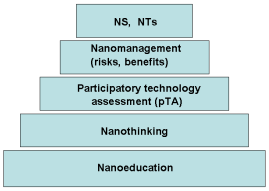
INNOVATIVE INFOTECHNOLOGIES FOR SCIENCE, BUSINESS AND EDUCATION

Vol. 1(10) 2011

CONTENTS

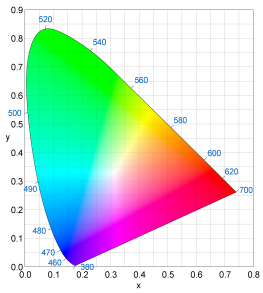
IITSBE, **Vol. 1(10)2011**

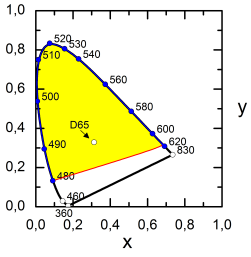
Section: New concepts



Section: Information Systems

Section: Advanced Programming





Nanothinking as an educational concept of the 21st century

Tamara Lobanova-Shunina, Yurij N. Shunin [a]
Department of Natural Sciences and Computer Technologies
Information Systems Management Institute, Ludzas 91, LV-1019, Riga, Latvia

Received 11 January 2011, accepted 1 February 2011

Abstract. Around the world, the pace, complexity and social significance of technological changes have been increasing. Striking developments in such areas as computer and communications technology, biotechnology and nanotechnology are finding applications and producing far-reaching effects in all spheres of business, government, society and the environment. However, the far-reaching social consequences are often not understood until after new technologies become entrenched. Historically this has resulted in important lost opportunities, significant social and environmental costs and channeling societal development down long-term unhealthy paths.

Citations: Tamara Lobanova-Shunina, Yurij N. Shunin. Nanothinking as an educational concept of the 21st century – *Innovative Infotechnologies for Science, Business and Education,* ISSN 2029-1035 – 1(10)2011 – Pp. 3-9.

Keywords: Nanothinking; Nanoeducation; Nanorisks.
Short title: Nanothinking and Nanoeducation.

Introduction

One decade into the 21st century, people and governments worldwide face decisions on daily basis involving complex scientific considerations or innovations in technology. Decisions small and large – whether they are policy-makers' votes on a climate bill, biotech corporations' considerations of potential product lines, consumers' choices of food purchases or educators' use of computers in the classroom – must incorporate a dizzying array of factors. The new participative democracy demands that citizens be asked to make judgments, and even vote, on subjects about which they know very little – the desirability of cloning animals and human beings, creating novel biological organisms, manipulating matter at an atomic scale, nuking your enemies, eugenics, genetic engineering (GE), genetically modified (GM) foods, nano-products, and other great moral and economic questions of the day. Therefore, educational systems have to produce a steep increase in citizens' intellectual potential in order to provide sane answers to such deep philosophical questions, previously the domain of university researchers.

Educational environment is becoming a new supercomplex system with a constantly changing intellectual pattern. It has been predicted that today's school-leavers will have many careers – not just jobs, over their lifetimes, and that more than 50 % of the jobs they will be doing do not yet exist. But one thing is certain – they will be knowledge jobs, intellectually more demanding and almost certainly involving interaction with technologies far more sophisticated than those existing at present. Mindpower is replacing manpower.

Still, the structure of our universities has changed very little in the past fifty years; they are still organized in the traditional fields with little or no horizontal structures. In materials science, as in many other fields, much of the most exciting discovery potential is located at the boundaries between traditional disciplines. Already today, many novel multifunctional nanomaterials, advanced nanodevices, new nano-based products and processes are designed and developed by team efforts of materials scientists working with chemists, biologists, physicists, information technology experts, and engineers. It is thus apparent that we need to create new types of universities, so called virtual universities, which have 'departments without walls' [1-3].

1. Nanotechnology as the imperative for educational redesign

Rapid technological changes have dramatically altered our educational needs. The simplest explanation for the current need of educational change is that we, as society, have outgrown our educational systems disseminating core knowledge and building basic skills. With the advent of the information age, and now the beginning of new technologies age, the educational model of today no longer meets our societal needs. In fact, it is limiting the ability of teachers and students to adapt to the 21st century.

Nanotechnology is an exciting area of scientific research and development that is truly multidisciplinary. Nanotechnology originates from the Greek word *nano* which means *dwarf*. A nanometer is one billionth (10^{-9}) part of a meter, which is tiny, only the length of ten hydrogen atoms, or about one hundred thousandth of the width of a hair!

Nanotechnology is not really anything new. In one sense, it is the natural continuation of the miniaturization revolution that we have witnessed over the last decade.

[a]Corresponding author, email: *jurijs.sunins@isma.lv*

It is necessary to point out that millionth of a meter (10^{-6} m) tolerances in engineered products have become commonplace. A good example of the application of nanotechnology is a mobile phone, which has changed dramatically in a few years – becoming smaller and smaller, while paradoxically growing cleverer and faster – and cheaper!

What is new, though, is the multidisciplinary approach and the ability to 'see' these entities and to manage them. Although scientists have manipulated matter at the nanoscale for centuries, calling it physics or chemistry, it was not until a new generation of microscopes were invented in the late 1980s in IBM, Switzerland that the world of atoms and molecules could be visualized and managed. Now biologists can discuss steric effects of cell membranes with chemists, while physicists provide the tools to watch the interaction *in vivo* - infrared (IR) microscopes to study molecular systems up to single molecules and and X-ray microscopes to study atomic structures and to handle even single atoms.

In simple terms, *Nanoscience* can be defined as the study of phenomena and manipulation of materials at atomic, molecular and macromolecular scales, in order to understand and exploit properties that differ significantly from those on a larger scale. It is not really a new field, but a different way of looking at all fields. Its development will require the expertise of all scientists – from engineers to ecologists. *Nanotechnology* can be defined as design, engineering, production and application of structures, devices and systems by controlling shape and size at a nanometer scale..

A concise definition is given by the US National Nanotechnology Initiative: 'Nanotechnology is concerned with materials and systems whose structures and components exhibit novel and significantly improved physical, chemical, and biological properties, phenomena, and processes due to their nanoscale size. The goal is to exploit these properties by gaining control of structures and devices at atomic, molecular and supramolecular levels and to learn to efficiently manufacture and use these devices' [4]. This term can be applied to many areas of research and development – from medicine to manufacturing, to renewable energy, transport, computing, and even to textiles and cosmetics.

At the nanoscale, the properties of a material may change. For example, hardness, electrical conductivity, thermal conductivity, colour or chemical reactivity of minuscule particles of materials are related to the diameter of the particle. They demonstrate new and unusual properties that are not obvious in the bulk material.

This is because a nanoparticle has a large surface area in relation to its size, and is consequently highly reactive. This is exemplified by the fine grained materials that we use in our daily lives, such as flour, which can become explosive in some circumstances.

Specific functionalities, therefore, can be achieved by reducing the size of the particles to $1 \div 100$ nm. Particles at the nanoscale are below the wavelength of visible light, and therefore cannot be seen. It can be difficult to think of and imagine exactly the invisible world of atoms and molecules to get a greater understanding of how it will affect our lives and the everyday objects around us.

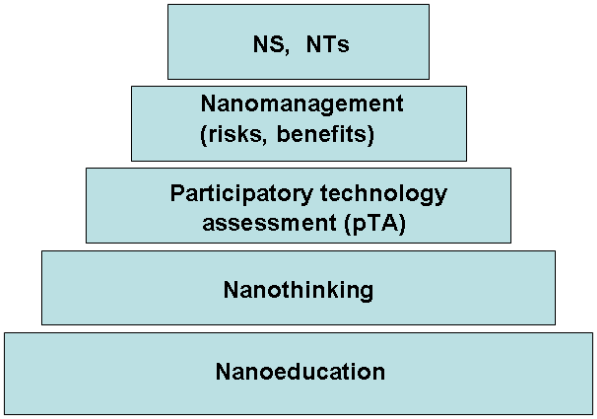


Fig. 1. Nanochallenges hierarchy.

But the areas where nanotechnologies are set to make a difference are expanding alongside with the challenges they pose to society. Challenges in nanotechnologies can be presented in their hierarchical priorities (Fig.1).

Nanochallenges comprise such basic areas as: i) nanoeducation; ii) nanothinking; iii) participatory technology assessment (pTA); iv) nanomanagement (incorporating risks and benefits).

2. Nanoeducation

During the past 10 years, we have seeded many ideas into the global consciousness to stimulate preparing our students for their future. The world is changing but our education matrix remains in the Industrial version of reality. We are not even close to understanding, nor preparing our students for these major changes they will face in the next few decades. *Nanoeducation* - is the new foundation for a new way of thinking, for the integration of all disciplines to expand our next generation students' knowledge base and prepare them for a very different future in a global society enhanced by all of the integrated science research now in process.

The rise of a highly networked global knowledge economy is changing the interface between scientists, researchers and the general public as consumers of new technologies, new materials and devices. Nanoeducation has to contribute to closing the gap between public rhetorical value and nanotechnologies practice on decision-making policy.

Many companies throughout Europe and the world report problems in recruiting the types of graduates they need, as many graduates lack the skills to work in a modern economy. For Europe to continue to compete alongside prestigious international institutions and programmes on nanomaterials, it is important to create educational institutions which would provide a top-level education and the relevant skills mix and would cover education, training, sciences and technologies for research and have strong involvement by European industry. The elements for such a high level education are supposed to be as following:

i) multi-disciplinary skills;
ii) top expertise in nanomaterials science and engineering;
iii) literacy in complementary fields (physics, chemistry, biology);
iv) exposure to advanced research projects;
v) literacy in key technological aspects; exposure to real technological problems;
vi) basic knowledge in social sciences, culture, management, ethics, foreign languages;
vii) literacy in neighbouring disciplines: international business, law, IT, etc;
viii) interlinkages between education, research and industrial innovation: students will be ready for what research and development will provide;
ix) sharing of post-docs, PhD and MS (masters) students to foster the mobility of permanent researchers and professors between different institutions to create 'team spirit'.

Companies, universities, governments, research organizations and technical societies must all strive to define their roles in this partnership. The 'output' will be graduates with a new way of thinking, skillful manipulators, synthesizers and creators of new knowledge excellently equipped to solve future complex problems and to work collaboratively.

3. Nanothinking

Data saturation that accompanies the 'new technologies age' has fostered an ever-increasing interdependency between people. The pace of expected adaptation is accelerated to a pace that exceeds individuals' abilities to accommodate. Being on the receiving end of technologies deluge serves to undermine people's confidence and sense of personal responsibility giving rise to the sense of helplessness that many people feel as the world enters the 'age of interdependency'. Nanothinking can serve as the antidote to the sense of helplessness since it is a concept for seeing the 'structures' that underlie complex processes, for a much better understanding how our organism works, and for discerning how to foster health, safety and the surrounding environment. If we do not understand ourselves, we will not be able to change our life for the better.

Nanothinking is a comprehensive systems thinking which offers a language that begins by restructuring the way how we think. It is a dynamic concept where practitioners continually engage in a process of 'seeing wholes' – a perspective that pays attention to the interrelationships and patterns of influence between constituent parts to foster the dissolution of compartmentalization of science and the corresponding compartmentalization of the mind. Nanothinking can be defined as the understanding of a nanophenomenon within the context of a larger whole. *To think nanoscalely* – means to put things into a nanoscale context and to establish the nature of their relationships.

Contemporary top-level education envisions causing students think systemically – integrating not only macro-, micro- but also the nano- scale. Nanothinking can be defined as 'visualizing matter, structures and processes at the nanoscale'.

Public thinking can be formed and improved through sustained and carefully crafted dialogue, which has to be integrated into educational communication practice. Educational communication has to contribute to developing a new way of thinking – the systemic thinking, with the main strategy – 'how to think' rather than 'what to think'. It is the privilege of a liberal university not to give the right answers to students but to put the right questions.

Educational communication, as human communication in general, can be defined (according to a German sociologist Niklas Luhman) in terms of interactive construction of meaning/thinking. Thus, it can be presented as the unity of three components:

i) *information* - provided by teachers possessing knowledge;
ii) *utterance* - by means of language;
iii) *understanding* - a kind of created 'identical' thinking.

To this unity is added the acceptance or rejection of the receiver to continue communication and interaction (Fig. 2).

Anthony Giddens, a famous British sociologist, points out that people are always to some extent knowledgeable about what they are doing. Because people are reflexive and monitor the ongoing flow of information, activities, and conditions, they adapt their actions/ways of thinking to their evolving understanding.

As a result, knowledge changes human activities/ways of thinking, thus, shaping our consciousness (Fig.3). Language, in this respect, can act as a constraint on action/way of thinking, but at the same time, it also enables action by providing common frames of mutual understanding [1].

Consciousness is not inherited or static. It rather becomes a reflective project - an endeavour, which we continuously work out and reflect on. It is not a set of observable characteristics of a moment, but becomes an account of a person's life.

4. Participatory technology assessment (pTA)

The development of a new way of thinking envisions bringing the practice of *participatory technology assessment* (pTA) into alignment with the realities of the 21st century technology – to create a 21st century educational model.

The ability to create novel biological organisms, manipulate matter at an atomic scale, or intervene significantly (and possibly irreversibly) in the earth's climate system raises a host of ethical, social, legal and environmental questions that will require broad public discourse and debate.

Scientists and researchers engaged in nanoscience and nanotechnology research and development constitute a relati-

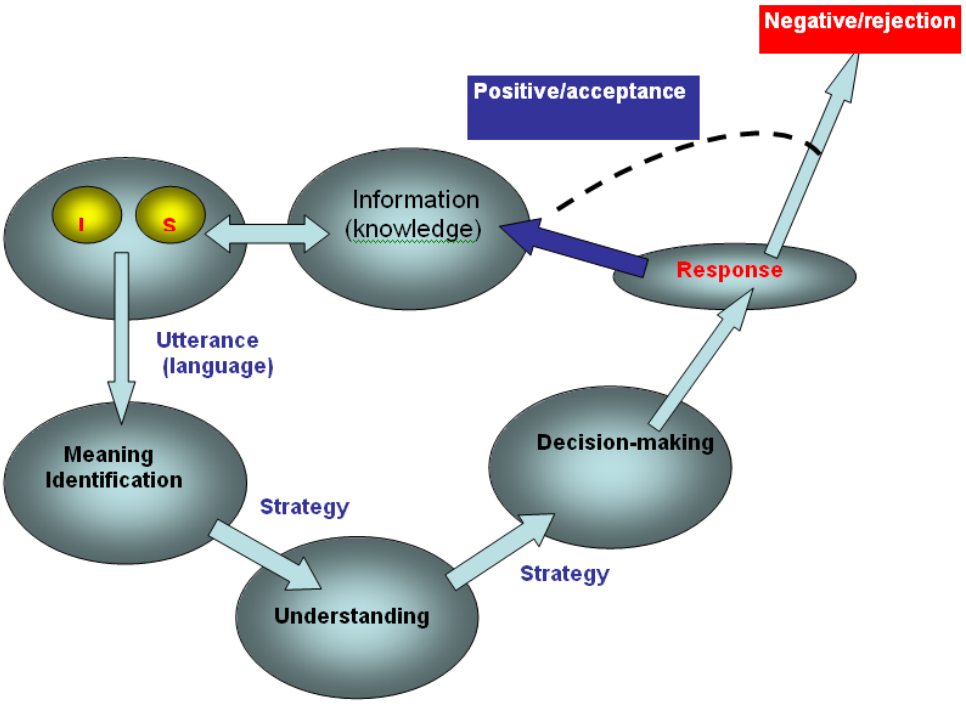


Fig. 2. Language and understanding as major determinants for shaping consciousness.

vely small group compared to the general public. However, the outcome of their work – innovative materials, devices and technologies have a strong impact on the life of the whole human society.

Nanotechnology applications are being developed in nearly every industry, including electronics and magnetics, energy production and storage, information technology, materials development, transportation, medicine and health. There are currently more than 600 consumer nano-products incorporating engineered nanoparticles on the market including food and beverages, dental fillers, toothpaste, optics, electronics, clothing, wound dressings, sporting goods, dietary supplements, and cosmetics.

In the future, mechanical 'microbes' injected into an organism may be able to combat disease-causing bacteria and viruses, remove cancerous cells or dispense medicines. Microscopic robots may be able repair, or even assemble complex devices or remove harmful substances from the environment.

Technology assessment (TA) is a practice intended to enhance societal understanding of the broad implications of science and technology. This creates the possibility for citizens of the world of preparing for – or constructively influencing – developments to ensure better outcomes.

Participatory technology assessment (pTA) enables the general public/laypeople, who are otherwise minimally represented in the politics of science and technology, to develop and express informed judgments concerning complex topics, as well as, to make informed choices.

Since applications of nanotechnology will quickly penetrate all sectors of life and affect our social, economical, ethical and ecological activities, the general public's acceptance is compulsory for further developments in the field of nanotechnology and its applications. This acceptance will be influenced by the low level of public awareness of many innovations in science, and especially, in nanotechnology. This is mainly due to the unpredictability of their properties at the nanoscale and the fragile public confidence in technological innovation and regulatory systems.

Consequently, it is of the utmost importance to educate the public, and to disseminate the results of nanotechnology development in an accurate and open way so that the general public will eventually transform their way of thinking to accept nanotechnology. In this endeavour, educational institutions have a pivotal role in developing pTA practices by following factors:

i) educating public (pupils, students) about science and technology;
ii) informing the public about the benefits and risks of nanomaterials and nanoproducts;
iii) evaluating, minimising, and eliminating risks associated with the manufacturing and use of nanomaterials and nanotechnology enabled products (risk assessment);
iv) exchanging with public authorities for the risk management of nanotechnologies.

In the process, pTA deepens the social and ethical analysis of technology, complementing the expert-analytic and stakeholder-advised approaches. The Internet and interactive TV capabilities can help pTA be more effective and cost-efficient and would also align with the policy-makers' initiatives to make them more transparent, accessible and responsive to popular concerns.

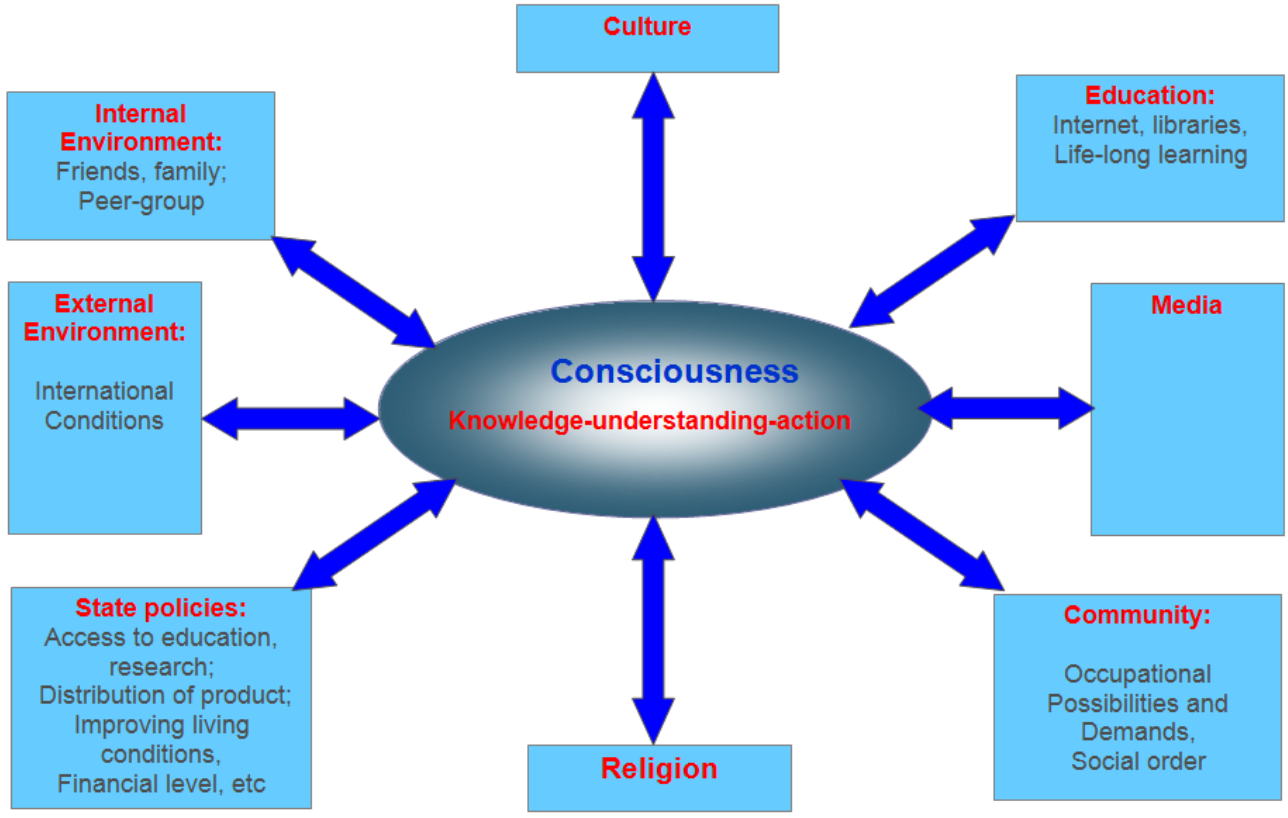


Fig. 3. Factors influencing consciousness development.

5. Nanomanagement

Nanotechnology is a radically new approach to manufacturing. It will affect so many sectors that failure to respond to the challenges will threaten the future competitiveness of a large part of the economy.

As nanotechnology has emerged from the laboratory into industrial manufacture and commercial distribution, the potential for human and environmental exposure, and hence risk, has become both reality and priority.

The research into health, safety and the environmental implications of nanotechnology lacks strategic direction and coordination. As a result, researchers are unsure about how to work safely with new nanomaterials, nano-businesses are uncertain about how to develop safe products, and public confidence in the emerging applications is in danger of being undermined.

Nanotechnology presents both an unprecedented challenge and unparalleled opportunity for risk management. Existing risk management principles are inadequate, given pervasive uncertainties about risks, benefits and future directions of this rapidly evolving set of technologies. The health implications of nanoparticles are unknown, the ramifications may be profound, and only a lengthy and extensive research effort can assess the safety implications with any certainty.

Yet the public, driven by heuristics such as *Affect and Availability*, is likely to stigmatize and reject this technology unless effective and credible risk management processes can be put into practice quickly. Because traditional command and control regulation will be unable to fill this need, innovative approaches that are incremental, flexible and decentralized should be developed to fill the risk management gap.

Scientific and technological innovation now requires accompanying innovations in management mechanisms that place an emphasis on public engagement. In its turn, public policy has to be grounded on understanding the risks and benefits of new technologies to have practical impact on decision-making.

One of the most pertinent examples of a multi-stakeholder approach to voluntary nanotechnology regulation is the Foresight Institute, which was organized explicitly to provide a forum for public discussion of the risks and benefits of nanotechnology and to 'pave the way' for its societal acceptance. Institute members include scientists, engineers, business people, investors, ethicists, policy makers and lay people as well as firms. Thus, the organization represents a broad spectrum of stakeholders, interests and opinions to be at the forefront of public discussions of nanotechnology risks and benefits [2].

Some engineered nanoparticles, including carbon nanotubes (CNT), although offering tremendous opportunities also may pose risks which have to be addressed sensibly in order that the full benefits can be realized. We have all already learned how to handle electricity, gas, steam and even cars, aeroplanes and mobile phones in a safe manner because we

need their benefits.

The same goes for engineered nanoparticles. Mostly they will be perfectly safe, embedded within other materials, such as polymers. There is some possibility that free nanoparticles of a specific length scales may pose health threats if inhaled, particularly at the manufacturing stage. Industry and government are very conscious of this, are funding research into identifying particles that may pose a hazard to health or the environment, and how these risks may be quantified, and minimized over the whole lifecycle of a given nanoparticle.

There is no doubt that nanotechnology has great potential to bring benefits to society over a wide range of applications, but it is recognized that care has to be taken to ensure these advances come about in as safe a manner as possible.

We need to manage nanotechnologies making our life more intellectual, comfortable and safe.

6. Bringing the spirit of nanotechnology into the classroom: Pilot study

With the aforementioned in mind, we launched a pilot study at Information Systems Management Institute (ISMA, Riga, Latvia) in different groups of students comprising Information Technologies, Management, Tourism, and Design departments as well as international students enrolled in ISMA on the ERASMUS student exchange programme.

We have undertaken a set of researches into the nature of students' intellectual potential development in order to elicit their general knowledge of some basic scientific notions and their understanding of the utilitarian value of some scientific phenomena. The study envisioned providing the necessary knowledge, understanding and support to our students to be successfully introduced to the technologically empowered environment of today's life, to adjust and adapt in it.

The purpose of the pilot study was primarily to work with the delivery of the questionnaires and interview questions to determine what was required to elicit the quantity and quality of data needed to respond powerfully to the research question. As a result of four pilot undertakings – a fluid conversation with students, an interview, a questionnaire with a feedback analysis – a level of intimacy and trust was created that supported the gathering of quantity and quality data.

Our mission has a focus on introducing nano science curriculum into classrooms. In order to encourage students and teachers to understand the importance of this scale of science, they need to see that *size matters* in the unseen world of nature. This introduction to the *unseen size of nature* can stimulate curiosity and a desire to learn more about their world through study with advanced microscopes that lead to an interest in chemistry, biology, physics, information technologies and other sciences.

The results of the study make us conclude that students' general knowledge of basic disciplines is rather restricted, sometimes rather obscure or fluid. What is more discouraging, the research has established that students do not possess the systemic vision of the sciences and the world. Their knowledge is compartmentalized – they are unable to relate physics to chemistry, to biology, etc. Hence is their low level of awareness of many innovations in science, and especially, in developments in the field of nanotechnology and its applications. This is mainly due to the inability to imagine the world at the nanoscale level. Hence is the fragile confidence in technological innovation and regulatory systems.

There might be objective and subjective reasons for the situation observed. Most higher education teachers feel that the knowledge students gain at secondary school is not sufficient for a higher education institution. But most importantly, our educational programs are structured in the way that perpetuates the myth that knowledge exists in separate compartments, as if there were no relationship between physics, chemistry, biology; between language and literature, and art, and history, and in so doing, encourages a similar compartmentalization of the mind. At the same time, the main problem area mentioned concerns the link between theoretical knowledge and students' envisioning their utilitarian value.

In any case, this is an alarming signal, which demands a critical analysis of the adequacy of the educational materials, the methods of teaching and research and other components of the educational practice.

Conclusions

The European context is a stage for developing new relationships, new ideas, new discoveries and new people. It is the stage for European-wide educational and scientific exchange and success for those individuals who can engage their intellectual and emotional potentials in scientific research and development, talents for openness and flexibility in order to exchange innovative thoughts, ideas, approaches and strategies.

This research is not attempting to solve the problem. The intent is to highlight the possibilities available through systemic, integral education to shape up and manage students' intellectual potential development, which offers a powerful philosophical, theoretical and practical approach to educating new generation specialists capable of providing the solutions to many long-standing medical, social and environmental problems. These ideas are all leading to what is termed 'disruptive' solutions, when the old ways of making things are completely overtaken and discarded. New solutions demand new ways of thinking.

The new paradigm of contemporary technologically advanced society brings to the agenda a new paradigm of higher education. This new paradigm envisages that higher education practitioners become pluriskilled, transdisciplinary mediatots promoting constructive solutions to innovative unprecedented problems of the day.

References

1. Lobanova T., Shunin Yu. The Systemic Approach to Language Acquisition as a Way to European Linguistic Unity. – *Proceedings of the 12th CiCe Conference*, Barcelona, Spain. (2010) 306-312.
2. Greenwood M. Thinking Big about Things Small: Creating an Effective Oversight System for Nanotechnology. – Washington DC., Woodrow Wilson International Center for Scholars, 2007.
3. Shunin Yu.N., Zhukovskii Yu.F., Burlutskaya N., Lobanova-Shunina T., Bellucci S. Novel Nanoelectronic Devices Based on Carbon Nanotubes and Graphene. – Programme and Theses of the 4th international conference *Innovative information technologies*, IIT-2010, Vilnius, Lithuania. (2010) 6-7.
4. Nanotechnology White Paper, Science Policy Council. – Washington: Nanotechnology Workgroup, US Environmental Protection Agency (EPA).

System approach to planning and implementation of enterprice data warehouse

Viktor Gopeyenko [a] , Maksim Levchenko
Department of Natural Sciences and Computer Technologies
Information Systems Management Institute, Ludzas 91, LV-1019, Riga, Latvia

Received 1 February 2011, accepted 23 February 2011

Abstract. Nowadays the importance of information exchange, accumulation and processing in modern business is increasing. It is hard to imagine a big company not having tens, hundreds or even thousands of terabytes full of accumulated data. Server systems of large multinational corporations demand significant costs to maintain and apply huge amounts of the accumulated data. The question is: how are these amounts of data formed, are they really necessary, and how can they be disposed? Therefore, in the long run company's management is to face these questions because the storage of information requires certain expenses, which increase depending on the growing amount of data.

Citations: Viktor Gopeyenko, Maksim Levchenko. System approach to planning and implementation of enterprice data warehouse – *Innovative Infotechnologies for Science, Business and Education,* ISSN 2029-1035 – 1(10)2011 – Pp. 10-12.

Keywords: Online transaction processing; OLTP; data warehouse; DWH; business intelligence; BI; data mining; systems analyst; decision support system; DSS; online analytical processing; OLAP; relational online analytical processing; ROLAP; multidimensional online analytical processing; MOLAP; structured query language; SQL; Oracle BI; OBI; Microsoft SQL Server.
Short title: System approach.

Introduction

In order to answer these questions it is essential to understand the nature of information and its content. Everything is simple when concerning emails, working papers and files: if they are relevant and valuable, they are stored, otherwise they are deleted. In turn, *online transaction processing* (OLTP) systems contain data with completed transactions and directories necessary for the formation of these transactions. From the point of view of OLTP system the concept of transaction includes any information, generated by any system used by an enterprise for its daily activities. As an example we can consider the situation with the sale of wares in a shop. The system captures the information of each item of goods that/which pass through checkout counter. Thus, a huge amount of transactions is recorded in OLTP data storage, which accumulates information from all the supermarkets of the chain. Having clarified the source of the formation of data, it is necessary to consider its content. As a rule, initially all sales contains the minimum of information which is useful for an analysts. For example, such minimum information could be: the date and the place of purchase, the name of the product, the amount of the sale and tax, the number of wares sold.

Having minimum information, it is worth considering how beneficial its systematization and processing can be. A quite serious analytical system can be designed on the base of the aforementioned example. This analytical system could reflect the dynamics of sales concerning the items of wares, compare the current amounts of sales with the amounts of sales of the previous periods, calculate the average purchase amount, the amount of tax paid, and other useful information. In practice, the OLTP systems of big enterprises store much more useful information, which enables monitoring warehouse stocks, logistics, perform financial analysis and accounting, develop marketing activities, etc [1,2].

1. Methods

Let's consider information as data after its representation in comprehended, ordered and focused on *business intelligence* (BI) instruments form.

As a rule, in the majority of organisations, there are several different groups of users that need various representation of the information. *Report readers* prefer to receive answers to the questions in a form of the same report that is well familiar to them. Their approaches to the analysis and to the information requirements are usually the same. Usually, individual help is required to this group of users to in order to start working. The other group of users which could be de-

[a]Corresponding author, email: *viktors.gopejenko@isma.lv*

fined as *information browsers* needs a slightly longer training to develop the means to work with data, however having the corresponding means they are able to prepare simple reports independently. Generally, they do not require powerful tools for the construction of the reports.

Advanced users prefer to build their own queries and reports. The possibilities of aggregated values search are important to them, as in this case they may quickly prepare reports without going deep into the details of the data model. Power users constantly put forward new demands for the data storage. They make considerable impact on all aspects of *data warehouse* (DWH) activity and aspire to have the information concerning data units. This group can be distinguished by the most intense usage of metadata repository and is the most interested in data mining.

One of the errors made by the developers of many DWH projects is an attempt to select the only software as a standard for all users. Usually it is better to select several instruments for different DWH users. As a result, the price of these programs will be significantly lower than time costs of system analysts.

The means of data access might also be concerned to various groups. It might be the general purpose means to manipulate multidimensional data arrays or more complex *decision support system* (DSS). Data mining means are used to find regularities in the data. Data visualization means represent data in a way, that helps to reveal these regularities.

Applications that use data from DWH and represent information to end-users are concerned with the means of *online analytical processing* (OLAP) group [3]. The main reason of OLAP usage for the query processing is the working speed. OLAP creates a DWH relation snapshot and restructures it into a multidimensional model for queries. The declared time for query processing in OLAP is approximately 0.1% from similar queries in the relational database. This means that use data from the relational databases are classified as *relational online analytical processing* (ROLAP) instruments. ROLAP is an alternative to the *multidimensional online analytical processing* MOLAP technology. While both ROLAP and MOLAP analytic tools are designed to allow the analysis of data through the use of a multidimensional data model, ROLAP differs significantly since it does not require the pre-computation and storage of information. Instead, ROLAP tools access the data in a relational database and generate *structured query language* (SQL) queries to calculate information at an appropriate level when the end-user requests it. The access means that use multidimensional databases are concerned with *multidimensional online analytical processing* (MOLAP) group. MOLAP is an alternative to the ROLAP technology. While both MOLAP and ROLAP analytic tools are designed to allow data analysis through the use of multidimensional data model, MOLAP differs significantly as it requires the pre-computation and storage of information in the cube.

To create an analytical system an idea is needed as an integral part of a final product. The correct selection of BI instrument is the other integral part of the successful development of the project. In the modern world various analytical solutions are available from multiple vendors, the most recognized among them are Oracle, Microsoft, SAP, SAS, IBM [4]. Globally they can be divided into two groups.

1. *Off-the shelf* solutions are installed on clients servers and adjusted to their needs.
2. *Development kits* provide a development environment only and make it possible to create custom solutions.

Oracle Business Intelligence Enterprise Edition (Oracle BI-EE) can serve as an example of a off-the-shelf solution, which is necessary to integrate in the existing software environment and to adjust to the requirements of each client. Oracle BI is a business intelligence platform which delivers a full range of analytic and reporting capabilities. Designed for scalability, reliability, and performance, Oracle BI-EE delivers contextual, relevant and actionable insight to everyone in an organization, resulting in improved decision-making, better-informed actions, and efficient business processes.

2. Realization

Microsoft SQL Server 2008 including SQL *Server Analysis Services* (SSAS), SQL *Server Integration Services* (SSIS), SQL *Server Reporting Services* (SSRS) can be considered as an example of the development instrument, which is a quite powerful tool to create customized solutions. It is a powerful and reliable data management system that delivers a rich set of features, data protection, and performance for embedded application clients, light Web applications, and local data stores. It is designed for easy deployment and rapid prototyping.

The data storage creation is a long-term and quite an expensive enterprise. When buying the data storage, analytical system or any other project, the customer wants to acquire not a set of algorithms and ready procedures, but a product that allows minimizing its expenses by optimizing working processes and the quantity of workplaces. Correctly adjusted system is able to substitute a large number of workers, freeing human and financial resources. In such a way we get to the first and, probably, the most important stage in the creation of any system.

The first stage of any system creation is its projecting. The customer will not pay anyone for unnecessary options that do not satisfy his requirements. The system must be carefully planned and organized, thereby responding to the requests of the customer. At this stage a plan is created which is a logical structure of the system. This plan is coordinated with the customer and sooner or later after a number of long and painful negotiations is approved.

The next stage is defining the necessary configuration of the hardware and software. This stage is also important since

without the skilfully chosen hardware and software base the created system will not be stable. As a consequence the periodic unforeseen faults in the system are possible. As a result the customer will bear losses which might be addressed to the developer of the system. Usually the contracts accompanying such projects clearly specify such nuances.

Then when everything is approved, we have the action plan; it is time to realize it. At first it is necessary to choose the team of developers, select the specialists with appropriate skills. In practice, the necessary number of specialists with certain (sometimes, specific) skills is not always available. Large companies with a large number of employees in different branches organize multinational groups. Smaller companies which, as a rule, have smaller capabilities either search for corresponding personal on the labour-market or join with other developers. As large companies have enough financial resources for their project realization they, usually, cooperate with other large companies.

When the team of developers is created and the responsibilities between them are distributed, it is possible to start the creation of the project. If the customer already possesses any data set, a part of array which is enough to develop software and to perform the test operations is taken from this data set. Then, as a rule, the end product is created as modules. This is a long-term process that can take a long period of time depending on the desired results. Usually, the average duration of the project is defined by the time period of one or two years. A large project may take two years or even more.

Of course, after the creation of the system it should be tested. This is an important step in its creation as a poorly working project is useless. During testing, the system is checked according to many parameters. The main testing parameters are the correct operations of mechanisms which form the system as well as the tolerance to stress loads. Paradigma *correctness of operations* means that all mechanisms (interfaces, services etc) should work without any significant faults and system overloads in any cases. *Tolerance to stress* loads is a relative concept, although it might be clearly measured. Initially the system tolerance is checked at the stress level defined by the customer (e.g. the number of transactions or connections). Then its working capabilities are checked at double load. And finally, the testing is performed at the maximum load level in order to understand when the system becomes inoperative.

The last and the final stage is the system support. Usually, after the creation of a large-scale project, the customer desires to get not only the warranty of its operability but also to get its support. Generally it is a long-term cooperation which is specified for several years ahead.

Special attention should be devoted to the security question during the project development. The best approach to such approach is formulated as follows: *The confidential information is the information that is not defined as non-confidential information*. The Developer company should thoroughly protect clients' data preventing any possibilities of the information leak. Usually, there is a set of clear instructions, the main idea of which is that in order to evade the problems it is better not to create them initially.

Conclusion

The choice of the software is one of the key features during the planning of the analytical system, as the functionality, licence costs, development and support costs, development perspectives depend exactly on the software choice.

The implementation of an analytical system at an enterprise is a rather expensive activity that requires involvement of significant financial and human resources. If the approach is correct, the return of investment is high and fast. A correctly adjusted system can perform a large set of different operations repeatedly, the fulfilment of which could require the work of several departments.

Due to the automation of the calculations the cost of data processing is significantly reduced. Other advantages are as follows:

i) a significant increase in calculation speed;
ii) an increase in the availability of data (the portal of reports can be seen in the Internet);
iii) an opportunity to obtain information in real time (during the report formation);
iv) an improvement of security (the data is available only to those staff members who have access).

Consequently, the presence of the data array and the lack of its usage strategy create prerequisites for the OLAP solution planning.

In the modern world, where in order to be successful in business it is necessary to be faster, wiser and more flexible than your competitors, exactly OLAP is capable to help to achieve success.

References

1. John M. Coe. The Fundamentals of Business-to-Business Sales & Marketing. – New York: McGraw-Hill, 2003.
2. Daniel Amor. The E-Business (R)evolution, 1st edition. – New York: Prentice Hall, 1999.
3. Hanson Ward A. Principles of Internet-marketing. – Cincinnati: South-Western College Publishing, 2001.
4. Eric Sperley. Enterprise Data Warehouse: Planning, Building, and Implementation, volume 1. – Upper Saddle River: Hewlett-Packard Company, 1999.

Modelling of multicolour light source using distributed computing network. 1. Statement of the physical and computational problems

Aistė Ambrazevičiūtė, Rimantas Vaicekauskas [a]
Department of Informatics, Faculty of Mathematics and Informatics
Vilnius University, Naugarduko 24, Vilnius, Lithuania

Received 29 January 2011, accepted 20 February 2011

Abstract. Light-emitting diodes (LEDs) are widely applied in conventional lighting. We investigate white light sources made of three primary coloured LEDs with respect to colour rendition ability. The spectral power distribution of LED is simulated by a function of a particular shape that depends on wavelength and intensity values, therefore searching for the optimal characteristics of compound light source becomes a computational problem. The aim of this work is to create an efficient parallel algorithm allowing us to find a trichromatic light source that minimizes colour distortions.

Citations: Aistė Ambrazevičiūtė, Rimantas Vaicekauskas. Modelling of multicolour light source using distributed computing network. 1. Statement of the physical and computational problems – *Innovative Infotechnologies for Science, Business and Education,* ISSN 2029-1035 – 1(10)(2011) 13-20.

Keywords: LED lighting; Colour rendering; Global optimization; Distributed computing; Message passing interface; MPI.
PACS: 07.05.Tp; 42.66.Ne; 02.60.Pn.
Short title: Modelling of light source - 1.

Introduction

Solid-state semiconductors *Light Emmiting Diodes* (LEDs) are compact and very energy-efficient sources: luminous efficency exceeds 200 lm/W, emission duration exceeds up to 100000 hours [1]. LED's are non-toxic and natural wear-resistant. Furthermore, they emit light in a wide variety of spectra - IR, visible or UV depending on their composition. LEDs operating in visible spectrum emit close-to-monochromatic light, so, by combining several of them and mixing their emission, it is possible to obtain the light emission of any colour and intensity. These features allow creation of high quality and efficiency lighting that is nature friendly as well.

In our case, the aim is to compound a white light source (sometimes called the *RGB* source) whose distortion of illuminated surface colours would be as low as possible. As the measure of the ability of light source to accurately reproduce colours of illuminated surface we will use general *Colour Rendering Index* (CRI). According to CRI calculation procedure briefly described in Ref. [2], colour shifts for 8 standard samples are taken into consideration by illuminating the samples with a reference light source and the test light source. CRI is defined as an average of eight particular colour rendering indices. Higher CRI value means better colour rendition (i.e., lower colour distortion). The light source identical to reference source has the highest CRI value 100.

This work is aimed:

i) to develope the mathematical model which describes spectral distribution of the light source consisting of several coloured LEDs and
ii) to relate physical characteristics of the light to the psychophysical colour related quantities such as CRI.

The searching problem can be defined as mathematical optimization problem, which takes many input variables. To quickly find optimal solution, advantages of parallel computations using the programming tools like *Message Passing Interface* (MPI) can be taken into account.

It should be noted that particular approximate solutions (CRI-optimal RGB LED systems) are already known [3]. For example, RGB system with selected wavelengths of LEDs near to 460 nm, 540 nm and 610 nm may have CRI value close to 90 [3]. However, in contrast to trial-error or randomized optimization techniques used elsewhere, we apply the *branch and bound algorithm* (assuming the objective function is the Lipschitz) that guarantees to find a global optimum with a desired accuracy. We will also discuss related computational problems and their solutions.

[a]Corresponding author, email: *Rimantas.Vaicekauskas@mif.vu.lt*

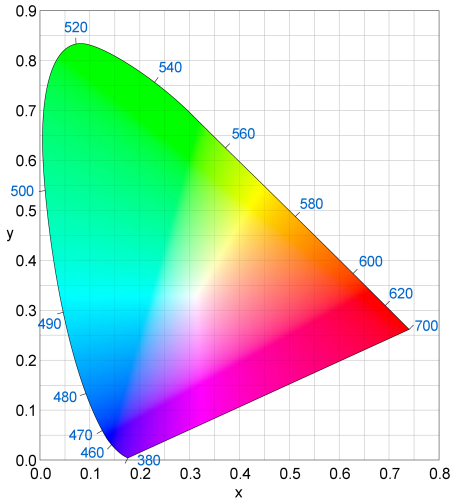


Fig. 1. CIE 1931 colour space chromaticity diagram. Adapted according to Ref. [4].

1. Relationship between light and colour

1.1. Colour mixing

Consider we have a set of LEDs emmiting in visible region with the SPD $S(\lambda, \lambda_0)$, where $\lambda \in$ [360÷830] nm is a wavelength and $\lambda_0 \in$ [360÷830] nm is a main LED parameter - peak wavelength.

Let's select three λ_0 values, in other words, three LEDs emitting the radiation of the following (central) wavelengths: λ_{01}, λ_{02}, λ_{03}. The resulting source of light after combining all three spectra is noted as the following equation where c_1, c_2, $c_3 \geq 0$ are coefficients describing the ratio between separate LEDs intensities.

$$S(\lambda) = c_1 S(\lambda, \lambda_{01}) + c_2 S(\lambda, \lambda_{02}) + c_3 S(\lambda, \lambda_{03}) \quad (1)$$

Our task is to find such combination of LEDs described by parameters λ_{01},λ_{02}, λ_{03} for a specific white light source (for example, standard D_{65} illuminant) that the CRI would be maximized. In other words, optimization problem has to be solved. The following conditions (constraints) should be satisfied: i) chromaticity of the reference source has to be exactly the same as the chromaticity of the resulting source; and ii) luminuous flux have to coincide.

Since every λ_{01},λ_{02}, λ_{03} may have a value from interval [360÷830] nm, the search space is a cube, intensities c_1, c_2, c_3 are determined by solving a colour mixing equation for three partial sources [3].

1.2. Colorimetry

In order to productively mix the emmision of LEDs of different colours, we need some basics knowledge about colours and their perception.

The reason why colours exist is purely based on a physical phenomenon known as visible spectrum electromagnetic radiation, although the perception of colours is only physiological and psychological phenomena. It is based on three types of cone cells that have different spectrum sensitivity, excitation and processing of information in our brains [5]. Therefore, the assumption was made, that the three primary colours - red, green, blue - are sufficient to create any colour.

1.3. Colour definition

Colour perception represents a research subject in the field of biology, psychophysics and photochemistry. Mathematically defined colour space as digital colours can be characterized by chromaticity coordinates (x, y), created in 1931 by the International Commission on Illumination (CIE), so called *CIE 1931 XYZ colour space* [6], [7]. Chromaticity coordinates (x, y) of a source consisting of n sources that are defined by power distributions $S_i(\lambda)$, are calculated from tristimulus values X_i, Y_i, Z_i where $a(\lambda)$, $b(\lambda)$, $c(\lambda)$ represent the *colour matching functions*, and the integration interval is presented in range [360÷830] nm.

$$X_i = \int a(\lambda) S_i(\lambda) d\lambda \quad (2)$$

$$Y_i = \int b(\lambda) S_i(\lambda) d\lambda \quad (3)$$

$$Z_i = \int c(\lambda) S_i(\lambda) d\lambda \quad (4)$$

$$x = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} (X_i + Y_i + Z_i)} \quad (5)$$

$$y = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} (X_i + Y_i + Z_i)} \quad (6)$$

x and y represent the chromaticity coordinates of compound lighting source. Fig. 1 shows the CIE 1931 colour space chromaticity diagram. All visible colours are bound by a horseshoe-shaped figure. In case of our problem, it is worth mentioning that in CIE 1931 colorimetric system differences between colours are not equal - the same distance at different coordinates cause different change in colour perception.

To evaluate colour differences we transform colours into uniform colour spaces, for example, *CIE 1960* and *CIE 1976* [8]. The standard technique is CIE 1964 [5]. In order to calculate CRI, the so called CIE 1960 colour space (CIE UCS) is used as an auxiliary colour space. The colour coordinates (u, v) of CIE UCS are defined as a transformation:

$$u = \frac{4x}{12y - 2x + 3} \quad (7)$$

$$v = \frac{6y}{12y - 2x + 3} \quad (8)$$

The colour rendering procedure, described in CIE 1995 [2], is based on eight standard and six extra colour samples. The difference between colour when illuminated by a reference light source (it is chosen to have same colour temperature as the test light source) and test light source is calculated for every sample. Based on these rendering numbers, CRI is

calculated. The average value of CRI on 8 standard colour samples is called a general CRI, which equals 100 points if test light source does not make any difference in colour. This system uses colour samples from A. H. Munsell system - see Table 1.

Munsel system represents behaviour of sample colour, where colour is defined by hue, value and chroma. Index 7,5R6/4 7,5 describes the sample: hue is red, value is 6 and chroma is 4 [5].

Ref. [3] presents a research of solid state semiconductor emitters consisting of several LEDs: two, three or more, when LEDs of different colour are used. As it was shown, the required number of primary LEDs depends on the efficiency and general CRI parameters.

In particularly, light source of decent efficiency and colour rendering characteristics can be made of $3 \div 4$ primary LEDs. Light source with 5 or more LEDs yields negligible benefit in improving colour rendering, however, its spectrum is almost continuous and that may be used in order to get a specific lighting.

Table 1. Munsell colour indexes.
1-8 for colour rendering. [4]

Sample N	Munsell label	Colour at daylight
1	7,5R6/4	Light greyish red
2	5Y6/4	Dark greyish yellow
3	5GY6/8	Strong yellow green
4	2,5G6/6	Moderate yellowish green
5	10BG6/4	Light bluish green
6	5PB6/8	Light blue
7	2,5P6/8	Light violet
8	10P6/8	Light reddish purple
9	4,5R4/13	Strong red
10	5Y8/10	Strong yellow
11	4,5G5/8	Strong green
12	3PB3/11	Strong blue
13	5YR8/4	Light yellowish pink (skin)
14	5GY4/4	Moderate olive green (leaf)

2. Computational aspects

2.1. Review of algorithms

Various optimization methods and techniques may be applied in our case. The main focus is on *branch and bound optimization* method. We take into account that our objective function meets the Lipschitz condition: the change in CRI is linearly bound by change of parameters. This method enables us to find regions without a global minimum or maximum and ignore them.

The feasible region is gradually divided to smaller subregions (branching), evaluated separately by using the value of Lipschitz constant. This subregion is further partitioned if optimal solution may be found in it (or discarded otherwise).

In the final step, by reviewing the remaining region, the lowest or biggest value of object function is the global minimum or maximum, respectivelly, with error bounds defined before the computation.

In order to use the branch and bound algorithm effectively, several partitioning strategies were analysed as well as the point selection in a region, that would be analysed most efficiently. A method was created allowing prediction whether there are any solutions in a segment, when the solution in investigated point is not found. Trying to keep the computation time of problem solving as low as possible, a new objective - parallelization of computational algorithms was taken into consideration. By analysing branch and bound algorithm parallelization methods, a new method was proposed and its efficiency in respect to a speedup was experimentally confirmed.

Parallel branch and bound algorithm for solving optimization problem was designed and implemented using the message passing interface library (MPI). MPI ensures effective and platform-independent message passing, so MPI programs written in C, C++ or Fortran, may be moved from one PC to another without any hassle.

Simple scheme of centralized data exchange was chosen for implementation of parallel algorithm. It is based on *master-slave* communicative structure, where only the master node communicates with slave nodes. In case of intensive data exchange, communication network may become overloaded. In that case, centralized scheme may become ineffective and modification of the scheme by decentralizing the communication might be taken into consideration.

Several modifications of the parallel algorithm were implemented and analysed in order to find the optimal data exchange rate. In every case the workload was distributed dynamically. The only difference was how often a process communicates with a master process. It was observed, that new candidates to the optimal values are found rarely during the optimization problem solving. A more common appearance of candidates was observed only in several segments of search interval. Conclusion was made that *master-slave* communicational scheme is sufficient for this problem.

2.2. Hardware

The algorithms were tested on a SGI Altix 4700 supercomputer, installed at faculty of Mathematics and Informatics, Vilnius University. SGI distributed computations network is made of 16 blocks, each consisting of 2 dual-core Intel Itanium 2 model 9020 (Montecito) 1.4 GHz CPUs - 64 cores in total. Algorithms were evaluated by speed and efficiency characteristics. Speed factor indicates the increase in computational speed gained by using a parallel algorithm [10]. Efficiency is based on the workload of cores used.

2.3. Evaluation of colour rendering

Procedure of the evaluation of colour rendering is described in Ref. [2]. According to it, the actions needed to take in order to calculate CRI of our sample light source are as follows.

1. Spectral distribution. Spectral power distribution of the sample $S_k(\lambda)$ is found and by using it, coordinates of colour (u_k, v_k) in CIE UCS space may be found, as well as the correlated colour temperature if it is needed.

2. Normalization. Spectral distribution of power $S_k(\lambda)$ is normalized in such way, that luminance Y_k = 100.

3. Light source. Reference light source with power distribution of $S_r(\lambda)$, colour coordinates (u_r, v_r), emission flux normalized in a way that luminance Y_r = 100 is selected according to the correlated colour temperature of the tested light source.

4. Reflected spectra. Power distributions of reflected light from colour samples $S_k(\lambda)\rho_i(\lambda)$ and $S_r(\lambda)\rho_i(\lambda)$ are calculated for both light sources: tested and reference. Coordinates of luminance Y_{ki}, Y_{ri} and colour (u_{ki}, v_{ki}) and (u_{ri}, v_{ri}) are calculated for both spectra.

5. Transformation. Transformation of colour coordinates (u_{ki}, v_{ki}) is applied using the following scheme.

$$u^*_{ki} = \frac{10.872 + 0.404 \cdot C_q - 4 \cdot D_q}{16.518 + 1.481 \cdot C_q - D_q}; \quad (9)$$

$$v^*_{ki} = \frac{5.20}{16.518 + 1.481 \cdot C_q - D_q}; \quad (10)$$

$$C_q = \frac{c_r \cdot c_{ki}}{c_k}; \quad (11)$$

$$D_q = \frac{d_r \cdot d_{ki}}{d_k}; \quad (12)$$

$$c = \frac{4 - u - 10 \cdot v}{v}; \quad (13)$$

$$d = \frac{1.708 \cdot v + 0.404 - 1.481 \cdot u}{v}. \quad (14)$$

6. Differences in luminance and colour. Differences in luminance and colour are calculated in a CIE 1964 (W*U*V*) colour space for every sample ΔW, ΔU, ΔV by expression G:

$$\Delta W^*_l = 25 \cdot [\sqrt[3]{Y_{ki}} - \sqrt[3]{Y_{ki}}] \quad (15)$$

$$\Delta U^*_l = 13 \cdot [G \cdot (u^*_{ki} - u_r) - G \cdot (u^*_{ri} - u_r)] \quad (16)$$

$$\Delta V^*_l = 13 \cdot [G \cdot (v^*_{ki} - v_r) - G \cdot (v^*_{ri} - v_r)] \quad (17)$$

$$G = 25 \cdot \sqrt[3]{Y_{ki}} - 17 \quad (18)$$

7. Differencies for every sample. Difference in colour is calculated for every sample:

$$\Delta E_i = \sqrt{(\Delta U^*_l)^2 + (\Delta V^*_l)^2 + (\Delta W^*_l)^2} \quad (19)$$

8. Test. CRI is evaluated for every test sample:

$$R_i = 100 - 4.6 \cdot \Delta E_i \quad (20)$$

9. Averaging. Average CRI is calculated in order to find the general CRI.

2.4. Optimization tasks

Usually, minimum or maximum value of object function has to be found in order to find a solution of the optimization problem. If the derivative of object function is unknown, other information about the function may be used [9]. Random search methods may be applied too.

Search of global minimum. The search for global minimum is described in Ref. [9]. Local *hill climbing method* may be applied in search for global minimum, because, by default, it is a local minimum too. The initial point has to be a part of the area of attraction of global minimum. This point is often found by random prediction - if points are generated by a continuous distribution in a permitted area, the probability p of one random guess being in the attraction area of the global minimum is presented as equation where G and V are hyperspace of the attraction area and permitted area of the global minimum, respectively. Probability p_n represents situation that 1 out of n attempts will be in this area.

$$p = \frac{G}{V} \quad (21)$$

$$p_n = 1 - (1 - p)^n \quad (22)$$

Another technique of searching for global minimum is called *random local minimum enumeration*. Local minimization is usually performed in search method instead of hill climbing. It is sufficient to have a local minimization method combined with the generation of initial points in order to implement the search method. This technique is valid only if the function has just a couple of local minimums with similar-sized attraction areas. Otherwise, the result of hill climbing can appear to be just a point of local minimum, especially if it has a wide attraction area.

Cluster search method. The general idea of this method is to detect clusters of points on the descending curves. In order to do that, a set of initial points is generated. Local descend computation is started after few steps in order to move the points from their initial state. The resulting points are analysed and if a cluster of points is found, we can conclude that several descending processes move towards the same minimum, a point with the best value of object function is selected. Afterwards, local descend is continued for this point, while all remaining processes are cancelled. After the analysis of clustering of points, only a specific number of processes may be cancelled and a new set of points is generated. The process is repeated until several analysis in a row do not locate any new point clusters. This method is pretty effective, since local descends to a minimum are located quite rapidly and every but one processes are cancelled. In order to avoid unnecessary local minimums, search method instead of hill climbing is applied. As specified in the Ref. [9], experimental comparison between this and other methods proved high efficiency of this method.

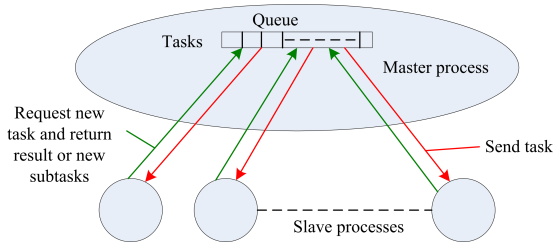


Fig. 2. Scheme of creation of centralized dynamic schedule. Adapted according to Ref. [12].

Branch and bound method. A classical approach for optimization problems is *branch and bound method*. It concentrates around looking for subintervals that do not contain the global maximum or minimum and cancellation of them.

This results in the permanent division of initial search range. While analysing every part separately, it is decided whether optimal solution exists in it or not. If it does, the top or bottom boundary is extended and the range is divided further, otherwise it is discarded. The efficiency of branch and bound method relies on the initial division algorithm. Several templates of these algorithms exist, their purpose being simplification of implementation of a specific branch and bound algorithm. Examples of such templates are as follows: BOB, BOB++, PICO, PPBB, PPBBB [10].

One of parallel implementations of branch and bound method was described in Ref. [10]. MPI libraries were used to exchange data. The main feature of this template is the division of search ranges across processing cores.

The amount of these ranges may be equal to, or higher than the number of cores. Every processor checks its range whether optimum solution exists in that range. Processors may communicate and share their bounds by applying data exchange protocols. Also this template is supplemented to a load balance feature, enabling reallocation of tasks during the execution in order to keep the workload of system balanced. Tools enabling use of Lipschitz criteria are implemented too.

3. Parallel computation

Parallel computations help reduce time it takes to solve a problem, although, if it is applied incorrectly, it may take much longer. In order to succesfully apply parallelization, such routines must be constructed according to advices in Ref. [11].

1. Dividing the complex task to smaller independent tasks.
2. What task size should be chosen?
3. How many nodes should be used for fastest run?
4. How to assign tasks between nodes?

The specificity of the problem as well as data exchange between nodes for parallel computation needs to be taken into consideration before dividing the problem to separate tasks.

The cost of message passing between nodes has to be discussed. Nodes pass messages in order to exchange data in parallel computations. If a message constitutes of n bytes, the cost of that message passing may be defined as:

$$T = \alpha + \beta n. \tag{23}$$

where α is a sum of the preparation of message (setting up contents of message, assigning overhead, flags, find out the optimal route in a network) and the duration of message travelling through the media. β is the time of transferring one byte of data [10]. In order to reduce the cost of message passing, the messages should be grouped together to keep the ratio ξ as low as possible.

$$\xi = \frac{\alpha}{\beta n} \to 0 \tag{24}$$

3.1. Creation of a schedule

Process of scheduling constitutes of assigning nodes to specific tasks. Schedule is defined by vector S:

$$S = (s_0, s_1, ..., s_{p-1},), \tag{25}$$

where s_j is a set assigned to j-th node. Schedule is titled as correct in case if all following conditions are satisfied [11].

1. Every task is assigned to only one set s_j.
2. Execution time for every task is later than execution time for any predecessor task.

According to the division of tasks, these techniques may be presented for scheduling.

3. Static scheduling.
4. Dynamic scheduling using centralized or decentralized manager.

Static scheduling. Static scheduling is the simpliest technique of scheduling. It may be used only in case when all tasks are known. Using this technique, problem is divided into smaller tasks and every task is assigned to different processor. The work is done, when every core returns its result, although the drawback of this technique is the case when some tasks take longer to complete and a part of nodes will be idle until every single one will finish the computation.

Dynamic scheduling. Dynamic scheduling allows to assign tasks to nodes in a optimized way [12]. Also, it enables creation of schedule without knowing every possible task. Sometimes they appear as a result of computing another task (see Figs. 2-3). Centralized dynamic scheduling has a master node that assigns tasks.

The tasks are sorted according to their "weight" and sent to slave nodes in small groups. Upon completion, a node sends back the result to a master and asks for another task. The process of scheduling end together with the solution - when queue is empty and every node has returned its result.

This technique is applicable when tasks are *heavy* and the amount of nodes is moderate. Decentralized schedule is a modified version of aforementioned dynamic schedule. It is used when the amount of nodes is large and they request a new task more often.

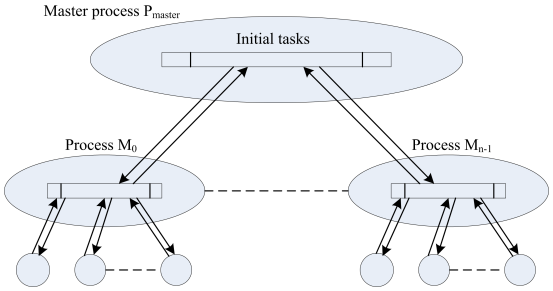


Fig. 3. Management scheme of decentralized dynamic schedule. Adapted according to Ref. [10].

In order to cope with that, master process initiates several management processes. Such management processes receive tasks and then distribute them independently, receiving and storing results in the meantime. Afterwards, every management process returns their results back to master process. After the master process finishes processing of all results, the scheduling is over [12]. When using the decentralized dynamic schedule, an extra task is to choose the right amount of management processes.

3.2. Optimization of schedule

Task of optimal scheduling. The point of optimal scheduling is to achieve the fastest possible execution of specific algorithm. In order to create the optimal schedule, many points have to be taken into consideration:

i) the problem itself;
ii) the cost of communication between nodes;
iii) the difficulty of separate tasks;
iv) the amount and characteristics of nodes.

Also, it has to be noted, that workload and characteristics of nodes may change in time. The optimization of schedule is NP-complete and in order to save time, heuristics are applied resulting with a decent schedule. The examples of scheduling algorithms are as follows:

i) algorithms of list division;
ii) critical path algorithms;
iii) graph analysis heuristics;
iv) Monte Carlo method and its modifications;
v) adaptive algorithms;
vi) genetic algorithms;
vii) combined task distribution algorithms.

The main idea of list division algorithm is as follows: the list of tasks is arranged and every free node picks up a task that has not started yet. If there are no such tasks, the node in question remains idle until a task appears. Critical path algorithms try to determine tasks that would have direct impact on time of finding a solution for whole problem. The main idea of Monte Carlo algorithm is to make computations with only a small set of randomly chosen data, for example, nodes are arranged by their computation rate and tasks are assigned to them with corresponding difficulty [12], ignoring the fact that workload on a node is a dynamic characteristic. Various modifications of Monte Carlo method exists [13].

Dynamic scheduling. A method of scheduling depending on cost of communication, cost of task completion, relations between tasks and computer characteristics is presented in a Ref. [14].

The main idea on the scheduling presented in the reference is the minimization of cost and maintaining balance between computer workload on a system. One of propositions is to arrange grouping of tasks by intensity of communication between them and make sure that similar tasks communicate only between themselves in the same group. The grouped tasks should be assigned to nodes for execution before any other tasks.

Nodes have to be arranged by their workload before having a task assigned to them. The workload levels are as follows: low workload, moderate workload and overload. When the workload levels are known, tasks can be assigned to them according to "weight" of the task and the workload of a processor to keep system balanced. In addition to this method strictly defined conditions such as priorities, resource availability, real time characteristics and dynamic resource distribution on grid are implemented. Virtual jobs are created by grouping tasks using this technique. Every of them constitute the queue of jobs that should communicate between each other. Because they are assigned to the same resource if possible, the communication cost between them is reduced as much as possible.

3.3. Parallelization of algorithm

Parallel algorithm can be implemented depending on memory type used: either shared or distributed memory. Threads have a common variable allowing them to watch each others progress and results in case of the shared memory. In order to see that in case of distributed memory, threads have to pass messages to each other. The cost of them has already been reviewed. In order to maximize the efficiency of problem solving, optimal message passing scenario has to be chosen, including the occasion and reason why the messages need to be passed, because communication cost is quite significant.

***Master-slave* communication.** *Master-slave* structure defines a type of communication between processes or threads when one process or thread is responsible to initial assignation of jobs for slaves and retrieval of their results. There are two base variations of this algorithm: the amount of slaves can be static or dynamic. This work presents *master-slave* communicational structure, evaluates it and then the decision whether there is a need to decentralize it is made.

Case of shared memory. In case of shared memory *master-slave* communication, one thread, known as master, starts and then launches slave threads and equally assigns tasks to them. After that, master thread waits at a point synchronized by every slave until it receives every result from slave threads.

Table 2. Main operations of semaphores.

Operation	Value	Description
S_p	false	to block on s
S_p	true	$s = s - 1$
S_v	true	to unblock on s
S_v	false	$s = s + 1$

Then it ends its work or repeats the process recursively.

Synchronization of two data elements access needs to be taken care of in shared memory case. First element is a tree. In its leaves the search range is fixed and master assigns a part of general search range that is in a leaf to a slave. The second element describes the best solution at a given moment.

Synchronization needs to be ensured in case if: i) the best solution changes; ii) processes associated with task allocation take place. Semaphores may be used for synchronization. Their operations are presented in Table 2. We have used the described technique for semaphore manipulation. For example, three semaphores - $NumberOfTasks$, $ksProtection$, $ksProtection2$ are presented in Table 3.

Case of distributed memory. In that case one process is responsible for task assignation to the other processes in *master-slave* communication scheme. Master process:

i) prepares initial list of tasks;
ii) while optimal solution is not found, accepts requests for new task, operation *receive* is performed;
iii) while optimal solution is not found, returns task or end sign, operation *send* is performed;

Slave process:

i) requests for a new task, operation *send* is performed;
ii) receives a new task and current maximum or an end sign, operation *receive* is performed;
iii) performs reduction of search range operation;
iv) performs division operation;
v) forms new tasks;
vi) forms new maximum value;
vii) transmits max value and new tasks, operation *send* is performed;

Table 3. Additional operations of semaphores.

Semaphore	Function or operation for semaphore
	Retrieve jobs
$NumberOfTasks$	P()
$csProtection$	P()
	Assigned job
$csProtection$	V()
	Insert job
$csProtection2$	P()
if (best)=true	renew_results
$csProtection2$	V()
	Insert job
$NumberOfTasks$	V()

Evaluation of parallel algorithm. Time needed for the parallel algorithm to complete problem solving depends on many indicators, such as specifics of a problem, partitioning of a problem to smaller tasks and the amount of these tasks, amount of processor cores, technique of task allocation between processors, computer characteristics, computer workload at given moment. Several criteria are used to evaluate the parallel algorithms. One of them is a speedup factor S_p:

$$S_p = \frac{T^*}{T_p}. \tag{26}$$

The speedup factor indicates the speed of parallel algorithm using p processors where T^* is time required to complete an assignment using the best known serial algorithm.

T_p is titled as the time to complete same assignment using parallel algorithm on p nodes [10, 13]. Parallel algorithm may also be evaluated by processor workload usage efficiency E_p, where p is amount of processor cores [15]. It is known that the efficiency of parallel algorithms decreases by increasing the amount of processor cores.

$$E_p = \frac{S_p}{p} \tag{27}$$

This characteristic of an algorithm is known as *scalability* and it defines the required increase in a problem by increasing the amount of nodes used. Scalability of an algorithm is better, when the required increase in a problem size is as small as possible in order to keep the computations effective by adding more processor cores.

MPI. *Message-Passing Interface* (MPI) is a *de facto* standart for implementation of message passing in parallel algorithms. The purpose of MPI is to define a flexible, effective and platform-independent message passing in distributed computing environment. The computer programm using this standard may be moved from one system to another without any problems [16]. MPI defines only function names, parameters and purpose, but does not describe implementation. Commercial, non-commercial and manufacturer-bound MPI versions exist. Computing cluster uses non-commercial MPI version called LAM/MPI.

Conclusions

A mathematical model has been built in order to find out a light source made of several multi-colour LEDs with adjustable wavelengths having the highest general colour rendering index. The methods for a optimization problems were discussed. In order to reduce the search span, branch and bound optimization method was introduced. The usage of MPI in parallelization of optimization algorithm was investigated. Balancing of CPU load is required to have an effective parallel algorithm. This is achieved by solving computation queueing problem. The effectiveness and scalability of the parallel algorithm is the main goal. The solution will be checked on VU MIF distributed computation cluster.

References

1. J. Y. Tsao, H. D. Saunders, J. R. Creighton, M. E. Coltrin, and J. A. Simmons. Solid-state lighting: an energy-economics perspective – *J. Phys.* D43 (2010) 354001.
2. Method of measuring and specifying color rendering properties of light sources – *CIE Publication* 13.3 (1995).
3. A. Žukauskas, R. Vaicekauskas, F. Ivanauskas, R. Gaska, and M. S. Shur. Optimization of white polychromatic semiconductor lamps – *Appl. Phys. Lett.* 80(2) (2002) 234-236.
4. CIE 1931 color space. <http://en.wikipedia.org/wiki/CIE_1931>, retrieved 2010 05 10.
5. Wyszecki, G. and Stiles, W. S. Book. Color Science: Concepts and Methods, Quantitative Data and Formulae – New York:Wiley, 2000.
6. Commission internationale de l'Eclairage proceedings. – Cambridge: Cambridge University Press, 1931.
7. Smith Thomas, Guild John . The C.I.E. colorimetric standards and their use – *Transactions of the Optical Society* 33(3)(1931–1932) 73–134.
8. Yoshi Ohno. CIE Fundamentals for Color Measurements. – Gaithersburg, USA, Optical Technology Division National Institute of Standards and Technology (2000).
9. A. Žilinskas. Matematinis programavimas (in lith.) – Kaunas : *VDU leidykla*, 2000.
10. B. Wilkinson, M. Allen. Parallel Programming: Techniques and Applications Using Networked Workstations and Parallel Computers – USA (2005).
11. M. Baravykaitė, R. Čiegis. An implementation of a parallel generalized branch and bound template. – *Mathematical modelling and analysis: the Baltic journal on mathematical applications, numerical analysis and differential equations*. ISSN 1392-6292 – 12(3) (2007). p. 277-289.
12. R. Čiegis. Lygiagretieji algoritmai (in lith.) - Vilnius :*Technika*, 2001.
13. S. Chantaravarapan, A. Gunal, E. J. Williams. On using Monte Carlo methods for scheduling. – Dearborn, MI, USA, Production Modeling Corp. (2004).
14. L. Liu, Y. Yang, W. Shi, W. Lin, L Li. A dynamic Cluster Heuristic for Jobs Scheduling on Grid Computing Systems. – *IEE* 2006.
15. A. Igumenov, T. Petkus. Analysis of Parallel Calculations in Computer Network. – *Information Technology And Control* 37(1) (2008) 57-62.
16. William Gropp, Ewing Lusk and Anthony Skjellum. Using MPI. Portable Parallel Programming with the Message Passing Interface. 2nd Edition. – *Scientific and Engineering Computation*. MIT (1999).

Modelling of multicolour light source using distributed computing network. 2. Parallel implementation and experimental results

Aistė Ambrazevičiūtė, Rimantas Vaicekauskas [a]
Department of Informatics, Faculty of Mathematics and Informatics
Vilnius University, Naugarduko 24, Vilnius, Lithuania

Received 29 January 2011, accepted 20 February 2011

Abstract. Light-emitting diodes (LEDs) are widely applied in conventional lighting. We investigate white light sources made of three primary coloured LEDs with respect to colour rendition ability. The spectral power distribution of individual LED is simulated using a Gaussian function with selectable peak wavelength. The optimization problem of finding a trichromatic source having maximum colour rendering index was defined and solved using parallel branch and bound method. We have implemented and examined several centralized data exchange schemes on a parallel computing cluster. Speed and efficiency of the proposed algorithms were revealed.

Citations: Aistė Ambrazevičiūtė, Rimantas Vaicekauskas. Modelling of multicolour light source using distributed computing network. 2. Parallel implementation and experimental results – *Innovative Infotechnologies for Science, Business and Education,* ISSN 2029-1035 – 1(10)(2011) 21-29.

Keywords: LED lighting; Colour rendering; Global optimization; Distributed computing; MPI.
PACS: 07.05.Tp; 42.66.Ne; 02.60.Pn.
Short title: Modelling of light source - 2.

Introduction

Simulating physical phenomenon requires an accurate physical model and computational power. Our previous publication [1] was aimed to develop a mathematical model which describes the spectral distribution of the light source consisting of several coloured *Light-emitting diodes* (LEDs). This problem was solved relating physical parameters of light to the psychophysical colour quantities such as CRI. General approaches - how to quickly find optimal solution - were investigated in respect of applying distributed computing model - *Message Passing Interface* (MPI) technique, which requires effective management of distributed computing resources.

A problem of optimization investigated in this work concentrates on maximizing the CRI of a modelled light source of 3 LEDs with adjustable peak wavelenght. Though tabular functions are used for the analysis of the light source, analytical methods based on derivatives of the objective function may not be applied to solve this optimization problem. Trial and error method is very inefficient, so it cannot be used either. A similar problem presented in Ref. [2] was solved by using the stochastic *hill climbing* technique, although this method does not guarantee that local and global maximum values coincide.

[a]Corresponding author, email: *Rimantas.Vaicekauskas@mif.vu.lt*

1. Spectral distribution of light source

To have a model of compound light emmiter consisting of several LEDs the primary emitters are described at first.

1.1. Gaussian shaped LED emission

Radiation of light source can be described using *spectral power distribution* (SPD). The spectra of primary monochrome LEDs were approximated by Gauss-shaped distribution $f(x, \mu, \sigma)$ (see Fig. 1)

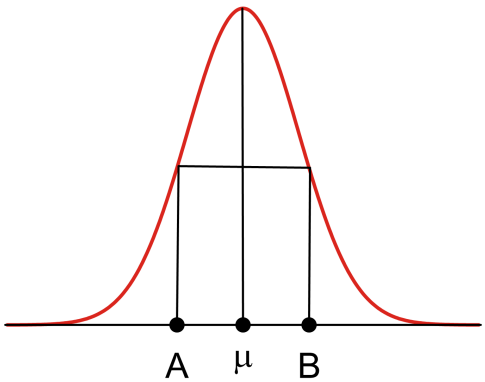


Fig. 1. Spectral power distribution of LED radiation as intensity function on wavelength. Half-width - σ=30 nm.

SPD $f(x,\mu,\sigma)$ expresses an behaviour where μ and σ represent the wevelength of peak and half-width, respectivelly, and x is functional parameter - wavelength. Due to physical circumstances - emmision stability of LEDs, we simplify the function $f(x,\mu,\sigma)$ by choosing a constant value of σ. This distribution $f(x,\mu)$ will depend only on the parameter of peak position μ.

$$f(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad (1)$$

$$f(x,\mu) = 15 \cdot \sqrt{\frac{\log(e)}{\log(4)}} \cdot \exp\left[\log 8.0 \cdot \left(\frac{x-\mu}{30}\right)^2\right] \quad (2)$$

We can then use this function $f(x,\mu)$ as a template to generate various SPDs. In the computation experiments following distribution function was employed.

If the height of peak in gauss distribution is h, we get the width of the distribution at half magnitude $\cdot h$ as a half-width - μ=30 nm. That is an average value for high-brightness LEDs made by using common (AlInGaP, InGaN) technologies and operating at typical modes. In our case, the peak position value λ_0 belongs to the visible wavelength range [360÷830] expressed in nm.

1.2. Mixing of spectral components

By mixing two different light sources with different spectra, a third light source is made and it is called a resulting light source. The resulting spectrum depends on the ratio the initial sources were mixed. At first, the method describing the spectrum of the resulting light source has to be introduced. It has to be known in order to be able to evaluate the light source. The most convenient way to describe the spectrum of a resulting light source is when all initial light sources are normalized. Normalization of a light source defined by distribution of spectral power $S(\lambda,\lambda_0)$ is performed by finding tristimulus values X, Y, Z and division of the spectral power distribution by the sum of these values. Spectral power distribution of normalized light source is described as follows:

$$S_n(\lambda,\lambda_0) = \frac{S(\lambda,\lambda_0)}{X+Y+Z}. \quad (3)$$

It may be noted, that colour coordinates describing the colour of the light source remain the same after normalization of the light source, so these parameters used in colour mixing equations may be found at any time. Brief description of these parameters is in the next section. Spectrum of resulting light source, after normalization the initial sources, may be found using $S_{sum}(\lambda)$ expression. All further actions are performed using virtual source of such type.

$$S_{sum}(\lambda) = \sum_{j=1}^{3} c_j \cdot S_n(\lambda,\lambda_{0j}) \quad (4)$$

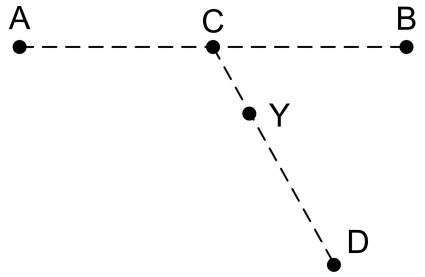


Fig. 2. Initial points A, B, D resulting in a point Y.

1.3. Obtaining colour mixing proportions

The resulting light source of LED mixture has to have particular colour coordinates. This is achieved by deciding, whether it is possible to have a resulting light source of required colour, and, if so, what light source intensities should be chosen. Three initial light sources and a required resulting light source have to be marked in a colour space. The initial sources are marked with points A, B, D and the resulting source is marked with Y - see Fig. 2.

If we need to get point Y(y_1, y_2), with points A(a_1, a_2), B(b_1, b_2) and D(d_1, d_2) given, we have to find at first the point C(c_1, c_2).

$$C(c_1,c_2) = \eta A(a_1,a_2) + (1-\eta)B(b_1,b_2) \quad (5)$$

Since $\eta \leq 1$, point C may be defined as follows:

$$c_1 = \eta a_1 + (1-\eta)b_1; \quad (6)$$

$$c_2 = \eta a_2 + (1-\eta)b_2. \quad (7)$$

Once we have the coordinates of point C(c_1, c_2), we may find the point Y(y_1, y_2) :

$$Y(y_1,y_2) = \mu D + (1-\mu)C; \quad (8)$$

$$Y(y_1,y_2) = \mu D + (1-\mu)\eta A + \Omega; \quad (9)$$

$$\Omega = (1-\mu)B - (1-\mu)\eta B; \quad (10)$$

where $\eta \leq 1$ and $\mu \leq 1$. In order to simplify the expression, we perform a substitution in order to receive resulting equation.

$$\gamma = (1-\mu)\cdot\eta; \quad (11)$$

$$Y = \mu D + \gamma A + (1-\mu-\gamma)B. \quad (12)$$

If we want to get a light source Y(y_1, y_2) from the three known sources A(a_1, a_2), B(b_1, b_2) and D(d_1, d_2) (with colour coordinates (a_1, a_2), (b_1, b_2) are (d_1, d_2), respectively) the parameters of intensity of a source, defined by γ and μ have to be found first. From an equation above, we get:

$$y_1 = \mu d_1 + \gamma a_1 + (1-\mu-\gamma)b_1; \quad (13)$$

$$y_2 = \mu d_2 + \gamma a_2 + (1-\mu-\gamma)b_2; \quad (14)$$

$$y_1 = \gamma(a_1-b_1) + \mu(d_1-b_1) + b_1; \quad (15)$$

$$y_2 = \gamma(a_2-b_2) + \mu(d_2-b_2) + b_2. \quad (16)$$

We apply Cramer's rule in order to solve this system with two unknown variables.

$$\gamma = \frac{\det \gamma}{\det}; \qquad \mu = \frac{\det \mu}{\det}; \tag{17}$$

$$\gamma \cdot (a_1 - b_1) + \mu \cdot (d_1 - b_1) = y_1 - b_1; \tag{18}$$

$$\gamma \cdot (a_2 - b_2) + \mu \cdot (d_2 - b_2) = y_2 - b_2. \tag{19}$$

Rearrange the system we retrieve the solutions from the values of $\det$, $\det \gamma$ and $\det \mu$:

$$\det = (a_1 - b_1)(d_2 - b_2) - (a_2 - b_2)(d_1 - b_1); \tag{20}$$

$$\det \gamma = (y_1 - b_1)(d_2 - b_2) - (y_2 - b_2)(d_1 - b_1); \tag{21}$$

$$\det \mu = (a_1 - b_1)(y_2 - b_2) - (a_2 - b_2)(y_1 - b_1). \tag{22}$$

This means that virtual source Y(y_1, y_2) could be constructed using sources A, B and D representing by colour components c_1, c_2, c_3, respectively.

$$c_1 = \gamma; \tag{23}$$

$$c_2 = 1 - \mu - \gamma; \tag{24}$$

$$c_3 = \mu. \tag{25}$$

It is necessary to point out that the chromaticity point of light source Y may only be a result of mixing if it belongs to a triangle ABD bound by chromaticities A, B and D. This can be checked by solving three inequalities, or simply checking if c_1, c_2, $c_3 \geq 0$ meet. This equality is enough to know that point Y exists and can be found:

$$c_1 + c_2 + c_3 = 1. \tag{26}$$

1.4. Obtaining of total spectrum

Spectra mixing technique presented in section 1.2 works only with sources, whose sum of tristimulus values X, Y and Z equals 1, this does not apply in general. In our case, this condition is not satisfied for LEDs sources. Their max value of a spectrum equals one, but half-width is equal to 30 nm. To get the required properties on a resulting LED, intensity parameters of the initial LEDs need to be transformed. Procedure of transformation could be formulated as follows.

1. Tristimulus values X_i, Y_i, Z_i are calculated for our LEDs, where i = 1..3.

2. Colour components c_1^*, c_2^*, c_3^* are calculated by using technique presented earlier.

3. Normalization of each LED characteristics (i = 1..3) is performed using following equation.

$$LED_i^* = \frac{LED_i}{X_i + Y_i + Z_i} \tag{27}$$

4. Total spectrum of normalized light source LED_{Sum}^* is retrieved from LED_j^* and colour components c_j (j=1÷3).

5. Max value max_{Sum} is found in spectrum LED_{Sum}^*.

$$c_i^* = \frac{c_i}{X_i + Y_i + Z_i} \cdot \frac{1}{max_{Sum}} \tag{28}$$

6. c_i^* values are normalized, to get their total value equal 1.

$$c_1^* + c_2^* + c_3^* = 1 \tag{29}$$

Integration of functions defined in table. Integration is inevitable when performing various operations with LEDs. The integration is performed on functions with values known only in particular points. These values are described in tables, so, in order to integrate such function, it's value has to be approximated to a value known. The way of approximation depends on the occasion, which in our case is the most simple approximation algorithm - discontinuous step function, which may be integrated by a sum of all rectangular areas.

1.5. Obtaining of General Colour Rendering index.

The light source, resulting from a set of separate LEDs and satisfying all requirements, has to be evaluated. The evaluation of a light source is received by computing general CRI value. In case of our problem, the resulting sample emitter has a specific chromaticity (like reference light source). Lets assume the D_{65} emitter as a standard emitter. The correlated colour temperature C_{CT} depends only on a colour parameters x,y [3]:

$$C_{CT} = g_1 \cdot \beta^3 + g_2 \cdot \beta^2 + g_3 \cdot \beta + g_4; \tag{30}$$

$$g_1 = -437; g_2 = 3601; g_3 = -6861; g_4 = 5513.31; \tag{31}$$

$$\beta = \frac{x - 0.3320}{y - 0.1858} \tag{32}$$

Procedure of colour rendering evaluation may be based on spectrum of the emitter D_{65} itself. We normalize the parameters of both sources in a way, that their luminance Y would be equal to 100. Then, in order to evaluate the resulting light source, we calculate parameters X, Y, Z for every initial source and every function of reflection $\rho_i(\lambda)$:

$$X_{j,i} = \int_{360}^{830} a(\lambda) \cdot S(\lambda) \cdot \rho_i(\lambda) d\lambda \tag{33}$$

$$Y_{j,i} = \int_{360}^{830} b(\lambda) \cdot S(\lambda) \cdot \rho_i(\lambda) d\lambda \tag{34}$$

$$Z_{j,i} = \int_{360}^{830} c(\lambda) \cdot S(\lambda) \cdot \rho_i(\lambda) d\lambda \tag{35}$$

where subscript index i stands for reflection function and index j is for sample or a reference emitter. ΔE_i may be found using previously found values of $X_{j,i}$, $Y_{j,i}$, $Z_{j,i}$. The general CRI is found from an average of ΔE_i values. If our sample source gets a CRI value of 100, the computation can be stopped, because it is the perfect emitter.

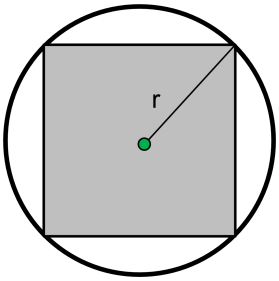


Fig. 3. Example of discardation of square range.

Due to reflection functions, reference light source and the colour are constant, our sample source depends only on these variables $\lambda_{01}, \lambda_{02}, \lambda_{03}$, the general CRI is a function of them: $R_a(\lambda_{01}, \lambda_{02}, \lambda_{03})$.

2. Branch and bound technique

As the basic questions regarding emitter modelling have been answered, the optimization method has to be decided. There are a lot of different techniques, but branch and bound method will be used in further solution of our problem.

2.1. Discarding of search range

It is useful to know when creating the branch and bound algorithm, that $R_a(\lambda_{01}, \lambda_{02}, \lambda_{03})$ satisfies Lipschitz condition (because the first partial derivatives are bounded in the search range).

$$| f(a) - f(b) | \leq L | a - b | \tag{36}$$

L is called a *Lipschitz constant*. It means that for every point c from interval [a÷b], value of a function satisfies these inequalities:

$$f(c) \leq f(a) + L | a - b |; \tag{37}$$

$$f(c) \geq f(a) - L | a - b | . \tag{38}$$

Therefore, satisfying of mentioned condition proposes the candidates to minimum and maximum, although minimum and maximum values could be distributed in next intervals. Accuracy may be increased by dividing them to smaller intervals and discarding these that may not possibly contain global optimal value.

Let's assume, that object function satisfies Lipschitz condition. That allows us to predict the distance, where optimal value can not be found, from our current point. This is a circle in 2D case and a sphere in 3D. Being inconvenient to divide the search range to spheres, it is divided to cubes inscribed in these spheres.

We can see that from one-dimensional case, this procedure differs in such way, that some points remain undiscarded, although it is known that they do not contain the optimal value.

2.2. Division of search range

In our case, the search range is continuous three-dimensional cube. This technique definining the search order and points where object function computation is performed, has to be chosen. Independently from technique, the search range should be divided to smaller ranges. Search range can be divided into a set of strictly identical elements or a set of elements depending on results of computations.

2.3. Division of search range to variable parts

Search range may be divided to a set of elements, where every element depends on the results of computation. Using this method, all search range can be divided into 8 separate parts of cuboid (see Fig. 4) and values may be calculated only in one point of a cuboid boundary. Once it is obvious that particular area may be discarded, the remaining range is divided into 7 parts and every one of them is computed as well.

Octree is sufficient for such range division, but known information for data division is not used efficiently. In addition to that, the same point which is partially computed would be examined again in next computation, so finding the next global maximum may take a while. In order to use our known info more effectively, it is better to examine middle range point instead of a boundary and divide the search range into 27 parts instead of 8 according to Fig. 5.

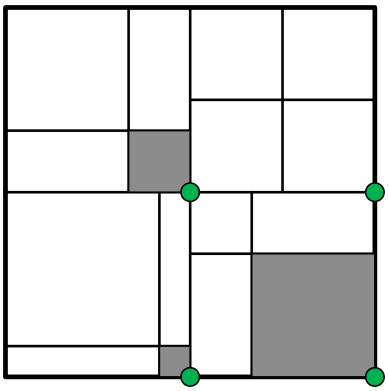


Fig. 4. Example of division of a 2D search range into 4 parts. Circles mark points where the value of function is calculated, grey rectangles mark discarded parts from a search range, white rectangles mark parts to be computed.

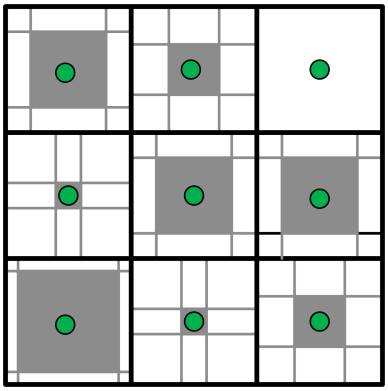


Fig. 5. Example of division of a 2D range into 9 parts, where the mid-range point is examined. Remarks are identical to Fig. 4.

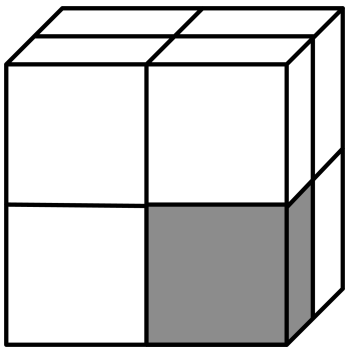


Fig. 6. Example of discrete division of search range into 8 parts.

To manage separate parts of a search range divided this way, a tree where every treetop would have no more than 27 children values. More difficult structure is applied in the second case, but ranges bigger up to 8 times may be discarded, it is also possible that new max values would be found more quickly using the second technique.

2.4. Division of search range into sized parts

Search range may be divided into a predefined amount of even-sized parts, as it is represented by a net in Fig. 6. A point is chosen in every part systematically and, if it is seen that the part can be discarded - it is done so. Otherwise, the part is further divided into even smaller parts and every of them is computed as well. This way, discrete range is received instead of continuous. The advantage of this technique is notice of size of parts that the range is divided into, although it may take by far much more divisions to discard certain areas.

Decision was made that in order to effectively use known information, the mid-point of an area should be investigated by using the dynamic partition technique. To manage it effectively, a tree where every node has no more than k children where k is the amount of parts the range is divided at every step.

Partition of a search area to variable parts enables the reduction of search range every time, so the range would constitute of a set of different sized search areas.

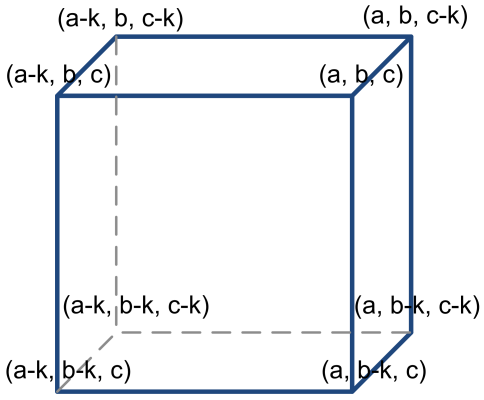


Fig. 7. Boundaries of search area in 3D case.

In order to assign the computation to processors more efficiently after parallelization, it is needed to even out these areas. In order to achieve it, this technique may be transformed to even out the areas or more sophisticated techique might be used. This problem is avoided by using the second division technique where every part is even, though it may be less effective in serial computation because the search range is divided into bigger parts. The second technique was chosen after analysis of every case to make it more effective in parallel computation case.

2.5. Evaluation of search range without solution

In a process of problem solution, quite often occurs a case, when the solution is not at the examined point. It is needed to find out whether the solution exists in the whole search area, and how to evaluate it if it does exists. A closer look has to be taken into the examined area. In 3D case, the area is a cube, represented in Fig. 7.

In order to evaluate whether a point representing the combination of emitters exists in this cube, a set bound by the cube has to be moved to the colour diagram. Separate task was visualised using tridimensional cube model. Fig. 8 represents the search area of point D_{65} by Planck curve.

The most convenient way to be sure that a point belongs to a set accessible with our current light sources is fixing two points, named a^*, b^* and checking whether a solution exists in a set of points c^* through c^*-k. Every possible combination should be checked in this way. An array holding colour characteristics for every wavelength is the most convenient way to analyse variants presented in Table 3.

Importance of Lipschitz constant. Once a proper point is found, its distance from the center of a cube should be evaluated and its minimal value may be found by using the Lipschitz constant. Algorithm calculating the maximal difference between objective function value of light source (when wavelength of one of the initial LEDs changes by 0.5 nm) was used to define the value of Lipschitz constant L, which is essential for the search results.

The resulting value is bigger than it was expected - L=40.9127. This means, that in the worst case, when one LED changes by 1 nm, the CRI may increase or decrease by a value equal to $(2 * L)$=81.8254 nm.

3. Description of algorithms

3.1. Serial algorithm

In serial execution case, the program does not need a structure of data storage. The program execution is performed as follows: data for initial computations is retrieved, the initial search area is formed and, using the recursion, depth search is performed. Operations presented in Table 4 are performed on every program execution step.

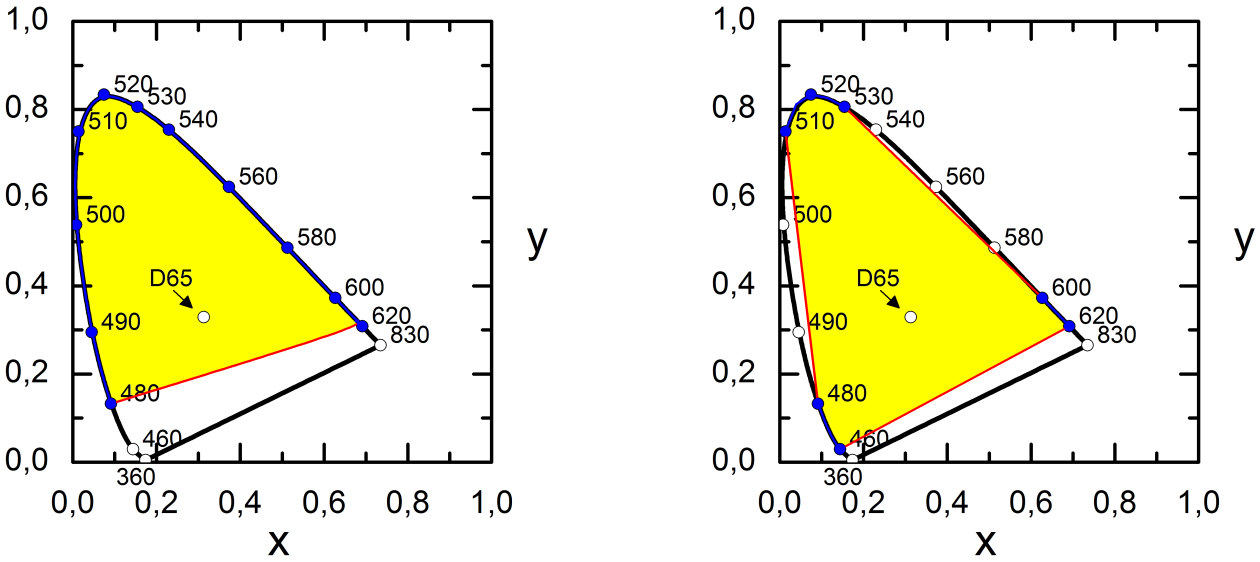


Fig. 8. Colour area bound by a different search cube:
left) a=620, b=620, c=620, k=140;
right) a=620, b=530, c=480, k=20.

3.2. Parallel algorithm

The *master-slave* communication structure is adopted for these computations. This structure defines the way of communication when one process assigns the jobs for the other processes, known as slaves, and retrieval of their results. This communication structure is centralized and the effectiveness in a case of a specific problem depends on the level of communication between the processes.

Four programs implementing parallel computations were investigated. They differ in the way of communication between processes. In the case of the first program, slaves communicate with the master process only to request a job or return a result. In the second program, this is defined by a system of exchanging the information about new or current global max value. In the third program, slaves inform the master process about a new max value found, and, if it is not found for a specific amount of cycles, a slave requests the master for the newest search results. The fourth program is similar to the third one, except the last job which is divided furthermore.

Output data depends on two parameters - Lipschitz constant L and constant of minimal cube border length m. While increasing of Lipschitz constant L, the search velosity decreases, but the accuracy of search increases.

Algorithms were tested in accordance with the following parameters: Lipschitz constant L=81.8254, minimal cube border length m=0.5 nm, reference light source - D65, the initial search range is divided into 125 parts. Additional data is required for the program to operate: colour matching functions, standard 8 - reflection functions, spectrum of a reference light source.

4. Testing and evaluating

In order to evaluate a parallel algorithm, it is essential to know the amount of time it takes for serial algorithm to find a solution - the time amount of time found experimentally was 19231 seconds. The solution was the best combination of emitters defined by wavelengths: 462.996; 540.473; 610.973. The CRI value of the trichromatic system with these wavelengths is 89.1242. The result agrees with condition and outlooks presented in Ref. [2]. Fig. 9 represents the original result: spectral power distributions of standard D65 source and CRI-optimal trichromatic light source.

Parallel algorithm - case of minimal communication. Test results are presented in Table 5a. The initial search range is divided into 125 parts.

Table 3. Search variants for separate task.
Selecter parameters:a=620, b=530, c=480, k=20.

a	b	c
620	530	480÷460
620	510	480÷460
600	530	480÷460
600	510	480÷460
480	530	620÷600
480	510	620÷600
460	510	620÷600
460	530	620÷600
620	480	530÷510
620	460	530÷510
600	480	530÷510
600	460	530÷510

Table 4. Steps of the search algorithm.

1.	Emitter spectra are generated.
2.	Decision whether it is possible to obtain a reference chromaticity from a combination of LEDs is made. If it is true:
2.1.	total spectrum is generated and its CRI is calculated;
2.2.	if the CRI value is higher than any previous, it is memorized along with the point.
3.	If the specific light source can not be composed from current LED combination, a search defining whether there is a point in search area defining the initial emitters suitable for the needed light source, is performed:
3.1.	if a point exists, CRI is evaluated at that point;
3.2.	if the value is higher than any previous, it is memorized as well as the point. Distance from the center of a search range is performed and by using the Lipschitz condition, the lower boundary of a mid-point in cube is evaluated.
4.	If a solution exists and the border of cube is equal to or higher than the determined lower boundary, we make a decision whether another solution may exist in the cube by using the Lipschitz condition. If it may exist:
4.1.	we divide the search cube into 8 parts;
4.2.	perform the search for the first part;
4.3.	if the conditions are still met, the search is performed for the second part and so on.

Master process assigns jobs with initial maximum value to slaves and idles until slave returns his results and, if there are any jobs left, it assigns a new job to a slave. Slaves perform a job similar to the serial algorithm. The highest efficiency was achieved with 4 processors, slight increase in performance is also seen when using 12 nodes. Speedup characteristics are constantly increasing, although the rate drops rapidly, until 20 working nodes were reached.

Parallel algorithm - case of internal (extra) communication. Test results are presented in Table 5b. The initial search range is divided to 125 parts. Master process assigns tasks with initial max value to slaves. Master idly waits for reports on process: it may be either new max value found, or the assigned job complete. Then, if needed, master process updates it's data concerning the max value and transmits it to slave. In a second case, master assigns a new job to slave or, if there are none left, sends him an end message. Slave processes execute a slightly modified parallel algorithm. They arrange the initial data and create a structure storing the search ranges. After that, they evaluate the mid-point of cube. If a new max value is found during that, it is sent to master process. After that, slave process waits for an update of currently highest max value. Once slave finishes its work it sends a message to master process and either gets a new job or an end message.

In this case, the highest efficiency value is received using 4 nodes aswell, the second way of increase in efficiency is seen when node count reaches 12. The highest point of speedup value is achieved when using 20 nodes and after that, even a drop is noticeable. It is possible, that it is a consequence of higher level of communication between nodes.

Parallel algorithm - case of moderate communication. Test results are presented in Table 6a. This algorithm is analogical to one used in the second case, except the slave process has an extra counter, which is used to synchronize the max value with the master process.

Table 5. Test results of parallel algorithms: a) case of minimal communication; b) case of internal (extra) communication.

a)				b)			
Node count	Completion time	Speedup	Efficiency	Node count	Completion time	Speedup	Efficiency
2	15204.80	1.265	0.632	2	15776.45	1.219	0.609
3	8192.25	2.347	0.782	3	8384.12	2.294	0.765
4	5842.66	3.291	0.823	4	5795.79	3.318	0.830
5	4753.81	4.045	0.809	5	4766.80	4.034	0.807
6	4254.51	4.520	0.753	6	4115.99	4.672	0.779
7	3810.73	5.047	0.721	7	3921.33	4.904	0.701
8	3577.95	5.375	0.672	8	3542.58	5.429	0.679
9	3242.96	5.930	0.659	9	3211.48	5.988	0.665
10	2958.71	6.500	0.650	10	2892.39	6.649	0.665
12	2378.63	8.085	0.674	12	2278.54	8.440	0.703
15	2151.00	8.941	0.596	15	2082.21	9.236	0.616
20	1974.89	9.738	0.487	20	1827.99	10.520	0.526
25	1975.22	9.736	0.389	25	1944.87	9.888	0.396
30	1986.74	9.680	0.323	30	1901.59	10.113	0.337
35	1884.70	10.204	0.292	35	1923.72	9.997	0.286

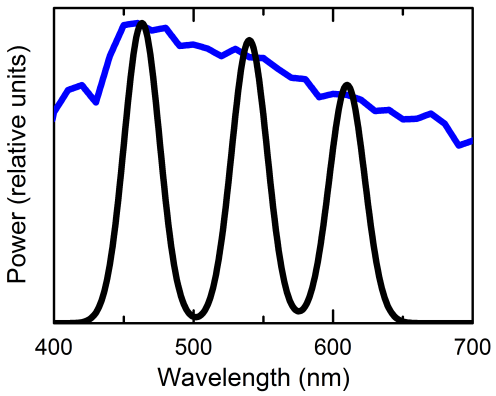


Fig. 9. Spectral power distributions of standard D65 (blue) and CRI-optimal trichromatic (black) light sources.

If the slave does not find a new max value in a specific amount of cycles (20 in this case), it sends a message to master process requesting an update on current highest max value. In this case, similarly to the others, the highest efficiency is achieved using 4 nodes, the next increase, although slighter one is seen at 12 nodes. The max speedup value is reached with 30 nodes.

Parallel Algorithm - 2nd case of moderate communication. Test results are presented in Table 6b. This algorithm is similar to the 3rd one, except the last job is divided to smaller jobs. This is used in order to reduce the idle time on nodes.

In this case, an increase in efficiency is visible until node count does not exceed 6. There is a possibility, that in order to increase the efficiency with more than 6 nodes, several last jobs should be divided instead of just the last one. As observed previously, the highest efficiency is achieved with 4 nodes and slight increase is seen at 12 nodes. The growth in speedup drops drastically, when node count reaches 25.

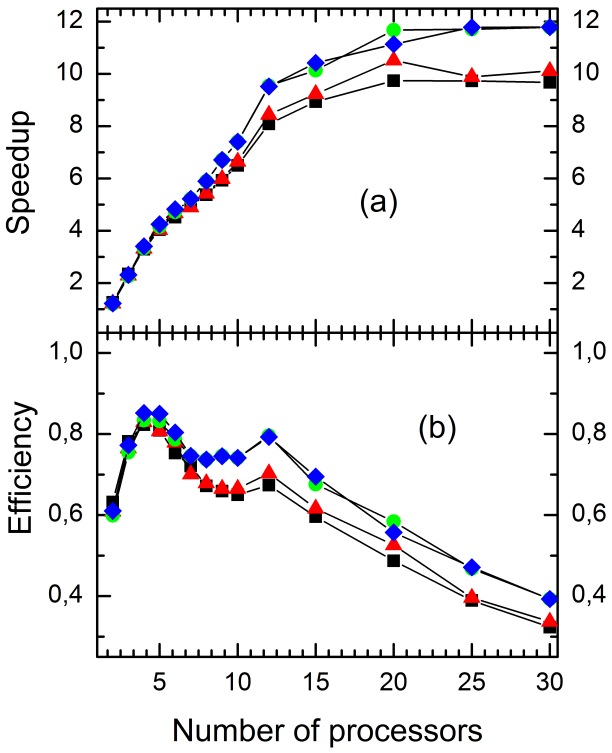


Fig. 10. Algorithm speedup dependence (a) and algorithm efficiency dependence (b) on number of CPU.
Red triangle - internal communication; Black square - minimal communication; Green circle - regular communication; Blue diamond - regular communication, case 2.

Comparison of algorithms. Fig. 10 visualizes speedup dependence and efficiency dependence on number of CPU.

Table 6. Test results of parallel algorithms, regular (constant) communication: a) 1st case; b) 2nd case.

a)				b)			
Node count	Completion time	Speedup	Efficiency	Node count	Completion time	Speedup	Efficiency
2	16063.44	1.197	0.599	2	15751.05	1.221	0.610
3	8490.68	2.265	0.755	3	8307.36	2.315	0.772
4	5768.01	3.334	0.834	4	5644.77	3.407	0.852
5	4624.90	4.158	0.832	5	4523.91	4.251	0.850
6	4073.22	4.721	0.787	6	3988.68	4.821	0.804
7	3683.34	5.221	0.746	7	3684.91	5.219	0.746
8	3259.90	5.899	0.737	8	3261.36	5.897	0.737
9	2866.05	6.710	0.746	9	2867.01	6.708	0.745
10	2595.43	7.410	0.741	10	2595.88	7.408	0.741
12	2013.87	9.549	0.796	12	2020.50	9.518	0.793
15	1896.29	10.141	0.676	15	1845.22	10.422	0.695
20	1646.92	11.677	0.584	20	1726.12	11.141	0.557
25	1642.23	11.710	0.468	25	1632.17	11.783	0.471
30	1632.18	11.782	0.393	30	1631.44	11.788	0.393
35	1632.45	11.780	0.337				

It is obvious that speed of computations increases the most until the node count does not exceed 12. Also, as observed more than 20 nodes for these computations should not be used either, because the speedup increase is slight or even negative. During parallel computations efficiency maximum value is achieved with 4 nodes, slight increase is seen at 12 nodes as well.

We can see in the chart, that the curves of efficiency for every algorithm are almost identical until node count reaches 4. The highest efficiency is achieved by a program of regular communication with the division of the last job. In the case, when we want to get the solution faster by using more nodes, it is best to use 12 nodes and one of the regular communication programs. In addition to that, it might be useful to try to divide several of the last jobs to reduce the idle time even more.

Conclusions

The optimization problem was successfully solved and the speedup of parallel algorithm was investigated experimentally. It was found that the magnitude of the Lipschitz constant has huge affect on the overall computing time. The most "interesting" ranges with positive CRI value have much lower Lipschitz value, therefore in order to reduce the computation time, several Lipschitz values could be used: for positive and negative CRI values. Also, a technique could be developed to discard negative search areas more efficiently, but ensuring that none of the possible optimal points were discarded.

In this work, technique enabling modelling of multicolour light sources was examined. A branch and bound algorithm was adapted for solving this problem and software, enabling the search of optimal solution on a distributed computer network, was created. Several parallel algorithms using centralized data exchange scheme were created and tested. There was no need for a decentralized data exchange scheme as communication between processes is not frequent (new max value is found rarely).

VU MIF distributed computer network was used to test these algorithms and recommendations for such algorithm performance on such computing network were given. During the research, the node count providing the best efficiency was determined: the highest efficiency is achieved with 4 nodes, but there is always an increase in efficiency at around 12 nodes. This may be useful, when the problem has to be solved quickly by increasing the amount of nodes, to maintain efficiency. In all cases, using more than 20 nodes should be avoided, because the speedup increase is minor or even negative, therefore the efficiency drops significantly.

References

1. Aistė Ambrazevičiūtė, Rimantas Vaicekauskas. Modelling of multicolour light source using distributed computing network. 1. Statement of the physical and computational problems – *Innovative Infotechnologies for Science, Business and Education,* ISSN 2029-1035 – 1(10)(2011) 13-20.
2. Žukauskas A., Vaicekauskas R., Ivanauskas F., Gaska R. and Shur M. S. Optimization of white polychromatic semiconductor lamps – *Appl. Phys. Lett.* 80 (2002) 234–6.
3. Wyszecki, G. and Stiles, W. S. Color Science: Concepts and Methods, Quantitative Data and Formulae – New York: Wiley, 2000.