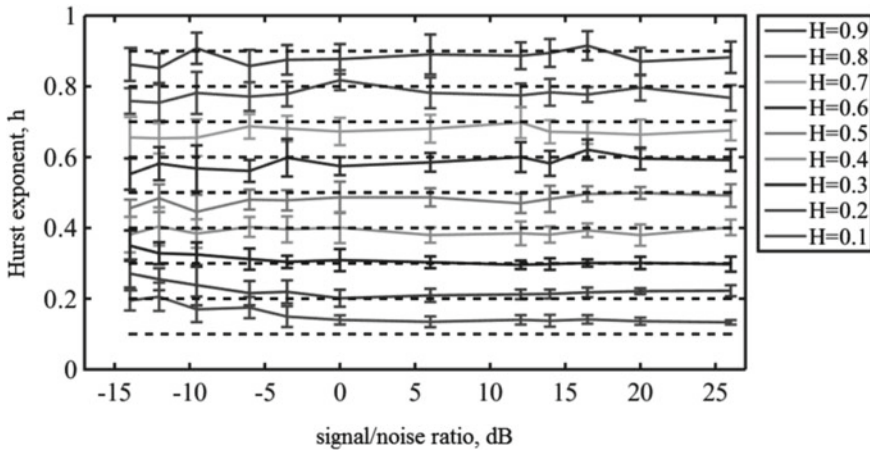
Fig. 7.4 Hurst exponent evaluation by MFCCA in case of white noise addition

approaches do not have a limitation on the stationarity and noisiness of the several signals recorded simultaneously.

In the article, we have evaluated the accuracy of the MFDFA and MFCCA approaches with the use of model time series of fractal Brown motion for certain Hurst exponent values in the range from 0.1 to 0.9. First, it was shown that estimations of the Hurst exponents by the MFDFA approach have an error less or equal to 3%, while for the integrated time series the error is less than 7% on average. These results indicate the high accuracy of the studied multifractal approaches. Moreover, estimations of the noised time series were obtained. The MFCCA have shown relatively low sensitivity to the additional noise, the error was less than 10% for a Hurst exponent higher than 0.2.

Analysis of MFDFA and MFCCA approaches by the model signals have shown that such approaches can be applied to the assessment of short-term biomedical signals of the multichannel Radiophysical complex MRTHR [14].

References

1. Y.E. Moskalenko, (ed. Chazov E.I). *Cerebral Circulation* (Meditsina, Moscow, 1992), p. 496
2. A.C. Guyton, J.E. Hall, *Textbook of Medical Physiology* (Saunders Elsevier, Philadelphia, PA, 2011), p. 1048
3. R.F. Schmidt, G. Thews, *Human Physiology,* vol. 1 (Springer, London, 2005), p. 323

4. V.S. Kublanov, Microwave radiation as interface to the brain functional state, in *Proceedings of the International Conference on Biomedical Electronics and Devices BIODEVICES* (Barcelona, 2013), pp. 318–322

5. V.S. Kublanov, V.I. Borisov, A.Y. Dolganov, The interface between the brain microwave radiation and autonomic nervous system, in *7-th Annual International IEEE EMBS Conference on Neural Engineering* (Montpellier, 2015), pp. 922–925

6. J.W. Kantelhardt, S.A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, H.E. Stanley, Multifractal detrended fluctuation analysis of nonstationary time series. Physica A **316**(1–4), 87–114 (2002)

7. L. Krištoufek, Multifractal height cross-correlation analysis: a new method for analyzing long-range cross-correlations. Europhys. Lett. **95**(6) (2011). article number 68001

8. B. Podobnik, H.E. Stanley, Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series. Phys. Rev. Lett. **100**(8) (2008). article number 084102

9. W-.X. Zhou, Multifractal detrended cross-correlation analysis for two non-stationary signal. Phys. Rev. E, **77** (2008). article number 066211

10. C. Stan, M.T. Cristescu, B.I. Luiza, C.P. Cristescu, Investigation on series of length of coding and non-coding DNA sequences of bacteria using multifractal detrended cross-correlation analysis. J. Theor. Biol. **321**, 54–62 (2013)

11. D.C. Ghosh, S. Dutta, S. Chakraborty, Multifractal detrended cross-correlation analysis for epileptic patient in seizure and seizure free status. Chaos, Solitons Fractals **10**, 1–10 (2014)

12. V.S. Kublanov, V.I. Borisov, A.Y. Dolganov, Multifractal formalism in study of the role of autonomic regulation in formation of intrinsic electromagnetic radiation of the brain. Biomed. Eng. **1**(50), 30–34 (2016)

13. J.B. Bassingthwaighte, G.M. Raymond, Evaluation of the dispersional analysis method for fractal time series. Ann. Biomed. Eng. **23**(4), 491–505 (1995)

14. V.S. Kublanov, V.I. Borisov, Biophysical evaluation of microwave radiation for functional research of the human brain. IFMBE Proc. **65**, 1045–1048 (2017)

Chapter 8
Methods of Autonomous Indoor 3D Navigation

Check for updates

M. Osipov and Yu. Vasin

8.1 Introduction

Indoor orientation is an important applied area and is much in demand in the market of navigation services. The internal structure of such buildings as large business and shopping centers, hotels, railway stations, airports, etc., is a maze with many floors, corridors, and rooms where orientation can be extremely difficult. In such buildings, it is often difficult to find a particular office or shop and to determine escape routes in case of emergency. Indoor orientation is also widely used in the rapidly growing geo-context advertising market. Since this area is very promising from the commercial point of view, large corporations like Google, Sony, Apple, Samsung, Nokia, etc., invest hundreds of millions of dollars in the development, implementation, and promotion of indoor navigation systems [1].

Currently, visual navigation remains a traditional way to navigate indoors. It is a system of signs and nameplates that serves as a guide in the building where such signs are placed. This approach is simple to implement, but it has some significant drawbacks: a high information load on users and complexity. The use of computer vision in navigation tasks would significantly automate this process.

Computer vision methods have been used in location assessment and robot navigation tasks for over 10 years. The robot's movement is monitored by analyzing the video stream from the camera mounted on it. This approach is called visual odometry and is based on a sequential calculation of the camera motion during the period between the new and the previous frame. Unlike robots, humans cannot provide constant smooth camera movement when they walk (in particular, to maintain a constant

M. Osipov (✉) · Yu. Vasin
Lobachevsky Nizhni Novgorod State University, Nizhny Novgorod, Russia
e-mail: osipovmp@mail.ru

Ural Federal University, Yekaterinburg, Russia

distance to the surface and tilt angle). Therefore, visual odometry techniques are not effective in human navigation.

Indoor use of such navigation systems as GPS-GLONASS that are traditional for this niche is very difficult, since the walls and floors pose an obstacle to satellite signals. At present, different approaches are proposed to solve the problem of indoor autonomous navigation. These approaches are based on the use of various data transmission networks (Wi-Fi, Bluetooth, FM radio, UWB), Earth's magnetic field for positioning in buildings, micromechanical inertial sensors (MEMS) (accelerometers, angular velocity sensors), RFID tags, and QR codes [2, 3]. However, these approaches have significant limitations in terms of complexity of their implementation, accuracy, reliability, and cost.

8.2 Statement of the Problem

This paper proposes and explores a research area whose objectives are to develop virtual reality methods; to produce 3D models of buildings and structures as well as 3D models of premises with their infrastructure and built-in reality; to develop models, methods, and algorithms for autonomous indoor navigation with extensive use of 3D modeling to determine one's location; to work out optimal routes and to monitor one's movement along a given route.

In modern navigation systems, graphic information is represented for the user in the form of a 2D representation (projection) of the floor plan, which cannot give the user the real spatial information.

3D models accurately convey the shape and color of the objects being modeled. It is difficult to distinguish a photorealistic 3D model of the object from a photograph of this object. 3D models of the internal structure objects of buildings and structures allow the user to adjust their location by comparing the actual environment with a 3D visualization of the calculated (estimated) location. This feature can significantly improve the positioning system reliability without using any additional equipment.

The use of 3D visualization in conjunction with the traditional positioning systems allows navigation in low (restricted) visibility. This is especially important when organizing rescue operations in emergency situations, in combating crime and terrorism, etc.

8.3 Methods for Solving the Problem

To solve this problem, we have researched and developed some models, structures, and algorithms to optimize the collection, creation, and visualization of 3D models of buildings and structures with floor plans and infrastructure. We have also developed some methods and algorithms for autonomous navigation using 3D modeling and computer vision.

The development of 3D models involves a highly labor-intensive process of gathering the required information based on the original graphic materials (floor plans of buildings and structures), photographs of objects and textures, as well as of interior space, and the semantic information. The process of 3D modeling involves a large proportion of manual labor. A characteristic feature of the urban area is a large number of repeating standard solutions and objects. By means of creating and classifying a renewable library of 3D models and textures of standard objects, it is possible to automate significantly the process of production of 3D models of floor plans.

To increase the level of automation of data collection and creation of 3D models of buildings and floor plans with their infrastructure, information and terminological support has been developed along with the object-oriented DBMS, knowledge base management system [4] to work with 3D information, and the following methods and technologies have been developed: automated production of digital floor plans for their subsequent use in 3D modeling; automated collection of information about geospatial objects on small platforms; storage, and classification of a library of unique 3D models and textures using the attribute information obtained from the GIS-TERRA database; automated process of preparation of textures using photographs.

The visualization process in navigation systems must have a sufficient degree of realism [5] (Fig. 8.1). In many cases, a highly detailed 3D model cannot be visualized on a modern computer (particularly on small platforms) in real time. Therefore, further optimization of the process of 3D scene visualization is required. For effective visualization of digital 3D models, we have developed and designed: methods and data structures for efficient 3D visualization; adaptive methods for presentation of spatially distributed data for automatic modeling and interactive visualization of digital terrain models; effective methods and algorithms for clipping of invisible areas of the scene during visualization; multidimensional data structures that allow the indexing of a 3D scene for efficient and optimal visualization; data uploading/downloading in real time [6].

In this work, we have implemented an algorithm for selecting the optimal path through the connectivity graph. Methods have been developed for the construction and presentation of connectivity graphs of rooms, corridors, stairways, elevators, etc., based on digital floor plans with special patterns of graph arcs and nodes, algorithms have been implemented for optimal route plotting, subject to specified constraints. We have also developed search methods for characteristic beacon points based on 3D modeling and built-in virtual reality.

The tracking of the movement along the chosen route is performed in accordance with the readings of MEMS sensors. The fact that there is no need to install additional equipment on the premises is an important advantage offered by the use of MEMS sensors in navigation. However, such sensors in home appliances have insufficient precision. The cumulative error amounts to about 10% of the distance traveled, thus such sensors are unreliable for determining one's location. The growth in the cumulative error of the sensors is due to the fact that the calculation of the new position is based on the value of the previous one. Therefore, when using micromechanical inertial sensors, their readings need to be constantly adjusted. With the help of the

Fig. 8.1 The visualization process in navigation systems

floor plan of the building and its 3D visualization, it is possible to compensate for the accumulation of the error in the readings of MEMS sensors.

Planning structures of buildings and structures convert the path of motion into a polyline. The accelerometer counts the number of steps walked. By using the readings of the magnetometer with a gyroscope, one can reliably determine the moment of turning in the path of movement. Thus, the error in the readings of MEMS sensors accumulated between turns will be constantly "zeroed out" at each turn (Fig. 8.2).

At the same time, sections of the route between turns may be quite long amounting to tens of meters. In the process of traversing such sections, a significant error may be accumulated. However, since the route section between the turns is in the line of sight, the user can adjust their location in accordance with the 3D visualization of the current position calculated with the help of the sensors. To carry out this adjustment, a button is provided for stopping the sensor.
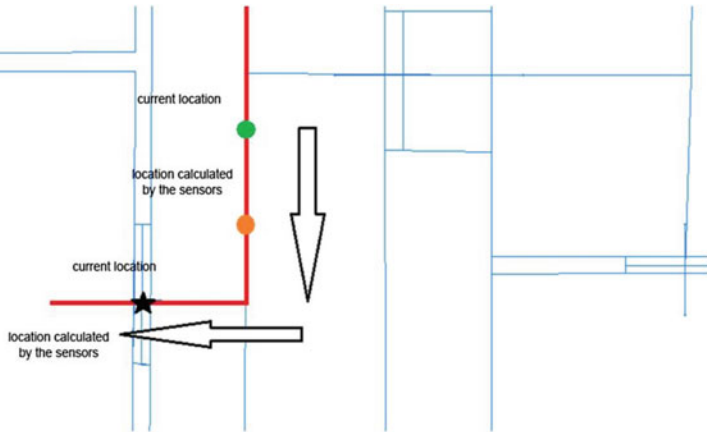
Fig. 8.2 The error in the readings of MEMS sensors accumulated between turns will be constantly "zeroed out" at each turn

For situations where it is impossible to make location adjustment using 3D visualization, the possibility is envisaged to determine (adjust) one's location using computer vision techniques [7]. For this purpose, in advance, a database is created of images of flat information objects pertaining to the premises (pictures, names, and numbers of rooms, evacuation plans, fire safety signs, etc.) with coordinate referencing of their location (Fig. 8.3).

Thus, when location adjustment is required, the user makes a video of the nearest flat information object using a mobile device with a built-in camera. Next, using

Fig. 8.3 Images of flat information objects pertaining to the premises

content-based image retrieval (CBIR) methods, image elements obtained when making the video are compared with the database of information object images [8]. If the original image is found, then, using the information on the location of the object stored in the database one can determine the location and orientation of the camera that produced the video, and therefore, the location and orientation of the user. The disadvantage of this approach is the fact that information objects are not unique. However, as the approximate location of the user at the time of adjustment is known, the search is confined to a certain neighborhood and repeating objects are left out.

By combining different approaches, it is possible to compensate for disadvantages of individual methods and to obtain more reliable results.

The proposed approach is implemented in the form of a client-server system (Fig. 8.4). The client part implements a user interface that allows the start and end points of the path to be determined. The client part also contains MEMS sensors whose readings are sent to the server. All calculations, including the determination of the path of motion and 3D visualization, are performed on the server. Content-based image search is also performed on the server due to memory and computational resources requirements. The image of the surrounding situation at the current moment of time is formed on the server and is sent to the client part with the indication of the optimal path of motion to be shown on the user's display.

Effective hierarchical algorithms for adaptive compression and decompression have been developed, as well as algorithms for entropy coding with context-based modeling of raster images, thus enabling on-the-fly transfer of images of a 3D model of floor plans and a 3D model of the area surrounding the user via communication channels and their visualization on small platforms. Experimental testing of the proposed models and algorithms has confirmed the effectiveness of the proposed approach.

8.4 Conclusions

One of the key elements in the problem of navigation is the information about the route to the selected object that can be understood intuitively. 3D visualization can significantly simplify the user's progress along the chosen route.

In this paper, we have proposed and developed a new approach to solving the problem of autonomous indoor navigation based on the development of virtual reality methods, producing 3D models of floor plans and extensive use of 3D models created in the process of solving various navigation tasks.
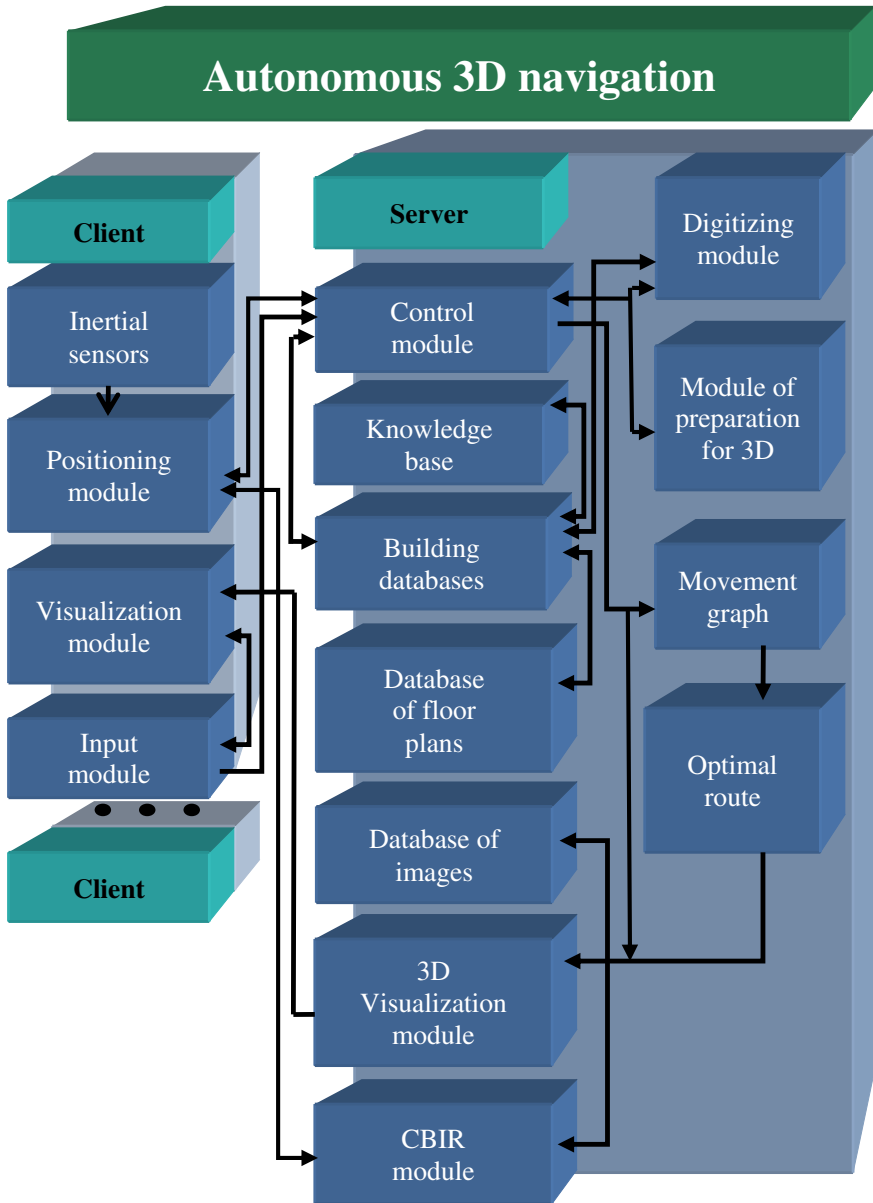
Fig. 8.4 Architecture of interaction between the client and server parts

References

1. Official website. In-location Alliance. http://www.in-locationalliance.com. free
2. C. Lukianto, H. Sternberg, Overview of current indoor navigation techniques and implementation studies, in *Bridging the Gap between Cultures* (Marrakech, 2011)
3. Continuous Indoor Positioning: GNSS, Wi-Fi, and MEMS Dead Reckoning. http://gpsworld.com/continuous-indoor-positioning-gnss-wi-fi-and-mems-dead-reckoning/. free
4. Yu.G. Vasin, Yu.V. Yasakov, GIS Terra: A graphic database management system. Pattern Recognit. Image Anal. **14**(4), 579–586 (2004)
5. Yu.G. Vasin, M.P. Osipov, T.N. Tomchinskaya, Development of interactive virtual models of the urban landscape of the historical center of Nizhni Novgorod. Pattern Recognit. Image Anal. **21**(2), 351–353 (2011)
6. Yu.G. Vasin, M.P. Osipov, S.V. Muntyan, E.A. Kustov, Procedural modeling and interactive 3D visualization of objects of the internal structure of buildings and facilities. Pattern Recognit. Image Anal. **25**(2), 278–280 (2015)
7. M. Serrão, S. Shahrabadi, M. Moreno, J.T. José, J.I. Rodrigues, J.M.F. Rodrigues, J.M.H. du Buf, Computer vision and GIS for the navigation of blind persons in buildings. Univ. Access Inf. Soc. **14**(1), 67–80 (2014)
8. R. Huitl, G. Schroth, S. Hilsenbeck, F. Schweiger, E. Steinbach, Virtual reference view generation for CBIR-based visual pose estimation. ACM Multimed. 93–996 (2012)

Part III
Simulation

Chapter 9
Investigation of Money Turnover in the Computer Agent-Based Model

O. M. Zvereva

9.1 Introduction

There are three main simulation paradigms; they are as follows: discrete event modeling, system dynamics and agent-based modeling (ABM). Discrete-event simulation is used for development of stochastic, dynamic models where simulation state variables change at discrete points in time. System dynamics is an approach aimed to understand the behaviour of complex systems over time using stocks and flows, internal feedback loops and time delays. Agent-based (sometimes called Individual-based) modelling is used to model a system as a collection of communicating entities called the agents.

The third mentioned modelling paradigm has gained increasing attention over the past decade. Macal and North [1] proposed some explanation for this fact. The first reason is that the present-day systems that one needs to analyze have been becoming more and more complex and having various type of interdependences. The second consideration is a consequence of the first—nobody can model these systems adequately. '…We are beginning to be able to relax some of… assumptions and take a more realistic view of these… systems through ABM' [1]. The third and the fourth reasons are related to modern computational knowledge and power, because of which ABM is possible in the real-world scale and can solve the real-world problems. Some modellers contend that ABM 'is a third way of doing science' and could augment traditional deductive and inductive reasoning as discovery methods.

According to [2] every agent-based model consists of the following basic elements:

1. Agents with their attributes, goals and rules of behaviour;
2. Relations between the agents;

O. M. Zvereva (✉)
Ural Federal University, Yekaterinburg, Russia
e-mail: OM-Zvereva2008@yandex.ru

3. Environment: agents communicate through the environment and in some cases can interact with this environment as well.

There are some generic agent characteristics: an agent is autonomous, reactive, pro-active and social.

Agent's autonomy means that the agent encapsulates some state, which is not accessible to other agents. Reactivity is considered as the agent's ability to respond in a timely fashion to changes that occur in the environment in which the agent is positioned. Agent's pro-activeness is realized as its property to exhibit the goal-directed behaviour by taking the initiative. Agent's social ability indicates its competence to communicate with other agents using some kind of a language.

There is no established agent classification because ABM is a rather new paradigm and has been rapidly developing; many new agent types appear every now and then. Some efforts aimed at the development of the agent taxonomy are discussed in [3] in the context of economics.

Agents can greatly differ from each other and their behaviour can be different as well. They can behave in a random way (Zero Intelligence Agents), or they can learn and store their knowledge. But even a simple agent-based model can exhibit complex behavior patterns and provide valuable information about the dynamics of the real-world system which it emulates.

Bonabeau in [4] indicated benefits of ABM over other simulation techniques: ABM captures emergent phenomenon, ABM provides a natural description of a system's behavior, and ABM is a flexible one.

Emergent phenomenon results from the interactions of individual entities. In the engineered model, one determines micro parameters (agent's attributes, rules of interactions) and can measure system macro parameters. Nobody can predict these macro parameters' values, and sometimes they appear to be really amazing. It is rather common to make a project in the 'top-down' manner (especially in the field of software engineering), but, in the case of ABM, another 'bottom-up' approach is used.

In many cases, ABM is the most natural for describing and simulating a system composed of acting and interacting entities. Such systems are common in different scientific fields: e.g. physics [5], chemistry [6], biology [7, 8], history [9], etc. But one can argue that ABM is geared towards social and economic system modelling.

Luhmann and many other sociologists considered society to be a set of communications in its essence [10], and ABM is the most fitting technique for a social communications system simulation. In [11] ABM is even called 'the right mathematics for the social sciences'. In the economic science, one can reveal the mainstream called the agent-based computational economics, which is based on the ABM paradigm (a lot of materials is presented in [12]).

ABM flexibility can be observed along multiple dimensions. In most agent-based models, one can easily (without significant code change) increase the quantity of agents. You can, with minimal effort, change agent's rules of behaviour, include some new elements into their environments, etc.

9.2   Communication Model Specification

A communication process is a process which runs in systems from various scientific domains. Active system entities try to interact with each other exchanging something that has a value for them [10] with the predefined purpose and using a kind of a channel. In a social system, knowledge (or sense) is the value subject for exchange, and, in an economic system, money and products exchanges could be observed.

It was decided to create a model with active productive entities which are involved in economic communications. In this system, every agent has a very clear goal: it tries to receive all the resources necessary for manufacturing the specified product volume; in order to succeed in this activity, this agent also tries to sell its own already manufactured product. According to communication process definition [13], every system member is involved in such type of process with the other system members. There is an environment (economics), which delivers money supposed to be means for exchanges. Thus, we have three main components of an agent-based system: a set of agents, an environment, relations between the agents and with the environment as well. In this case, it becomes evident that the most suitable modelling paradigm is the agent-based one.

In order to build a valid model, we have to determine what questions our model will answer, what problems it will help to solve. The main question is—'Under what conditions a productive system will be a sustainable one?' It means that a system can function in a long run: in its every lifecycle it must create a kind of material background for the succeeding lifecycle. This issue is detailed in [14].

Trying to investigate the problem of system sustainability, several additional pieces of knowledge about money turnover were received, namely, we got to know how much money it was necessary to introduce into the system to support its sustainable functioning, and how this money volume has influenced time necessary to carry out all communications, what way of behaviour the agents had to choose to optimize the communication time. Also with the help of the model, some experiments were held to prove (or eject) the hypothesis of virtual, or internal, money usability, their opportunity to play the role of real money.

The main idea for a sustainable productive system model was that some kind of equilibrium should be adopted as its quantitative basis. As this equilibrium, Leontief's static intersectoral equilibrium may be proposed, and his famous 'input-output' model [15] has become an algorithmic basis for the engineered model (Communication Model).

In the Communication Model, there are N agents, where the ith agent is characterized by its product volume (x_i), its vector of demands in products of another system agents ($\vec{W}_i = (w_{ik})_{k=1}^{N}$), and its final demand volume (y_i) that can be spent for meeting its non-product requirements after all communication acts.

As for the system as a whole, it can be described by the following macro parameters: ($\vec{X}(t) = (x_i)_{i=1}^{N}$) which is a system product volume vector; ($\vec{Y}(t) = (y_i)_{i=1}^{N}$) which is a system final demands vector; and ($W_{NxN} = (\vec{w}_i)_{i=1}^{N}$) which is a communication matrix (matrix of mutual payments, as it will be shown further).

For every agent in the model, it is true that:

$$x_i = a_{1i} \cdot x_1 + a_{2i} \cdot x_2 + \cdots + a_{Ni} \cdot x_N + y_i \tag{9.1}$$

In accordance to Leontief's model [15], the technological matrix ($A_{N \times N}$) can be obtained, being calculated on the basis of $W_{N \times N}$, where the matrix element (a_{ki}) determines the k-th agent's product volume which is consumed by the i-th agent for its single product unit manufacturing.

For the system in a whole, the following vector equation can be written:

$$\vec{X} - A\vec{X} = \vec{Y} \tag{9.2}$$

As was mentioned above, economic environment provides money to support exchanges in the system, to determine the money volume Money supplement coefficient (K) was proposed. The money volume circulated in the system is estimated as the multiplication of coefficient K and the total product volume in the system. Every agent receives a sum of money on its account, the initial value of agent's money (m_i) is estimated in the direct proportion to its product volume ($m_i = Kx_i$). Every agent tries to communicate with the other system agents and uses money from its account to provide these communications. Money in this model plays its common role means of exchanges.

An agent can behave in various ways. These different kinds of agent's behaviour were named 'strategies'. In every communication act, it is necessary to determine the partner (or the partners), which product is necessary to receive, and the volume of this exchange. Thus, these strategies differ from each other according to the following three parameters: the number of partners, the rule of the partner choice, and the exchange volume. Thus, strategy and money volume also have become the essential agent's characteristics.

The communication process has to last as long as any communication is possible. Communication stage is divided into exchange cycles (the modelling time is measured in exchange cycles). During every exchange cycle, every agent (chosen in a random way) receives a chance to communicate only once, i.e. to process a single exchange operation in the way assigned by its strategy.

In order to engineer the Model according to the revealed specifications, it is necessary to choose an information toolkit. This problem is discussed in the next section.

9.3 Modeling Toolkit Choice

In the last few years, the ABM community has taken a giant step in developing practical agent-based modeling toolkits that enable individuals to engineer significantly sized and complex applications. There are several surveys which try to evaluate and compare these toolkits. In [16], the authors have evaluated more than 50 frameworks,

in [17, 18] only five modelling frameworks are under discussion, but they might be considered as the most popular ones, they are as follows: Swarm, Java Swarm, Repast, MASON and NetLogo.

In the first survey [16], five widely used characteristics have become the basis for the comparison: complexity of interface and modelling language; operating system required to run the toolkit; type of the license governing the platform (free or proprietary); primary domain for which the toolkit is intended (is it a multipurpose instrument or is intended to model systems of the specific type, i.e. social, biological, economic, etc.); degree of support available to a toolkit user (e.g. documentation quality).

The toolkit language and interface are really important issues. According to this issue, all toolkits can be divided into two groups: in the first group general-purpose programming languages are used such as Java, C++, Python; and the frameworks of the second group have their own modelling languages. Modellers also can be divided into two groups: the first group consists of those who are professionals in the domain area (sociologists, biologists, chemists, etc.), and the second group includes professionals in programming. The first group members are more concerned with the modelling framework ease of use, the degree of programming skills required, and the existence of friendly interfaces to manage simulations. They are not addicted to coding, and it is likely that they will choose something from the second group of modelling toolkits. If the modeller is a professional in programming, he/she will choose a toolkit based on his/her 'favourite' language.

As the exception only proves the rule, we have chosen the toolkit with its special modelling language. Being professionals in IT area, we have decided that we would gain interesting new experience along with a new language implementation. The NetLogo [19] modelling environment was chosen as an instrument for the Communication model development and later simulation. This toolkit has its own modelling language, which is considered to be a Logo dialect extended to support agents.

Considering the main NetLogo characteristics (the latter four from the list above), it runs on different Windows versions (e.g. Windows 7/8/10), and also on Mac OS X 10.4, or newer. It needs Java Virtual Machine (JVM), or Java Runtime Environment (JRE) being preinstalled. NetLogo is a free open source system (it is under the terms of the GNU General Public License). This tool is oriented towards education as its primary specialization, but now it has become a powerful tool widely used in various scientific domains, and it is declared to be geared specifically towards the social science domains.

To discuss the degree of the user support, it is necessary to mention: the well-organized documentation built in the environment (tutorial and vocabulary), the excellent example models library, the well-organized Web site with links to the third party extensions which individuals have developed to fulfill specialized needs.

In [17, 18] NetLogo is one of the toolkits under discussion, this fact can be considered as an additional evidence of its popularity. The survey described in [17, 18] has delivered an interesting approach to modelling toolkits discussion and evaluation. The authors have engineered a set of 16 simple template models named in a whole as the 'Stupid Model'. The template models differ in their complexity: the first

model is the basic and the simplest artifact, the subsequent versions incrementally add features that are commonly used in real models. While realizing these models in every framework, it becomes evident how successful and effective modelling process can be, how much effort will be necessary for its completing. According to this survey, one could create all the following models with the help of NetLogo. This proves that our model, which specification was discussed in this paper, with no doubt, could be engineered in NetLogo.

Coding realized in NetLogo has some peculiarities, the first and sounding rather strange is that the modelling language is not an object-oriented but is a procedure-oriented one. It is a common fact that agents are often associated with objects (in the sense of object-oriented programming). In the NetLogo language, we cannot find objects as class instances at all.

All produced code is divided into two parts. One part is collected in the form of procedures under the 'Code' tab of the environmental window, and some code is produced through dialog windows of different GUI controls such as buttons, sliders, monitors, inputs, plots, etc. This approach has the evident disadvantage: there is no single place where one can look to see the whole program code.

There are two obligatory procedures in every NetLogo model. The first one is called 'to setup' and determines all the initial values and primary conditions for further simulation. The second one is called 'to go', and this procedure starts the simulating process itself.

The following lines contain a code example for the beginning of the 'setup' procedure taken from the model code:

```
to setup
clear-all
set-plot-pen-mode 1
set sp_ready []
  ;;set the current directory
  set-current-directory "D:\\models"
;;open the file with input data
file-open user-file
set v_num branches - 1
let temp 360 / branches
;;depict agents in the circle
crt branches [
      set heading temp * who
      set label who
      fd 10]
      …
end
```

One more interesting fact might be mentioned: in the modelling language's basic version, there are no arrays, which are to be the common structures in most of well-known programming languages. Instead of arrays, one can use lists, which have some peculiarities to be compared with arrays. It is more complex to set a single value in the list. For example, the following command set to 0 (zero) the element with number 'who' in the list named 'res_list'.

```
set res_list replace-item who res_list 0
```
On the contrary, it is easier to process the whole list in comparison with an array. For example, following command creates a copy of the list:
```
set res_list_01 res_list
```
In the last NeLogo language versions, it has become possible to use arrays as language extension, but one must point it out explicitly.

9.4   Simulation Results

Screenshot of the Communication Model is shown in Fig. 9.1. As was already discussed, this model was intended to solve different problems. The main goal was declared as to find out the conditions of economic system sustainability. It was shown that if the system's characteristics were in accordance with the Leontief's equilibrium (9.2), then all the communications would be done successfully and agents would be ready to begin a new production cycle.

Although money volume and agent exchange strategy do not affect the final successful result, they greatly influence the overall communication time and the character of money circulation in a system.
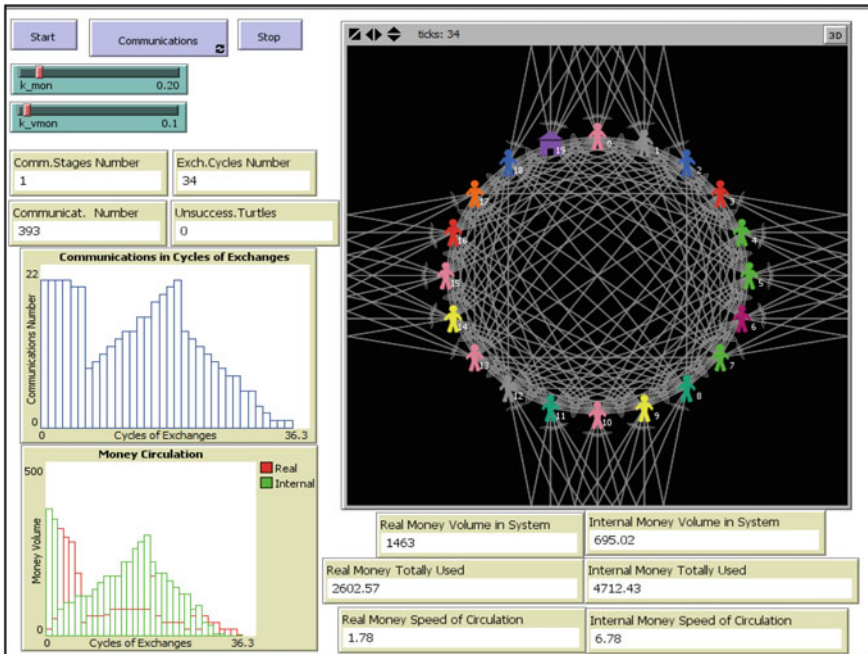
Fig. 9.1 Communication model screenshot
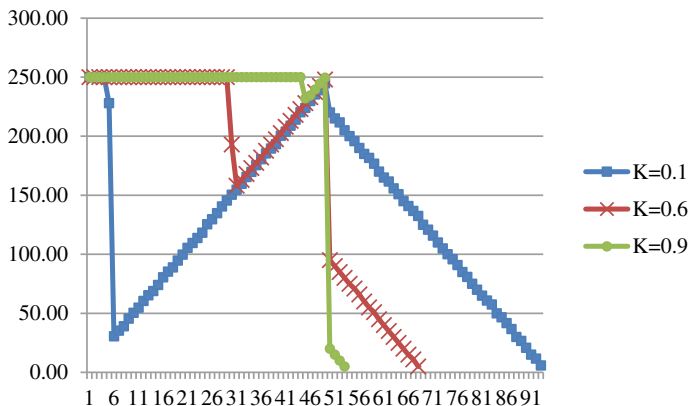
Fig. 9.2 Money turnover in the model with 50 agents with money supplement coefficients $K = 0.1$, $K = 0.6$ and $K = 0.9$

Communication time depends on the money volume in a system. The form of this dependence one can comprehend from Fig. 9.2, it becomes evident that the more money is circulated in a system, the shorter the period of time that is necessary to fulfill all the communications.

Figure 9.2 summarizes the results of experiments in the model with 50 agents with the same strategy used. In case of $K = 0.1$ (the money volume introduced into the system is equal to 10% of overall product value in it), all communications are completed while 94 cycles (in average); in case of $K = 0.6$, it takes 69 cycles, and, in case of $K = 0.9$, it takes only 54 cycles (the minimum value is 49 cycles in case of $K = 1.0$).

One can also find some specific features of the money turnover in the model. There are 3 main time periods: the first one—when all exchanges are successful and are fulfilled in the whole possible volume; the second one—the decrease of communications which is resulted in money circulation drop; and the third one—when remained communications are done by degrees.

The second period could be understood as a crisis period, and the third one is the way out of the crisis. In terms of money supplement, one can reveal that the less money in a system circulates the deeper and earlier the crisis occurs, and the longer time is necessary to leave this crisis state. In case of $K = 0.1$, the crisis starts in the 6th cycle and lasts till the 22nd cycle (15 cycles), while in case of $K = 0.9$, the crisis starts in the 46th cycle and lasts not more than 2 cycle (in case of $K = 1.0$ there is no crisis at all).

While simulations, one more phenomenon was revealed—the phenomenon of egoism. It emerges when a single agent changes its strategy and chooses the less optimal one. This agent as a single producer wins, but a system as the whole loses: it takes more time to complete all communications.

To understand possible roles of virtual money [20], a new version of the model was engineered, in which a new agent type was modelled, this type agent was called the External agent.

This new version model has proved to be a model of an opened type system. In the real life, we observe only opened type economic systems, since real economic systems are interrelated with each other, and local economics are involved into the global economic. In the new model one agent starts to play a role of an external environment, and exchanges with this agent simulate exchanges with the external world (i.e. import and export operations). One can find this agent in Fig. 9.1, it is depicted by the house icon.

In this model, two types of money exist: the internal, or virtual, money used only for internal exchanges, and the external, or real, money used for exchanges with the External agent (and also for internal exchanges in the case of the virtual money lack). The virtual money volume is in direct proportion to the preset virtual money supplement coefficient (K_vmon), and to the product volume in the system (the same rule is proposed for the real money volume, the corresponding coefficient is denoted as K_mon).

After simulations, we can postulate that virtual money can play a role of the real money and improve the communication process. It is clear from Fig. 9.3, where one can find diagrams of virtual, real and total money exchanges in the system with different volumes of real and virtual money.

From the structural point of view, there are internal exchange chains, they are of a closed type, virtual money circulates along these contours, and this accelerates the system communication process. Increasing the virtual money supplement coefficient value, one could shorten the communication time (degree of shortening depends on the agent strategy).

9.5 Conclusions

This paper tries to prove that the agent-based technology could deliver not only interesting, but rather useful results having been implemented for investigating a local economic system. Using the Communication Model, it was revealed that a productive system with industrial agents can become a sustainable one, i.e. work stably for a long time, if this system is based on the Leontief's intersectoral equilibrium. Money circulated in the model influences the communication process time and, moreover, this time depends on the agent behaviour (its exchange strategy).

Some phenomena were revealed during experiments, such as the crisis phenomenon and the phenomenon of egoism. They can hardly be predicted without this kind of simulation.

By exploring circular closed contours, it was found that virtual money introduction is a very useful act. It begins to work in accordance with the real one and significantly accelerates the communication process.
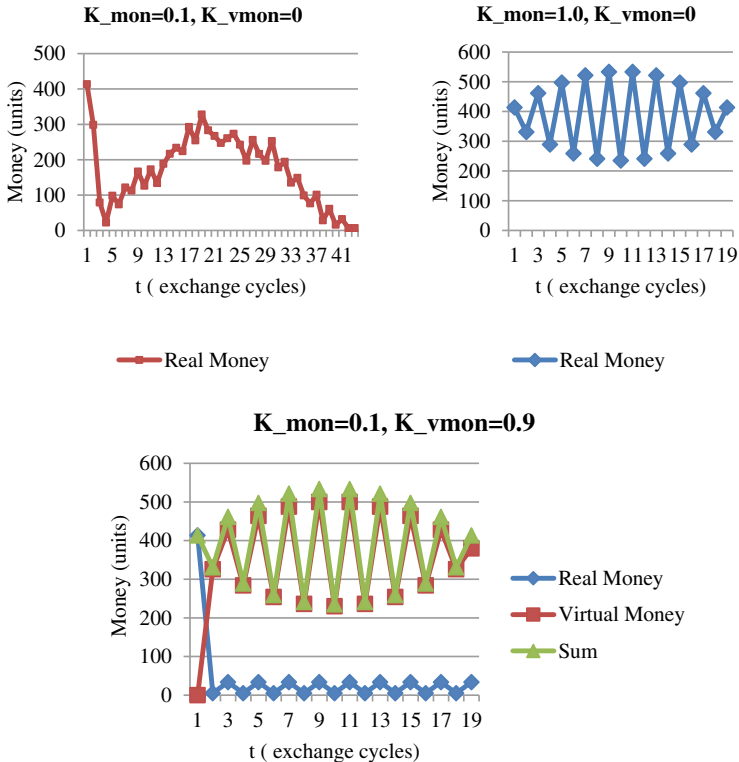
K_mon=0.1, K_vmon=0

K_mon=1.0, K_vmon=0

Real Money

Real Money

K_mon=0.1, K_vmon=0.9

Real Money

Virtual Money

Sum

Fig. 9.3 Money turnover in the model with 20 agents in different conditions of money supplement

References

1. C.M. Macal, M.J. North, Agent-based modeling and simulation, in *Proceedings of the 2009 Winter Simulation Conference*, pp. 86–98(2009)
2. C.M. Macal, M.J. North, Tutorial on agent-based modelling and simulation. J. Simul. **4**, 151–162 (2010)
3. S.-H. Chen, Varieties of agent-based computational economics: a historical and interdisciplinary perspective. J. Econ. Dyn. Control (2011)
4. E. Bonabeau, Agent-based modeling: methods and techniques for simulating human systems. PNAS **99**(suppl. 3), 7280–7287 (2002)
5. A. Troisi, V. Wong, M.A. Ratner, An agent-based approach for modeling molecular self-organization. Proc. Natl. Acad. Sci. USA **102**(2), 255–260 (2005). www.pnas.org/cgi/doi/10.1073/pnas.0408308102. Accessed 12 Jan 2018
6. E. Tatara, I. Birol, F. Teymour, A. Cinar, Agent-based control of autocatalytic replicators in networks of reactors. J. Comput. Chem. Eng. **29**, 807–815 (2005)

7. K.J. Mock, J.W. Testa, An Agent-Based Model of Predator-Prey Relationships Between Transient Killer Whales and Other Marine Mammals (University of Alaska Anchorage, Anchorage, AK). www.math.uaa.alaska.edu/orca/. Accessed 11 Jan 2018
8. V.A. Folcik, G.C. An, C.G. Orosz, The basic immune simulator: an agent-based model to study the interactions between innate and adaptive immunity. Theor. Bio. Med. Model. **4**(1), 39–56 (2007)
9. M. Gavin, Agent-based modeling and historical simulation. Digit. Hum. Q. **8**(1) (2014)
10. N. Luhmann, *Social Systems. Sketch of the General Theory* (Science, St. Petersburg, 2007), 668 p
11. P.L. Borrill, L. Tesfatsion, Agent-based modeling: the right mathematics for the social sciences, in *Working Paper No 10023*, July (2010)
12. L. Tesfatsion, Agent-based computational economics. www.econ.iastate.edu/tesfatsi/ace.htm. Accessed 19 Jan 2018
13. V.B. Kashkin, *Intoduction To Communication Theory: A Tutorial* (Flinta, Moscow, 2013), 224 p
14. D.B. Berg, O.M. Zvereva, S. Akenov, «Economic Microscope»: the agent-based model set as an instrument in an economic system research, in *ICNAAM 2016 Conference Proceeding,* Vol. 1863, Paper No 050006, 21 July 2017
15. V.V. Leontief, *Essays in economics. Theories, theorizing, facts and policies* (Politizdat, Moscow, 1990), 415 p
16. C. Nikolai, G. Madey, Tools of the trade: a survey of various agent based modeling platforms. J. Artif. Soc. Soc. Simul. **12**(2) (2009)
17. S.L. Railsback, S.F. Lytinen, S. Jackson, Agent-based simulation platforms: review and development recommendations. Simulation **82**, 609–623 (2006)
18. S.L. Lytinen, S.F. Railsback, The evolution of agent-based simulation platforms: a review of NetLogo 5.0 and ReLogo, in *Proceedings of the Fourth International Symposium on Agent-Based Modeling and Simulation* (2011). http://www2.econ.iastate.edu/tesfatsi/NetLogoReLogoReview.LytinenRailsback2012.pdf. Accessed 21 Jan 2018
19. U. Wilensky, NetLogo (Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL, 1999). http://ccl.northwestern.edu/netlogo/. Accessed 20 Jan 2018
20. B.A. Lietar, *The Future of Money. Creating New Wealth, Work and Wiser World* (KRPA Olymp: AST: Astrel, Moscow, 2007), 493 p

Chapter 10
Application of the Process Parameters Analysis Tree for the Melting Process Stability Control

Konstantin Aksyonov, Anna Antonova and Vasiliy Kruglov

10.1 Introduction

At present, most large industrial enterprises implement manufacturing execution systems and enterprise resource planning systems, digitize the metallurgical production, realize material tracking and introduce analytical analysis of historical data about production and orders. The problem of fully digitized production is solved by a metallurgical enterprise information system (MEI-system) [1–3] developed in conjunction with Ural Federal University.

One of the system tools intended for the work in real time is a process parameters analysis tree implemented in the models integration module. The tree is an AND-OR graph with four types of nodes [4]: source (mandatory node), result (mandatory node), formula node (designed to calculate parameter values) and If-then node (intended for tree branching). The process parameters analysis tree is used to solve the problems of dispatching and reassignment of products between orders based on the diagnostics of the process parameter values deviations. Interaction between the tree and external information systems is provided by the data exchange module that functionally corresponds to the class of Enterprise Services Bus [5]. The input parameters values for the parameters analysis tree come through the data exchange module from external information systems; for example, directly from the process control system sensors that control the process parameters. As a result of the process parameters analysis tree execution, the one output parameter is calculated, which through the data exchange module is broadcast to all subscribers including external information systems and other modules of the MEI-system, for example, the module for creating models or automated personnel workstation.

We consider the application of process parameters analysis tree for control of the melting process stability on a continuous casting machine.

K. Aksyonov (✉) · A. Antonova · V. Kruglov
Ural Federal University, Yekaterinburg, Russia
e-mail: wiper99@mail.ru

10.2 Problem Statement

When performing the melting operation, it is necessary to maintain low silicon [Si] content in the melting to ensure specific mechanical properties of the steel. Deviations in the silicon content are not allowed since in this case the forecast of the mechanical properties of steel will become difficult. Technologists monitor the content of silicon in steel using a Statistical Process Control (SPC) chart. The SPC chart is a tool based on a statistical analysis of the characteristics of the process under study [6]. The control is carried out by technologists along the green boundaries of the permissible range of variation in the relative content of silicon in the melting: the upper permissible boundary (*UPB*) and the lower permissible boundary (*LPB*). Deviations (shifts beyond *UPB* and *LPB*) should be analysed for causes and the causes in turn should be eliminated.

The control chart at the technologist workplace is constructed on the results of analysis of accumulated statistics of the studied process in any convenient instrument that supports performance of statistical calculations. The basis for constructing a control chart of the analysis of the relative silicon content in the melting is a vector of values of the silicon relative content in the melting in percentage *x*[Si]. This vector is formed via adding by technologist into the tool for creating control chart the values of [Si] obtained from the process control system at certain intervals.

An example of a control chart built in MS Excel to solve the considered problem is shown in Fig. 10.1. The upper graph illustrates the change in the values of the vector *x*[Si] when compared with the green permissible boundaries of the changes in the silicon content (*UPB* and *LPB*). The upper graph also contains red statistical
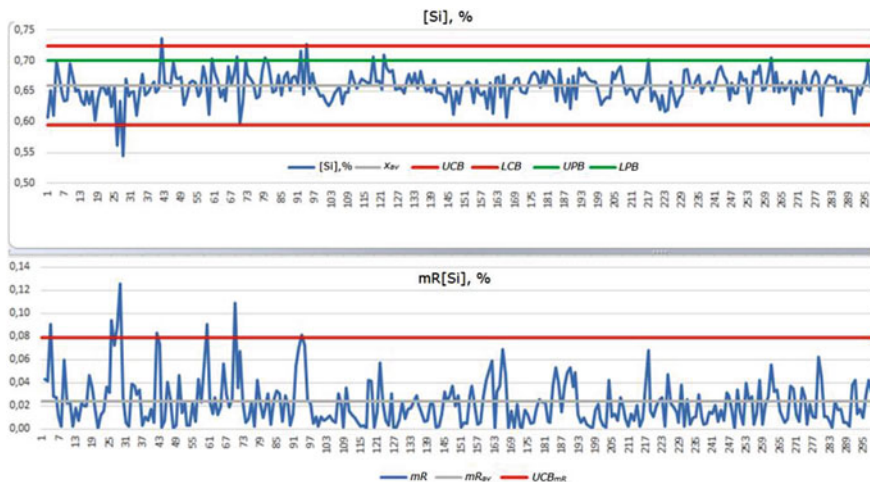
Fig. 10.1 Control chart example

control boundaries: the upper control boundary (*UCB*) and the lower control boundary (*LCB*); the boundaries of the *LPB* and *LCB* coincide. The gray line on the upper graph shows the average value of silicon content in the melting in percentage x_{av}.

The lower graph illustrates the change in the values of the vector mR[Si] containing the calculated values of the difference between the neighboring values of [Si]. The lower graph also has a red upper control boundary (UCB_{mR}) of the vector mR[Si]. The gray line on the lower graph shows the average value of the difference between adjacent values of the silicon content mR_{av}.

During the analysis of the control chart, the technologist performs the following algorithm. The melting process is considered as statistically unstable when the value of the relative silicon content [Si] in the melting exceeds the upper or lower control boundary, or when the value of the difference between adjacent values of [Si] exceeds the upper control boundary. This incident requires additional quality control of the resulting slabs. The melting process is considered as unacceptable when the value of the relative content of silicon [Si] in the melting exceeds the upper or lower permissible boundary. In this case, the technologist launches processes to find the causes of the incident, develops measures to address the causes of the incident and implements the measures developed.

10.3   Application of the Process Parameters Analysis Tree

We consider the application of the process parameters analysis tree to control the stability and permissibility of the casting process on a continuous casting machine.

The algorithm for performing calculations in the process parameters analysis tree is shown in Fig. 10.2 and Table 10.1.

The following parameters are used in Table 10.1 to calculate the stability of the process.

- x_i is the current relative silicon content in the melting (is the input parameter of the process parameters analysis tree).
- x_{i-1} is the previous relative content of silicon in the melting; by default at the time of the tree launch, this parameter is equal to 0.61 percentage.
- S_x is the cumulative sum of the silicon values; by default, at the time of the tree launch this parameter is equal to 0.
- N is the number of the accounted silicon values; by default it is equal to 0.
- mR_i is the range between neighbouring values of x.
- S_{mR} is the cumulative sum of the mR_i; by default it is equal to 0.

The values of the relative silicon content in the melting from the sensors of the automated process control system are automatically fed to the input of the tree via the data exchange module at specified intervals. A set $\{R_j\}$ is formed at the tree output, where R is an indicator of statistical stability and admissibility of the process studied, j is the leaf number of the tree, $j = 1, …, 8$. The calculation of R is carried
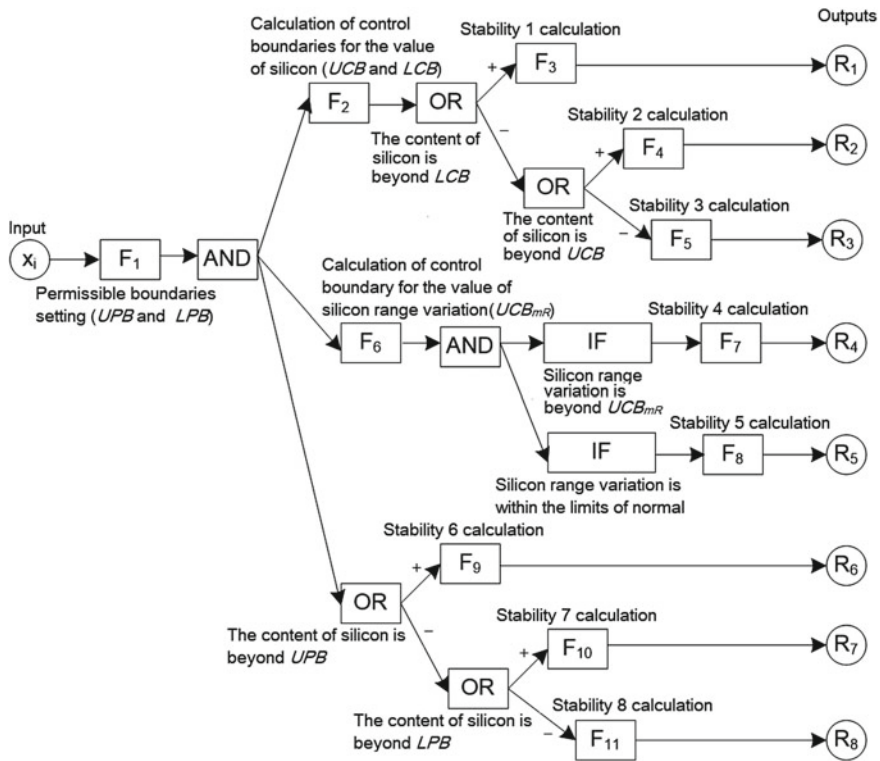
Fig. 10.2 Analysis tree structure

out according to the following formula: $R = \max\{R_j\}$. The parameter R can take the following values (see Table 10.1).

- "0" when the process is statistically stable and permissible ("green traffic light");
- "1" when the process is statistically unstable due to the output of the relative silicon content in the melting beyond the upper control boundary ("yellow traffic light");
- "2" when the process is statistically unstable due to the output of the relative silicon content in the melting beyond the lower control boundary ("yellow traffic light");
- "3" when the process is inadmissible due to the output of the relative silicon content in the melting beyond the upper permissible boundary ("red traffic light");
- "4" when the process is inadmissible due to the output of the relative silicon content in the melting beyond the lower permissible boundary ("red traffic light").

The structure of the model of analysis tree of the parameter of relative silicon content in the melting is shown in Fig. 10.3. The model has been implemented in the module for creating models of the MEI-system. In the upper right corner of the figure, the form for describing the rules of the node "Calculation of control boundary for the value of silicon range variation (UCB_{mR})" is shown.

Table 10.1 Algorithm for tree calculation

Node	Name of tree node	Calculations in the node
F_1	Permissible boundary setting (UPB and LPB)	$UPB = 0.7$ $LPB = 0.5$ $N = N + 1$
F_2	Calculation of control boundaries for the value of silicon (UCB and LCB)	$S_x = S_x + x_i$ $x_{av} = S_x/N$ $UCB = x_{av} + 2.67 \cdot x_{av}$ $LCB = x_{av} - 2.67 \cdot x_{av}$
F_6	Calculation of control boundary for the value of silicon range variation (UCB_{mR})	$mR_i = \max(x_i, x_{i-1}) - \min(x_i, x_{i-1})$ $S_{mR} = S_{mR} + mR_i$ $mR_{av} = S_{mR}/N$ $UCB_{mR} = 3.27 \cdot mR_{av}$
IF	The content of silicon is beyond LCB	$x_i < LCB$
IF	The content of silicon is beyond UCB	$x_i < UCB$
IF	Silicon range variation is beyond UCB_{mR}	$mR_i > UCB_{mR}$
IF	Silicon range variation is within limits of normal	$mR_i \leq UCB_{mR}$
IF	The content of silicon is beyond UPB	$x_i > UPB$
IF	The content of silicon is beyond LPB	$x_i < LPB$
F_3	Stability 1 calculation	$R_1 = 2$ $x_{i-1} = x_i$
F_4	Stability 2 calculation	$R_2 = 1$ $x_{i-1} = x_i$
F_5	Stability 3 calculation	$R_3 = 0$ $x_{i-1} = x_i$
F_7	Stability 4 calculation	$R_4 = 1$ $x_{i-1} = x_i$
F_8	Stability 5 calculation	$R_5 = 0$ $x_{i-1} = x_i$
F_9	Stability 6 calculation	$R_6 = 3$ $x_{i-1} = x_i$
F_{10}	Stability 7 calculation	$R_7 = 4$ $x_{i-1} = x_i$
F_{11}	Stability 8 calculation	$R_8 = 0$ $x_{i-1} = x_i$

After the development of the tree model, for each models parameter of the type "input" and "output" it is necessary to connect: the source-parameter of the data exchange module with the input parameter of the tree model, the receiver-parameter of the data exchange module and the output parameter of the tree model.

To perform the analysis of statistical stability and feasibility of the melting casting in real time, it is necessary to launch the developed tree model in the models integration module of the MEI-system. The models integration module interacts in real time
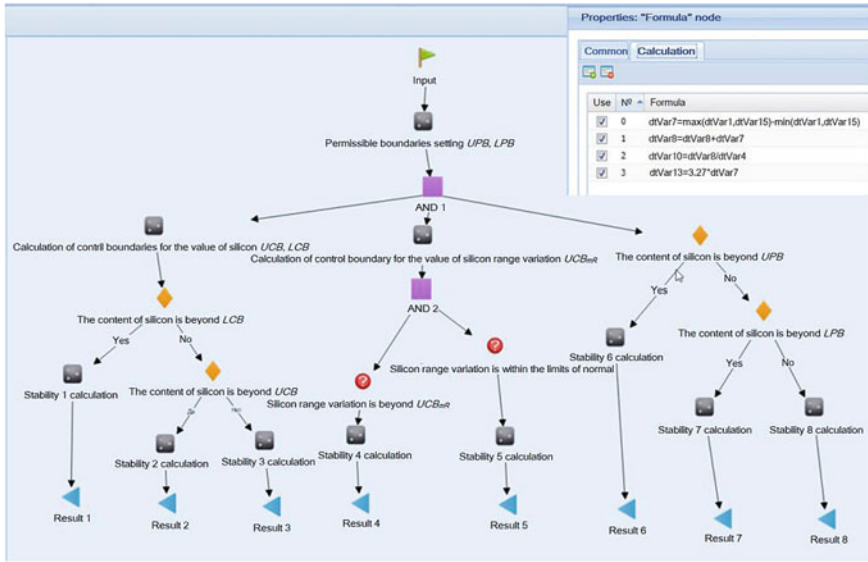
Fig. 10.3 Tree model structure

with the data exchange module, receives and transmits parameters to it. Parameters are further transmitted by the data exchange module to external information systems including automated workplace of technologist.

When the tree model is executed in the module for models integration of the MEI-system, the technologist observes in their workplace a graph of the current silicon content change [Si] and an indication in the form of a traffic light signalling the stability and feasibility of the melting process on the continuous casting machine (Fig. 10.4). A green traffic light signal does not require any additional action by the technologist. A yellow signal of a traffic light requires additional quality control of
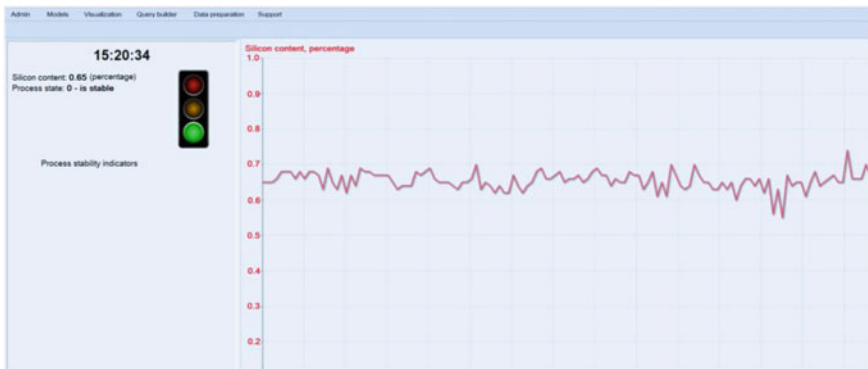
Fig. 10.4 Example of an indication

the resulting slabs. In the event of a red traffic signal, the technologist launches the processes of searching for the causes of the incident, develops measures to eliminate the causes of the incident, and implements the measures developed.

As a result of comparing the application of the process parameters analysis tree in the structure of the MEI-system and control charts, the following conclusions have been drawn to the solution of the problem of control of statistical stability and permissibility of the melting casting process on the continuous casting machine. The advantage of control charts is the visibility of the information provided from the point of view of the trained technologist. The drawbacks of control charts include the orientation towards the implementation of calculations with accumulated statistics as a result of which the decision time is late for the time of collection, preparation and analysis of the process flow statistics.

The advantages of the process parameters analysis tree include:

- performing calculations in real time, which increases the speed of decision-making;
- automatic exchange of production and design parameters between the modules of the MEI-system and external information systems with a configurable interval of exchange with an accuracy of tenths of a second;
- visualization in real time changes in the values of analysed process parameters;
- use of intuitive colour indication of the stability of the analysed process (traffic light as indicator).

10.4 Conclusions

A tool for analysing the parameters of the observed process in real time is presented in the paper. This is the process parameters analysis tree. The analysis tree is designed to solve the problems of dispatching and reassigning units of products between orders based on diagnostics of deviations of process parameters values. The analysis tree is implemented in the subsystem of the metallurgical enterprise information system. The execution of the tree in real time is carried out in the models integration module of the MEI-system. The connection of production parameters (at the input of the tree) and calculated parameters (at the output of the tree) to the enterprise external information systems and other modules of the MEI-system is carried out by means of the data exchange module with the automated systems of the enterprise.

The process parameter analysis tree has been applied for the stability control problem of melting process for a continuous casting machine. Based on the results of the problem solution, the analysis tree has been compared with control charts as the most common tool for performing statistical analysis of the process parameters. The advantages of using the tree for analysing the process parameters for solving the problem of operational control of the stability of the process are revealed. The advantages are primarily due to the execution of the analysis tree in real time as well

as the automatic exchange of tree data with external systems of the enterprise and the modules of the MEI-system.

Future work is related to the application of the tree for analysis of process parameters when solving the problems of reassignment of products between orders based on analysis of deviations in product parameters from given ranges.

References

1. A. Borodin, Y. Kiselev, S. Mirvoda, S. Porshnev, On design of domain-specific query language for the metallurgical industry, in *Proceedings of the 11th International Conference BDAS: Beyond Databases, Architectures and Structures: Communications in Computer and Information Science* (2015), pp. 505–515
2. K. Aksyonov, E. Bykov, O. Aksyonova, A. Antonova, Development of real-time simulation models: integration with enterprise information systems, in *Proceedings of the Ninth International Multi-Conference on Computing in the Global Information Technology* (2014), pp. 45–50
3. A. Borodin, S. Mirvoda, I. Kulikov, S. Porshnev, Optimization of memory operations in generalized search trees of PostgreSQL, in *Proceedings of the International Conference: Beyond Databases, Architectures and Structures* (2017), pp. 224–232
4. K. Aksyonov, A. Antonova, Application of the process parameter analysis tree for technological process analysis, in *Proceedings of the 2018 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT)* (2018), pp. 212–215
5. D. Chappell, *Enterprise Service Bus* (O'Reilly, 2004)
6. W. Shewhart, E. Deming, *Statistical Method from the Viewpoint of Quality Control* (Courier Corporation, 1986)

Chapter 11
Artificial Neural Networks as an Interpolation Method for Estimation of Chemical Element Contents in the Soil

A. Buevich, A. Sergeev, D. Tarasov and A. Medvedev

11.1 Introduction

Estimation and simulation of distribution of contaminants in deposited media is an important issue of environmental studies in habitat quality estimation research [1]. There are two groups of methods of estimation: deterministic and geostatistical. One of the most widely used methods among the geostatistic techniques is kriging. This method minimizes the variance of the estimation errors and shows considerable advantages in the prediction of the soil pollution compared with deterministic interpolation methods such as polynomial methods [2–4]. However, efficiency of the kriging application depends on the actual spatial irregularity of the distribution of modelled contaminants, which under conditions of heterogeneous environment makes the application of interpolation method ineffective.

In recent years, a significant number of publications are devoted to the application of alternative methods in environmental modeling, such as the Land Use Regression and ANN. These methods, in contrast to kriging, allow taking into account various factors influencing the spatial distribution of contaminants in the deposited media. Studies show the successful application of the ANN modeling for the prediction of the global and local climate change [5], groundwater pollution [6], as well as to the study of the behavior and prediction of the spread of animals [7, 8]. Some publications [9, 10] are dedicated to the comparison of various modeling methods (multiple regression analysis, nonlinear dynamic algorithms) and ANN which is pointed out as preferable. Some works [11, 12] are devoted to the search of ANN optimal structure, training algorithms, overcoming the drawbacks inherent in the method of modeling [13], and to the creation of the hybrid models based on several methods of modeling [14]. In this study, the ANN approach is applied to evaluate and

A. Buevich · A. Sergeev · D. Tarasov (✉) · A. Medvedev
Institute of Industrial Ecology, UB RAS, Yekaterinburg, Russia

Ural Federal University, Yekaterinburg, Russia

forecast the spatial distribution of chemical elements in the topsoil on the example of a small natural geological landscape area (1.0 m × 1.0 m) located away from sources of pollution.

11.2 Materials and Methods

The soil sampling site was selected in an undisturbed and natural area about 60 km away from the urban territories and had the form of a square with a 1 m side. The marked square was split into 100 cells. The soil cores were sampled to a depth of 0.05 m and then were dried. For analysis of the soil specimens, the X-ray fluorescence spectrometer INNOV X-5000 was chosen. The method of spectrometry does not require any complex preparation and expensive reagents, and allows one to analyse the specimens many times. The soil specimens were analysed in the "Mountain Plus" mode, which allows one to use beams with different power (10–50 keV) for the analysis. The exposition of each specimen was 60 s. Each specimen was exposed 10 times and the average value was calculated for further work.

A multilayer perceptron of the ANN type with the Levenberg-Marquardt learning algorithm was selected for this study. Such network consists of a plurality of sensor elements forming the input layer: one or more hidden layers of neurons, and one output layer. The Levenberg-Marquardt algorithm is considered as one of the best nonlinear optimization algorithms known to date. The interpolation methods were implemented using the Esri ArcGIS geostatistical software, while the ANN was trained using a neural network toolbox in MATLAB.

At the first step, the network model for constructing the distribution of each element contained in the specimen was chosen. Coordinates of the measured points were submitted to the input of the network. At the output, concentrations of each separate element were received as a result of the X-ray fluorescence analysis. For searching the optimal network, a subroutine was created in MATLAB. The subroutine iterates the network models with a given architecture until the model with the highest correlation coefficient for each element will be found. If the correlation coefficient is lower than the predetermined value ($r = 0.8$), the network settings are changed and the models are iterated again. A previous study [1] showed that the optimal ANN model for a given area must contain three hidden layers. Therefore, only the total number of neurons in the hidden layers and the number of neurons in each layer were changed. As a result, a neural network with the highest value of the correlation coefficient for each element was constructed.

At the second step, concentration distributions for each element were built by the Esri ArcGIS geostatistical software based on the data of X-ray fluorescence analysis. Then, the sample was randomly separated using the "create subset" function of the Geostatistical Analyst in the Esri ArcGIS on two sub-samples, namely, the test and the training ones in the ratio of 30/70. The training sub-sample (70 specimens) was used as the learning set. Then the concentration value of each element was returned by the ANN as the test sub-sample (30 specimens). Three validation indices were

used to evaluate the performance of interpolation: the mean absolute error (MAE), root mean square error (RMSE), and relative root mean square error (RRMSE). These indices were computed from the expressions:

$$MAE = \frac{\sum_{i=1}^{n}|p_i - o_i|}{n},$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(p_i - o_i)^2}{n}},$$

$$RRMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{p_i - o_i}{o_i}\right)^2},$$

where p_i is the estimated concentration (ANN), o_i is the measured concentration, and n is the number of validation sites.

11.3 Results and Discussion

Table 11.1 summarizes the main statistics of the concentration of the measured elements in the study area. Concentrations of all measured elements were positively skewed (excepting Si and Ni) and showed both leptokurtic (Cr, Ti, V, Mn, Fe, Zr) and platykurtic distributions (Si, K, Ni). Table 11.2 shows the best network architecture for each element with the maximum value correlation coefficient obtained by computer simulations. Table 11.3 shows the accuracy assessment indices of the estimated concentrations of the elements at validation sites for the ANN. The MAE for ANN, which indicates the extent, to which the process leads to error, was low for all modeled elements for the model built on all 100 specimens. The RMSE was low with ANN for all modeled elements in both prediction models. The RRMSE was almost identical for all modeled elements.

A trained ANN is suitable for the construction of the spatial distribution of chemical elements both as an interpolator and as a prognostic model. The ANN gives prediction models comparable in accuracy for all modeled elements for the site with homogeneous medium. Moreover, the ANN has more functionality and allows predicting the spatial distribution and availability of sources in a heterogeneous environment of urbanized areas. Figure 11.1 shows the surface distribution of chromium in the study area made by the trained ANN.

Table 11.1 Descriptive statistics of the sample

	Si	K	Ti	Fe	V	Cr	Mn	Zr	Ni
Min	395940	10196	1978	1389	683.3	384.3	189.7	31.9	68.7
Max	461050	31715	7090	5696	892.4	573.8	469.4	350.5	103.2
Mean	429500	18897	3319	3055	774.5	460.3	295.6	119.5	86.75
SD	13403	5625	854.7	799.1	38.98	34.57	47.38	64.02	6.976
Coefficient of variation	0.03	0.30	0.26	0.26	0.05	0.08	0.16	0.54	0.08
Skewness	−0.032	0.589	1.21	0.489	0.285	0.477	0.819	1.089	−0.117
Kurtosis	2.609	2.373	5.625	3.288	3.406	3.328	4.11	4.144	2.505
Median	438540	17607	3219	2972	773	457.5	289.1	110.4	87.4

Table 11.2 ANN structures and correlation coefficients

Multilayer perceptron built on 100 specimens

	Si	K	Ti	Fe	V	Cr	Mn	Zr	Ni
Structure[a]	22-16-22	20-16-20	22-16-20	20-16-20	20-16-20	20-16-20	20-12-16	20-16-20	22-16-20
Max CC[b]	0.81	0.93	0.90	0.91	0.89	0.86	0.87	0.91	0.80

Multilayer perceptron built on 70 training specimens

Structure	22-16-18	22-16-18	22-16-18	22-16-18	22-16-18	22-16-18	22-16-18	22-16-18	22-16-14
Max CC	0.81	0.87	0.86	0.90	0.82	0.87	0.87	0.92	0.89

[a]Structure 22–16–22 means that the ANN has 3 hidden layers with the following numbers of the neurons: 22 in the first layer, 16 in the second, and 22 in the third layer

[b]Max CC is the maximum correlation coefficient achieved for a given element by computer simulation

Fig. 11.1   Predicted Chromium concentration maps using **a** ANN (built on 100 specimens), **b** ANN (built on 30 test specimens)

11.4   Conclusions

The assumption that the trained ANN will give prediction models comparable in accuracy with other methods even for a homogeneous medium is completely confirmed in the present study for all modelled elements. The authors believe that in a heterogeneous environment of urbanized areas, neural network models will be able to identify sources of exposure. Moreover, such models allow one to determine the intensity of element concentrations, to demonstrate the characteristic geochemical range of their emissions and to predict their impact on the environment. In addition [14], the application of a hybrid ANN-Kriging model in which the residuals of ANN are estimated by ordinary kriging improves the accuracy of the distribution's forecast for the chemical elements in the topsoil.

References

1. A.P. Sergeev, E.M. Baglaeva, A.V. Shichkin, Case of soil surface chromium anomaly of a northern urban. Atmos. Pollut. Res. **1**, 44–49 (2010)
2. R. Webster, M. Oliver, *Geostatistics for Environmental Scientists*. (Wiley, Chichester, 2001)

3. X.M. Liu, K.L. Zhao, J.M. Xu, M.H. Zhan, B. Wang, F. Si, Spatial variability of soil organic matter and nutrients in paddy fields at various scales in southeast China. Environ. Geol. **53**, 1139–1147 (2008)

4. L. Worsham, D. Markewitz, N. Nibbelink, Incorporating spatial dependence into estimates of soil carbon contents under different land covers. Soil Sci. Soc. Am. J. **74**, 635–646 (2010)

5. L. Zelin, P. Changhui, W. Xiang et al., Application of artificial neural networks in global climate change and ecological research. Chin. Sci. Bull. **55**, 3853–3863 (2010)

6. R. Shaker, L. Tofan, M. Bucur et al., Network modelling approach applied to dobrogea, Romania. J. Environ. Prot. Ecol. **11**, 337–348 (2010)

7. A.J. Tracey, J. Zhu, R.K. Crooks, Modeling and inference of animal movement using artificial neural networks. Environ. Ecol. Stat. **18**, 393–410 (2011)

8. J.M. Watts, S.P. Worner, Comparing ensemble and cascaded neural networks that combine biotic and abiotic variables to predict insect species distribution. Ecol. Inf. **3**, 354–366 (2008)

9. G.B. Sahoo, S.G. Schladow, J.E. Reuter, Forecasting stream water temperature using regression analysis, artificial neural network, and chaotic non-linear dynamic models. J. Hydrol. **378**, 325–342 (2009)

10. S. Helama, N.G. Makarenko, L.M. Karimova et al., Dendroclimatic transfer functions revisited: Little ice age and medieval warm period summer temperatures reconstructed using artificial neural networks and linear algorithms. Ann. Geophys. **27**, 1097–1111 (2009)

11. R.C. Tosh, D.G. Ruxton, The need for stochastic replication of ecological neural networks. Philos. Trans. R. Soc Biol. Sci. **362**, 455–460 (2007)

12. S.L. Ozesmi, C.O. Tan, U. Ozesmi, Methodological issues in building, training, and testing artificial neural networks in ecological applications, in *3rd Conference of the International-Society-for-Ecological-Informatics (ISEI)*, vol. 195 (Rome, Italy, 2006), pp. 83–93

13. M. Gevrey, I. Dimopoulos, S. Lek, Two-way interaction of input variables in the sensitivity analysis of neural network models, in *3rd Conference of the International-Society-for-Ecological-Informatics (ISEI)*, vol. 195 (Rome, Italy, 2006), pp. 43–50

14. D. Fuqiang, Z. Qigang, L. Zhiqiang, W. Xuemei, L. Gangcai, Spatial prediction of soil organic matter content integrating artificial neural network and ordinary kriging in Tibetan Plateau. Ecol. Ind. **45**, 184–194 (2014)

Chapter 12
Five-Module Concept of Modernization of Urban Passenger Transport Management

S. Trofimov, N. Druzhinina and O. Trofimova

12.1 Introduction

Passenger networks of megapolises are actively developing, and there is not enough substantiation of the proposed projects for their reorganization. To solve this problem, it is necessary to identify the correspondence matrix of the passenger traffic, build criteria for assessing the quality of the transport network and develop algorithms for comparing alternative projects.

12.2 Pre-processing Module

In the data pre-processing module, the network topology is analysed. Data on the routes and the movement schedule are combined into clusters. A cluster is created for each stopping point and consists of the stages coming from this stop. A stage contains the start time, the number and the direction of the route, the type of route and so forth. Thus, the cluster consists of schedules chopped into pieces. Clusters can be interconnected by special stages, for example, pedestrian stages. Clusters allow one to aggregate information about schedules for different types of transport.

S. Trofimov (✉) · N. Druzhinina · O. Trofimova
Ural Federal University, Yekaterinburg, Russia
e-mail: tsp61@mail.ru

12.3 Mathematical Model of the Transport Network, the Algorithm and Software Implementation for Finding the Optimal Route

In the second module, the passenger is provided with a new information service: the construction of an optimal route with transfers taking into account a previously strictly approved schedule. The optimal route is based on the given start and end points of the trip and the start time. The optimality criteria can be the duration of the trip, the cost of the trip taking into account the transfers, the level of comfort, etc. In addition to the utilitarian profit of optimal routes, this module plays a decisive role in the proposed concept. It is reasonable to assume that a passenger intuitively or knowingly chooses the optimal routes for their trips. Therefore, when changes in the *UPT* network one can predict future passenger traffic.

For the approbation of algorithms, a model of the *UPT* network has been developed, which is close to the real conditions of the metropolis [1]. The model megacity is a rectangular system of streets with given dimensions. The number of rectilinear and circular routes of the *UPT* is specified. The network topology and the routing schedule are automatically generated.

The mathematical model of the *UPT* network is determined by the following sets: X is the set of stops; P is the set of ranges where the range p_{ij} is the directed section of the road between the stops i, j of X; the set of trajectories $T(n)$ for each of the N routes, that is, the sequence of bound ranges; a set of schedules for the movement of a mobile unit (*MU*) $G(n, t_{start})$. The simulation of each element of all sets is performed using a random number generator with user-defined parameters. The distribution parameters are chosen so that the results are as close as possible to the actual transport network. The trajectory of a route is an ordered sequence of ranges

$$T(n) = \left\{ P_{1,2}; \ldots; P_{m(n)-1,m(n)} \right\}, \tag{12.1}$$

where $m(n)$ is the number of stops in route n, including the start and end stops. Trajectories can contain repeated ranges. The time sequence of the route is specified by the time of start t_{start} of the *MU* n from the initial stop and the arrival times to the next stops

$$V(n, t_{start}) = \{t_1 = t_{start}, t_2(n, t_{start}), \ldots, t_m(n)(n, t_{start})\}. \tag{12.2}$$

The route schedule is defined as the set of the route path and the time sequence of the route

$$G(n, t_{start}) = \{T(n), V(n, t_{start})\}. \tag{12.3}$$

The route of the *UPT* has several traffic schedules within 24 h. An algorithm is developed for searching for the passenger's optimal route with a given start time

[1]. The algorithm is based on the principle of dynamic programming. We calculate the time spent on traffic and on waiting for transport, taking into account transfers. Suppose at a stop A at the initial time t_{start} there is a passenger who needs to arrive at a stop B in the minimum time. The passenger's way can contain transfers. The stop is the center of the cluster with outgoing ranges. The stop label $M(i)$ is the set of the route and the time of arrival at the given stop for the given schedule.

$$M(i) = \{G(n, t_{start}), t_i(n, t_{start})\}. \tag{12.4}$$

The stop label allows us to determine on which route and from which previous stop the passenger arrived at stop i. A label with an earlier arrival time overwrites the old label. The labels restore the passenger's way in the opposite direction: from stop i to the initial stop A. Stop A does not have a label. Arrangement of stop labels is carried out according to the wave principle: first, stops from A are considered at a radius equal to one stage, then two stages, etc. The criterion for ending the algorithm is: none of the labels has changed when viewing all stops of the next and previous radius; if the final stop B has labeled with the arrival time t_{end}, then viewing of the stops originating from other labeled stops with more than t_{end} time is not performed.

A summary screen of the simulation example is shown in Fig. 12.1. Stops coincide with crossroads and are numbered from 1 to 400. The optimal trajectory is coloured. The colour changing (digits) indicates a transfer.

In this paper, we demonstrate a possibility of application of this algorithm not for a single person, but for the entire population of the city. This allows one to get the experimental data and to calculate the new statistics quality for *UPT*.

The simulation results show that the algorithm can be applied to the real transport network with real schedules.

Fig. 12.1 The result of searching for the optimal path
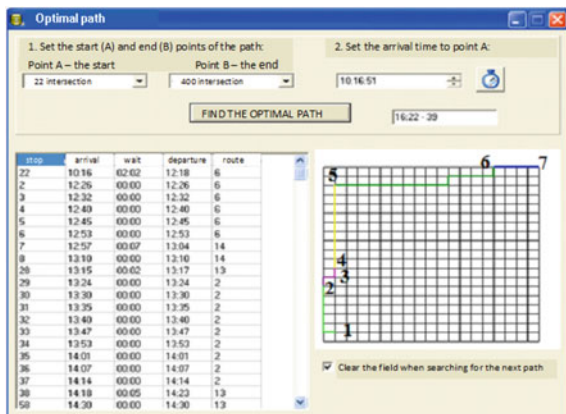
12.4   Identification of the Correspondence Matrix of a Urban Passenger Traffic on the Basis of the Navigation System

The third module includes the identification of the correspondence matrix of passenger traffic using information from the electronic payment system [2–4]. In this case, passenger transfers are glued together in one route. Based on the analysis of all one-day transactions of this passenger, the starting point of its return way is determined, which, most likely, coincides with the end point of its direct path. Thus, transfers are not included in the correspondence matrix.

Each stop of the *UPT* route network of a real megacity is represented as a node of a topological grid with known coordinates (x, y). The passenger's route (based on passenger traffic) is represented by a set of coordinates of the start (x_{start}, y_{start}) and finish (x_{end}, y_{end}) points of the route and the start time of the movement (t_{start}). The correspondence matrix of passenger traffic allowed to determine the passenger traffic density function in the 5-dimensional space of points $(x_{start}, y_{start}, t_{start}, x_{end}, y_{end})$. The estimation of passenger traffic density of the whole megacity is based on passenger traffic density estimated for one passenger, but for a long time, for example, during the year. The passenger is identified by the *E*-card number. A significant amount of information about one passenger allows us to restore incomplete or distorted data on the routes of this passenger. As a result, the annual individual density of passenger traffic (*AIPT*) is constructed. To obtain a one-day *AIPT*, the annual matrix is normalized by dividing by the number of days in a year. The total one-day correspondence matrix is constructed as the weighted sum of one-day *AIPTs*, in which weights are the activity of the movements of an individual passenger.

Thus, the one-day travel of an individual passenger is not a single trip of one person, but their fractional trips along all the trajectories that they made during the year.

12.5   Movement of Passenger Traffic Taking into Account the Current Schedule of the UPT

In the fourth module, we assume that passenger traffic with a known correspondence matrix moves along the optimal routes.

It is assumed that the fractional passenger flow, built on the basis of a correspondence matrix and the density of passenger traffic, moves along the optimal routes. The element of a fractional passenger traffic is the weighted passenger route. The set of constructed optimal routes with different weights is the movement of the passenger traffic.

This approach allows constructing integral estimates of the *UPT* network. The quantitative estimates of the schedule, the topology of the routes of the *UPT* network, are compared. Such assessments may include:

- Boundary indicators: minimum time or minimum number of transfers for which the passenger is guaranteed to get from one arbitrary stop of the megacity to any other, subject to the schedule for the routes;
- average parameters: average waiting time by a passenger at a stop, average length of a passenger's route, average travel time, average number of transfers, average traffic load.

These indicators can be obtained both on working days and on weekends.

As a rule, passengers of the *UPT* do not have the ability to influence the scheduling or routes, but on the contrary they are forced to adapt to them. For more convenient (optimal) use of the *UPT*, it would be useful for the passenger to know how the path and time of its passage will change as the initial conditions of the trip change. The sensitivity of the optimal solution is analysed for changes in the parameters of the initial model [1].

In Fig. 12.2 shows the analysis of the sensitivity of the optimal path time with a change in the start time. Sensitivity is the derivative of this graph. The real schedule for tram and trolleybus traffic in Yekaterinburg was used in the calculations. Initial data: departure point *A*: stop "str. Enthusiasts"; destination *B*: stop "str. Peace". The results of the analysis show that in the interval from 15 h 32 min up to 21 h 32 min, the minimum travel time from stop *A* to stop *B* varies from 30 to 50 min. High sensitivity is a disadvantage of *UPT*, it contributes to increasing the workload of the *UPT* on optimal trajectories.

The system allows city's authority to evaluate various assessments in case of upgrading the *UPT* network and to justify the transition to a new configuration of the *UPT* network, which is relevant for modern megacities.

The implementation of the first four modules makes it possible to formulate a new concept of a system for informing passengers about the work of the *UPT*.

Fig. 12.2 Analysis of dependence of the optimal path on the start time
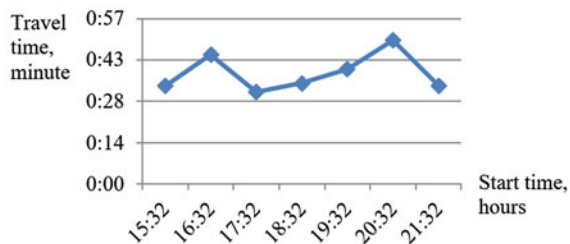
12.6 Mobile Web-Application "Informator" for Finding the Best Way for a Public Transport Passenger, Taking into Account the Current Traffic Schedule

The fifth module is focused on the development of the information system based on the new concept of the passenger information system about possible options for its current route to the *UPT*. In the master's thesis Bernhardt M.A. developed a prototype of the mobile web application "Informator" [5]. In real time, the application informs the passenger about the transfers and transitions to other branches, including the temporary routes from the cluster database (Fig. 12.3). If the *MU* leaves the
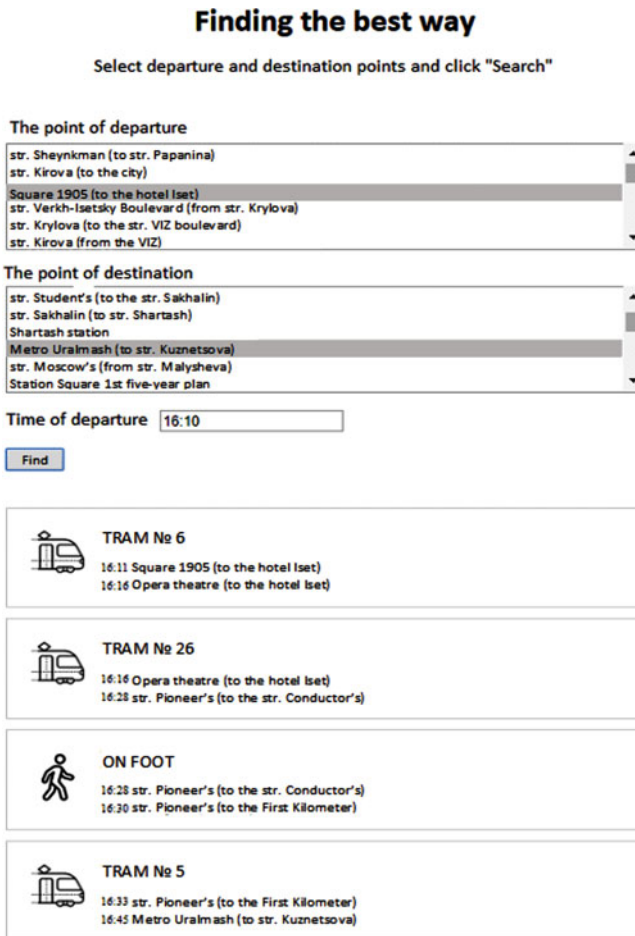
Fig. 12.3 The optimal passenger route: "Square 1905 (to Iset)"–"Metro Uralmash (to str. Kuznetsova)", departure time: 16:10, time of arrival: 16:45

route (or when the schedule is changed), information on previously planned stages is removed from the clusters and new temporary stages are added to the clusters located on the changed route. Changes in clusters will affect the choice of the optimal route for the passenger in online mode.

The program is designed to find the best way for the passenger of the *UPT*, taking into account the current schedule for the movement of route vehicles. However, the program does not allow to solve a multi-criteria problem: to choose from a set of all optimal routes, the optimal route according to another criterion, for example, with a minimum number of transfers. The current changes in the work of the routed transport (traffic jams, road accidents, arrivals to depots) should be made in real time in the schedule, which is not currently implemented by the dispatch services. For pedestrian crossings between stops, only the average pedestrian parameters are taken into account. The main advantage of the program: the passenger has a guaranteed opportunity to find the best route, taking into account the current schedule at any time of the day, and guests of the city have the opportunity to navigate through unfamiliar routes.

12.7   Conclusions

A modular concept of improving the work of urban passenger transport was developed and implemented. In the first module, the raw data on the network topology and the traffic schedule is converted into an intermediate database, convenient for implementing the algorithm for finding the optimal passenger path in the second module. The algorithm uses dynamic programming method, complicated by the presence of the time factor. The third module involves assessing the correspondence matrix of passenger traffic using electronic e-cards. The passenger's transfers as they move from the start to the end point of the path are glued into one route. In the fourth module, it is supposed that passenger routes in practice are optimal. This enables us to calculate the quantitative assessment of the quality of the transport network topology and route schedules. This module uses the correspondence matrix, resulting in the third module. In the last module, we construct a web-application, which allows a passenger to find out optimal paths taking into account the actual schedule.

References

1. S.P. Trofimov, N.G. Druzhinina, O.G. Trofimova, Algorithms for constructing optimal paths and statistical analysis of passenger traffic, in J. Phys. Conf. Series, **944** (2018). 012117. https://doi.org/10.1088/1742-6596/944/1/012117
2. N.G. Druzhinina, S.P. Trofimov, O.G. Trofimova, The formation of transaction traffic at the urban electric transport, in *The certificate of state registration of computer programs*, 2014661819, The Federal service for intellectual property: Moscow, Russia (2014)
3. N.G. Druzhinina, S.P. Trofimov, O.G. Trofimova, Estimation of distribution of inter-district passenger traffic of city electric transport. Softw. Products Syst. **2**(110), 147–153 (2015)
4. N.G. Druzhinina, K.V. Shirtladze, O.G. Trofimova, Statistical analysis of passenger traffic using the example of an electronic payment system, in *The certificate of state registration of computer programs,* 2016661213, The Federal service for intellectual property: Moscow, Russia (2016)
5. S.P. Trofimov et al., Search for the optimal path of the passenger of urban electric transport, using the schedule, in *The Certificate of State Registration of Computer Programs, 2018662684,* The Federal service for intellectual property: Moscow, Russia (2018)

Part IV
Antennas and Microwaves

Chapter 13
Green's Functions for Multilayer Cylindrical Structures and Their Application to Radiation, Propagation and Scattering Problems

S. Daylis and S. Shabunin

13.1 Introduction

For electromagnetic wave radiation, propagation, and scattering problems both analytical and numerical methods are used. Analytical methods make possible the physical interpretation of obtained results but they are often rather limited by the shape and structure of qualified objects. Numerical methods allow one to perform the analysis of complicated structures of any shape, structure, and materials but often do not allow the interpretation the observed effects. The analytical approach can be used to build correct solutions and provide optimal design [1–5]. Usually, they need fewer processing time. Thus, it is unreasonable to underestimate analytical approaches.

The method of Green's functions is appropriate to use in electromagnetic problems with boundaries aligned with the coordinate surfaces of the adopted coordinate system. Being analytical, the method of Green's functions allows the physical interpretation of simulated results.

The steepest descent method is beneficial for far field calculations in radiation and diffraction problems. There is no need to waste processor time for complicated procedures of numerical integration over the spectral variables. The Green's function for the structure under investigation stores internally all structural properties and does not require field matching at media boundaries. Fields are not evaluated over the entire region of definition since calculations are performed only within the limited subregion of interest.

S. Daylis
National Instruments – AWR Group, El Segundo, USA

S. Shabunin (✉)
Ural Federal University, Yekaterinburg, Russia
e-mail: shab313@ya.ru

In this paper, the spectral domain approach to extract Green's functions for multilayered cylindrical magneto-dielectric structure is presented and suggested for solution of problems of electromagnetic radiation, propagation, and scattering.

13.2 Spectral Domain Full-Wave Approach for Multilayer Cylindrical Structures

The spectral domain full-wave approach suggested in [6] uses the rigorous solution of the Maxwell equations for radially non-uniform structures with arbitrary electrodynamic properties and any distribution of impressed electrical and magnetic currents (Fig. 13.1). This method with a matrix approach is suitable for MATLAB. It is easy to simulate layered structures, impedance surfaces, and coating layers including metamaterials. In short, the fundamentals of our approach are the following.

Maxwell's equations in any homogeneous layer for a given distribution of impressed electrical and magnetic currents give rise to inhomogeneous equations for longitudinal (z-oriented) components of electric and magnetic field. Expanding E_z and H_z in the cylindrical region into Fourier integral series representation, one obtains inhomogeneous Bessel equations. The vector field components can be represented as integral transforms of respective spectral components:

$$\dot{E}_i(r, \varphi, z) = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} e^{-jm\varphi} \int_{-\infty}^{\infty} \dot{E}_{imh}(r)\, e^{-jhz} dh\ , \tag{13.1}$$

$$\dot{H}_i(r, \varphi, z) = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} e^{-jm\varphi} \int_{-\infty}^{\infty} \dot{H}_{imh}(r)\, e^{-jhz} dh\ . \tag{13.2}$$

In each layer we can express the E and H components with reference to the axis z in terms of the longitudinal components from Eqs. (13.1) and (13.2) as follows:

Fig. 13.1 Multilayer cylindrical structure
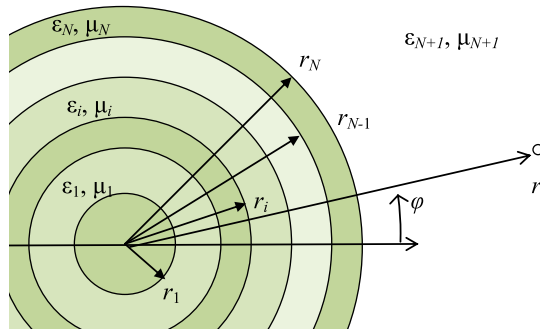
$$\dot{E}_{\varphi mh} = -\frac{mh}{\gamma_i^2}\frac{1}{r}\dot{E}_{zmh} + \frac{j\omega\mu_i}{\gamma_i^2}\frac{d\dot{H}_{zmh}}{dr}, \quad \dot{H}_{\varphi mh} = -\frac{j\omega\varepsilon_i}{\gamma_i^2}\frac{d\dot{E}_{zmh}}{dr} - \frac{mh}{\gamma_i^2}\frac{1}{r}\dot{H}_{zmh},$$

$$\dot{E}_{rmh} = -\frac{jh}{\gamma_i^2}\frac{d\dot{E}_{zmh}}{dr} - \frac{\omega\mu_i}{\gamma_i^2}\frac{m}{r}\dot{H}_{zmh}, \quad \dot{H}_{rmh} = \frac{\omega\varepsilon_i}{\gamma_i^2}\frac{m}{r}\dot{E}_{zmh} - \frac{mh}{\gamma_i^2}\frac{d\dot{H}_{zmh}}{dr}.$$

For a radially layered structure, the model with equivalent lines is used. Modal voltages in the equivalent E-line are associated with the spectral field components E_{zmh} and modal currents in the equivalent H-line are associated with the spectral field components H_{zmh}. The modal currents in the equivalent E-line are associated with the spectral field components $H_{\varphi mh}$ and modal voltages in equivalent H-line are associated with spectral field components $E_{\varphi mh}$ Eq. (13.3). In the sequel, indices are omitted.

$$V_{mh}^E = \dot{E}_{zmh}^e, \ I_{mh}^H = \dot{H}_{zmh}^h, \ I_{mh}^E = -k_0 r\dot{H}_{\varphi mh}^e, \ V_{mh}^H = k_0 r\dot{E}_{\varphi mh}^h \qquad (13.3)$$

Expressions of the spectral domain longitudinal electric and magnetic field components will be given later.

Wave propagation in each layer can be described by the 4-port transmission matrix $[C]^i$ (Fig. 13.2).

There is no coupling between the E- and H-lines inside the homogeneous medium of each layer and the respective transmission matrix is relatively simple

$$[C_i] = \begin{bmatrix} [C_{iE}] & [0] \\ [0] & [C_{iH}] \end{bmatrix}, \qquad (13.4)$$

where $[0]$ is a null matrix of order 2. The two-port layer network of the equivalent radial E-line is described by transmission matrix $[C_{iE}]$ as follows:
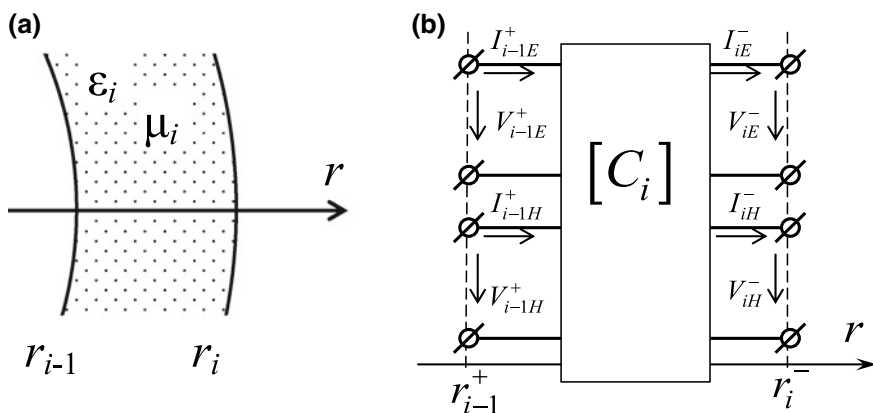
Fig. 13.2   A magneto-dielectric layer (**a**) and its transmission matrix (**b**)

$$\begin{bmatrix} V_{i-1E}^+ \\ I_{i-1E}^+ \end{bmatrix} = [C_{iE}] \begin{bmatrix} V_{iE}^- \\ I_{iE}^- \end{bmatrix}, \tag{13.5}$$

and another two-port network for the equivalent radial H-line is described by the transmission matrix $[C_{iH}]$

$$\begin{bmatrix} V_{i-1H}^+ \\ I_{i-1H}^+ \end{bmatrix} = [C_{iH}] \begin{bmatrix} V_{iH}^- \\ I_{iH}^- \end{bmatrix}, \tag{13.6}$$

where V_{i-1}^+, I_{i-1}^+, V_i^-, and I_i^- are the modal voltages and currents at the boundaries of the ith layer. The distribution of fields inside the layer is defined by radial variation of the model currents and voltages. Green's function of inhomogeneous differential equations for these currents and voltages can be constructed from two independent solutions of homogeneous equations with specified boundary conditions at radial coordinates of layer's boundaries, namely, $\vec{V}_{E,H}$, $\overleftarrow{V}_{E,H}$, and $\vec{I}_{E,H}$, $\overleftarrow{I}_{E,H}$. Here, the cap arrow direction points to the boundary with specified boundary conditions. A similar technique is adopted in [7]. Equivalent circuits are defined via directional currents, voltages, admittances, and impedances.

The transmission matrices of the ith layer for boundary conditions specified at the radial coordinate $r < r_{i-1}$ are

$$\left[\overleftarrow{C}_{iE}\right] = \begin{bmatrix} C_{2m}(\gamma_i r_{i-1}, \gamma_i r_i) & -j\frac{Z_0 \beta_i}{\varepsilon_i'} \frac{1}{k_0 r_i} S_{2m}(\gamma_i r_{i-1}, \gamma_i r_i) \\ -j\frac{\varepsilon_i' k_0}{Z_0 \beta_i} r_{i-1} S_{1m}(\gamma_i r_{i-1}, \gamma_i r_i) & \frac{r_{i-1}}{r_i} C_{1m}(\gamma_i r_{i-1}, \gamma_i r_i) \end{bmatrix}, \tag{13.7}$$

$$\left[\overleftarrow{C}_{iH}\right] = \begin{bmatrix} \frac{r_{i-1}}{r_i} C_{1m}(\gamma_i r_{i-1}, \gamma_i r_i) & -j\frac{\mu_i'}{Y_0 \beta_i} k_0 r_{i-1} S_{1m}(\gamma_i r_{i-1}, \gamma_i r_i) \\ -j\frac{Y_0 \beta_i}{\mu_i'} \frac{1}{k_0 r_i} S_{2m}(\gamma_i r_{i-1}, \gamma_i r_i) & C_{2m}(\gamma_i r_{i-1}, \gamma_i r_i) \end{bmatrix}. \tag{13.8}$$

For boundary conditions specified at the radial coordinate $r > r_i$ the transmission matrices are:

$$\left[\vec{C}_{iE}\right] = \begin{bmatrix} C_{2m}(\gamma_i r_{i-1}, \gamma_i r_i) & j\frac{Z_0 \beta_i}{\varepsilon_i'} \frac{1}{k_0 r_i} S_{2m}(\gamma_i r_{i-1}, \gamma_i r_i) \\ j\frac{\varepsilon_i' k_0}{Z_0 \beta_i} r_{i-1} S_{1m}(\gamma_i r_{i-1}, \gamma_i r_i) & \frac{r_{i-1}}{r_i} C_{1m}(\gamma_i r_{i-1}, \gamma_i r_i) \end{bmatrix}, \tag{13.9}$$

$$\left[\vec{C}_{iH}\right] = \begin{bmatrix} \frac{r_{i-1}}{r_i} C_{1m}(\gamma_i r_{i-1}, \gamma_i r_i) & j\frac{\mu_i'}{Y_0 \beta_i} k_0 r_{i-1} S_{1m}(\gamma_i r_{i-1}, \gamma_i r_i) \\ j\frac{Y_0 \beta_i}{\mu_i'} \frac{1}{k_0 r_i} S_{2m}(\gamma_i r_{i-1}, \gamma_i r_i) & C_{2m}(\gamma_i r_{i-1}, \gamma_i r_i) \end{bmatrix}. \tag{13.10}$$

In Eqs. (13.7)–(13.10) the fundamental Cauchy functions are

$$C_{1m}(x, y) = \frac{y\pi}{2} \left[J_m(y) N_m'(x) - N_m(y) J_m'(x) \right], \tag{13.11}$$

$$C_{2m}(x, y) = \frac{y\pi}{2} \left[N_m'(y) J_m(x) - J_m'(y) N_m(x) \right], \tag{13.12}$$

$$S_{1m}(x, y) = \frac{y\pi}{2}\left[N'_m(y)J'_m(x) - J'_m(y)N'_m(x)\right], \tag{13.13}$$

$$S_{2m}(x, y) = \frac{y\pi}{2}\left[N_m(y)J_m(x) - J_m(y)N_m(x)\right], \tag{13.14}$$

where J_m, N_m are Bessel and Neumann functions, respectively. In special cases including but not limited to diffraction problems, replacing of the Neumann functions by the Hankel functions may serve better with Eqs. (13.11)–(13.14) respectively updated.

In Eqs. (13.7)–(13.10) $\gamma_i = \sqrt{\varepsilon_i \mu_i - h^2}$, $\beta_i = \gamma_i / k_0$, $Z_0 = Y_0^{-1} = 120\pi$ Ohm, and k_0 is the wavenumber of the free space.

Unlike flat and spherical structures, the boundary conditions at the layer's boundaries in the cylindrical structures cause energy exchange between E and H waves similar to coupling that occurs in the coupled lines. Taking into account the coupling of E and H-lines, the four-port boundary network with transmission matrix $[\Gamma_i]$ of order 4 is added (Fig. 13.3) in the following form:

$$\begin{bmatrix} V^-_{iE} \\ I^-_{iE} \\ V^-_{iH} \\ I^-_{iH} \end{bmatrix} = [\Gamma_i] \begin{bmatrix} V^+_{iE} \\ I^+_{iE} \\ V^+_{iH} \\ I^+_{iH} \end{bmatrix}, \quad [\Gamma_i] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -N_i \\ N_i & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \text{ where } N_i = \frac{mh}{k_0}\left(\frac{1}{\gamma_i^2} - \frac{1}{\gamma_{i+1}^2}\right).$$

The outer and inner space boundaries are modelled with the equivalent terminal impedances and admittances (Fig. 13.4).

For example, for the free outer space (Fig. 13.4) the terminal impedance and admittance are calculated as follows:

$$\vec{Z}^+_{NH} = jZ_0 \frac{\mu'_{N+1}}{\beta_{N+1}} k_0 r_N \frac{H^{(2)'}_m(\gamma_{N+1}r_N)}{H^{(2)}_m(\gamma_{N+1}r_N)},$$
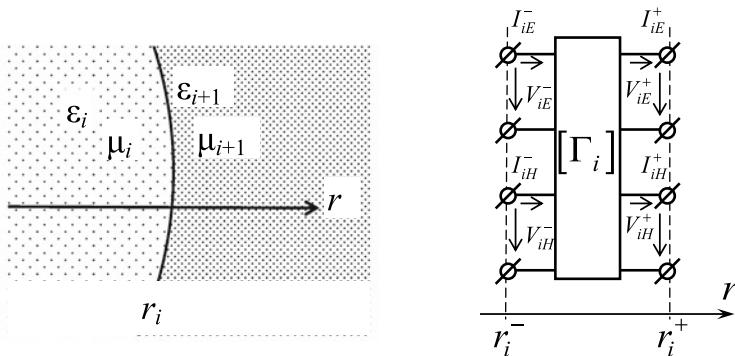
Fig. 13.3 A boundary between layers and its transmission matrix
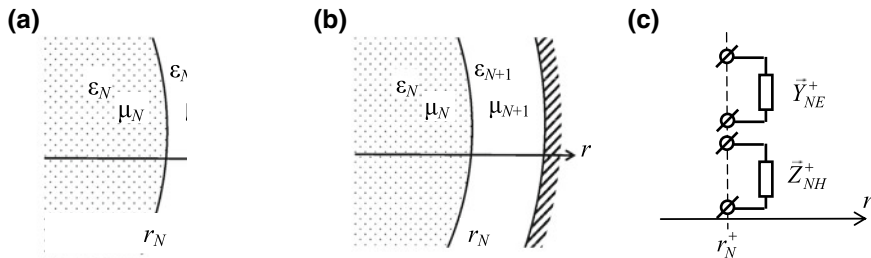
Fig. 13.4 Terminal loads (**c**) for free outer space (**a**) and conducting surface (**b**)

$$\vec{Y}_{NE}^{+} = jY_0 \frac{\varepsilon'_{N+1}}{\beta_{N+1}} k_0 r_N \frac{H_m^{(2)'}(\gamma_{N+1} r_N)}{H_m^{(2)}(\gamma_{N+1} r_N)},$$

where $H_m^{(2)}(\gamma_{N+1} r_N)$ and $H_m^{(2)'}(\gamma_{N+1} r_N)$ are the Hankel function and its derivative.

If the outer cylindrical surface is an ideal conductor (as in the case of cylindrical waveguide) then the terminal impedance and admittance are

$$\vec{Y}_{NE}^{+} = jY_0 \frac{\varepsilon'_{N+1}}{\beta_{N+1}} k_0 r_N \frac{C_{1m}(\gamma_{N+1} r_N, \gamma_{N+1} r_{N+1})}{S_{2m}(\gamma_{N+1} r_N, \gamma_{N+1} r_{N+1})},$$

$$\vec{Z}_{NH}^{+} = jZ_0 \frac{\mu'_{N+1}}{\beta_{N+1}} k_0 r_N \frac{S_{1m}(\gamma_{N+1} r_N, \gamma_{N+1} r_{N+1})}{C_{2m}(\gamma_{N+1} r_N, \gamma_{N+1} r_{N+1})}.$$

The terminal admittances for the case of impedance cylindrical surface are

$$\vec{Y}_{NE}^{+} = k_0 r_N / Z_S, \quad \vec{Z}_{NH}^{+} = k_0 r_N Z_S.$$

For lossy conductors the surface impedance is $Z_S = \sqrt{\frac{\omega \mu_0}{2\sigma}}(1 + j)$, where σ is material conductivity.

In most cases, the innermost boundary of a cylindrical layered structure is a conducting surface. Assuming the single magneto-dielectric layer enclosing the ideal conducting surface $r = r_0$ (Fig. 13.5), the terminal impedance and admittance can be easily calculated.

For an ideal conducting surface covered with a dielectric layer the equivalent loads are

$$\overleftarrow{Y}_{1E} = jY_0 \frac{\varepsilon'_1}{\beta_1} k_0 r_1 \frac{C_{1m}(\gamma_1 r_1, \gamma_1 r_0)}{S_{2m}(\gamma_1 r_1, \gamma_1 r_0)}, \quad \overleftarrow{Z}_{1H} = jZ_0 \frac{\mu'_1}{\beta_1} k_0 r_1 \frac{S_{1m}(\gamma_1 r_1, \gamma_1 r_0)}{C_{2m}(\gamma_1 r_1, \gamma_1 r_0)}.$$

If there is a dielectric cylinder at the center of the structure, the loads for electric and magnetic lines are
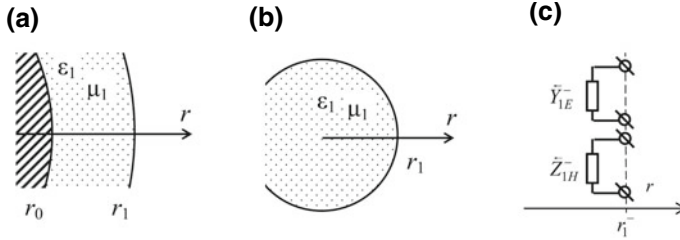
Fig. 13.5 Terminating loads (**c**) for a conducting cylinder with a coat (**a**) and dielectric cylinder (**b**)

$$\overleftarrow{Y}_{1E}^{-} = -jY_0 \frac{\varepsilon_1'}{\beta_1} k_0 r_1 \frac{J_m'(\gamma_1 r_1)}{J_m(\gamma_1 r_1)}, \quad \overleftarrow{Z}_{1H}^{-} = -jZ_0 \frac{\mu_1'}{\beta_1} k_0 r_1 \frac{J_m'(\gamma_1 r_1)}{J_m(\gamma_1 r_1)}$$

For the impedance surface, the boundary loads are

$$\overleftarrow{Y}_{1E}^{-} = \overleftarrow{Y}_{1E}^{+} = -k_0 r_0 / Z_S, \quad \overleftarrow{Z}_{1H}^{-} = \overleftarrow{Z}_{1H}^{+} = -k_0 r_1 Z_S$$

Thus, all possible variations of the internal and external boundaries are taken into account.

To handle excitation, radiation, propagation, and diffraction problems in multi-layered structures, we use translation matrices derived as products of transmission matrices of layers and boundaries. We use various configurations to construct translation matrices. In the figures below, the dashed line indicates the radial cross-sections between which translation is performed.

To translate fields from r_1 to r_i , we need to define matrices $[L_i]$ or $\left[\tilde{L}_i\right]$ depending on position of r_1 with respect to r_i (Figs. 13.6 and 13.7). We used matrices $[l_i] = \left[\overleftarrow{C}_i\right][\Gamma_i]$ to calculate $[L_i]$ and $\left[\tilde{L}_i\right]$.

If the terminal loads at the right-hand side of the structure are specified (at the external boundary), then we need to translate V and I from r_N to r_i ; in this case, we define translation matrices $[R_i]$ or $\left[\tilde{R}_i\right]$ in accordance with the position of r_N with respect to r_i (Figs. 13.8 and 13.9). We used matrices $[r_i] = \left[\overrightarrow{C}_i\right][\Gamma_i]$ to calculate $[R_i]$ and $\left[\tilde{R}_i\right]$.

The inverse matrices $[F_i] = [L_i]^{-1}$, $\left[\tilde{F}_i\right] = \left[\tilde{L}_i\right]^{-1}$, $[B_i] = [R_i]^{-1}$, and $\left[\tilde{B}_i\right] = \left[\tilde{R}_i\right]^{-1}$ allow one to translate modal currents and voltages in the inverted order.

The suggested technique allows one to design fast algorithms for calculation of all field components because all the interlayer boundary conditions are included at the stage of analytical formulations.

Fig. 13.6 Translation of modal currents and voltages between r_1^- and r_i^+
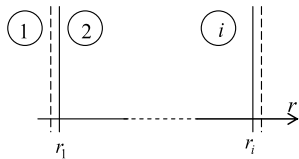
$$
\begin{bmatrix} \bar{V}_{1E}^- \\ \bar{I}_{1E}^- \\ \bar{V}_{1H}^- \\ \bar{I}_{1H}^- \end{bmatrix} = [L_i] \begin{bmatrix} \bar{V}_{iE}^+ \\ \bar{I}_{iE}^+ \\ \bar{V}_{iH}^+ \\ \bar{I}_{iH}^+ \end{bmatrix}
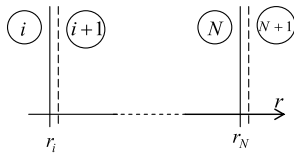$$

$$
[L_i] = \begin{cases} [\Gamma_1]\prod_{j=2}^{i}[l_j], & N \geq 2,\ 2 \leq i \leq N, \\ [\Gamma_1], & N = 1. \end{cases}
$$

Fig. 13.7 Translation of modal currents and voltages between r_1^- and r_i^-
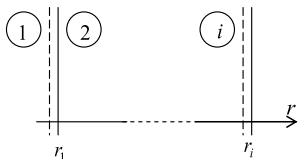
$$
\begin{bmatrix} \bar{V}_{1E}^- \\ \bar{I}_{1E}^- \\ \bar{V}_{1H}^- \\ \bar{I}_{1H}^- \end{bmatrix} = [\tilde{L}_i] \begin{bmatrix} \bar{V}_{iE}^- \\ \bar{I}_{iE}^- \\ \bar{V}_{iH}^- \\ \bar{I}_{iH}^- \end{bmatrix}
$$

$$
[\tilde{L}_i] = \begin{cases} [\Gamma_1]\prod_{j=2}^{i-1}[l_j], & N \geq 3,\ 3 \leq i \leq N, \\ [\Gamma_1][\bar{C}_2], & N = 2;\quad [1], N = 1. \end{cases}
$$

Formulations for input admittances and impedances at various radial cross-sections are:

$$
\overleftarrow{Y}_{NE}^+ = \frac{F_{21}^N + F_{22}^N Y_{1E} + \left(F_{24}^N + F_{23}^N Z_{1N}\right)\overleftarrow{Y}_{1HE}^C}{F_{11}^N + F_{12}^N Y_{1E} + \left(F_{14}^N + F_{13}^N Z_{1N}\right)\overleftarrow{Y}_{1HE}^C}, \tag{13.15}
$$

$$
\overleftarrow{Y}_{1HE}^C = -\frac{F_{31}^N + F_{32}^N Y_{1E} + \left(F_{41}^N + F_{42}^N Y_{1E}\right)Z_{NH}}{F_{34}^N + F_{33}^N Z_{1H} + \left(F_{44}^N + F_{43}^N Z_{1H}\right)Z_{NH}}, \tag{13.16}
$$
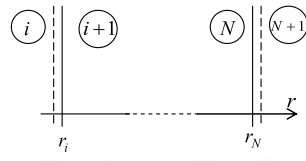
Fig. 13.8 Translation of modal currents and voltages between r_i^+ and r_N^+

$$
\begin{bmatrix} \vec{V}_{iE}^+ \\ \vec{I}_{iE}^+ \\ \vec{V}_{iH}^+ \\ \vec{I}_{iH}^+ \end{bmatrix} = [R_i] \begin{bmatrix} \vec{V}_{NE}^+ \\ \vec{I}_{NE}^+ \\ \vec{V}_{NH}^+ \\ \vec{I}_{NH}^+ \end{bmatrix}
$$

$$
[R_i] = \begin{cases} \displaystyle\prod_{j=i+1}^{N} [r_j], & N \geq 2,\ 2 \leq i \leq N, \\ [1], & i = N. \end{cases}
$$

Fig. 13.9 Translation of modal currents and voltages between r_1^- and r_N^+

$$
\begin{bmatrix} \vec{V}_{iE}^- \\ \vec{I}_{iE}^- \\ \vec{V}_{iH}^- \\ \vec{I}_{iH}^- \end{bmatrix} = [\tilde{R}_i] \begin{bmatrix} \vec{V}_{NE}^+ \\ \vec{I}_{NE}^+ \\ \vec{V}_{NH}^+ \\ \vec{I}_{NH}^+ \end{bmatrix}
$$

$$
[\tilde{R}_i] = \begin{cases} \displaystyle [r_i]\prod_{j=i+1}^{N} [r_j], & N \geq 2,\ 1 \leq i \leq N-1, \\ [\Gamma_N], & i = N. \end{cases}
$$

$$
\overleftarrow{Z}_{NH}^C = \frac{\left(F_{31}^N + F_{32}^N Y_{1E}\right)\overleftarrow{Z}_{1EH}^C + F_{33}^N Z_{1H} + F_{34}^N}{\left(F_{41}^N + F_{42}^N Y_{1E}\right)\overleftarrow{Z}_{1EH}^C + F_{43}^N Z_{1H} + F_{44}^N}, \tag{13.17}
$$

$$
\overleftarrow{Z}_{1EH}^C = -\frac{F_{23}^N Z_{1H} + F_{24}^N + \left(F_{13}^N Z_{1H} + F_{14}^N\right)Y_{NE}}{F_{21}^N + F_{22}^N Y_{1E} + \left(F_{11}^N + F_{12}^N Y_{1E}\right)Y_{NE}}, \tag{13.18}
$$

$$
\vec{Y}_{1E}^- = \frac{\tilde{R}_{21}^1 + \tilde{R}_{22}^1 Y_{NE} + \left(\tilde{R}_{23}^1 Z_{NH} + \tilde{R}_{24}^1\right)\vec{Y}_{NHE}^C}{\tilde{R}_{11}^1 + \tilde{R}_{22}^1 Y_{NE} + \left(\tilde{R}_{13}^1 Z_{NH} + \tilde{R}_{14}^1\right)\vec{Y}_{NHE}^C}, \tag{13.19}
$$

Fig. 13.10 On the definition of transformation ratios
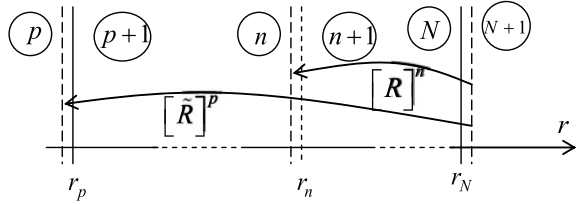
$$\vec{Y}_{NHE}^{C} = -\frac{\tilde{R}_{31}^{1} + \tilde{R}_{32}^{1}Y_{NE} + \left(\tilde{R}_{41}^{1} + \tilde{R}_{42}^{1}Y_{NE}\right)Z_{1H}}{\tilde{R}_{33}^{1}Z_{NH} + \tilde{R}_{34}^{1} + \left(\tilde{R}_{43}^{1}Z_{NH} + \tilde{R}_{44}^{1}\right)Z_{1H}}, \tag{13.20}$$

$$\vec{Z}_{1H}^{-} = \frac{\left(\tilde{R}_{31}^{1} + \tilde{R}_{32}^{1}Y_{NE}\right)\vec{Z}_{NEH}^{C} + \tilde{R}_{33}^{1}Z_{NH} + \tilde{R}_{34}^{1}}{\left(\tilde{R}_{41}^{1} + \tilde{R}_{42}^{1}Y_{NE}\right)\vec{Z}_{NEH}^{C} + \tilde{R}_{43}^{1}Z_{NH} + \tilde{R}_{44}^{1}}, \tag{13.21}$$

$$\vec{Z}_{NEH}^{C} = -\frac{\tilde{R}_{23}^{1}Z_{NH} + \tilde{R}_{24}^{1} + \left(\tilde{R}_{13}^{1}Z_{NH} + \tilde{R}_{14}^{1}\right)Y_{1E}}{\tilde{R}_{21}^{1} + \tilde{R}_{22}^{1}Y_{NE} + \left(\tilde{R}_{11}^{1} + \tilde{R}_{12}^{1}Y_{NE}\right)Y_{1E}}, \tag{13.22}$$

where impedances $\overleftarrow{Z}_{1EH}^{C}, \vec{Z}_{NEH}^{C}$ and admittances $\overleftarrow{Z}_{1HE}^{C}, \vec{Y}_{NHE}^{C}$ are the E- and H-coupling impedances and admittances at the cross-sectional positions r_1 and, r_N respectively.

Formulation of the input impedances and admittances from Eqs. (13.15), (13.17), (13.19), and (13.21) for arbitrary radial cross-sectional positions allows one to compute fields in the selected layer avoiding field computations over the entire space. This is also beneficial for speeding up electrodynamic problems.

To provide normalization of modal current and voltages in different layers with respect to the reference value at the selected boundary (Fig. 13.10), we use voltage transformation ratio $T_{n,p}^{V} = \vec{V}_{E}(r_n)\big/ \vec{V}_{E}(r_p)$ and current transformation ratio $T_{n,p}^{I} = \vec{I}_{H}(r_n)\big/ \vec{I}_{H}(r_p)$.

Formulations for the transformation ratios are the following:

$$T_{n,p}^{I} = \frac{\tilde{R}_{44}^{n} + \tilde{R}_{43}^{n}Z_{NH} + \left(\tilde{R}_{42}^{n}Y_{NE} + \tilde{R}_{41}^{n}\right)\vec{Z}_{NEH}^{C}}{\tilde{R}_{44}^{p} + \tilde{R}_{43}^{p}Z_{NH} + \left(\tilde{R}_{42}^{p}Y_{NE} + \tilde{R}_{41}^{p}\right)\vec{Z}_{NEH}^{C}},$$

$$T_{n,p}^{V} = \frac{\tilde{R}_{11}^{n} + \tilde{R}_{12}^{n}Z_{NH} + \left(\tilde{R}_{13}^{n}Y_{NE} + \tilde{R}_{14}^{n}\right)\vec{Y}_{NHE}^{C}}{\tilde{R}_{11}^{p} + \tilde{R}_{12}^{p}Z_{NH} + \left(\tilde{R}_{13}^{p}Y_{NE} + \tilde{R}_{14}^{p}\right)\vec{Y}_{NHE}^{C}}.$$

13.3 Calculation of Longitudinal Spectral Components of Electrical and Magnetic Field

Previously, we presented formulations for the transversal field components based on the known longitudinal components. Below, we discuss how to evaluate these longitudinal components. Let us consider a source point located in a layer p and an observation point located in a layer n.

Spectral densities of longitudinal components of the electrical and magnetic fields can be defined as follows:

$$
E_{zmh}(r) = -\frac{1}{2\pi} \int\limits_{V'} \left[\Phi_1^E\left(r, r', \varphi', z'; m, h\right) + \Phi_2^E\left(r, r', \varphi', z'; m, h\right) \right] e^{jm\varphi'} e^{jhz'} dv',
$$

(13.23)

$$
H_{zmh}(r) = -\frac{1}{2\pi} \int\limits_{V'} \left[\Phi_1^M\left(r, r', \varphi', z'; m, h\right) + \Phi_2^M\left(r, r', \varphi', z'; m, h\right) \right] e^{jm\varphi'} e^{jhz'} dv'.
$$

(13.24)

Integrations in Eqs. (13.23) and (13.24) are carried out over area V' of source point $\mathbf{r}'$ with coordinates r', φ', z'.

The concrete form of the functions Φ_1^E, Φ_2^E, Φ_1^M, Φ_2^M depends on the structure under investigation and orientation and type of impressed currents. The functions Φ_1^E and Φ_1^M belong to direct (either E mode to E component or H mode to H component) parts of the Green's functions while the functions Φ_2^E and Φ_2^M belong to coupling (either E mode to H component or H mode to E component) parts of the Green's functions.

In Eqs. (13.23) and (13.24) we have

$$
\Phi_1^E = \omega\varepsilon_p \frac{k_0}{k_p^2 - h^2} \frac{T_{n_1p}^V}{\overset{\leftrightarrow}{Y}_{pE}} V_E(r, r_n) \left[V_E\left(r', r_p\right) R_p^{(1)} + \frac{d V_E\left(r', r_p\right)}{dr'} R_p^{(2)} \right], \quad (13.25)
$$

$$
\Phi_2^E = \omega\mu_p \frac{k_0}{k_p^2 - h^2} \frac{T_{n_1p}^I}{\overset{\leftrightarrow}{Z}_{pH}} Z_{nEH}^C V_{EH}(r, r_n) \left[I_{EH}\left(r', r_p\right) R_p^{(3)} + \frac{d I_{EH}\left(r', r_p\right)}{dr'} R_p^{(4)} \right],
$$

(13.26)

$$
\Phi_1^M = \omega\mu_p \frac{k_0}{k_p^2 - h^2} \frac{T_{n_1p}^I}{\overset{\leftrightarrow}{Z}_{pH}} I_H(r, r_n) \left[I_H\left(r', r_p\right) R_p^{(3)} + \frac{d I_H\left(r', r_p\right)}{dr'} R_p^{(4)} \right], \quad (13.27)
$$

$$
\Phi_2^M = \omega\varepsilon_p \frac{k_0}{k_p^2 - h^2} \frac{T_{n_1p}^V}{\overset{\leftrightarrow}{Y}_{pE}} Y_{nHE}^C I_{HE}(r, r_n) \left[V_{HE}\left(r', r_p\right) R_p^{(1)} + \frac{d V_{HE}\left(r', r_p\right)}{dr'} R_p^{(2)} \right]
$$

(13.28)

The current functions $R_p^{(i)}$ account for all sorts of impressed electric and magnetic current components and are defined as follows:

$$R_p^{(1)} = \frac{k_p^2 - h^2}{j\omega\varepsilon_p} J_z(r', \varphi', z') + \frac{jm}{r'}\left(\frac{h}{\omega\varepsilon_p} J_\varphi(r', \varphi', z') - M_r(r', \varphi', z')\right),$$

$$R_p^{(2)} = \frac{h}{\omega\varepsilon_p} J_r(r', \varphi', z') + M_r(r', \varphi', z'),$$

$$R_p^{(3)} = \frac{K_p^2 - h^2}{j\omega\mu_p} M_z(r', \varphi', z') + \frac{jm}{r'}\left(\frac{h}{\omega\mu_p} M_\varphi(r', \varphi', z') + J_r(r', \varphi', z')\right),$$

$$R_p^{(4)} = \frac{h}{\omega\mu_p} M_r(r', \varphi', z') - J_\varphi(r', \varphi', z').$$

The structure of expressions for normalized currents and voltages used in Eqs. (13.25)–(13.28) is dictated by relative positions of the source and observe points.

Let us consider two layers numbered p and q, where $p < q$ (Fig. 13.11). In expressions below, parameters of the layer containing the observation point have index n and parameters of the layer containing source point have index i.

The following notations are adopted:

$$n = \begin{cases} p, & r_{p-1} < r \le r_p, \\ q, & r_{q-1} < r \le r_q; \end{cases} \quad i = \begin{cases} p, & r_{p-1} < r' \le r_p, \\ q, & r_{q-1} < r' \le r_q; \end{cases}$$

$$r_n = \begin{cases} r_p, & n = p, \\ r_{q-1}, & n = q; \end{cases} \quad r_i = \begin{cases} r_p, & i = p, \\ r_{q-1}, & i = q. \end{cases}$$

For the *E*-mode voltage and *H*-mode current at the layer with source point

$$V_E(r', r_i) = \overleftarrow{V}_E(r', r_p), \quad I_H(r', r_i) = \overleftarrow{I}_H(r', r_p), \quad r' < r, \text{ Fig. 6, } a, c, e, g,$$
$$V_E(r', r_i) = \overrightarrow{V}_E(r', r_{q-1}), \quad I_H(r', r_i) = \overrightarrow{I}_H(r', r_{q-1}), \quad r' > r, \text{ Fig. 6, } b, d, f, h.$$

For the E-mode voltage and H-mode current at the layer with observe point

$$V_E(r, r_n) = \overrightarrow{V}_E(r, r_{q-1}), \quad I_H(r, r_n) = \overrightarrow{I}_H(r, r_{q-1}), \quad r' < r, \quad \text{Fig. 6, } a, c, e, g,$$
$$V_E(r, r_n) = \overleftarrow{V}_E(r, r_p), \quad I_H(r, r_n) = \overleftarrow{I}_H(r, r_p), \quad r' > r, \quad \text{Fig. 6, } b, d, f, h.$$

Modal currents and voltages induced on the *E*-line by *H*-line (lower index *EH*) and modal currents and voltages induced on the *H*-line by *E*-line (lower index *HE*) are specified as
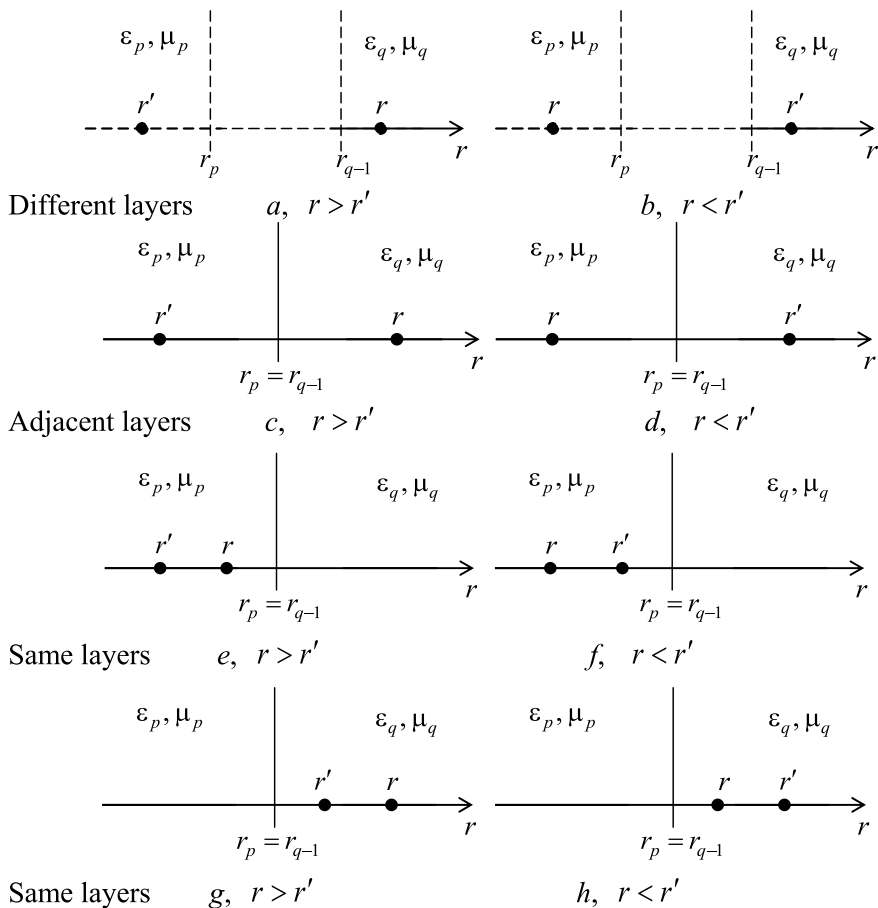
Fig. 13.11 Options of relative locations of source and observation points

$$I_{EH}(r', r_i) = \begin{cases} \overleftarrow{I}_H(r', r_p), & r' < r_{q-1}, \quad \text{Fig. 11}, a, c, e, f, \\ \overrightarrow{I}_H(r', r_{q-1}), & r' > r_{q-1}, \quad \text{Fig. 11}, b, d, g, h; \end{cases}$$

$$V_{EH}(r, r_n) = \begin{cases} \overrightarrow{V}_E(r, r_{q-1}), & r > r_{q-1}, \quad \text{Fig. 6}, a, c, e, f, \\ \overleftarrow{V}_E(r, r_{q-1}), & r < r_{q-1}, \quad \text{Fig. 6}, b, d, g, h; \end{cases}$$

$$V_{HE}(r', r_i) = \begin{cases} \overleftarrow{V}_E(r', r_p), & r' < r_{q-1}, \quad \text{Fig. 11}, a, c, e, f, \\ \overrightarrow{V}_E(r', r_{q-1}), & r' > r_{q-1}, \quad \text{Fig. 11}, b, d, g, h; \end{cases}$$

$$I_{HE}(r, r_n) = \begin{cases} \overrightarrow{I}_H(r, r_{q-1}), & r > r_{q-1}, \quad \text{Fig. 11}, a, c, e, f, \\ \overleftarrow{I}_H(r, r_{q-1}), & r < r_{q-1}, \quad \text{Fig. 11}, b, d, g, h. \end{cases}$$

In the expressions above normalized currents and voltages are defined using the fundamental Cauchy functions (see Eqs. (13.11)–(13.14)):

$$\overleftarrow{V}_E(x, r_p) = C_{2m}(\gamma_p x, \gamma_p r_p) - jZ_0 \frac{\beta_p}{\varepsilon'_p} \frac{1}{k_0 r_p} \; S_{2m}(\gamma_p x, \gamma_p r_p)\overleftarrow{Y}_{pE},$$

$$\overrightarrow{V}_E(y, r_{q-1}) = C_{2m}(\gamma_q y, \gamma_q r_{q-1}) + jZ_0 \frac{\beta_q}{\varepsilon'_q} \frac{1}{k_0 r_{q-1}} \; S_{2m}(\gamma_q y, \gamma_q r_{q-1})\overrightarrow{Y}_{q-1\,E},$$

$$\overleftarrow{I}_H(x, r_p) = C_{2m}(\gamma_p x, \gamma_p r_p) - jY_0 \frac{\beta_p}{\varepsilon'_p} \frac{1}{k_0 r_p} \; S_{2m}(\gamma_p x, \gamma_p r_p)\overleftarrow{Z}_{pH},$$

$$\overrightarrow{I}_H(y, r_{q-1}) = C_{2m}(\gamma_q y, \gamma_q r_{q-1}) + jY_0 \frac{\beta_q}{\varepsilon'_p} \frac{1}{k_0 r_{q-1}} \; S_{2m}(\gamma_q y, \gamma_q r_{q-1})\overrightarrow{Z}_{q-1\,H}.$$

Input impedance and admittance of the terminated E- and H-lines used in Eqs. (13.25)–(13.28) are directed to the left or right side of the reference points r_p and r_{q-1} for various positions of the point r with respect to the reference point:

$$\overleftarrow{Z}_{pH}, \overleftarrow{Y}_{pE} = \begin{cases} \overleftarrow{Z}_{pH}^-, \overleftarrow{Y}_{pE}^- & r < r_p, \\ \overleftarrow{Z}_{pH}^+, \overleftarrow{Y}_{pE}^+ & r > r_p. \end{cases} \quad \overrightarrow{Z}_{q-1\,H}, \overrightarrow{Y}_{q-1\,E} = \begin{cases} \overrightarrow{Z}_{q-1\,H}^+, \overrightarrow{Y}_{q-1\,E}^+ & r > r_{q-1}, \\ \overrightarrow{Z}_{q-1\,H}^-, \overrightarrow{Y}_{q-1\,E}^- & r < r_{q-1}. \end{cases}$$

$$\overleftrightarrow{Y}_{pE} = \overleftarrow{Y}_{pE} + \overrightarrow{Y}_{pE}; \quad \overleftrightarrow{Z}_{pH} = \overleftarrow{Z}_{pH} + \overrightarrow{Z}_{pH}$$

Coupling (between modal lines) admittances and impedances

$$Y_{nHE}^C, \; Z_{nEH}^C = \begin{cases} \overrightarrow{Y}_{q-1\,HE}^C, \; \overrightarrow{Z}_{q-1\,EH}^C & r < r_{q-1}, \quad \text{Fig.11}, a, c, e, f, \\ \overleftarrow{Y}_{pHE}^C, \; \overleftarrow{Z}_{pEH}^C & r > r_{q-1}, \quad \text{Fig.11}, b, d, g, h. \end{cases}$$

Admittances and impedances are always defined at the reference boundary of the layer that contains an observation point.

13.4 Conclusion

The method based on the layered-structured Green's function for radiation, propagation, and scattering problems is described. All kinds of magneto-dielectrics including materials with negative index of refraction may be taken into account. The suggested method is very effective for different electrodynamic problems such as dipole, patch, or slot antenna radiation from a coated conducting cylinder, wave propagation in layered circular waveguides and coaxial lines, and electromagnetic wave scattering on layered cylinders.

References

1. C. Tokgöz, G. Dural, Closed-form Green's functions for cylindrically stratified media. IEEE Trans. Microw. Theory Tech. **48**(1), 40–49 (2000)
2. J.R. Wait, General solution for excitation by slotted aperture source in conducting cylinder with concentric layering. IEEE Trans. Microw. Theory Tech. **35**(3), 321–325 (1993)
3. Z. Xiang, Electromagnetic dyadic Green's function in cylindrically multi-layered media. IEEE Trans. Microw. Theory Tech. **44**(4), 614–621 (1996)
4. V.B. Erturk, Efficient computation of surface fields excited on a dielectric-coated circular cylinder. IEEE Trans. Antennas Propag. **48**(10), 1507–1515 (2000)
5. M. Thiel, Dyadic Green's function of multiplayer cylindrical closed and sector structures for waveguide, microstrip-antenna, and network analysis. IEEE Trans. Microw. Theory Tech. **50**(11), 2576–2579 (2002)
6. S. Dailis, S. Shabunin, Microstrip antennas on metallic cylinder covered with radially inhomogeneous magnetodielectric, in *Proceedings of 9th International Wroclaw Symposium on Electromagnetic Compatibility, Part 1*, pp. 271–277 (1988)
7. L. Felsen, N. Marcuvitz, *Radiation and Scattering of Waves* (Prentice-Hall Inc., Englewood Cliffs, 1973)

Chapter 14
Compact Topologies of Microstrip Ring Coupler

D. A. Letavin

14.1 Introduction

The modern stage of radioelectronic development is characterized by increased systems complexity, including an increase in the density of elements soldering. In this regard, questions of miniaturization are actual not only for basic functional elements of the scheme, but for matching circuits, including the microstrip ring coupler.

Today the microstrip ring coupler is widely used as a power divider and a combiner in balanced mixers, phase shifters, and others and has many different embodiments. For example, it can be realized on the basis of symmetrical equivalent circuits [1], artificial transmission lines [2], fractal structure lines [3], using quasi-lumped elements [4], distributed capacitor [5]. The widest range of the miniaturization techniques of microstrip couplers is observed in [6].

However, their implementation usually requires an additional calculation, which, incidentally, does not guarantee the reduction of coupler geometric dimensions or the improvement its frequency characteristics.

The main goal of this work is finding the optimal design of microstrip ring coupler based on maintaining the overall length of lines and allowing as much as possible minimization of the geometric dimensions of the device without a significant change in its frequency characteristics.

D. A. Letavin (✉)
Ural Federal University, Yekaterinburg, Russia
e-mail: d.a.letavin@urfu.ru

14.2   Coupler Design

As the main instrument of reducing the microstrip ring coupler area, it is proposed to use "snake" bends instead of straight quarter-sections. Such bends allow not only reducing the distance between the inputs, but also utilizing effectively the space inside it. Generally, the "snakes" form can be arbitrary, so the topologies quantity is almost endless. The most successful embodiments are given in this paper.

Figure 14.1 shows a comparison between the electrical lengths of a quarter-wave line and the corresponding line with changed form. The electrical and physical line lengths remained the same.

The microstrip rectangular ring topology presented in Fig. 14.2 was chosen as a basis for optimizing. The substrate material is FR-4 with a permittivity $\varepsilon = 4.4$, the dielectric loss tangent tg$\delta = 0.02$ and thickness h = 1 mm. This design, made to operate at a frequency F = 1 GHz, has an area equal to 3951.7 mm^2, determined by positions of maximum protruded points, i.e. by size of quarter-sections in this case.

Fig. 14.1 Electrical lengths of microstrip line segment (solid line) and microstrip with "snake" form (dotted line)
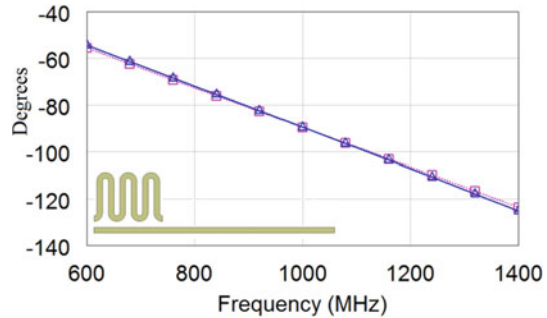
Fig. 14.2 The initial topology of microstrip ring coupler for optimization
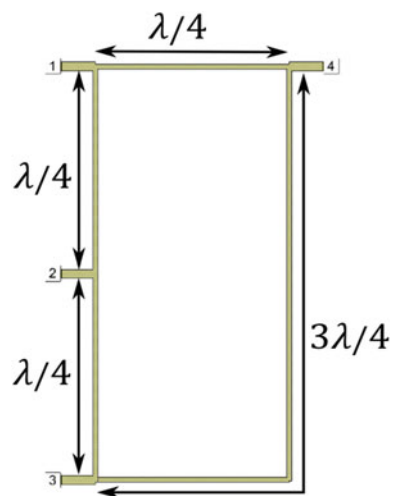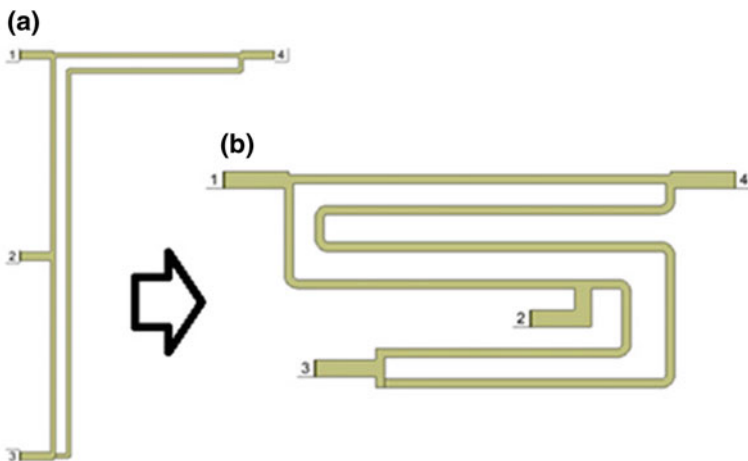
Fig. 14.3 Two steps of microstrip ring coupler modification

The bandwidth determined by the isolation level −20 dB is 320 MHz. However, in the present embodiment, 3649 mm² space inside the ring is not used.

At the first step of optimization the long section was bent inside the structure, as it is shown in Fig. 14.3a. Frequency dependencies remained unchanged. But according to the previously introduced definition, the useful area is not reduced. So for the full miniaturization of the occupied area, the obtained topology was bent like a snake (Fig. 14.3b). This topology allows reducing the geometrical dimensions of the coupler in comparison with initial design. The simulation results for the developed device are shown in Fig. 14.4.

The obtained computer simulation results show that this reduced microstrip coupler has an operation band of 323 MHz for 20 dB isolation (the initial design has a
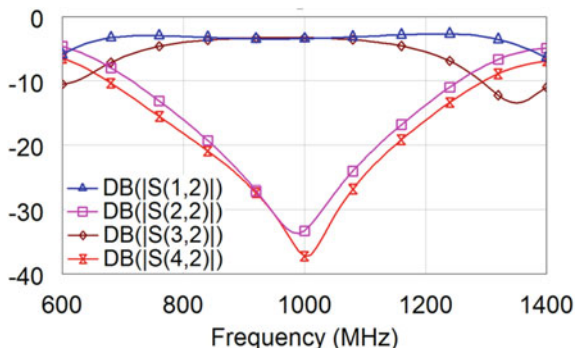
Fig. 14.4 Performance of developed device

bandwidth of 320 MHz). The geometrical coupler dimensions reduced from 3951.7 to 1099.15 mm^2, i.e. by 72.2%.

The next topology is realized by using the simplest "U"-bending of microstrip line segments (Fig. 14.5a). This implementation allowed reducing the geometric dimensions to 2695.7 mm^2 (31.8%). However such topology has the following disadvantage: with reducing the length of microstrip line, the width is raised. It is possible to obtain a larger degree of miniaturization by minor modification. Then, it is necessary to replace one of bends inside the ring (Fig. 14.5b). This change has allowed reducing the physical size to 1818 mm^2 (54%). Figure 14.6 shows the simulated results of this

Fig. 14.5 The second embodiment of microstrip ring coupler, **a** before; **b** after
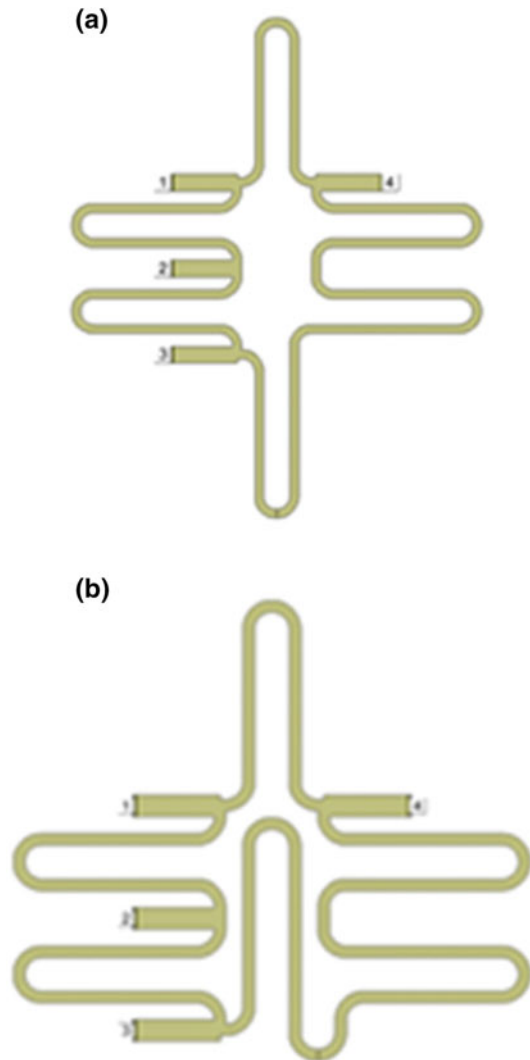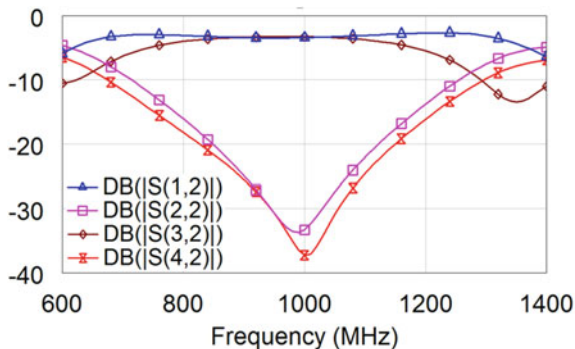
Fig. 14.6 The simulation results of the second developed topology

structure. It is seen that the center frequency was shifted by 10 MHz, and operation band, where S31 does not exceed -20 dB, is 308 MHz (12 MHz lower than the initial design).

The design shown in Fig. 14.5 can be improved due to the greater number of microstrip line bends. Dimensions in this case depend on the minimum clearance between adjacent conductive strips. In the considered case, the technologically realizable gap of 4 mm was chosen (because the mutual influence of two lines is less for this value). It allowed reducing the device area by 81.8% relative to the standard design (Fig. 14.7).
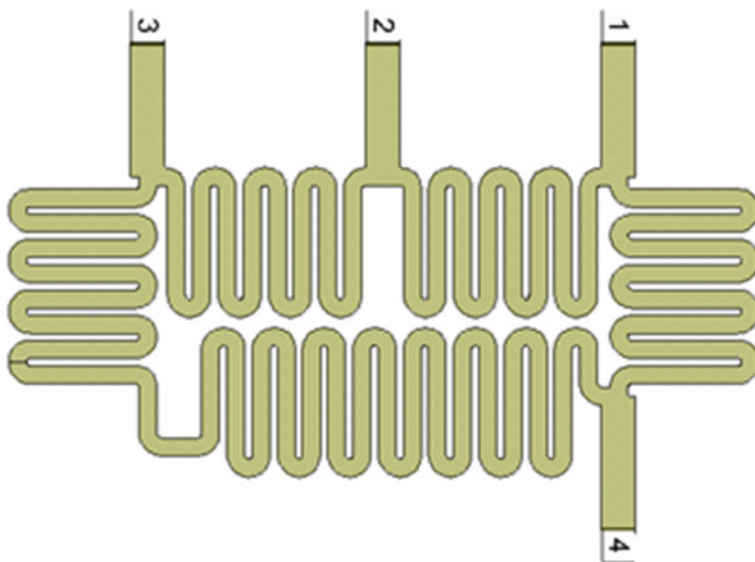
Fig. 14.7 The final variant of second topology
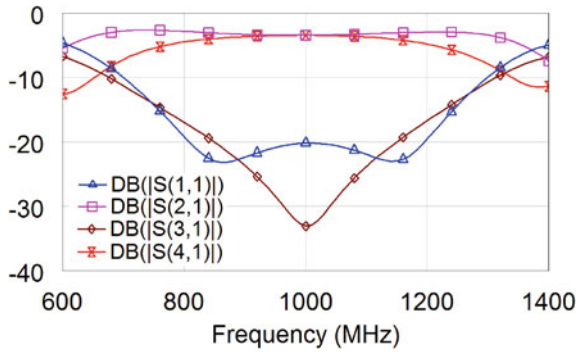
Fig. 14.8 The simulation results

The S-parameters dependencies obtained by the computer simulation are presented in Fig. 14.8. The proposed device operates on the frequency F = 1 GHz and has output power level −3.5 dB for each port. The operation band by the level of isolation −20 dB turned out to be equal to 300 MHz.

The latest design was implemented as a rectangle, which has a larger (3/4 wavelength) section collapsed inside the ring, as shown in Fig. 14.9. So the distance between adjacent ports is equal to quarter-wave line segment.

After these changes, the obtained structure still has a large unused area. Therefore the remaining microstrip lines were bent too with retaining the same length. The final design version is presented in Fig. 14.10.
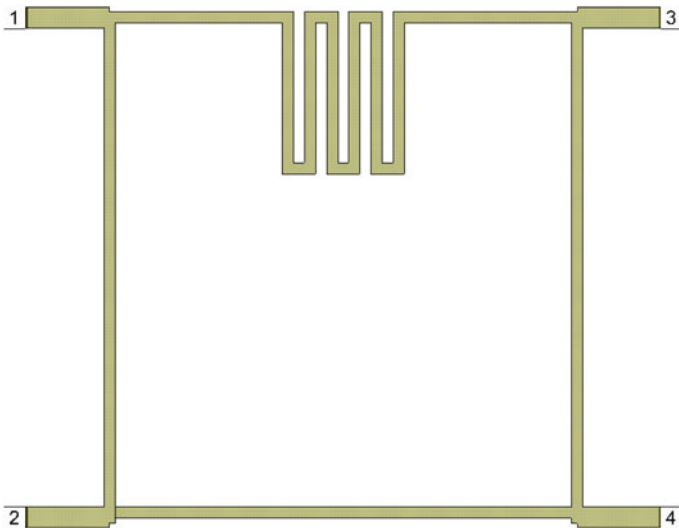
Fig. 14.9 The design of microstrip ring coupler
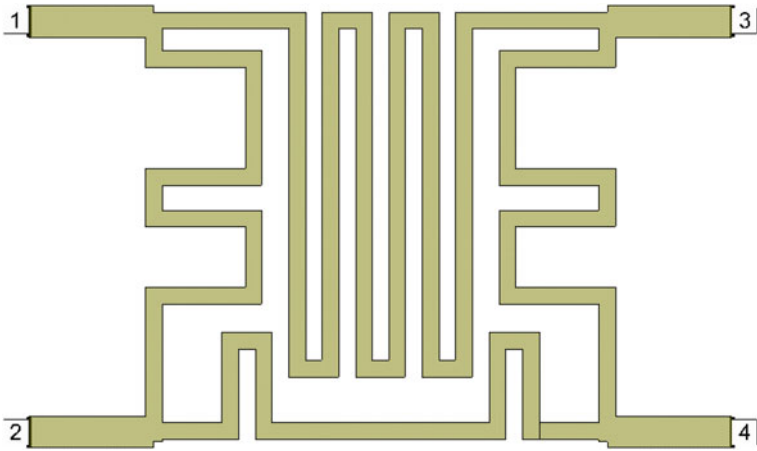
Fig. 14.10 The final design of microstrip ring coupler

Fig. 14.11 The simulation
results of compact microstrip
ring coupler
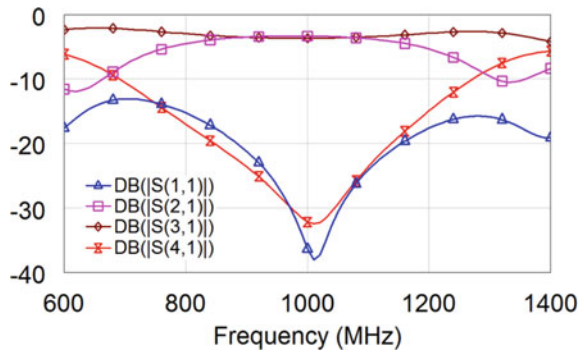
The results of computer simulation are given in Fig. 14.11. The central operation frequency is 1010 MHz. The power balance is 0.3 dB which is a negligible value. The operation frequency band by the level of isolation -20 dB is equal to 293 MHz. The geometric design dimensions are smaller by 81% in comparison with the initial structure.

14.3 Conclusion

Several versions of microstrip ring coupler miniaturization were proposed. Bending the microstrip lines without changing the overall length was chosen as the main method. This approach has shown its efficiency. It avoids the additional complexity of the mathematical apparatus.

The compact topologies of microstrip ring couplers operating at frequency F = 1 GHz were obtained. The results show that all designs effectively use the space inside the ring. The maximum gain of the resulting area is about 82%. Carried out computer simulation confirms the retention of the operation frequency band.

References

1. H.-R. Ahn, S. Nam, Compact microstrip 3-dB coupled-line ring and branch-line hybrids with new symmetric equivalent circuits. IEEE Trans. Microw. Theory Tech. **61**(3), 1067–1078 (2013)
2. K.W. Eccleston, S.H.M. Ong, Compact planar microstripline branch-line and rat-race couplers. IEEE Trans. Microw. Theory Tech. **51**(10), 2119–2125 (2003)
3. H. Ghali, T.A. Moselhy, Miniaturized fractal rat-race, branch-line, and coupled-line hubrids. IEEE Trans. Microw. Theory Tech. **52**(11), 2513–2520 (2004)
4. C.-W. Wang, T.-G. Ma, C.F. Yang, A new planar artificial transmission line and its applications to a miniaturized Butler matrix. IEEE Trans. Microw. Theory Tech. **55**(12), 2792–2801 (2007)
5. S.-C. Jung, R. Negra, F.M. Ghannouchi, A design methodology for miniaturized 3-dB branch-line hybrid couplers using distributed capacitors printed in the inner area. IEEE Trans. Microw. Theory Tech. **56**(12), 2950–2953 (2008)
6. D. Psychogiou, J. Hesselbarth, Comparing miniaturization techniques for microstrip 180° hybrid ring junctions, in *2010 Mediterranean Microwave Symposium (MMS)*, Aug 2010, pp. 29–32

Chapter 15
Analysis of Spherical and Geodesic Antenna Radomes by Green's Function Method

A. Karpov and S. Shabunin

15.1 Introduction

The primary function of any radomes is to protect an antenna system from the environment as rain, snow, wind and other atmospheric conditions. Antenna radomes must simultaneously meet various requirements such as mechanical strength, aerodynamic efficiency as well as the minimal transmission loss of electromagnetic waves in the frequency range [1]. It is known that radomes have an appreciable influence on properties of large electrical size antennas. Their main effects are: lowering antenna gain, sidelobes growing, boresight errors and depolarization [2].

It was shown that homogeneous radomes transmit well if the radome thickness is a multiple of half the wavelength in shell dielectric. The weight of monolithic dielectric panels may be too much for rigid antenna radomes and composite materials are more suitable in this case. Sandwich radomes that include low-density core materials and higher-density skin materials are popular because such radomes are more broadband than monolithic ones and also have a higher strength-to-weight ratio [2]. Layered structures are often used as radome walls for electrically large reflector antennas and antenna arrays.

Optimization of sandwich structures is required to reduce the radome impact on the antenna radiation in operating frequency bands; so, radome optimization may benefit from fast and efficient analysis algorithms.

The Green's function method is widely used for analysis of scattering and radiation electromagnetic problems due to its accuracy and efficiency [3–6]. In this paper, a relatively simple technique based on the Green's functions for multi-layer structures and the Huygens–Fresnel principle for spherical and geodesic radomes design is described. Instead of solving the diffraction formulation, we reduce the problem to the analysis of the radiation of equivalent electric and magnetic currents distributed over

A. Karpov · S. Shabunin (✉)
Ural Federal University, Yekaterinburg, Russia
e-mail: shab313@ya.ru; s.n.shabunin@urfu.ru

the illuminated radome surface. Electric losses in radome materials were taken into account. The advantage of the suggested method is that the reflected and transmitted waves are calculated simultaneously as waves emitted by equivalent currents at both sides of the shelter. We can use any kind of isotopic materials. Design of antenna radomes with metamaterials may be performed as well. The suggested approach is convenient for the antenna radome design if the antenna system operates in multiple frequency bands. Although the layered structures investigated by our method are flat, it is not so difficult to apply this approach to electromagnetic transparency evaluation for various incident angles and polarization needed for designing the spherical antenna radomes. For large antenna radomes, the framework structures of geometric patterns such as triangles and hexagons are used.

In this paper transmission and reflection coefficients as well as polarization properties of sandwich-type radome covers are under consideration. A lattice shell based on a grid of triangular flat elements on the surface of a sphere is analysed. Radiation patterns of a reflector antenna placed inside the layered radome are shown.

15.2 Sandwich-Type Structure Analysis as a Radiation Problem Solving

The problem is considered as follows: a reflected antenna or antenna array is placed inside a spherical or quasi-spherical radome (Fig. 15.1). The electrical field in the antenna aperture is calculated in some way, for example as described in [7]. Further, the aperture distribution is used for radiation patterns calculation. The amplitude and phase changed by the radome in the outside reference aperture deform the radiation pattern of the antenna. The distortions become much more if there is a space shift
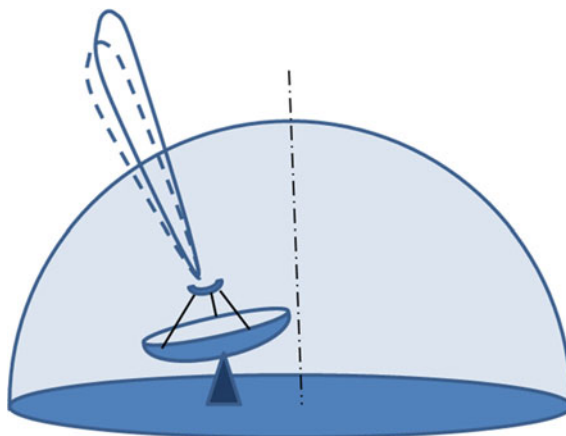
Fig. 15.1 Reflector antenna covered with a spherical radome

between the centre of the antenna and the centre of the spherical or quasi-spherical radome. These distortions must be considered for electrical and mechanical scanning antennas. Standard software for electrodynamic modelling as FEKO, HFSS and CST Microwave Studio requires a lot of resources for the considered problems. The way for simplifying the design of antennas with radomes is suggested in this paper.

We use the parallel and perpendicular electric field components relative to the incidence plane to describe any kind of incident wave polarization. Usually, such layered structures are mathematically modelled by the boundary value problem solution for forward and reverse propagating waves [2]. Matching tangential field components at each layer, the boundary reflection and transmission coefficients are derived from a set of equations. We replace this approach by the formulation of the radiation problem.

The reflected and transmitted waves generated by a layered dielectric structure are calculated as radiation of the equivalent electric and magnetic surface currents located at the illuminated side of the layered structure. The equivalent surface electric and magnetic currents for incident field are specified as

$$\mathbf{J} = \mathbf{n} \times \mathbf{H}, \quad \mathbf{M} = \mathbf{E} \times \mathbf{n}, \tag{15.1}$$

where $\mathbf{H}$ and $\mathbf{E}$ are incident wave components near the illuminated side of the radome, $\mathbf{n}$ is the surface normal.

The electrical field in the point of view r is calculated as follows

$$\mathbf{E}(\mathbf{r}) = \int\limits_{s'} \left[\bar{\bar{\Gamma}}_{11}(\mathbf{r}, \mathbf{r}') \, \mathbf{J}(\mathbf{r}') + \bar{\bar{\Gamma}}_{12}(\mathbf{r}, \mathbf{r}') \, \mathbf{M}(\mathbf{r}') \right] ds', \tag{15.2}$$

where $\bar{\bar{\Gamma}}_{11}(\mathbf{r}, \mathbf{r}')$ and $\bar{\bar{\Gamma}}_{12}(\mathbf{r}, \mathbf{r}')$ are the Green's tensor functions of the electric and magnetic type, respectively. Integration in Eq. (15.2) is performed over the illuminated surface area S' of the radome. For electrically large antenna radomes, the locally flat approximation may be used. The advantage of the suggested approach is in simultaneous calculation of the reflected and transmitted waves. The transmitted field is used for radiation pattern computing. The reflected field helps to define polarization isolation decreasing of antennas with orthogonal communication channels. The Green's functions in Eq. (15.2) are associated with electrical and geometric properties of the layered structure. This structure is modelled by the equivalent circuit method [8].

The integration surface area S' in Eq. (15.2) is associated with the incident waves region. The simplest case is for the infinite flat structure illuminated by the plane wave. The equivalent surface currents for the parallel polarization are specified as follows

$$M_x = -E_0 \cos\theta e^{-jk_0 x \sin\theta}, \quad J_y = E_0 Y_0 e^{-jk_0 x \sin\theta},$$

and for the perpendicular polarization we have

$$M_y = E_0 e^{-jk_0 x \sin\theta}, \quad J_x = -E_0 Y_0 \cos\theta e^{-jk_0 x \sin\theta},$$

where E_0 is the complex amplitude of the incident wave, θ is the incidence angle with respect to the surface normal, k_0 is the free-space wave number, $Y_0 = (120\pi)^{-1}$ Ohm^{-1} is the free-space wave admittance.

After substituting Eq. (15.1) into Eq. (15.2), the electric field components at the illuminated side at the plane $z = z_1$ are represented by rather simple expressions

$$E_x^t = E_0 \frac{2Y_0}{\overleftarrow{Y}^E(z_1) + \overrightarrow{Y}^E(z_1)} e^{-jk_0 x \sin\theta}, \tag{15.3}$$

$$E_y^t = E_0 \frac{2Y_0 \cos\theta}{\overleftarrow{Y}^M(z_1) + \overrightarrow{Y}^M(z_1)} e^{-jk_0 x \sin\theta}. \tag{15.4}$$

In Eqs. (15.3) and (15.4), the input directional admittances of electric and magnetic type are used as in [9]. The reference plane is at $z = z_1$. The input admittances to the right from the reference plane are calculated by recurrent formulas. If there is a free space at $z > z_1$, and so $\overleftarrow{Y}^{E,M}(z_1) + \overrightarrow{Y}^{E,M}(z_1) = 2Y_0$ the incident wave is expressed as

$$E_x^{inc} = E_0 \cos\theta e^{-jk_0 x \sin\theta}, \quad E_y^{inc} = -E_0 e^{-jk_0 x \sin\theta} \tag{15.5}$$

Subtracting Eq. (15.5) from Eqs. (15.3) and (15.4) we expressed the reflection coefficients by the following two formulas:

$$R_x = \frac{Y_0/\cos\theta - \overrightarrow{Y}^E(z_1)}{\overleftarrow{Y}^E(z_1) + \overrightarrow{Y}^E(z_1)} e^{-jk_0 x \sin\theta}, \tag{15.6}$$

$$R_y = \frac{Y_0 \cos\theta - \overrightarrow{Y}^M(z_1)}{\overleftarrow{Y}^M(z_1) + \overrightarrow{Y}^M(z_1)} e^{-jk_0 x \sin\theta}. \tag{15.7}$$

The transmitted field is calculated by Eq. (15.2) for the region $z > z_n$, where z_n is the coordinate location of the unilluminated radome side. The transmission coefficients for the plane $z = z_n$ are defined as follows:

$$T_x = \frac{2Y_0/\cos\theta}{\overleftarrow{Y}^E(z_1) + \overrightarrow{Y}^E(z_1)} e^{-jk_0 x \sin\theta} \prod_{i=2}^{n} T_i^E, \tag{15.8}$$

$$T_y = \frac{2Y_0 \cos\theta}{\overleftarrow{Y}^M(z_1) + \overrightarrow{Y}^M(z_1)} e^{-jk_0 x \sin\theta} \prod_{i=2}^{n} T_i^M, \tag{15.9}$$

where $\bar{Y}^E(z_1)$, $\vec{Y}^E(z_1)$, $\bar{Y}^M(z_1)$ and $\vec{Y}^M(z_1)$ are input directional admittances of the equivalent electric and magnetic circuits at the reference plate $z = z_1$ of the layered structure, $T_i^{E,M}$ are the transmission coefficients of the i-th layer in electric and magnetic equivalent circuits. In Eqs. (15.6)–(15.9) R_x, R_y, T_x and T_y are reflection and transmission coefficients for parallel and perpendicular electrical field components, respectively.

In Fig. 15.3, the transmission coefficients versus frequency data are shown for the perpendicular and parallel polarizations for different angles of wave incidence on the three-layer radome. The layers' numbers 2 and 4 according to Fig. 15.2 are fabricated from a solid dielectric with relative permittivity $\varepsilon = 3.3$. The layer number 3 has a
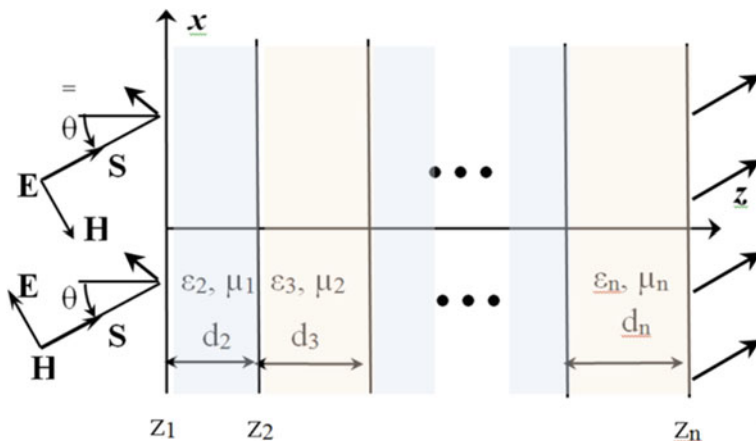
Fig. 15.2 An incident wave with perpendicular and parallel electric field vector near multilayer structure

Fig. 15.3 Transmission losses for parallel (red) and orthogonal (blue) polarized waves by a three-layer radome for different angles of incidence: $\theta = 0$ in black, $\theta = 15°$ in solid, $\theta = 30°$ in dash, $\theta = 45°$ in dash-dot
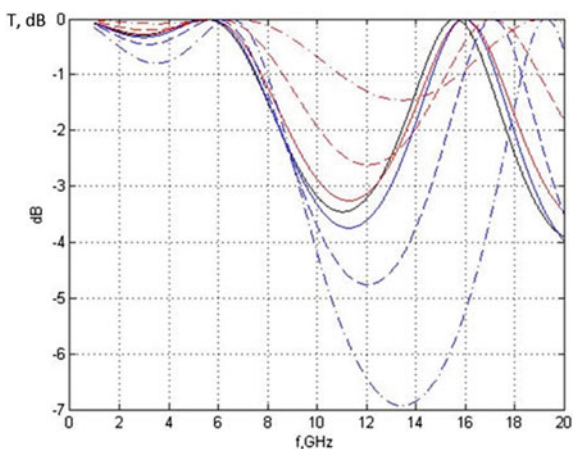
cellular structure with air cavities which measured relative permittivity is $\varepsilon = 1.2$. Thickness of each layer is $d_2 = d_4 = 2.5$ mm, $d_3 = 9$ mm.

Previously, we considered reflection and transmission coefficients for the electric field vector of the incident wave separately for two cases. The first one was for the vector **E** located in the incident plane and the second case was for the vector **E** orthogonal to the incident plane. If we take into account the phase shifts between these two orthogonal field components, the circular and elliptic incident wave interaction with the sandwich radome may be analysed.

15.3 Circular Polarized Wave Analysis for Sandwich-Type Radomes

It is well known that for the normal wave incidence the flat dielectric structure does not distort the circular polarization, but for other incidence angles, the relation between the orthogonal components and their phase shift is changed. The suggested method allows one to analyse the degradation of circular polarized electromagnetic waves caused by an inclined incidence. The amplitude and phase shift changing for the inclined incidence is shown in Figs. 15.4 and 15.5. Usually, the axial ratio of the transmitted wave is analysed. For the inclined wave incidence, the optimal frequencies for circular polarised wave transmission may change significantly (Fig. 15.6).

In Fig. 15.6 the axial ratio versus frequency data is shown for various angles of wave incidence on the three-layer radome as in Fig. 15.3. Up to frequency 7 GHz, the axial ratio is not more than 2.5 dB up to 60° incidence angle.

Fig. 15.4 The difference in transmission of the parallel and orthogonal wave components of the circular polarized wave for various incident angles: $\theta = 0$ in black, $\theta = 15°$ in solid, $\theta = 30°$ in dash, $\theta = 45°$ in dash-dot
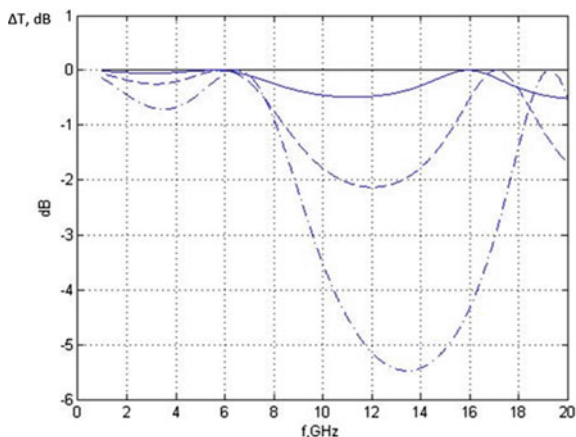
Fig. 15.5 The phase shift between the parallel and orthogonal wave components of the circular polarized incident wave for the three-layer radome for various incident angles: $\theta = 0$ in black, $\theta = 15°$ in solid, $\theta = 30°$ in dash, $\theta = 45°$ in dash-dot
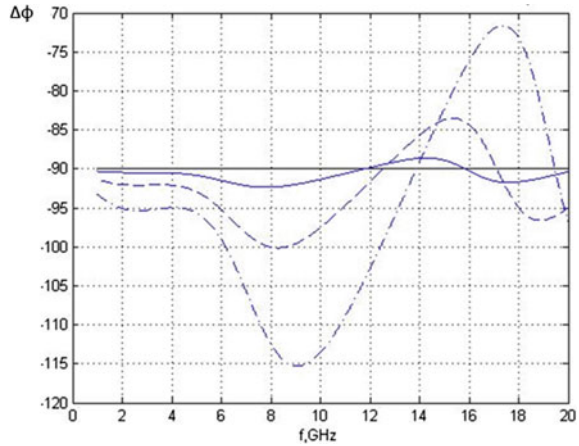
Fig. 15.6 The axial ratio versus frequency of the circular polarized incident wave and the three-layer radome for different angles of incidence: $\theta = 15°$ in black, $\theta = 30°$ in solid, $\theta = 45°$ in dash, $\theta = 60°$ in dash-dot
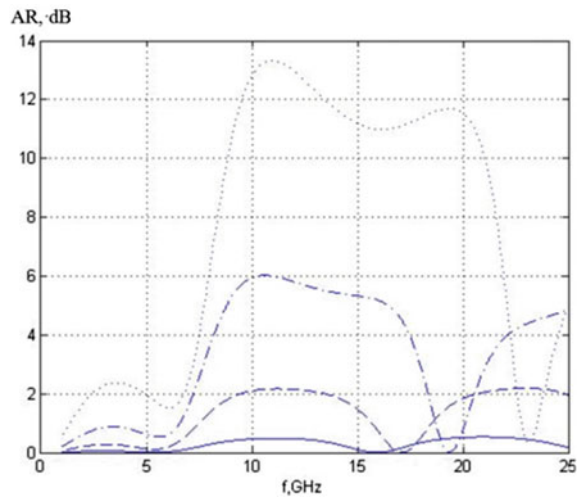
15.4 Large Antenna Radome with Reflector Antennas and Antenna Arrays

Analysis of the reflector antennas or antenna arrays covered with a large spherical radome is performed in the following way. Since the radius of spherical radome is a lot of wavelengths, the cover for any local region may be considered as flat, and the reflecting and transmitting coefficients are calculated as in Sect. 15.2.

At the first step, the electrical field near the inner radome surface is calculated by the Huygens source approximation method or another one. At the second step, the amplitude, phase and polarization distortion caused by the radome are derived. The far field is obtained by insertion of the amplitude and phase distribution over the

external equivalent antenna aperture. The phase delay and additional attenuation for different incident angles were calculated by the approach described in the previous sections. It is well known that antenna radome decreases the antenna gain because of the wave attenuation in the cover material and reflection from the cover. The additional gain reduction occurs due to distortion of a phase plane after transmission through an uneven thickness of the antenna cover. The suggested method allows one to take into account for all these mechanisms.

Radiation patterns of a 3.5-m Cassegrain antenna covered by a hemispherical three-layered radome with a radius of 3.5 m for a 15.5 GHz frequency band are shown in Fig. 15.7. The red line corresponds to the symmetrical location of the antenna. Introduction of 1.5 m displacement of the antenna aperture centre from the centre of the spherical dielectric radome turns radiation patterns unbalanced (blue line).

The boresight error (that takes place if there is a displacement of the reflector centre relative to the centre of spherical radome) is of 0.050° value for 1.5 m displacement. Small decreasing of directivity is observed. It is obvious that for electrically large antennas the boresight error will be more sensitive.

For large and rigid antenna radomes framework structures are used. A lattice shell based on a network of triangular flat elements on the surface of a sphere is applied very often. This kind of radomes is known as the geodesic antenna radomes. Usually, the aperture antenna or antenna array is located at the centre of the radome but in some cases, it may be a shift between antenna and radome centres (Fig. 15.8).

Fig. 15.7 Radiation pattern of the reflector antenna covered with a spherical radome at 15.5 GHz (centred in red, with 1.5 m displacement in blue)
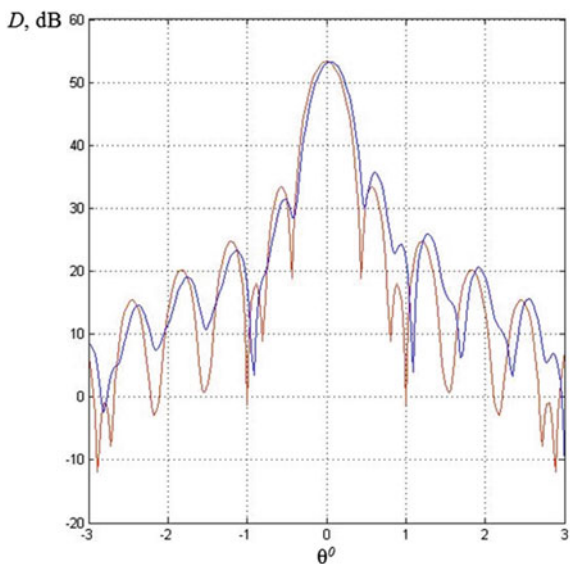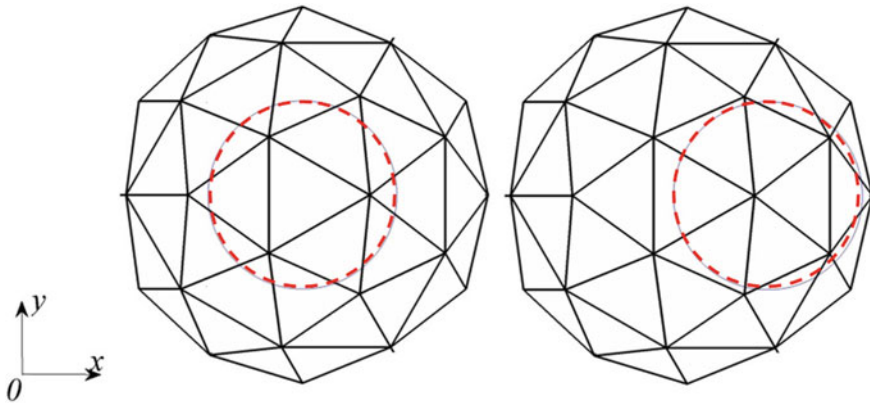
Fig. 15.8 Aperture antenna located in the centre and off-centre of the lattice shell based on a network of triangular flat elements

The radiation patterns of the 3.5-m circular aperture antenna without an antenna radome and covered by the three-layered radome with the radius 3.5 m formed by triangular elements for 15.5 GHz frequency band are shown in Fig. 15.9. A three-layer cover is used for triangular elements. Dielectric constant and thickness of each layer are the same as in Fig. 15.3.

Antenna directivity decreases by 0.5 dB for covered antenna in the 15.5 GHz frequency band. If there is a space shift between the antenna centre and the radome centre, the beam angular displacement is observed (Figs. 15.10, 15.11 and 15.12).

Fig. 15.9 Radiation pattern of the circular mirror antenna without the radome (red) and in the centre under the radome (blue) at frequency 15.5 GHz in x direction (Fig. 15.8)
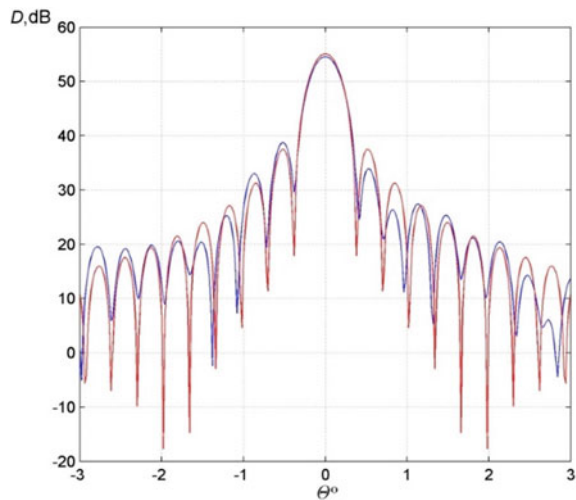
Fig. 15.10 Radiation pattern of the circular mirror antenna without the radome (red) and in the centre under the radome (blue) at frequency 15.5 GHz in y direction (Fig. 15.8)
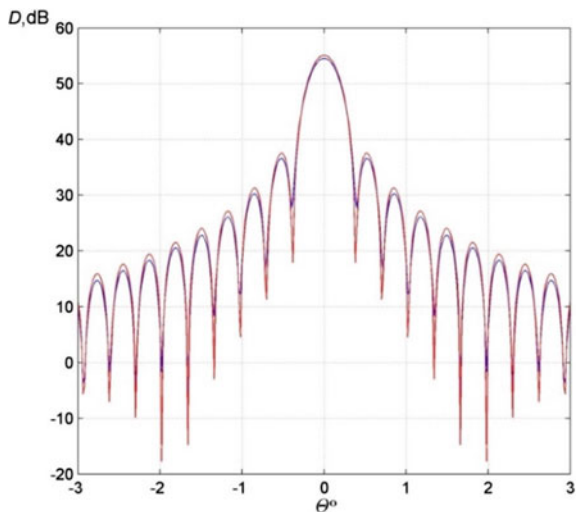
Fig. 15.11 Radiation pattern of the circular mirror antenna without the radome (red) and in the centre under the radome (blue) at frequency 6.0 GHz in x direction (Fig. 15.8)
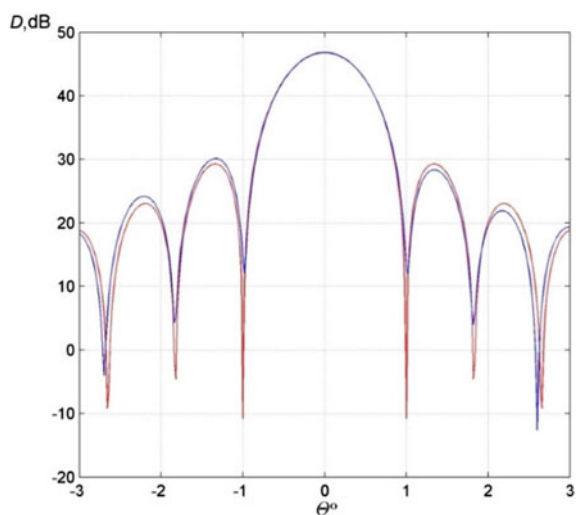
Despite of the radome under consideration having a spherical shape, the suggested method may be rather easily adapted for different kind of surfaces. So, radiation pattern degradation and antenna gain decreasing may be analysed for various radome shapes. Transparent frequency bands for sandwich antenna radomes may be determined as well. Moreover, optimization may be performed of layered dielectric structures for multiband telecommunication and radar systems.

Fig. 15.12 Radiation pattern of the circular mirror antenna without the radome (red) and under the radome with 1.5 m shift (blue) at frequency 15.5 GHz in x direction (Fig. 15.8)
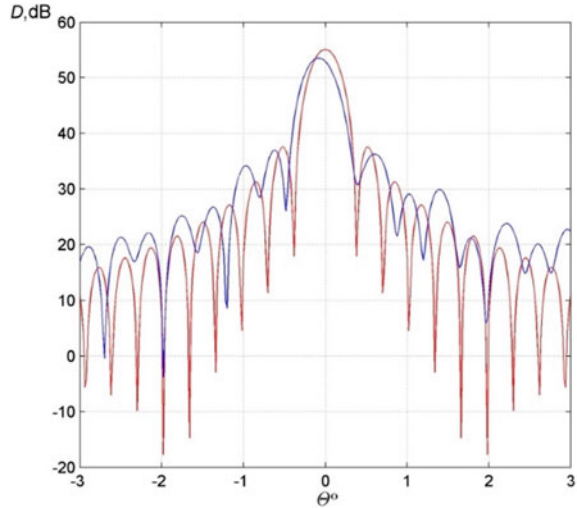
15.5 Conclusion

The Green's functions method was used for antenna radome analysis for obtaining the radiation of equivalent surface electric and magnetic currents. This method simplifies the determination of radiation and diffraction properties of antennas with flat and curved multi-layered covers. The method allows developing rather a compact algorithm and is appropriate for quick computer modelling electrically large antenna covers with multi-layered structure. The method is convenient for modelling, analysis and optimization of the antenna cover for the multi-frequency transmission.

References

1. L.N. Ridenour, *Radar Scanners and Radomes* (McGraw-Hill Book Company, New York, 1948)
2. D.J. Kozakoff, *Analysis of Radome-Enclosed Antennas* (Artech House Inc, Norwood, 2010)
3. L. Vegni, R. Cicchetti, P. Capece, Spectral dyadic Green's function formulation for planar integrated structures. IEEE Trans. Antennas Propag. **36**(8), 1057–1065 (1988)
4. L.-W. Li, P.-S. Kooi, M.-S. Leong, T.-S. Yee, Electromagnetic dyadic Green's function in spherically multilayered media. IEEE Trans. Microw. Theory Tech. **42**(12), 2302–2310 (1994)
5. Z. Xiang, Y. Lu, Electromagnetic dyadic Green's function in cylindrically multilayered media. *IEEE Trans. Microw. Theor. Tech.* **44**(4), 614–621 (1996)
6. S. Knyazev, L. Lesnaya, S. Shabunin, Green's functions of multilayered cylindrical structures and their application for radiation, propagation and scattering problems solving, in *Microwave and Optoelectronics Conference (IMOC), 2011 SBMO/IEEE MTT-S International*, pp. 748–752 (2011)

7. T.A. Milligan, *Modern Antenna Design*, 2nd edn. (Wiley-IEEE Press, New Jersey, 2005)
8. A. Karpov, S. Knyazev, S. Shabunin, A new technique for sandwich antenna radomes analysis, in *Antennas and Propagation Conference (LAPC)*, 2014 Loughborough, pp. 325–328 (2014)
9. L.B. Felsen, N. Marcuvitz, *Radiation and Scattering of Waves*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey (1973)

Chapter 16
Comparative Analysis of TE_{01} Mode Excitation Methods in Circular Waveguides

D. A. Letavin, V. Chechetkin and Yu. Mitelman

16.1 Introduction

A circular waveguide is a metal tube of circular cross-section. The number of waves existing simultaneously with the fundamental wave depends on the propagation conditions: only waves with a wavelength not exceeding the critical, can propagate in a circular waveguide. Numerous studies have shown that the axially symmetric magnetic waves have a unique feature, namely, the attenuation of such wave decreases with increasing frequency for the same diameter of the waveguide. This can be explained by the fact that these waves have longitudinal and transverse components of the magnetic field, which in turn excite the transverse and longitudinal currents, respectively, and for the TE_{0n} waves, unlike other types of waves, there are no longitudinal currents creating increased losses. The TE_{01} mode has the lowest attenuation among the axially symmetric magnetic waves TE_{0n}.

Due to the large difference in the attenuation, the TE_{01} mode was the most preferable for transmission of information over long distances and in the 40s of the 20th century, the first developments in the excitation of this mode begun in a number of countries such as the US, UK, USSR, France, Japan and Germany [1].

D. A. Letavin (✉) · V. Chechetkin · Yu. Mitelman
Ural Federal University, Yekaterinburg, Russia
e-mail: d.a.letavin@urfu.ru

Yu. Mitelman
e-mail: y.e.mitelman@urfu.ru

16.2   Comparative Analysis of Excitation Methods

The aim of this research was to carry out a comparative analysis of various methods of excitation of the TE_{01} mode in a circular waveguide. ANSYS HFSS was used for simulations of different designs and then, comparative analysis was made based on to the results of numerical simulation.

The design of transition from a rectangular waveguide with the TE_{10} mode to the circular waveguide with the TE_{01} mode in [2] is based on the principle of a smooth deformation of the cross-section shape of the waveguide.

In this structure, the input mode is TE_{10}. It passes through the conversion device, wherein the waveguide is gradually divided into two ones, which are tilted by 90° in opposite directions. Thus, on the inverter output at the point of contact of these two waveguides their electric fields are directed opposite to each other, which corresponds to the field structure of the TE_{20} mode. As a result, this section of the design becomes the rectangular waveguide mode converter from the TE_{10} to the TE_{20}. After that, the EM wave passes through the final part of the structure, in which the TE_{20} mode is converted into the TE_{01} through a smooth transition from rectangular to circular waveguide. This conversion is performed by rounding the side walls in such a way that their sizes approach the size of the circular waveguide. As a result, on the output of the device, the TE_{01} mode is formed with the electric field lines arranged in a circle and magnetic lines become radial, as shown in Fig. 16.1.

The model of the proposed design in HFSS is shown in Fig. 16.2. The simulation results are shown in Fig. 16.3. Analysis of the results shows that the device allows
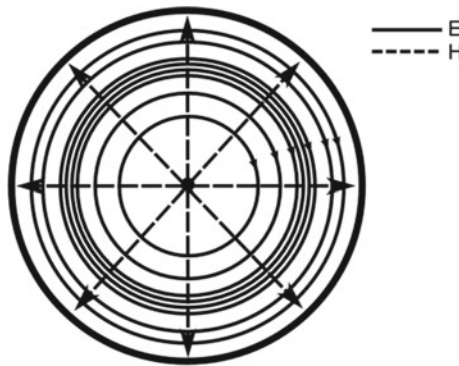

Fig. 16.1 Structure of the TE_{01} mode
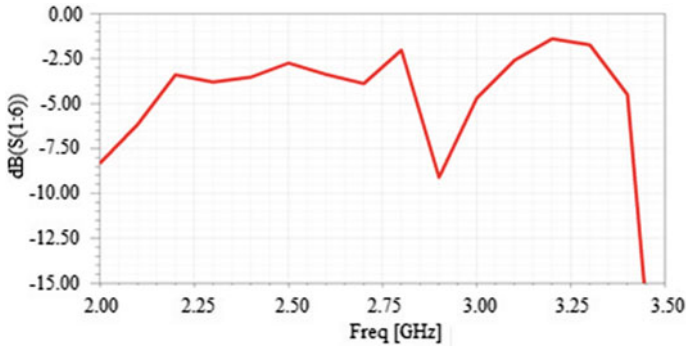

Fig. 16.2 Design based on the smooth deformation

Fig. 16.3 Transmission coefficient of the TE$_{01}$ mode

one to excite the TE$_{01}$ mode in a circular waveguide with a transmission coefficient better than −1.2 dB.

The disadvantage of this design is that the device is cumbersome due to the necessary length of the transition part. However, the smooth transition helps to obtain the desired purity of the TE$_{01}$ mode excitation in a wide frequency band.

In [3], a method is suggested for the TE$_{01}$ mode excitation in a circular waveguide by connecting both broad walls of a rectangular waveguide with the end of the circular waveguide via the waveguide connection. The waveguide connections can be made in the form of two waveguide segments bent at 90° angle with U-shaped or H-shaped cross-section. The lengths of the waveguides along the axial line are identical and equal to a quarter-wavelength on the average frequency. Using the U- and H-shaped cross-section waveguides allows avoiding the disadvantage of the rectangular waveguides, namely, expanding the frequency range by increasing the cut-off wavelength.

In this design, a pin is used for matching the rectangular and U-shaped waveguides. The model of the device is shown in Fig. 16.4, and the simulation results are in Fig. 16.5. The advantage of this method is that the usage of the U-shaped waveguide provides broadband operation of the device. However, it is necessary to match rectangular waveguide with the U-shaped one.

The method described in [4] is based on the TE$_{01}$ mode excitation using a rectangular waveguide with its narrow wall touching the short-circuited end of the circular waveguide. Contact between these two waveguides is carried out through the same holes in their common wall. In order to obtain a circular shape of the TE$_{01}$ mode, it is required to spread holes up to the half-wavelength in a rectangular waveguide, which leads to their antiphase excitation (Fig. 16.6). The geometrical dimensions of the waveguides are selected such that the cut-off frequency and, accordingly, the phase velocities of the TE$_{10}$ mode in the rectangular and the TE$_{01}$ mode in the circular waveguides were equal.
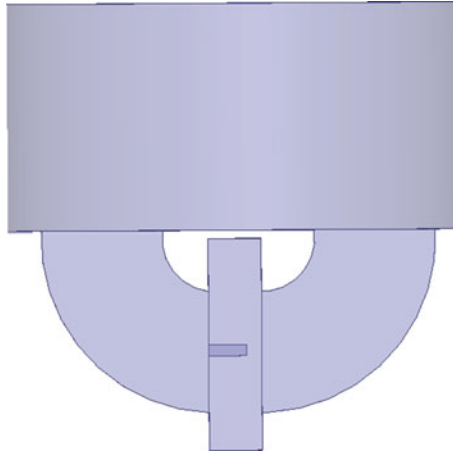
Fig. 16.4 The design, based on a U-shaped waveguide
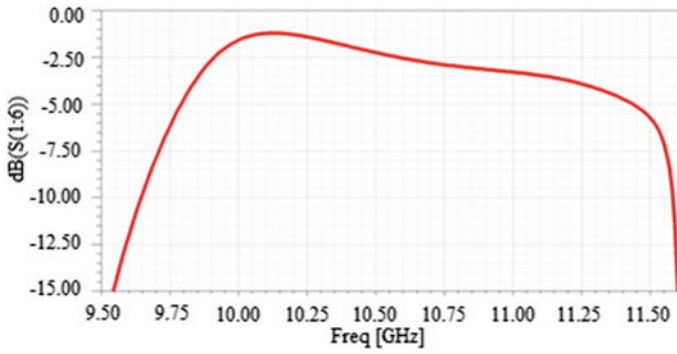

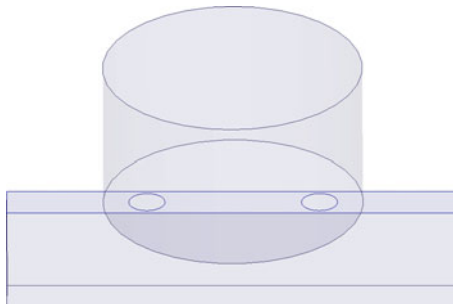**Fig. 16.5** Transmission coefficient of the TE_{01} mode

Fig. 16.6 The design based on the imposition of the narrow wall of the rectangular waveguide to circular waveguide
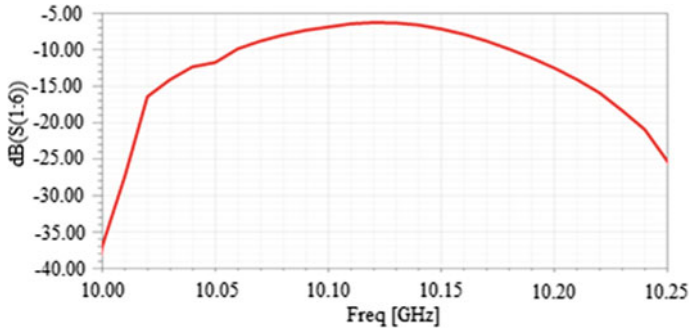
Fig. 16.7 Transmission coefficient of the TE$_{01}$ mode

The results (Fig. 16.7) show that the proposed method is inefficient, since the energy from the rectangular waveguide is passing through holes in its narrow wall. This leads to having in output a small fraction of input power in a narrow frequency band.

In [5], authors used a coaxial line segment mounted coaxially within the circular waveguide. The inner conductor is branched into multiple conductors, and each of them forms a sector-shaped loop in the plane perpendicular to the longitudinal axis of the waveguide (Fig. 16.8). The ring sections in the loop excite azimuthal components of the electric field in the circular waveguide, which are typical for the TE$_{0n}$ modes. Figure 16.9 shows the simulation results. The advantage of this device is in its compactness and the disadvantage is the difficulty of obtaining accurate dimensions of the sectors (and the installation location).

In [6], the application of two antiphase dipoles for the TE$_{01}$ mode excitation in a circular waveguide is considered. The operational principle is to excite dipoles with the antiphase voltage. The antiphase excitation is obtained with their opposite
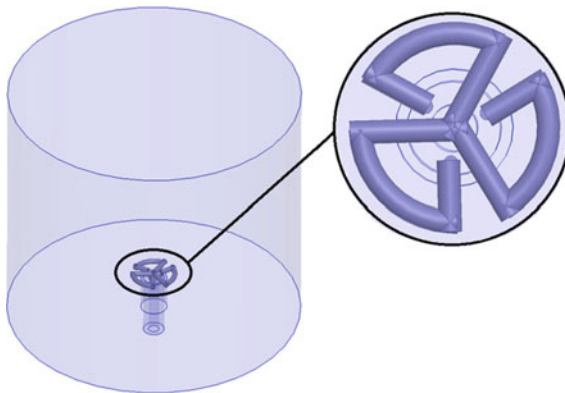

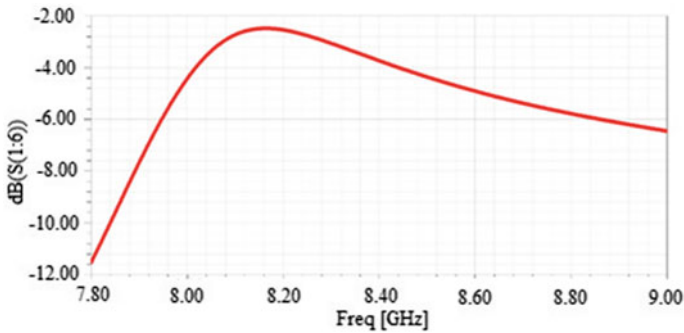**Fig. 16.8** The design based on the sector-shaped dipole

Fig. 16.9 The design based on the sector-shaped dipole

installation: the arm of the first dipole is connected to the inner conductor of the coaxial line, and the other one to the external and the second vice versa. This creates an enclosed electric field around the axis of the waveguide that is a condition for the propagation of the TE_{01} mode. This allows one to significantly expand the bandwidth, since it does not require an external phase shifter. The disadvantage of this method is the considerable difficulty of accurate setting of the dipoles inside the waveguide. Figure 16.10 shows the design of this excitation device, and Fig. 16.11 presents the results of numerical simulation.

The method proposed in [7] uses the conversion of the TE_{20} mode in a rectangular waveguide to the TE_{01} mode in the circular waveguide that are coupled with the end slots. The device works due to antiphase excitation of these slots and selection of geometrical dimensions of the rectangular waveguide in such a way that they satisfy the condition of equality of the critical frequencies, and consequently, equality of the phase velocities of the TE_{10} rectangular and TE_{01} circular modes. The end slots are
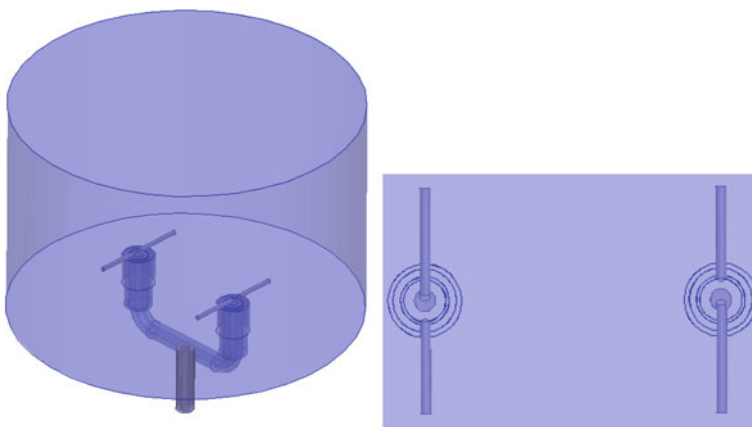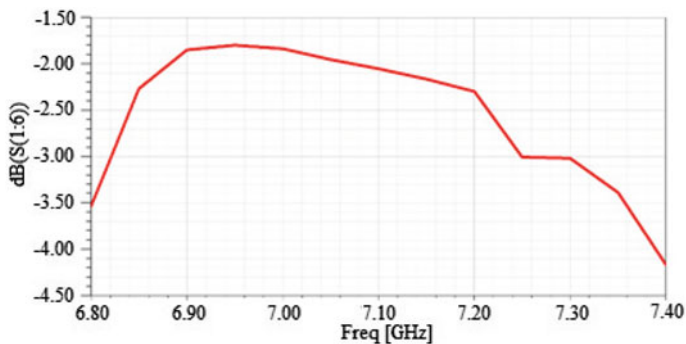

Fig. 16.10 The design based on two antiphased dipoles

Fig. 16.11 Transmission coefficient of the TE$_{01}$ mode

designed as half-rings. The advantage of this method is its simplicity, compact design, and high transmission coefficient, but narrow band. Figure 16.12 shows the design of such excitation device, and Fig. 16.13 shows the results of numerical simulation.
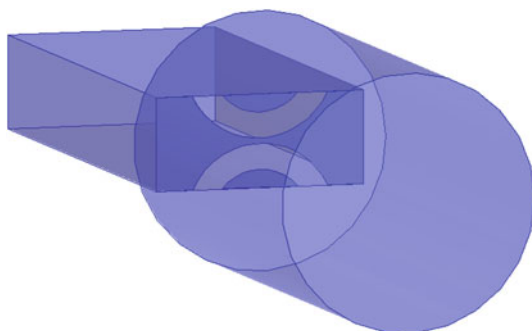

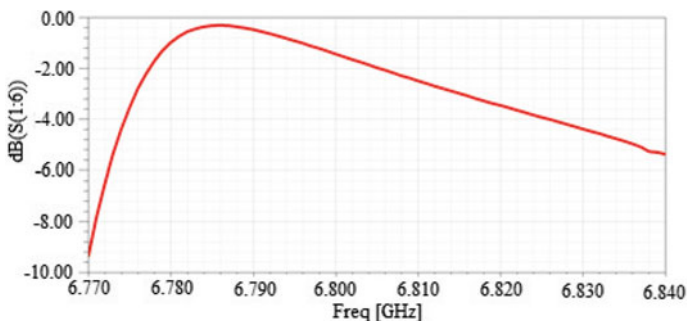**Fig. 16.12** The design, based on the stepped waveguide transition

Fig. 16.13 Transmission coefficient of the TE$_{01}$ mode

All of the methods mentioned above provide the excitation of the other modes in addition to the TE_{01} mode. Therefore, it is necessary to suppress lower and higher modes, for which the propagation conditions are agreed. In order to obtain the highest transmission coefficient of the TE_{01} mode, the following solutions can be used:

- using special mode filters to suppress spurious modes without suppressing the TE_{01} mode;
- applying resistive films to suppress parasitic modes;
- replacing solid walls with spiral or separated metal rings.

16.3 Conclusions

The article provides a brief overview of some existing methods of the TE_{01} mode excitation in a circular waveguide. It can be concluded that the most compact design is based on a coaxial cable inserted into the device because all the other designs require additional waveguides. For the case when it is necessary to obtain the maximum transmission coefficient, the device based on conversion of the TE_{20} mode in the rectangular waveguide to the TE_{01} mode in the circular waveguide through the end slots is the best choice. When the broadest bandwidth is needed, it is better to select a device with the U-shaped waveguide.

All these methods could be divided into two categories:

- excitation of the required mode is carried out by the smooth cross-sectional deformation;
- excitation of the required mode is carried out by using the opposite-coupling elements (slots, dipoles, etc.).

References

1. V.M. Dmitrachenko, Yu.M. Isayenko, Historic landmarks of telecommunication waveguide lines designing in USSR, in *Telecommunications: History and Modernity* (no. 3, 2007), pp. 21–27 [in Russian]
2. V.I. Lebedev, *Microwave Technologies and Devices* (Moscow, Vyssh. Shk. Publ., 1970) [in Russian]
3. B.B. Ramanuskas, Device for TE_{01} mode excitation in circular waveguide. Certificate of authorship USSR, no. 218253 (1968) [in Russian]
4. S.N. Sibircev, V.F. Matvejchuk, S.V. Poberezhnyj, N.A. Shhetkin, Device for TE_{01} mode excitation in cylindrical resonator. Certificate of authorship USSR, no. 232341 (1988) [in Russian]

5. B.-I.B. Ramanuskas, Device for TE$_{0n}$ mode excitation in circular waveguide. Certificate of authorship USSR, no. 1394280 (1968) [in Russian]
6. D.A. Letavin, Yu.E. Mitelman, The excitation device for H$_{01}$ mode in a circular waveguide based on antiphased dipoles, in *24nd International Crimean Conference Microwave and Telecommunication Technology (CriMiCo'2014)* (Sevastopol, 2014), pp. 561–562
7. E.E. Malov, Yu.E. Mitelman, The circular waveguide TE$_{01}$ mode converter based on aperture slot coupling with a rectangular waveguide, in *24nd International Crimean Conference Microwave and Telecommunication Technology (CriMiCo'2014)* (Sevastopol, 2014), pp. 576–577

Chapter 17
Features of Antenna-Applicator for Functional Studies of the Human Brain

Yuriy E. Sedelnikov, Vladimir S. Kublanov, Sergey A. Baranov and Vasilii I. Borisov

17.1 Introduction

In accordance with the Rayleigh–Jeans law, the electromagnetic radiation T_x of biological tissues in radio wavelengths, in general, varies in time t and depends on changes in the tissue absorption coefficient $\chi(\varepsilon, t)$ and on the thermodynamic temperature $T(t)$ according to

$$T_x \propto \chi(\varepsilon, t)T(t), \tag{17.1}$$

where ε is the permittivity.

The microwave radiometer, like any electronic device, uses an antenna for receiving *microwave* radiation. In the case of contact microwave radiometers, the antenna-applicator (AA) is used. The AA has a number of features due to the complexity of the brain when it is the object of study. The registered received AA signal in the antenna output T_A of the contact microwave radiometer is given by

$$T_A = \left(1 - \Gamma^2\right) \int_{4\pi} T_{br}(\Omega) F(\Omega) d\Omega, \tag{17.2}$$

where $T_{br}(\Omega)$ is the *brightness temperature*, $F(\Omega)$ is the *antenna pattern*, and $\left(1 - \Gamma^2\right)$ is the *reflection coefficient* that determines the effect of reflections on the border of body and AA.

Y. E. Sedelnikov
Kazan National Research Technical University, Kazan, Russia

V. S. Kublanov (✉) · S. A. Baranov · V. I. Borisov
Ural Federal University, Yekaterinburg, Russia
e-mail: kublanov@mail.ru

V. I. Borisov
e-mail: vi.borisov.official@gmail.com

In agreement with expression (17.2), the output signal and, respectively, the evaluation of brightness temperature depends on the parameters of the antenna's path. In order to minimize the measurement error of this estimation, the AA should provide good spatial selectivity that is determined by the characteristics of $F(\Omega)$ and good agreement of the AA with the brain tissues in the microwave band of the radiometer receiver path. For that microwave band, change of the reflected coefficient Γ from the boundary of AA and body will be minimal.

Note that increase of the AA "directional" properties is an urgent task. However, the results obtained in [1] showed that the losses in the medium cause expansion of the focal region. In the case of the medium, considerable damping the effect of focusing in the transverse direction almost disappears. At the same time for most biological tissues, the focal spot sizes are smaller compared to the omnidirectional applicator. Therefore, focusing improves the resolution only in the directions along the aperture.

Commonly, in contact microwave radiometers, the AAs are applied with dimensions close to the resonance [2]. In [3], AAs were described that are made on the basis of a circular waveguide filled with a dielectric with high dielectric constant. Such AAs allow one to measure brightness temperature of small objects. Compared to flat AAs, they have a larger depth of identification of thermal anomalies, smaller size and better resolution.

However, these advantages disappear in the case of studying objects, whose geometric characteristics are substantially larger than the length of the *microwave radiation* in the test environment. It is worth noting that there is an impact of the error estimation of brightness temperature of biological tissue on matching AA with brain tissues. This error is determined by the ratio of the radiation reflected $\left(1 - \Gamma^2\right)$ on the boundary of AA and body.

Particularly, this problem becomes acute during prolonged (for at least 5 min) monitoring of the functional processes of the brain. Depending on the design of the AA and its installation site on the head, a major impact on the radiation reflectance $\left(1 - \Gamma^2\right)$ can be made by the scalp, as well as, the formation of the "greenhouse" effect (changing heat exchange processes with the environment). This is a consequence of microcirculation alteration processes due to the mechanical action of AA, the gas exchange with the environment, etc. But it remains problematic how the AA fixing elements impact on the AA matching with the body, since the elements change the antenna diagram $F(\Omega)$.

Since the 70 s of the last century, developers of medical contact microwave radiometers used the design proposed by Ludeke et al. [4], in which the AA interface with the body allows one to reduce this impact at the expense of "additional noising". But, in our opinion, only one solution can be more correct, namely that which either excludes the factor mentioned above (that is impossible, in principle) or significantly decreases its influence on the estimation accuracy of brightness temperature.

The aim of this article is the development of an AA for a microwave radiometer with dimensions close to resonance, which provides the best conditions for the admission of the brain microwave radiation with long-term monitoring of functional changes in its tissues.

17.2   Analysis of Possible Ways to Implement the AA for Contact Microwave Radiometry

Among the most known AAs, there are three fundamentally different types of them, which can be roughly defined as AA of electric or magnetic type, or of the combined type. Such AAs combine the properties of electric and magnetic antenna types. Examples of electric type AA are the dipole antennas, of magnetic type (frame and slot) and combined type, the antennas in the form of open-ended waveguide.

In the analysis of the AAs characteristics, most interesting ones are their properties in the near-field $r < r$ (near-field) $= \lambda/2\pi$ and in the transition region of the antenna (where λ is its wavelength). These characteristics essentially depend on the type of the antenna radiator. In the far-field, the intensity of the electromagnetic field (EMF) does not depend on the type of transmitter and is wholly determined by the properties of the medium, the position of the observation point and the gain of the antenna [5].

Strictly speaking, any real antenna is not a transmitter of purely electric or magnetic type. We can only talk about the prevailing nature of it. Therefore, an accurate analysis of specific absorption (depending on the position of the observation point) is only possible when specifying the design of the antenna and the electrical properties of the medium. However, obtaining some estimates and corresponding recommendations is possible on the basis of classical results of solving the problem of radiation of EMF by a set of third-party sources.

The evaluation results obtained on this basis are valid for observation points separated from the position of third-party sources by the magnitude of the transverse dimensions of wire antennas. We use estimates for the well-known ratio (e.g. [3]) for the components of the electric and magnetic fields produced by electrical short vibrator ($l \ll \lambda$) and the small frame ($S \ll \lambda^2$) at an arbitrary point in space (where l is vibrator length, S is the frame contour area).

The values of electric $p_e = Il$ and magnetic $p_m = j\omega\mu lS$ torques, (where I is electric current, j is current density, ω is frequency, μ is magnetic permeability) should ensure the equality of the electric field strength in the far field for a short dipole and a small loop antenna. Comparison of specific absorbance values $|w_{abs}| = \sigma|E|^2$, ($\sigma =$ conductivity) depending on the type of antenna and the distance to the observation point allows us to give an answer to the question of preference or application of an elementary dipole framework radiometry problems. Typical results for frequency of 500 and 750 MHz are shown in Fig. 17.1.

Actual dimensions of the real AA are of the order of half a wavelength, while the above data are for electrically short antennas ($l \ll \lambda$). Similar estimates were obtained for these antennas. The electric type vibrator is considered as a totally formed electrically short antenna with a cosine distribution of the intensity of the excitation along the length of the vibrator.

Magnetic type AA consists of elementary magnetic radiators. Calculations of the specific values of the ratio between the absorbed power showed negligible difference (less than a few percent), unlike the data for the elementary sources (Table 17.1).
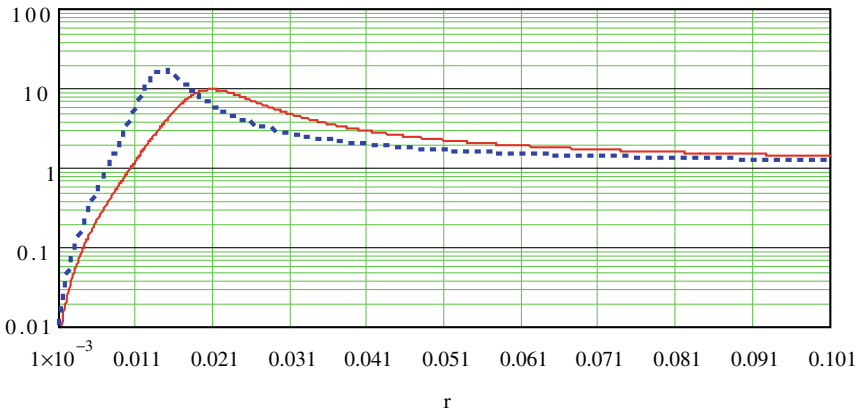
Fig. 17.1 The ratio of the absorbed power corresponding to the electric and magnetic radiators in a nonmagnetic medium type with the parameters $\varepsilon = 50$ and $\sigma = 1$ in the direction normal to the vibrator (in meters); for the frequency of 500 MHz (solid line) and 750 MHz (line in dots)

Table 17.1 Calculations of the specific values of the ratio between the absorbed powers

QD (50, 1, 31 × 10^8, r, 1.5)	Q (50, 1, 31 × 10^8, r, 1.5
0.922	0.910
9.623	9.468
5.148	5.137
3.050	3.048
2.301	2.301
1.94	1.94
1.732	1.732
1.597	1.597
1.503	1.503
1.435	1.435

The left column of Table 17.1 (QD) is the gain for the half-wave antenna; the right column (Q) is the gain for electrically short distances in the range of 0.01, …, 0.1 m; frequency 500 MHz, $\varepsilon = 50$ and $\sigma = 1$.

Thus, the estimates above are valid for the actual near-omnidirectional antennas of electric type i.e. dipole antenna and a magnetic type (slot antenna).

Unfortunately, in spite of evident advantages, magnetic type antennas could not be preferable for application. The reason is its impedance characteristics. The values of the input resistances of the electric and magnetic radiators of equal size are related by

$$Z_{mag} = \frac{Z_c^2}{Z_{elec}}, \tag{17.3}$$

where $Z_c = \sqrt{\frac{\mu}{\varepsilon - \frac{j\sigma}{\omega}}}$ is the characteristic impedance of the medium [6].

Since the active part of the input impedance of the half-wave dipole antenna in the air is of the order of 70 Ω, the slot antenna has 2 kΩ. In dissipative medium with higher impedance difference is partly levelled, but still remains significant. So, the achievement of a good antenna matching with the type of the magnetic type feeder ($Z_c = 50$ Ω) is not only difficult, but, also, possible in principle by less band width. Thus, the direct use of magnetic type antennas as a microwave applicator cannot be considered preferable.

Combined antennas, which joint features of the electric and magnetic radiators, are characterized by another important property: the reactive part of the input admittance has the opposite signs.

Therefore, at frequencies above the resonance frequency, there is a partial mutual compensation that provides improved matching of this antenna with a feeder in a wide frequency band. The largest effect occurs at equal ratio between "electric" and "magnetic" parts, which is achieved in the form of self-antenna structures [7], for example, in the form of endless four sectors 90°, two of which are metal and two are air.

Theoretically, the input impedance of such an antenna is purely active, constant, and equals Z_c that for typical biological environment is of the order of 50 Ω. Therefore, quite good results both in terms of the penetration depth and matching may be provided by flat dipole antennas, whose shoulders are formed as metal plates with a sector angle of about 90°.

Additionally, in biological media, such antenna broadband matching at frequencies from 500 to 1000 MHz provide:

- A significant (up to 1.5, …, 2.5 times) reduction of the absorbed power compared with dipole antenna in the surface layer thickness of up to 10, …, 15 mm;
- Increase in the specific absorption capacity of 75, …, 25% at a depth of about 20, …, 120 mm.

So, in this article we conducted research of this type of combined AA.

17.3 Results

17.3.1 Modelling the AA

Model calculations were carried out by the numerical hybrid method MOM/FEM (method of moments/finite element method) by the program of three-dimensional electromagnetic simulation FEKO (7.0) [8].

In this article, the model of the head is represented by a combination of plane-parallel dielectric structures. Electro-physical properties of these structures are defined by the corresponding characteristics of the scalp, cortical bone, cerebrospinal

Table 17.2 Geometrical and electro-physical parameters of the model

No. layer	Parameters of layers			
	Tissues	Thickness of layers, mm	ε	σ, Sm/m
1	Dry skin	2.0	42.316	0.81688
2	Cortical bone	0.5	12.606	0.12601
3	Cerebro-spinal fluid	2.0	69.006	2.355
4	Grey matter	5.0	53.558	0.87993
5	White matter	100.0	39.695	0.54592

Fig. 17.2 Flat wideband antenna
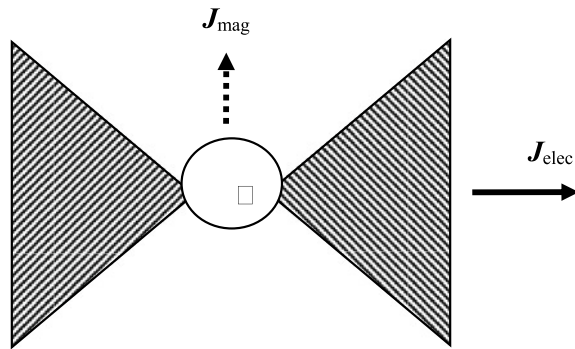
fluid, grey and white matter of the brain. Geometrical and electro-physical parameters of the structures are presented in Table 17.2 [9]. The simulation was performed in the frequency range 650–850 MHz (Fig. 17.2).

Among the variety of AAs, which are used to measure the microwave radiation of biological tissues, let us focus on one, the design of which is adapted to receive the radiation of brain tissue regardless of the size of the hair. For this purpose, in each vibrator, we housed conductive pins that provide contact of the antenna to the sculp. Figure 17.3 shows a version of the bow-tie AA with pins.

Dependence of the normalized power density EMF was built in the XZ cross-section (Fig. 17.4). Near the body-AA interface, the structure of the field is determined by the presence of the longitudinal electric field component E_r. Therefore, the maximum amplitude of the vibrator is directed along the axis.

Note that the field along the AA axis of vibrator is inversely proportional to the cube of the distance from the antenna. So, even the smallest heterogeneity between AA and reflection skin (such as hair, topography of the head, sweat drops and so on) can have an effect on $F(\Omega)$, and, therefore, on the reception properties of the AA.

Let us estimate the dependence of the EMF on the height of the pins. In Fig. 17.5, the EMF norm is intensity normalized to the value of the source field, and Z is length of the propagation.
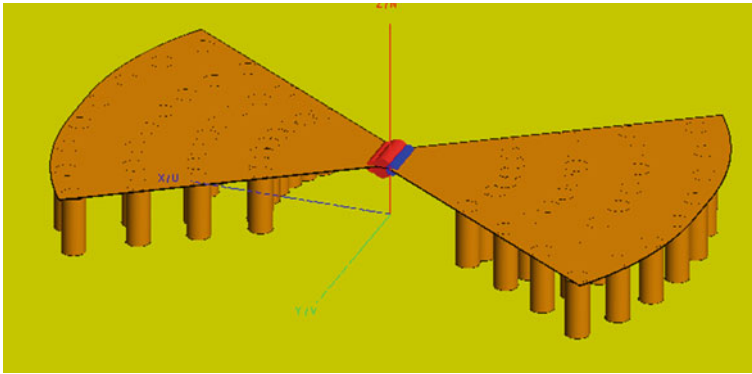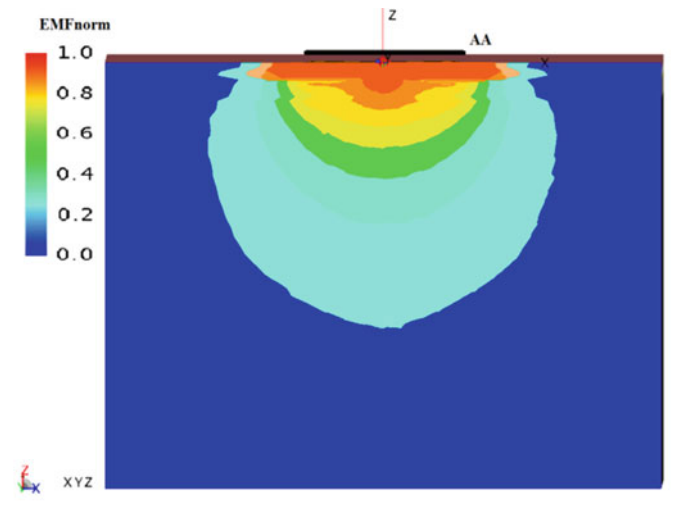
Fig. 17.3 Scheme of the modelling AA with pins

Fig. 17.4 Power density distribution EMF in the cross section XZ

The plots presented in Fig. 17.5 shows that the length of pins influences the field distribution in the reactive field region. This influence decreases in the transition region. An increase of the AA pins length reduces the characteristic impedance due to the formation of a thick vibrator (having a resistance of about 30–50 Ω) well-matched to the body. Thus, the presented scheme (Fig. 17.3) of the AA improves the accuracy of the brain microwave radiation measurement in functional studies.
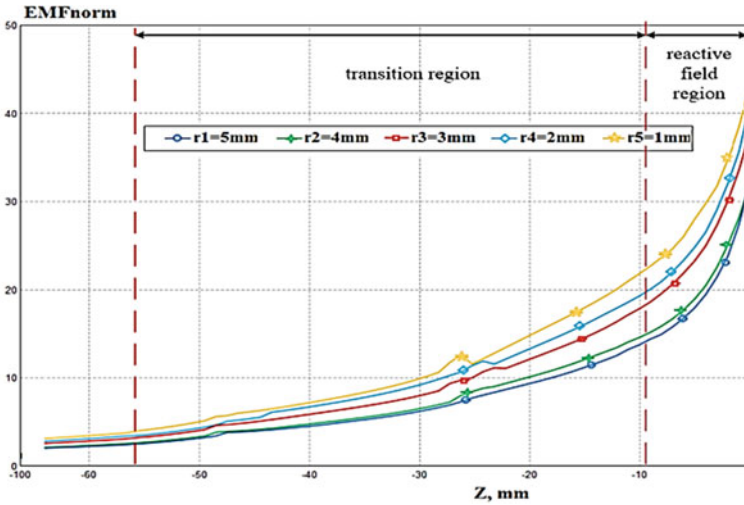
Fig. 17.5 Plots of the EMF norm propagation length for the various heights of the pins

17.3.2 Laboratory Results

The AA structure presented above (Fig. 17.3) was developed to decrease the influence of the mentioned inhomogeneities between the AA and head sculp. For this purpose, in each vibrator, we housed conductive pins that provide contact of the antenna to the sculp. A scheme of such modified AA is proposed in [10] and its prototypes AA are shown in Fig. 17.6. Such AA with pins allows one to provide good repeatability even considering hair on the head.

Laboratory tests of the AAs prototypes with and without contact pins were conducted by the vector analyser National Instruments PXIe-5630 [11]. Results of the tests are presented in Fig. 17.7.

Figure 17.7 shows that highest reflection coefficient changes during 30 min. The change (in modulus) for the AA without contact pins is 0.12. For the AA with contact pins, the largest change is 0.06.

17.4 Discussion and Conclusions

The results of modelling and laboratory studies suggest that as the conductivity of the tissues of the head becomes sufficiently high, the EMF decreases quickly and the characteristic impedance AA turns out to be sufficiently small.

Therefore, AA is better matched when the antenna is most closely adjacent to the surface of the head. This condition can be performed using pins, which are

Fig. 17.6 The prototypes AA without (**a**) and with (**b**) contact pins
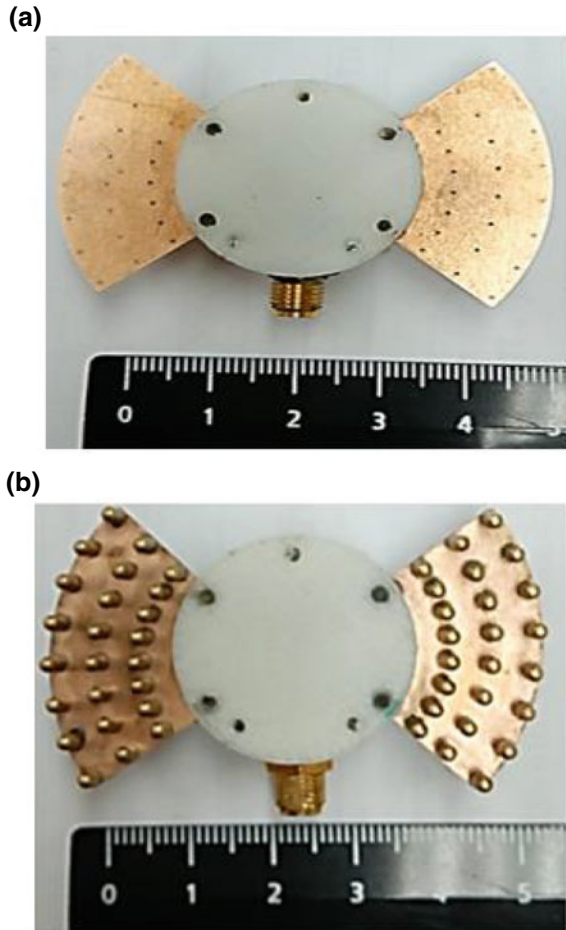
conditionally "thicken" vibrator. Simultaneously, the impact of hair on the matching of the AA and body is significantly reduced.

The results obtained allow us to draw the following conclusions:

1. A new design combined AA with contact pins is proposed. The methods of mathematical modelling of the investigated AA characteristics show that this AA can be used for long-term monitoring of functional processes in the brain tissue.
2. The results of the research are confirmed by laboratory tests of the AA prototypes. The experimental data confirm the possibility of using antennas combined with contact pins for monitoring the brain functional processes within 30 min. The variation coefficient of such an AA matching with brain tissues varies by less than half in comparison with the AA without pins.
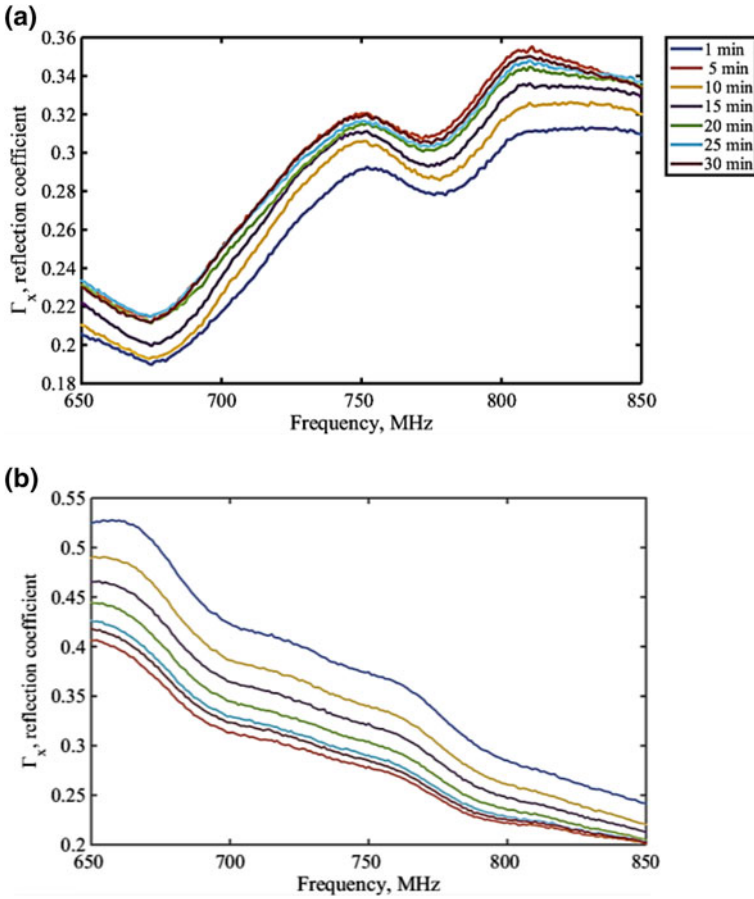
Fig. 17.7 Results of the laboratory tests, **a** AA with pins, **b** AA without pins

Thus, the use of combined AA with contact pins allows simplifying the maintenance of measurement invariance brain brightness temperature by matching AA-body and, consequently, improving the accuracy of measurement of the brain microwave radiation by monitoring functional processes in its tissues.

3. The use of pins is possible not only for the vibratory AA, but, also, for other types of AA. These studies will be conducted in the further work.

References

1. D.A. Vedenkin, O.V. Potapova, Y.E. Sedelnikov, Antennas, focused in the near radiated field zone. Features and technical application, in *Proceedings of 9th International Conference on Antenna Theory and Techniques*, (2013), pp. 560–565
2. F. Bardati, Time-dependent microwave radiometry for the measurement of temperature in medical applications. IEEE Trans. Microw. Theory Tech. **52**, 1917–1924 (2004)
3. S.G. Vesnin, M.K. Sedankin, Comparison of the microwave medical antennas. Biomed. Radioelectron. **10**, 63–74 (2012)
4. K.M. Luedeke, J. Koehler, J. Kanzenbach, A new radiation balance microwave thermograph for simultaneous and independent temperature and emissivity measurements. J. Microwave Power **14**, 117–121 (1979)
5. V.I. Petrosyan, Applicator antenna to resonance wave UHF/SHF radio spectroscopy natural formations. Biomed. Radioelectron. **8**, 36–42 (1999)
6. E.I. Nefyodov, S.M. Smolskiy, *Understanding of Electrodynamics, Radio Wave Propagation and Antennas*, (Scientific Research Publishing, 2013), 449 p
7. D. Valderas, J.I. Sancho, D. Puente, C. Ling, X. Chen, *Ultrawideband Antennas. Design and Applications*, (Imperial College Press, 2011), 194 p
8. D.B. Davidson, *Computational Electromagnetics for RF and Microwave Engineering*, (Cambridge University Press, 2005), 411 p
9. Italian National Research Council—Institute for Applied Physics, Dielectric Properties of Body Tissues in the Frequency Range 10 Hz–100 GHz, INRC (2012), http://niremf.ifac.cnr.it/tissprop/. Accessed 1 Sep. 2018
10. V.S. Kublanov, *Ustrojstvo dlya priema sobstvennogo radioteplovogo izlucheniya tela cheloveka*, [in Russian], RU Patent No. 2049424 (1995)
11. B.A. Panchenko, V.S. Kublanov, S.A. Baranov, V.I. Borisov, Y.E. Sedelnikov, Antenna for contact microwave radiometers for monitoring of the brain microwave radiation, in *Proceedings of International Applied Computational Electromagnetics Society Symposium*, (2017), pp. 1–2

Chapter 18
Multifractal Nature of the Brain Microwave Radiation Signals

Vasilii I. Borisov, Anton Yu. Dolganov and Vladimir S. Kublanov

18.1 Introduction

The Radiophysical complex MRTHR allows one to make a real-time record of the microwave radiation of the human brain and heart rate variability [1]. The brain microwave radiation is a result of the thermal Brownian motion. Therefore, it is of great interest to estimate the informational parameters using the methods of nonlinear dynamics based on concept of the fractional Brownian motion, namely, the theory of multifractal formalism [2].

In some studies, the possibility of the multifractal parameters estimation is described for functional investigations of long time series (TS) of the biomedical signals [3]. Also, the multifractal nature of heart rate variability (HRV) signals was shown for long functional studies [4].

Functional studies conducted with the MRTHR complex have a time limit of 15–30 min. This limit is defined by the technological features of the complex including the volume of the screened cabin [5].

Previously, the possibility of application of the model and real biomedical TS with the length of more than 1024 samples was shown [6]. This requirement is also satisfied for 5-min studied signals of the microwave radiation of the human brain.

In order to prove the possibility of fractal characteristic assessment for the physiologically significant time intervals of the considered biomedical signals one should evaluate the changes of the multifractal estimations for different time-frequency boundaries of considered TS.

V. I. Borisov · A. Yu. Dolganov · V. S. Kublanov (✉)
Ural Federal University, Yekaterinburg, Russia
e-mail: kublanov@mail.ru

V. I. Borisov
e-mail: vi.borisov.official@gmail.com

The aim of this article is to define possibilities of the multifractal detrended fluctuation analysis (MFDFA) method for estimating the fluctuation of the regulation processes of human brain based on TS recorded by two channels of the microwave radiation of the parietal areas of human brain.

18.2 Methodology

The algorithm and application features of the MFDFA method for the estimation of short-term TS are described in details in [7].

Firstly, the data points of the original biomedical signals are uniformly interpolated with the spline-interpolation method [8].

The main steps of the method include:

- the detrending procedure with second degree polynomial on non-overlapping segments

$$y_v(i) = \sum_{k=0}^{m} C_k i^{m-k};$$

(18.1)

- determination of the fluctuation functions

$$F_q(s) = \{\frac{1}{N_s} \sum_{v=1}^{Ns} [\frac{1}{s} \sum_{k=1}^{s} [y\{(v-1)s+i\} - y_v(i)]^2]^{q/2}\}^{1/q};$$

(18.2)

- estimation of the slope exponent in log-log plot

$$\log_2 F_q(s) = h(q) \cdot \log_2 s + Const;$$

(18.3)

- calculation of the scaling exponent

$$\tau(q) = q \cdot h(q) - 1;$$

(18.4)

- Legandre transform application to estimating the probability distribution of the spectrum $D(\alpha)$

$$D(\alpha) = q \cdot \alpha - \tau,$$

(18.5)

where $\alpha = \frac{d\tau}{dq}$ is the Hölder exponent.

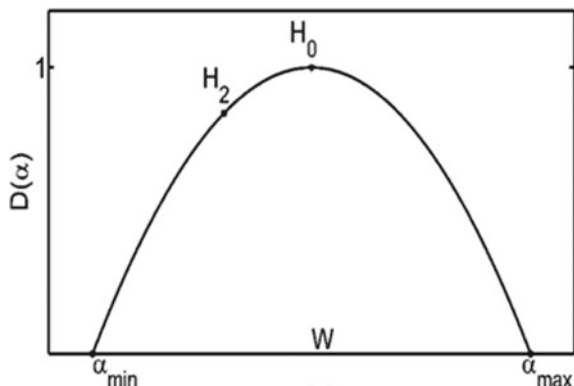**Fig. 18.1** The characteristic
values of the multifractal
analysis

Figure 18.1 represents the main parameters of the multifractal spectrum estimated by the MFDFA method [9].

The borders of the spectrum are defined as follows:

$$\frac{d\tau}{dq}\Big|_{q=-5} = \alpha_{\min}; \quad \frac{d\tau}{dq}\Big|_{q=5} = \alpha_{\max}; \tag{18.6}$$

where α_{min} represents the behavior of the smallest fluctuations in the spectrum, while α_{max} represents the greatest fluctuations.

In this case, the width of multifractal spectrum can be written as

$$W = \alpha_{\max} - \alpha_{\min}. \tag{18.7}$$

Here, W shows the variability of fluctuations in the spectrum.

The height of the spectrum $H_0 = \alpha|_{q=0}$ represents the most probable fluctuations in the investigated time window of the signal.

The generalized Hurst exponent (also known as correlation index) is defined as

$$H_2 = \alpha|_{q=2}. \tag{18.8}$$

For the study of the informational patterns of the microwave radiation of the human brain, the following time windows were chosen [10]: 1–10, 10–20, 20–30, 30–40, 40–50, 50–60, 60–70, 70–80, 80–90, 90–100 s. The lower limit is defined by the influence of the interpolation noise; the length N of TS defines the upper limit by the $N/3$ ratio [9].

As a test material, we use the data recorded in the Sverdlovsk Clinical Hospital of Mental Diseases for Military Veterans (Yekaterinburg, Russian Federation). The test group consisted of 20 psychologically healthy patients-volunteers with age of 18–20 years.

The research was implemented in two functional states: before the load (F, i.e., the functional rest) and during the passive orthostatic load (O, i.e., the functional load). Signals in both states were registered for approximately 5 min (300 s) with the modernized Radiophysical complex MRTHR [2, 5].

18.3 Results and Discussion

In Tables 18.1 and 18.2, the mean values of difference of the multifractal parameters between data of two channels of the microwave radiation in the functional states F and O are presented, respectively.

Table 18.1 Results of the statistical deviation multifractal parameters of the left and right channel under the state F

Time window(s)	H_2	H_0	W	α_{min}	α_{max}
1–10	−0.03	−0.17	0.80	−0.05	0.75
10–20	−0.02	−0.04	0.01	0.01	0.02
20–30	−0.01	−0.02	−0.14	−0.01	−0.19
30–40	−0.03	−0.08	−0.04	−0.02	−0.09
40–50	−0.08	−0.07	−0.02	−0.10	−0.08
50–60	−0.07	−0.04	−0.38	0.03	0.03
60–70	0.03	−0.10	−0.34	−0.01	−0.30
70–80	0.01	−0.04	−0.11	0.02	0.07
80–90	−0.10	−0.04	−0.12	0.13	0.39
90–100	−0.07	0.01	−0.43	−0.04	0.12

Table 18.2 Results of the statistical deviation multifractal parameters of the left and right channel under the state O

Time window(s)	H_2	H_0	W	α_{min}	α_{max}
1–10	−0.08	−0.16	−0.54	−0.23	−0.77
10–20	0.01	0.01	0.01	−0.01	0.01
20–30	0.02	−0.01	0.06	0.05	0.09
30–40	0.01	0.02	0.11	0.02	0.25
40–50	−0.11	−0.06	−0.08	0.06	−0.19
50–60	−0.08	0.04	−0.09	0.02	0.01
60–70	−0.01	−0.25	−0.43	0.11	−0.26
70–80	−0.21	−0.11	−0.34	0.31	0.04
80–90	−0.11	−0.06	−0.01	−0.23	−0.16
90–100	−0.42	−0.11	−0.33	0.40	−0.13

The values presented in the tables show several statistically independent time windows for multifractal assessment. The results, also, show that on average, the difference between H_2 and H_0 in the functional rest is minimal and less than difference between W and α_{max}. This allows one to use H_2 and H_0 as the markers of alteration processes during the functional load.

In previous studies, in accordance with the Guliaev and Godik hypothesis about parametric modulation of the own microwave radiation of the biophysical and biochemical processes in human body [11], it was shown that fluctuations of radiation with periods of 10–70 s have a physiological nature [12]. The results of multifractal analysis correspond to this hypothesis.

The multifractal analysis theory shows that the Hurst exponent has a critical value of 0.5, which represents the randomness of these processes [13]. For $H_2 > 0.5$, the TS are persistent. The increment of the persistent TS are is likely to keep trend of fluctuational changes. For $H_2 < 0.5$, the TS are anti-persistent. In Fig. 18.2, the values of H_2 for two functional states for one of the channel are given for mentioned time windows. The X-axis represents H_2 in the state F, and Y-axis represents H_2 in the state O.

The presented results show that:

- signals of the microwave radiation with 1–10 s are noisy;
- the periods of 10–40 s and 60–70 s have the lowest difference for both states and represent the anti-persistent nature;
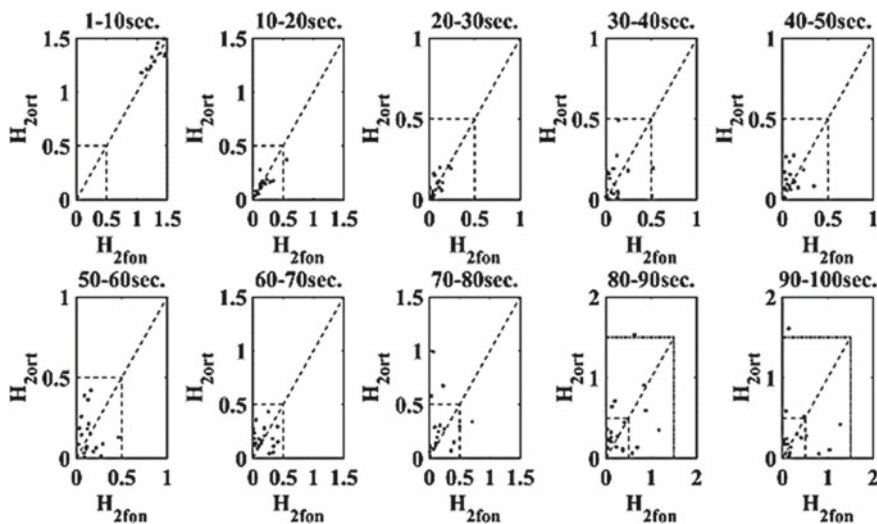- the periods higher than 70 s have mixed multifractal nature.

Fig. 18.2 Values of H_{2fon} and H_{2ort} for 10 time scale boundaries

Earlier, we have shown that the frequency spectrum of these fluctuations ranges from 0.05 to 0.025 Hz. This spectrum mainly reflects changes of the dielectric permeability in tissues in depth more than 10 mm and is a consequence of the humoral processes. In the band of frequencies below 0.025 Hz, the intensity of fluctuations of the brain microwave radiation is defined by the thermodynamic changes in its tissues that are stimulated by the metabolic processes [10, 12].

Based on the mentioned frequencies, the following time scale boundaries were chosen for statistical significance estimations of multifractal properties of the microwave radiation: $\delta(\varepsilon)$ as (20–40) s and T_{td} as (60–70) s.

The Bland-Altman criterion was used for comparison of multifractal estimates and obtained statistical significance of calculated multifractal characteristics for different time scale boundaries of the microwave radiation signals under two functional states [14]. Differences of H_2 and H_0 between two channels of the microwave radiation signals in scale boundaries $\delta(\varepsilon)$ and T_{td} in two functional states were considered.

Figures 18.3 and 18.4 show the results of the statistical significance of the indicators H_2 and H_0 under the states F and O, respectively.

The results obtained by the Bland-Altman criterion show that deviation of the multifractal parameters of the two channels has a statistical significance not less than 95%. The resulting estimates justify the limits of applicability of the MFDFA method for time boundaries from 10 to 70 s. This corresponds to two independent fluctuation periods of the microwave radiation of the brain, which have different physiological nature [15].
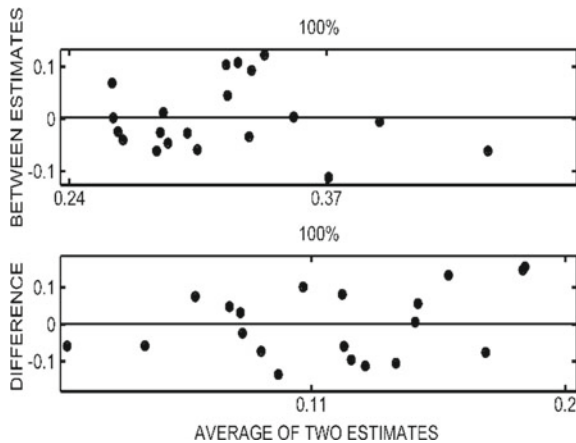

Fig. 18.3 The Bland-Altman plot for value calculated in $\delta(\varepsilon)$ (top) and T_{td} (bottom) boundaries
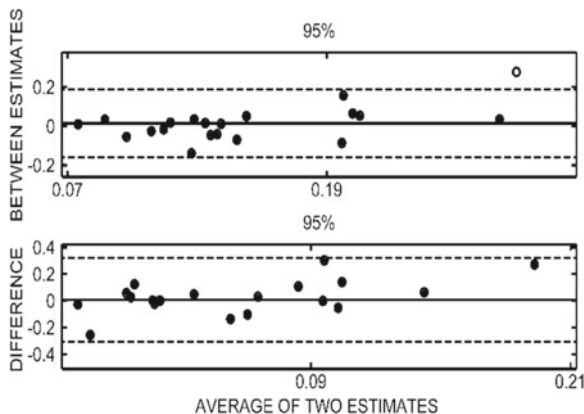
Fig. 18.4 The Bland-Altman plot for value H_2 calculated in $\delta(\varepsilon)$ (top) and T_{td} (bottom) boundaries

18.4   Conclusions

The application of the multifractal analysis to functional studies of the microwave radiation signals of the brain increases the possibilities of the modernized Radio-physical complex MRTHR for comparing the processes of its two information channels.

The results show that the time series of the microwave radiation signals of the brain have multifractal nature.

The obtained results allow one to conclude that there are time boundaries of the microwave radiation signals, which have invariant multifractal parameters during execution of functional loads for the group of relatively healthy patients. Clinical and physiological interpretation of each band separately or together requires additional knowledge from both clinical and other physiological systems.

Regarding the parameter H_2, obtained estimates show that a signal in different time boundaries has different behaviour of multifractal nature. This allows one to decrease the top time scale boundary to 70 s for signals of this type.

It was shown that the results obtained by the multifractal analysis are consistent with estimates obtained by other methods and researchers. The statistical significance of these results is high.

In other cases, nosology, functional, stress testing field periods of fluctuations of these signals (that reflect diagnostically relevant information in accordance with the patient's health status) may differ from those obtained in the present study.

References

1. V.S. Kublanov, Microwave radiation as interface to the brain functional state, in *Proceedings of the International Conference on Biomedical Electronics and Devices BIODEVICES*, Barcelona (2013), pp. 318–322
2. B.B. Mandelbrot, *Fractal geometry of Nature*, [in Russian], Institute of Computer Science: Moscow, 656 p. (2002)
3. D. Makowiec, A. Dudkowska, R. Gałąska, A. Rynkiewicz, Multifractal estimates of monofractality in RR-heart series in power spectrum ranges. Physica A: Stat. Mech. Appl. **388**(17), 3486–3502 (2009)
4. A.L. Goldberger, L.A.N. Amaral, J.M. Hausdorff, PCh. Ivanov, C.-K. Peng, H.E. Stanley, Fractal dynamics in physiology: alterations with disease and aging. Proc. Nat. Acad. Sci. USA **19**, 2466–2472 (2002)
5. V.S. Kublanov, Radiophysical system for examining functional state of a patient's brain. Biomed. Eng. **43**(3), 114–119 (2009)
6. V.S. Kublanov, V.I. Borisov, S.V. Porshnev, Features of using nonlinear dynamics methods for heart rate variability analysis. Biomed. Radioelectron. **8**, 30–37 (2014)
7. E.A.F. Ihlen, Introduction to multifractal detrended fluctuation analysis in Matlab. *Frontiers in Physiology*, **3**(141) (2012)
8. C.A. De Boor, *Practical Guide to Splines* (Springer: Berlin, 1978), p. 1978
9. J.W. Kantelhardt, (ed. Meyers, R.A.), *Fractal and Multifractal Time Series, Encyclopedia of Complexity and Systems Science* (Springer: New York, 2009) p. 10398
10. A.M. Syskov, V.I. Borisov, V.B. Parashin, V.S. Kublanov, Information analysis of radio brightness temperature fluctuations in brain tissues. Biomed. Eng. **46**(3), 100–103 (2012)
11. E.E. Godik, Y.V. Gulyaev, Functional imaging of human body. IEEE Eng. Med. Biol. **4**, 21–29 (1991)
12. V.S. Kublanov, JuE Sedelnikov, A.L. Azin, A.M. Syskov, The nature of fluctuations own electromagnetic radiation of the brain. Biomed. Radioelectron. **9**, 45–54 (2010)
13. Feder, J., *Fractals* (Plenum Press: New York 1989), p. 247
14. J.M. Bland, D.G. Altman, Statistical methods for assessing agreement between two methods of clinical measurement. Lancet **1**(8476), 307–310 (1986)
15. V.S. Kublanov, V.I. Borisov, T.G. Kopytova, Peculiarities of spectral and multifractal estimates of the brain microwave radiation. IFMBE Proc. **65**, 619–622 (2017)