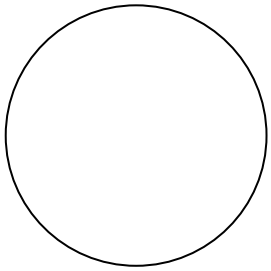
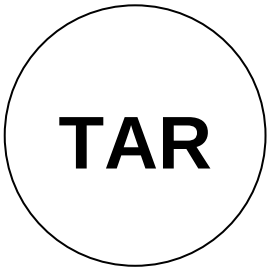




The IPSI BgD Transactions on Advanced Research

Multi-, Inter-, and Trans-disciplinary Issues in Computer Science and Engineering

A publication of
IPSI Bgd Internet Research Society
New York, Frankfurt, Tokyo, Belgrade
January 2007 Volume 3 Number 1 (ISSN 1820-4511)

Table of Contents:

Pearls of Wisdom by Nobel Laureates:

Another View on Computer Architecture
Wilson G., Kenneth 3
The Key to Innovation
Friedman, Jerome 4
Number and Organization of Primary Memory Objects in the Brain
De Gennes, Pierre-Gilles. 4

Invited Paper:

Flight Performance of Planetary Atmospheric Flight Airship (PLAS)
Fujii, Hironori A. 5

University of Belgrade Research Efforts:

Advances in Symbolic Simulation of Systems
Tošić, V., Dejan; and Lutovac, D., Miroslav 9
Specifying Sequent Calculi Rules for Managing Some Redundancies in Proof Search
Lutovac A., Tatjana 15
Accelerating Conjugate Gradient Solver: Temporal Versus Spatial Data
Korolija G., Nenad 21
The Pattern-Oriented Decision-Making Approach
Delibašić, A., Boris; and Suknović, B., Milija 26

***Regular Contributions*:**

Development of User-Friendly Didactic Climate Models for Teaching and Learning Purposes
Goyette, Stephane; Platteaux, Herve; and Jimenez, Francois 32
Knowledge Processing and Computer Architecture
Omerovic, S.; Tomazic,S.; Milovanovic, M.; and Torrents, D. 39
Development of a Biomechanical Knowledge System to Identify Brain Injuries in Emergency Department
Kou, Zhifeng; and Ziejewski, Mariusz 47
Literature Review of Water Demand
Milutinovic, Milan 55

The IPSI BgD Internet Research Society

The Internet Research Society is an association of people with professional interest in the field of the Internet. All members will receive this TRANSACTIONS upon payment of the annual Society membership fee of €100 plus an annual subscription fee of €1000 (air mail printed matters delivery).

Member copies of Transactions are for personal use only

IPSI BGD TRANSACTIONS ON ADVANCED RESEARCH

www.internetjournals.net

Another View on Computer Architecture

Wilson G., Kenneth (1982 Nobel Laureate)

The coming of the computer has created a revolution as profound as the change from the Middle Ages to the Renaissance. Many of the changes that took place around the time of the Renaissance - the invention of printing, the development of systematic experimental science, the invention of oil painting - have analogs today, made possible by the computer. We are moving from printed media communication, with time delays of a year or more for professional publications, to instantaneous communication via computer networks. Computers are revolutionizing the capability of scientific instruments. Supercomputers are enabling man to “see” phenomena not even accessible to experiment - from tomorrow’s weather, to the complete billion-year history of a star, to the deep interior of the earth. The ability of computers to sort information is giving mankind unprecedented capability to find needles in our rapidly growing haystack of knowledge.

In the past forty years, the power of computers has advanced by a factor of a million or so. Nevertheless, the computer revolution has only just begun. The technological opportunities for further advances seem almost limitless. Since the bit carries no weight or other mechanical burdens, one can expect the volume assigned to a single bit in processors, communications, and memory, to continue to shrink dramatically, vastly increasing the number of bits that can be handled at a time. The needs for computing power are likely to keep pace with any technological advances that come along, due to the many problems of exponential or close to exponential complexity that computers must deal with - from economic forecasting to probing the secrets of molecules.

Unfortunately, there is one constraint from the discipline of physics which is limiting and shaping computer architectures of today and into the future. There is a maximum speed with which bits can travel, namely the speed of light, and today’s computer designs already suffer from this limitation - forcing supercomputers to become smaller and smaller as their speed increases. The speed of light limitation is forcing architects to achieve new levels of processing capabilities mostly through parallelism rather than speed. As silicon chips (or whatever replaces silicon in future) become three-dimensional and the bit continues to shrink, the number of bits that can be processed in parallel could increase in spectacular fashion - is Avogodro’s number (the number of atoms in a few grams, or 10^{23}) out of reach? Clearly the challenge to computer architects is to harness the capabilities of bits processed in parallel for the benefits of man- and womankind.

Finally, I remind all readers already deep into the jargon of silicon circuits that the brain puts all silicon circuits that the brain puts all silicon to shame. The brain has cycle times of milliseconds, and a size smaller than even a desktop computer, yet it recognizes patterns, analyzes speech, and stores and sorts through databases, all at rates that are untouchable even by supercomputers. Its programming system is natural and user-friendly. Only its fault-tolerance does not meet engineering standards.

The Key to Innovation

Friedman, Jerome (1990 Nobel Laureate)

The development of Homo Sapiens has been a history of innovations, from the earliest crude tools to the modern technological society of today. The growth of science and technology has been exponential during the last century; and under the right circumstances, this rapid growth can be expected to continue.

The major innovations of the future - those that will shape the society of the future - will require a strong foundation of both basic and applied research. It is ironic that quantum mechanics, one of most abstruse conceptual frameworks in physics - one that was developed to explain atomic spectra and the structure of the atom, lies at the foundation of some of our most important technological developments, because it provided the understanding of semiconductors that was essential for the invention of the transistor.

Quantum mechanics thus contributed directly to the development of technologies that gave us world wide communication, computers with their applications to all phases of modern life, lasers with many diverse uses, consumers electronics, atomic clocks, and superconductors - just to mention a few. The internet and the world wide web, which are profoundly reshaping the way we communicate, learn, and engage in commerce, owe their origins in a deep sense to the physicists of the past who worked to understand the atom. In modern industrial nations, quantum mechanics probably lies at the basis of a sizable fraction of the gross national product. This is but one example, and there are many others in all areas of science that demonstrate this point.

It is clear that innovation is the key to the future and the human drive to understand nature is the key to future innovation. Society must do all that it can to preserve, nurture and encourage curiosity and the drive to understand.

Number and Organization of Primary Memory Objects in the Brain

De Gennes, Pierre-Gilles (1991 Nobel Laureate)

A memory area contains a large number ($N \sim 10^8$) of neurons, each of which is connected with many neighbors (number of efferents: $Z \sim 10^4$). But the connections are poor: the probability for one connection to be efficient is $p \sim 10^{-2}$. This is important: different memory objects must be independent.

We need to know how a definite memory object can be stored on a cluster of well connected neurons, and what is the statistics of these clusters. The average number M of neurons per cluster is contained within two limits: if M is too small, the memory is not faithful. If M is too large, the storage capacity is too small.

Various consequences of this picture have to be researched.

Flight Performance of Planetary Atmospheric Flight Airship (PLAS)

Fujii, Hironori A.; Kusagaya, Tairo; and Watanabe, Takeo

Abstract—*This paper studies flight performance of airship employed to observe scientifically the atmosphere of planet including Mars and Venus with little effort or fuel expenditure (Planetary Atmospheric Flight Airship: PLAS). The flight region of the planetary airship is determined taking into consideration the temperature and pressure on Mars and Venus, as well as basic limitations of airships. The performance of the atmospheric flight airship is studied on its longitudinal dynamical feature. Result of the study shows some interesting features of the airship flying on Mars and Venus.*

Index Terms— *Airship, Planet, observation*

1. INTRODUCTION

VENUS has been observed scientifically using balloons in 1985 as a result of cooperation between the Soviet Union and France (VEGA 1 & 2)[1]. These balloons had a diameter of 3.4 m and flew over one third of Venus in 46.5 hours at an altitude of 53.6 km. Future plans of ISAS/JAXA and NASA include a Mars airplane[2,3] and other planetary balloons[4,5], however, balloon is an only platform that has been successfully used.

The present study employs remotely piloted airships as a platform in the atmosphere for planetary observation[6-8]. The planet's surface topography, gravitational field, magnetic field, and atmospheric layer can be observed over an extended period of time with little effort or fuel expenditure. An observational platform can be located in a target planet's atmospheric layer from 50 m for a ground probe to as high as the altitude used for satellites. Data obtained from such platforms can be used to examine the general characteristics of the planet and its atmospheric layer and also to validate and to complement the data obtained from satellites or ground probes.

Fujii, Hironori A.; Kusagaya, Tairo; and Watanabe, Takeo
Department of System Design, Tokyo Metropolitan Institute of Technology 6-6, Asahigaoka, Hino, Tokyo 191-0065, Japan. E-mail: fujii@tmit.ac.jp.

Figure 1: Planetary Atmospheric Flight Airship (PLAS)

Planetary observations can be classified into 6 categories depending on the location of the observational platform. These six categories include the remote sensing of the target planet: (1) using telescopes; (2) using a space telescope that is located in another planet's orbit; (3) a spacecraft that is in an interplanetary orbit, such as Mariner 2 and 5; (4) from spacecraft[2] that are in the target planet's orbit, like Venera 15-16 and Pioneer Venus 1; (5) or direct observation at the time of descent under the target planet's gravitational field, such as an observation by Venera 1-7, or by flying a probe such as VEGA 1-2; and (6) or direct observation from a fixed point probe on the surface of the target planet, such as that performed by Venera 8-14 or the Rover Mars Pathfinder[3]. Study of a planet using planetary airships in its atmosphere is now being studied by ISAS[4,5], while NASA[9-12] is under study of the use of planetary balloons located in the atmosphere of the target planet. A team of Georgia Institute of Technology is studying the use of a Mars helicopter, "GTMARS[13]", while the University of Maryland the use of a Martian Autonomous Rotary-wing Vehicle (MARV)[14].

Information obtained using the methods described above is usually available either for a microscopic area over a short period of time or for a macroscopic area with low resolution. In the present study, an airship platform is presented of a subcategory of category (5) that include probes with "remote sensing or direct observation at the time of descent under the target planet's gravitational field or by flying a probe." One of the advantages of such an airship is that energy is not needed to produce the dynamic lift that would support the weight of the spacecraft. Instead, this lift results from the buoyancy of gas inside the airship. The airship is never in danger of falling, and its flight control permits a large observational area. Moreover, the speed of the spacecraft is slow enough so that the communication delay from Earth is covered.

The flyable region is already studied by the authors[6] based on the pressure and temperature in the atmospheres of Venus and Mars, and the required size and mass of the spacecraft.

2. FLYING CHARACTERISTICS OF PLAS

2.1 Review Stage

Papers can be submitted only electronically, as indicated on the web site. The longitudinal characteristics of PLAS is studied for flying characteristics under the change of flight atmosphere as the planet. The shape and dimensions of PLAS are fixed in the comparison between the flight atmosphere of Mars and Venus with Earth as a reference. The PLAS is modeled as shown in Fig. 2 and Table 1 with reference to the manned airship of type WDL-1. The dynamics of the model is numerically analyzed by the linear analysis employed with the small perturbation theory and a software MATLAB is employed in the dynamic analysis. The gravitational acceleration and atmospheric density are changed as the environment of the planet. Mass of PLAS is assumed to change in accordance with the change of atmospheric density. The changed mass is assumed to concentrate in the center of mass and the moments of inertia of the PLAS is assumed not to depend on atmospheric density.

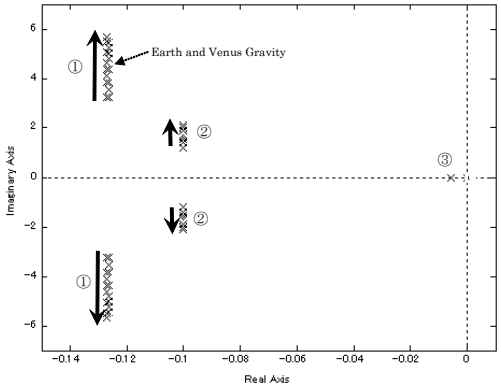


Figure 2: A model of PLAS

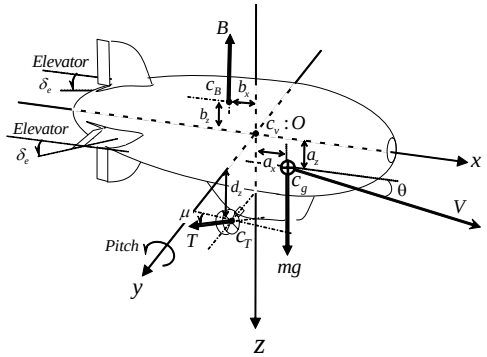


Figure 3: Root locus for the atmospheric density from 0.01kg/m^3 to 7.0kg/m^3 (gravity 1G, and flight velocity 0.5m/s)

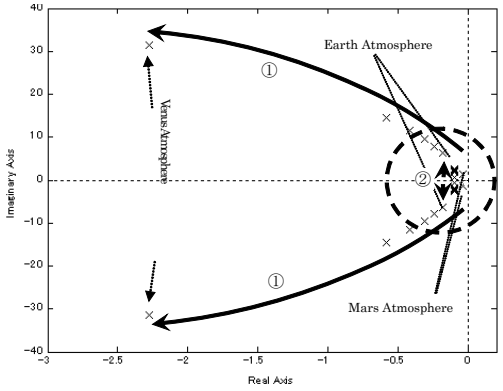


Figure 4: Root locus for the atmospheric density from 0.01kg/m^3 to 1.0kg/m^3 (gravity 1G, and flight velocity 0.5m/s)

Figure 3 shows the time response of the pitching motion for the PLAS in the atmospheric density from 0.01kg/m^3 to 1.0kg/m^3 with flight velocity, 0.5m/s, and a step input of 0.2N in thrust is applied for the motion. The atmospheric environment is earth and the period of motion is 4.8 sec. The short period mode is shown in the figure by (1), and the long period mode by (2), and the linear motion in flight by (3). It is seen the locus moves to bold arrow as atmospheric density increases. The short period mode is seen to be more stabilized and the frequency increases as the atmospheric density increases. The stability is seen insensitive for the atmospheric density but the frequency increases for the long period mode. The dotted circle is shown in Fig.4 enlarged. The atmospheric density is 7.0, 1.2, and 0.02 kg/m^3 for Venus in the altitude 35km, earth on the surface, and Mars on the surface, respectively.

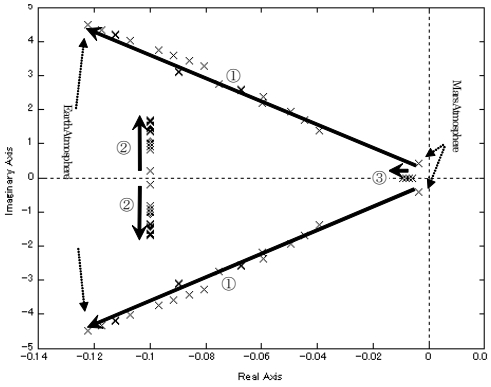


Figure 5: Root locus for the gravity acceleration from 0.5G to 1.5G (The air density 1.2g/m^3 and flight velocity 0.5m/s)

Table 1: Data for the atmosphere[15]

	Venus	Earth	Mars
Mean distance from Sun	1.082 x 10^8 [km]	1.496 x 10^8 [km]	2.279 x 10^8 [km]
Solar day	117 [days]	1 [days]	1.0287 [days]
Surface gravity	8.87 [m/s^2]	9.81 [m/s^2]	3.72 [m/s^2]
Mean atmospheric temperature(surface)	735 [K]	288 [K]	215 [K]
Mean atmospheric pressure(surface)	92x10^5 [N/m^2]	1x10^5 [N/m^2]	0.007x10^5 [N/m^2]
Composition of the terrestrial planet atmospheres	CO_2 : 96.5 %, N_2 : 3.5 %	N_2 : 77 %, O_2 : 21 %	CO_2 : 95 %, N_2 : 2.7 % , Ar : 1.6%
Mean molecular weight of atmosphere	43.4	28.96	43.5
Lapse rate of lower atmosphere	7.8 [K/km] (AGL 0 to60km) 8.6 [K/km] (AGL 60)	6.5 [K/km](AGL 0 to11km) 0.0 [K/km](AGL 11 to 20km) -1.0 [K/km](AGL 20 to 32km) [under ISA]	2.5 [K/km] (AGL 0 to40km)
Wind Speed	Surface : 1 to 1.5 [m/s] cloud deck : 100[m/s]	7 to10 [m/s]	Max.: 90 to100 [m/s]
Round trip of radio transport time from Earth	4.6 to 28 [min]	0 [min]	8.7 to 41.9 [min]

Figure 5 shows the time response of the pitching motion for the PLAS in change of the gravity acceleration from 0.5G to 1.5G with flight velocity, 0.5m/s, and a step input of 0.2N in thrust is applied for the motion. The atmospheric environment is earth and the period of motion is 4.8 sec. Dotted arrows point the short period modes for the earth with 9.81m/s^2 and Venus with 8.87m/s^2.

It is seen from the figure that the frequencies of motion increase but stability stays same for the change of the gravity acceleration. It is understood that the effect of gravity is not serious since the gravity balances with buoyancy and stability is not affected by the gravity.

It is concluded that the PLAS at the altitude 43km of Venus has the short period mode with increased stability and frequency of motion and the long period mode with identical stability and increased frequency of motion. It is also concluded that stability and frequency of motion decreases for both of the short and long period modes of the PLAS on the surface of Mars due to the effect of atmospheric density and gravity acceleration.

It may be recommended to apply similar analysis for unknown planets to obtain information of flight environment through the motion of PLAS.

3. CONCLUSION

This part can be broken in a as many sections and subsections as needed. The airship flying on the atmosphere of planet (PLAS) is studied in its flight performance with parameters of such atmospheric environment as atmospheric density and gravity acceleration. The analysis is applied to PLAS on Venus and Mars with Earth as a reference in which the atmospheric parameters are relatively known. Results of the study on the longitudinal performance have shown that the implementation is possible for the PLAS for the flight under the cloud of Venus.

It is recommended that the similar analysis can be applied for unknown planets to obtain information of flight environment through the motion of PLAS.

REFERENCES

[1] Kremnev, R. S., "VEGA Balloon System and Instrumentation," Science, 231, No.4744, 1986, pp.1408-1411.

[2] Venus Exploration Working Group of the Institute of Space and Astronautical Science, *Venus Mission Proposal,* Jan. 2001, Japan, ISAS, p 269 (in Japanese).

[3] Siebert, M.W., and Keith, Th.G. , "NASA Mars Pathfinder Mission," *Lubrication Engineering,* Vol.54, No.12, 1998, pp. 13-19.

[4] Izutsu, N. and Yajima, N., "Inflatable Venus Balloons at Low Altitude," *The Institute of Space and Aeronautical Science Report No.44*, ISAS, Sagamihara, Mar. 2002(in Japanese).

[5] Yajima, N., Izutsu, N., Honda, H., Goto, K., Sato, E., Imamura, T., Akazawa, K., and Tomita, N., "Extended Possibility of Planetary Balloons," *The Institute of Space and Aeronautical Science Report Special* Vol.44, Mar. 2002 (in Japanese).

[6] Kusagaya, T., Kojima, H. and Fujii, H.A.: Estimation of Flyable Regions for Planetary Airships, AIAA J. of Aircraft, 43(2006),pp.1177-1181.

[7] Kusagaya, T., "Unmanned Outdoor Blimp for Multi Remote Sensing," *Proceedings of 2nd International Airship-Conference*, Stuttgart Germany, July 1996, pp.23-34.

[8] Kusagaya, T., “Japan RPA ‘MAMBOW 3’,” *The Journal of the Airship Association*, No.112 pp.10, No.114, pp.14, 1996, UK.
[9] Landis, A. G.., LaMarre, C., and Colozza, A., “Atmospheric Flight on Venus,” 40th Aerospace Sciences Meeting & Exhibit, 14-17 January 2002, Reno Nevada, NASA TM-2002-0819(AIAA-2002-0819).
[10] Landis, A. G., “Exploring Venus by Solar Airplane,” STAIF Conference on Space Exploration Technology, Albuquerque NM, Febrary 11-15, 2001, AIP Conference Proceedings Volume 552, 2001, pp.16-18.
[11] Landis, A. G., “Solar Flight on Mars and Venus,” 17th Space Photovoltaic Research and Technology Conference, Cleveland OH, November 10-13, 2001, NASA Proceedings CP-2002-211831, pp.126-127.
[12] Neck, K., Balaram, J., Heun, M., Smith, S., and Gamber, T., “Mars 2001 Aerobot/Balloon System Overview,” AIAA International Balloon Technology Conference, 1997, AIAA 97-1447.
[13] Kondor, S. and Salinas, R., “Mars Exploration Rotorcraft: Georgia Tech Autonomous Rotorcraft System (GTMARS),” *Report of the American Helicopter Society Student Design Competition*, June 2000, pp.94.
[14] Datta, A., Roget, B., Griffiths, D., Pugliese, G., Sitaraman, J., Bao, J., Liu, L., Gamard, O., “Design of a Martian Autonomous Rotary-Wing Vehicle, ” *Journal of Aircraft*, Vol. 40, No. 3, 2003, pp.461-472.
[15] Shirley, J. H. and Fairbridge, R. W., “Encyclopedia of Planetary Sciences,” Chapman & Hall, London, 1997, pp.48-57, pp.432-455, pp.705-706, pp.887-905.

Advances in symbolic simulation of systems

Tošić, V., Dejan; and Lutovac, D., Miroslav

Abstract—*A framework and recent advances in symbolic simulation of discrete-time and continuous-time systems are presented. The role and application of symbolic analysis in modern engineering are highlighted. A software realization of a symbolic system simulator is presented and exemplified. Real-life application examples are presented in which systems are symbolically solved and simulated with Mathematica. We introduce an original approach to algorithm development, system design and symbolic processing that successfully overcomes some problems encountered in the traditional approach. Benefits of symbolic methods and the role of computer algebra systems are highlighted from the viewpoint of both academia and industry.*

Index Terms — electronic circuits, SALEC, SchematicSolver, symbolic simulation, systems

1. INTRODUCTION

COMPUTER-AIDED simulation of electronic circuits and systems is a mature field, as evidenced by the wide usage of simulation software tools. The great majority of currently available programs belong to the "numeric" category in the sense that their outputs are numbers. Symbolic analysis and simulation, on the other hand, is aimed at producing outputs as closed-form expressions that contain variables and numbers [1]-[7].

Successes in automating the design of basic analog building blocks such as opamps and comparators have almost uniformly employed an "equation-based" approach that substitutes analysis equations for simulation in order to predict the performance of an analog circuit. Symbolic analysis can be used to automatically generate a significant fraction of analysis equations needed to characterize a new circuit topology. Therefore, symbolic analysis is an important step forward in the development of CAD tools that aid in analog circuit design [2].

Symbolic analysis can provide many results which are simply not available from numeric simulation methods. Most importantly, they can provide explicit insight into the dominant behavior and properties of a circuit or system. Important application of the insight obtained from symbolic analysis is the development of the equations which are required in the use of optimization techniques to provide solution to particular design specifications. In addition, symbolic analysis can also be used in compiled-code evaluation for statistical analysis, and automated synthesis or failure diagnostics of systems [3].

The enormous increase in computing power of the present computers, combined with the development of new and more efficient analysis algorithms, allows for symbolic analysis of larger and more complex systems in shorter time than was possible before.

In this paper we present our recent advances in symbolic simulation of continuous-time and discrete-time systems. Section 3 introduces the basic terminology and definitions. Section 4 focuses on general application aspects of symbolic simulation. Section 5 highlights practical application issues of symbolic techniques. Section 6 introduces multirate systems symbolic simulation. Sections 7, 8, and 9 exemplify symbolic simulation by illustrative practical systems and the corresponding symbolic analysis.

2. PROBLEM STATEMENT

We focus our research on creating a framework for the symbolic analysis of circuits and systems that is suitable for research as well as industrial and educational applications.

Existing solutions have been focused on numerical system analysis, and their drawbacks are the inherent inability to give insight and analytic expressions for system characteristics.

Suggested solutions to the system analysis are based on new algorithms and original software implementation in *Mathematica*, which generate closed-form expressions for characteristics of microwave circuits and multirate systems.

3. DEFINITION OF SYMBOLIC SIMULATION

Symbolic simulation is a formal technique to calculate the behavior or a characteristic of a system (e.g. digital system, electronic circuit, or continuous-time system) with an independent variable (sample index, time, or frequency), the dependant variables (sample values, signals, voltages, and currents), and (some or all) the system elements represented by symbols.

The symbolic technique is complementary to numerical techniques (where the variables and the elements are represented by numbers) and qualitative analysis (where only qualitative values

Manuscript received December 15, 2006.

D. V. Tošić is with the School of Electrical Engineering, University of Belgrade, Serbia (e-mail: tosic@etf.bg.ac.yu), the contact person.

M. D. Lutovac is with the School of Electrical Engineering, University of Belgrade, Serbia (e-mail: lutovac@etf.bg.ac.yu).

are used for signals, such as increase, decrease or no change) [4].

A *symbolic simulator* is a computer program that receives the system description as input and can automatically carry out the symbolic analysis and thus generate the symbolic expression for the desired system characteristic.

Majority of the symbolic simulation research has concerned the analysis of linear systems in the frequency domain. For lumped, linear, time-invariant (LTI) systems, the symbolic transfer functions obtained are rational functions in the complex frequency variable (*s* for continuous-time systems and *z* for discrete-time systems) and the system elements that are represented by a symbol (instead of numerical value).

Depending on whether the complex frequency is the only variable, and whether all of the system elements are characterized by symbolic parameters, we have three levels of symbolic representation: (1) rational transfer function of the complex frequency with numerical coefficients, (2) partially symbolic transfer function, and (3) fully symbolic transfer function [1].

Majority of reported symbolic simulators [1]-[5] were implemented as compiled code generated with, for example, C or C++. Other simulators [8]-[27] were written as application packages (toolboxes) for computer algebra systems (CAS), such as *Mathematica* [28].

Implementing a symbolic simulator as a CAS toolbox has many advantages: (1) built in functions for the basic and advanced symbolic computations are highly optimized within the CAS, so the simulator relies on the most efficient code to be used for the analysis kernel, (2) the rich functionality of the CAS can be used for post-processing of the results generated by the simulator, (3) the supreme graphics/multimedia CAS support can be exploited for visualization, typesetting, or animation of the simulation results.

4. GENERAL APPLICATION OF SYMBOLIC SIMULATION

The value of symbolic analysis is well recognized in both industry and academia. In industry it has been used as an aid in the design of systems and circuits. In academic institutions it has been found useful as an instructional aid.

There are many reasons why one may be interested in symbolic simulation. A few of the more important ones are as follows [1]:

Frequency response calculation. Suppose that an accurate magnitude response curve is desired over a frequency range with a specified frequency step. With a numerical program such as SPICE [29] or Matlab Simulink [30], the same network or system will have to be analyzed many times. On the other hand, if we obtain the symbolic transfer function first as a rational function with real coefficients, then we need only evaluate it at different values of frequency – obviously a much simpler task.

Parameter iteration. For solving piecewise linear resistive networks the exhaustive segmentation combination method is conceptually simple and can produce all solutions. In this case, the use of a symbolic technique is a natural choice. If we first obtain symbolic expressions for the response, then it is only necessary to substitute the parameters for each segment combination into the expressions and check whether the solutions lie within the ranges defining the segments. If not, that particular segment combination does not yield a solution and we also proceed to another segment combination.

Sensitivity analysis. In the design of any system, it is important to know the effect on the network performance due to the variation of some element values. A precise measure of the effect can be expressed in terms of the sensitivity functions. These functions contain partial derivatives of the system response with respect to element values (system parameters), so symbolic computation becomes a natural approach: find symbolic system function, compute (symbolically) required derivatives, and determine (a closed-form expression for) the desired sensitivity function.

Filter design by optimization. In this approach of filter design, a reasonable network configuration is first proposed, and initial element values are selected from an approximate analysis. The actual frequency response is then calculated and compared with the specified response. The process is repeated until the error is minimized. It is seen that in this approach of network design, the response of the network has to be calculated at many frequency points, and at many different sets of element values. Obviously, if a symbolic transfer function can be obtained first, repeated evaluation of the transfer function will be much simpler job than repeated analysis of the network.

Insight. Symbolic transfer function can provide better insight than numerical solutions. By inspection of the symbolic transfer function, it might be immediately clear how a parameter (or an element value) contributes to the performance and behavior of the system. Without a symbolic transfer function, these conclusions could only be reached after the analysis of many numerical cases, and even then some degree of uncertainty still exists [1]-[8], [26].

Instructional aid. Beginning linear circuit courses and signals and systems courses contain many exercises that ask students to derive the expressions for the transfer function, input impedance, voltage gain, current gain, etc. It is very easy for students (in fact, even for the instructors!) to make minor “math errors” that lead to incorrect answers. Therefore, it would be very helpful to the students to have a tool that (symbolically) checks their answers for

correctness. Discovery of a wrong answer before handing in their work enables them to redo the problem and make corrections.

Symbolic analysis even has the potential to improve the training of young analog circuit designers and to guide more experienced through second-order phenomena such as distortion [2].

5. PRACTICAL APPLICATION EXAMPLES OF SYMBOLIC TECHNIQUES

Symbolic analysis is an intriguing topic in VLSI designs and it is crucial for the applications to the parasitic reduction and analog circuit evaluation [7].

For parasitic reduction, a huge amount of electrical parameters is approximated into a simplified RLC network. This reduction allows designers to handle very large integrated circuits. A symbolic analysis approach reduces the circuit according to the network topology. Thus, the designer can maintain the meaning of the original network and perform the analysis hierarchically.

For analog circuit designs, such as electrical filters or VLSI building blocks, symbolic analysis provides the relation between the tunable parameters and the characteristics of the circuit. Therefore, the analysis allows us to optimize the circuit behavior symbolically [6], [7].

Symbolic simulation of microwave linear circuits characterized by scattering parameters [14] has been found indispensable for generating analytic characterization of specific microwave networks (e.g., circuit models of microwave discontinuities) required by microwave software tools, such as *WIPL-D Microwave* [31].

Symbolic simulation has found successful application in the field of power engineering for computing the DC load flow in electric power systems [13].

Combinational networks were effectively analyzed symbolically by a simulator developed as a *Mathematica* toolbox [25].

Algorithm development can be greatly enhanced by symbolic techniques as reported in [26].

Analog Insydes is a *Mathematica* application package for modeling, analysis, and design of analog electronic circuits, tailored specifically for industrial applications [24].

SchematicSolver is a *Mathematica* application package that allows you to create symbolic representations of systems. It provides functionality for system drawing, solving, simulating, processing, and implementation [20]. The knowledge embedded in the representation can be used to generate implementation code or to analytically derive system properties, such as transfer functions or impulse responses.

SchematicSolver also automatically generates software implementations of linear and nonlinear discrete systems. This function can process symbolic samples: for a symbolic input sequence, you can compute the symbolic output sequence with both the system parameters and the states specified by symbols. Similarly, the transfer function of a complex multiple-input multiple-output (MIMO) system can be derived in terms of system parameters kept as symbols [23], [26].

Symbolic signal processing, an innovative feature of *SchematicSolver* not available in other software, brings you computation of transfer functions as closed-form expressions in terms of symbolic system parameters and can find the closed-form response of schematics. The derived result is the most general because all system parameters, inputs, and initial conditions (states) can be given by symbols [20].

The symbolic algorithm for the elliptic rational function was used to optimize the symbolic performance of analog and digital systems [21]. This optimization is not possible with traditional numeric algorithms. Specific application of this result is derivation of formulas for designing high-speed low-consumption systems known as quadrature mirror filter banks.

Moreover, we found a new function, known as *minimum-Q elliptic*, by symbolically optimizing the elliptic rational function [21]. Minimum-Q elliptic became a standard function in manufacturing integrated filters. In addition, again using symbolic optimization, we implemented a very efficient digital signal processing (DSP) system using programmable logic devices and very large-scale integrated circuits. By an efficient DSP system, we mean processing by multiplierless systems that consist of a small number of adders and binary shifters.

6. MULTIRATE SYSTEMS SYMBOLIC SIMULATION

Multirate systems are important constituents of the modern information communication technology (ICT) infrastructure, such as Internet or multimedia systems. These systems play a very important role in digital signal processing (DSP), and their design is a highly specialized field within ICT/DSP engineering [32].

We are currently developing a new concept of multirate systems simulation as outlined in [27]. Our original software package, being developed in *Mathematica*, is proposed for implementation and symbolic analysis of multiple-input multiple-output systems consisting of upsamplers, downsamplers, adders, multipliers, delays, and shift registers. Some initial works on our multirate symbolic analysis were presented in [20] and [26].

7. EXAMPLE CIRCUIT SIMULATION

Electric circuits are fundamental examples of continuous-time systems. Symbolic simulation of a lumped linear time-invariant electric circuit will be exemplified by the analysis of an active RC differentiator implemented with second

generation current conveyors (CCII) as shown in fig. 1.

Active RC differentiator and integrator networks are widely useful in the analog signal processing applications, such as signal generating, computing, process control, and many test instrumentation circuits.

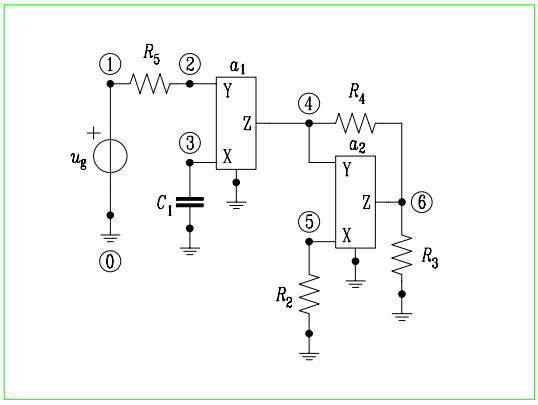


Figure 1. Differentiator with tunable time constant using current conveyors [33].

The schematic specification for the electric circuit of fig. 1 can be generated by DrawFilt [34] and is internally represented as a *Mathematica* symbolic object

```
numberOfNodes = 6
component[1] = {"V", "Ug", 1, 0, Ug}
component[2] = {"R", "R5", 1, 2, R5}
component[3] = {"C", "C1", 3, 0, C1}
component[4] = {"R", "R2", 5, 0, R2}
component[5] = {"R", "R3", 6, 0, R3}
component[6] = {"R", "R4", 4, 6, R4}
component[7] = {"CCII", "cc1", {4,0}, {2,3}, a1}
component[8] = {"CCII", "cc2", {6,0}, {4,5}, a2}
numberOfComponents = 8
```

which is stored in a file (e.g. Differentiator.m). The symbolic object contains all the details necessary for drawing, solving and simulating the circuit.

Simulation is carried out in the complex domain by SALEC, a *Mathematica* package for **S**ymbolic **A**nalysis of **L**inear **E**lectric **C**ircuits, which yields the analytic response as shown in fig. 2.

The circuit response can be post-processed in *Mathematica* to obtain a more suitable form of the symbolic expression that reveals the differentiator behavior, as shown in fig. 3.

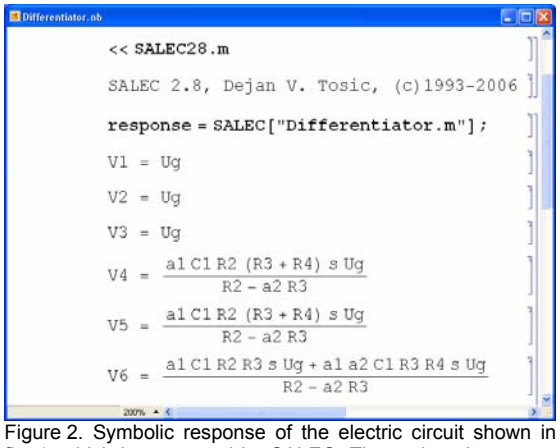


Figure 2. Symbolic response of the electric circuit shown in fig. 1, which is generated by SALEC. The node voltages are closed-form expressions in terms of element values (circuit parameters) and the Laplace variable (complex frequency).

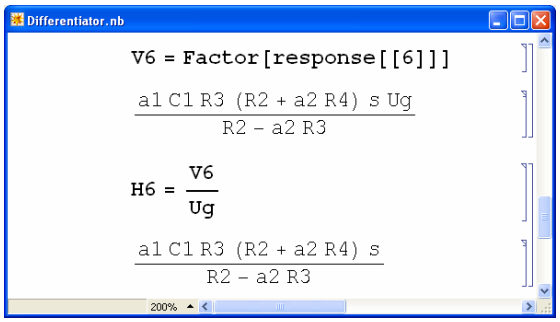


Figure 3. Refinement of the symbolic response that reveals the differentiator behavior: the transfer function is of the form $K\,s$, where K is a constant dependent on the element values, and s is the Laplace variable (complex frequency).

8. *Example Continuous-Time System Simulation*

The following example describes a typical use of CAS in control system analysis. Suppose that the published block-diagram of a system and the corresponding transfer function should be proved. Manual derivation can be tedious and error prone even for a motivated researcher. Alternatively, *SchematicSolver* [20] can be used to draw the system schematic, such as that shown in fig. 4, and to find the transfer function. The transfer function matrix of this three-input four-output MIMO system is shown in fig. 5.

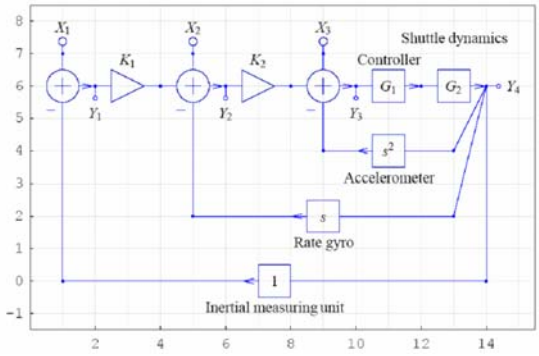


Figure 4. Simplified model of the pitch controller for the space shuttle.

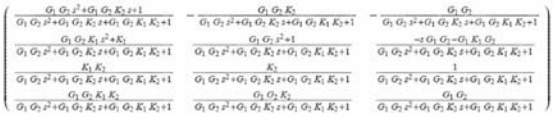


Figure 5. Transfer function matrix of the example multiple-input multiple-output (MIMO) continuous-time system shown in fig. 4.

9. *Example discrete-time system simulation*

SchematicSolver can be used to (a) draw the schematic of a discrete multiple-input multiple-output system, (b) compute the system transfer function directly from the schematic, (c) find the response for given input sequences, or (d) derive some properties of the system.

Consider a high-speed filter that can be used in multirate systems for interpolation and decimation with a factor of two. The automatically generated schematic of the filter is composed of all-pass filters and extra multipliers, and is shown in fig. 6 [23].

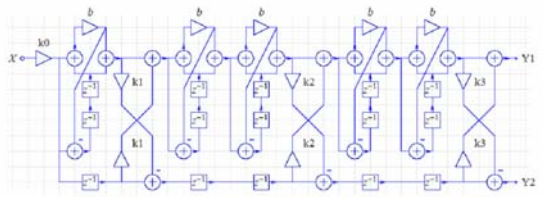


Figure 6. Power complementary high-speed filter suitable for multirate systems for interpolation and decimation with a factor of two.

Symbolic simulation by *SchematicSolver* [20] finds the partial transfer function for the second output as shown in fig. 7.

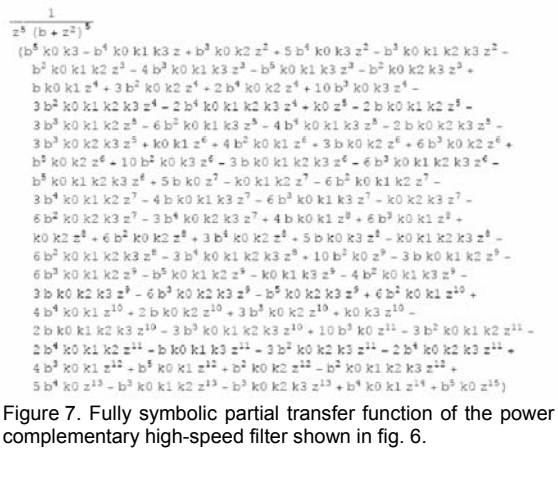


Figure 7. Fully symbolic partial transfer function of the power complementary high-speed filter shown in fig. 6.

Needless to say, derivation of this transfer function by hand is very time consuming and difficult (if possible at all) for more complex high-speed filters.

The filter of fig. 6 is a power complementary filter, that is, the partial transfer functions satisfy the power complimentary equation (condition):

$$H_1(z)H_1(1/z)+H_2(z)H_2(1/z)=1$$

This condition cannot be simply proved without a powerful computer algebra system (CAS), e.g., *Mathematica* [28], and an appropriate symbolic simulator, such as *SchematicSolver* [20]. Furthermore, *SchematicSolver* and *Mathematica* derive the exact analytic condition to be met by the filter coefficients [26]:

$$k0 \to \frac{1}{\sqrt{2}\,\sqrt{1+k1^2}\,\sqrt{1+k2^2}\,\sqrt{1+k3^2}}$$

Definitely, manual derivation of such a formula would be a fatiguing labor.

10. *Conclusion*

Contemporary trends to use very sophisticated algorithms combine expertise in many areas, such as multirate system design, control engineering, analog electronic circuit VLSI design, and signal processing. This trend caused programming to become a task of knowledge accumulation and an efficient human-machine interface.

This paper presented recent advances and the role of symbolic computations in modern engineering and signal processing. It provided illustrative application examples as appropriate to linear systems, modeling, and simulation.

System models were highlighted as visualized algorithms by means of block-diagram representations—schematics. The schematic was established as a symbolic object that contained all details for drawing, symbolic solving, simulating, and implementing the system. It was not seen as a static picture.

It was shown how computer algebra systems (CAS) analyzed the schematic as the symbolic object. The knowledge embedded in the schematic object was used according to the required task, such as, to generate the electric circuit response or to derive the transfer function.

A typical use of CAS in system analysis was illustrated by solving a continuous-time linear MIMO system; the transfer function matrix was determined directly from the schematic.

Transfer function matrix of a complex MIMO discrete system was derived in terms of system parameters kept as symbols. The important power complimentary property was proved, for a class of high-speed filters, for arbitrary symbolic system parameters. The proof involved manipulation of complex expressions that was practically impossible to perform by hand.

Benefits of symbolic methods and the role of CAS were highlighted from the viewpoint of both academia and industry.

REFERENCES

[1] Lin, Pen-Min, "Symbolic network analysis," *Elsevier*, Amsterdam, The Netherlands, EU, 1991.

[2] Gielen, G., Sansen, W., "Symbolic analysis of automated design of analog integrated circuits," *Kluwer Academic Publishers*, Norwell, MA, USA, 1991.

[3] Huelsman, L., Gielen, G., "Symbolic Analysis of analog circuits: Techniques and applications," *Kluwer Academic Publishers*, Norwell, MA, USA, 1993.

[4] Gielen, G., Wambacq, P., Sansen, W., "Symbolic analysis methods and applications for analog circuits: A tutorial overview," *Proceedings of the IEEE*, 1994, pp. 286-304.

[5] Fernández, F., Rodríguez-Vázquez, A., Huertas, J., Gielen, G., "Symbolic analysis techniques: Applications to analog design automation," *Wiley-IEEE Press*, 1997.

[6] Lutovac, D., Tošić, D., Evans, B., "Filter Design for Signal Processing using MATLAB and *Mathematica*," *Prentice Hall*, Upper Saddle River, NJ, USA, 2001.

[7] Qin, Z., Tan, S., Cheng, C., "Symbolic analysis and reduction of VLSI circuits," *Springer*, 2004.

[8] Riddle, A., Dick, S., "Applied electronic engineering with *Mathematica*," Addison-Wesley, Reading, MA, USA, 1995.

[9] Tošić, D. V., "SALECAS - a package for symbolic analysis of linear circuits and systems," in *Proc*. 4th International Workshop on Symbolic Methods and Applications to Circuit Design, 1996, pp. 227–230.

[10] Tošić, D. V., Hribšek, M. F., Reljin, B. D., "Generation and design of new continuous-time second order gain equalizers using program SALEC", *International Journal of Electronics and Communications* (AEÜ - Archiv für Elektronik und Übertragungstechnik), 1996, pp. 226-229.

[11] Tošić, D. V., Kovačević, B. D., Reljin, B. D., "Symbolic Analysis of Linear Dynamic Systems", *Control and Computers*, 1996, pp. 54-59.

[12] Tošić, D. V., "A contribution to algorithms in computer-aided symbolic analysis of linear electric circuits and systems," *Doctoral dissertation*, University of Belgrade, School of Electrical Engineering, Belgrade, Serbia, 1996.

[13] Škokljev, I. A., Tošić, D. V., "A new symbolic analysis approach to the DC load flow method", *Electric Power System Research Journal*, 1997, pp. 127-135.

[14] Tošić, D. V., Djordjević, A. R., Reljin, B. D., "Symbolic Analysis of Microwave Circuits", *Journal of Applied Electromagnetism*, 1997, pp. 37-45.

[15] Tošić, D. V., Lutovac, M. D., "Symbolic analysis of digital filters", *Académie Roumaine*, Revue Roumaine des Sciences Techniques, Série Électrotechnique et Énergétique, Bucarest, 1997, pp. 29-38.

[16] Lutovac, M. D., Tošić, D. V., "Symbolic computation of digital transfer function using MATLAB", in *Proc*. 23rd Int. Conf. Microelectronics, MIEL, 2002, pp. 651-654.

[17] Lutovac, M. D., Tošić, D. V., "Symbolic Signal Processing and System Analysis", *Facta Universitatis*, Series: Electronics and Energetics, 2003, pp. 423-431.

[18] Bakshee, I., "Control System Professional," *Wolfram Research*, Champaign, USA, 2003.

[19] DLR, Institut fur Robotik und Mechatronik, "PARADISE, Parametric Robustness Analysis and Design Interactive Software Environment," [online] http://www.dlr.de/rm/en/desktopdefault.aspx/tabid-482/admin-1/, 2004.

[20] Lutovac, M.D., Tošić, D.V., "*SchematicSolver* version 2," [online] http://www.schematicsolver.com, 2004.

[21] Lutovac, M.D., Tošić, D.V., "Elliptic Rational Functions," *The Mathematica Journal*, 2005, pp. 598-608.

[22] Palancz, B., Benyo, Z., Kovacs, L., ''Control System Professional Suite'', *IEEE Control Systems Magazine*, 2005, pp. 67–75.

[23] Lutovac, M. D., Tošić, D. V., "High-Speed Filter Design using *Mathematica*", in *Proc*. IEEE EUROCON 2005 – The International Conference on "Computer as a Tool", 2005, pp. 1626-1629.

[24] Fraunhofer ITWM, "*Analog Insydes* version 2.1," [online] http://www.analog-insydes.de, 2005.

[25] Tošić, D. V., Simić, S. K., "Analysis of Combinational Networks with *Mathematica*", *Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat.*, 2005, pp. 76–87.

[26] Lutovac, M. D., Tošić, D. V., "Symbolic analysis and design of control systems using *Mathematica*", *International Journal of Control*, Special issue on the use of computer algebra systems for computer aided control system design, 2006, pp. 1368–1381.

[27] Tošić, D. V., Lutovac, M. D., "Multirate systems simulation with *Mathematica*", in *Proc*. 14th Telecomm. forum TELFOR, 2006, pp. 588-591.

[28] Wolfram, S., The *Mathematica* Book, 5th Edition, *Cambridge University Press*, Cambridge, UK, 2003.

[29] Nagel, L. "SPICE2: A computer program to simulate semiconductor circuits," *Memorandum No. M520, University of California*, Berkeley, CA, USA, 1975.

[30] MATLAB Version 7, *MathWorks*, Inc., Natick, MA, USA, 2005.

[31] Kolundžija, B. M., Ognjanović, J. S., Sarkar, T. K., Šumić, D.S., Paramentić, M.M., Janić, B. B., Olćan, D.I., Tošić, D. V., Tasić, M. S., "WIPL-D Microwave: Circuit and 3D EM Simulation for RF & Microwave Applications," *Artech House*, Norwood, MA, USA, 2005.

[32] Mitra, S., "Digital Signal processing, A Computer Based Approach," *McGraw-Hill*, New York, USA, 2006.

[33] Liu, S., Hwang, Y., "Dual-Input Differentiators and Integrators with Tunable Time Constants Using Current Conveyors," *IEEE Trans. Instrumentation and Measurement*, 1994, pp. 650-654.

[34] Lutovac, M. D., Tošić, D. V., Huelsman, L. P. (editor), "DRAWFILT - Drawing Filter Realizations in MATLAB," *IEEE Circuits & Devices*, 2001, pp. 3-4.

[35] Tošić, D. V., "SALEC version 2.8," a *Mathematica* package for symbolic analysis of linear electric circuits, *University of Belgrade*, School of Electrical Engineering, Belgrade, Serbia, 2006.

Dejan V. Tošić is a professor in the School of Electrical Engineering at the University of Belgrade, Serbia. He has focused his research on creating a framework for the symbolic analysis of circuits and systems that is suitable for research as well as industrial and educational applications. He is developing automation tools for optimizing the design and synthesis of analog and digital systems. He is coauthor of the book "Filter Design for Signal Processing Using MATLAB and *Mathematica*," published by Prentice-Hall in 2001, and *SchematicSolver*, a *Mathematica* application for mouse-driven interactive drawing of systems and for solving and implementing systems.

Miroslav D. Lutovac (SM'99) is a professor at the University of Belgrade, Serbia. His research interests include the theory and implementation of analog and digital signal processing, and the symbolic analysis and synthesis of multiplierless and multirate digital systems. He is coauthor of the book "Filter Design for Signal Processing Using MATLAB and *Mathematica*," published by Prentice-Hall in 2001, and *SchematicSolver*, a *Mathematica* application for mouse-driven interactive drawing of systems and for solving and implementing systems.

Specifying Sequent Calculi Rules for Managing Some Redundancies in Proof Search

Lutovac A. Tatjana

Abstract – A central aspect of proof search is the identification and control over various forms of redundancies in the search space. We investigate systematic techniques for managing some redundancies in proof search in sequent calculi. This paper is a summary of some results of our investigation. In particular we have enriched inference rules with some additional information about status the search in order to preclude some redundant or useless choices which would otherwise be allowed in the standard sequent system. We have developed a method for detection of redundant and eliminable formulae from a given sequent proof and an algorithm for ensuring termination ie. for eliminating (infinite) loops during a backward sequent calculi proof search.

Key words: affine logic, backward proof search, linear logic, loops, redundant formulae, sequent calculus

1. *Introduction*

It is well known that logic programming may be thought of as the application of the techniques of mathematical logic to programming tasks. Logic programs may be considered as collections of formulas and their computation may be identified as *searching for proofs*: given a program $\mathcal{P}$ and a goal G we attempt to satisfy G by *searching for a proof* of $\mathcal{P} \rightarrow G$ using the inference rules of a given logic.

Proof search is the name given to the study of the construction, if possible, of proofs of a given logical assertion. In systems known as sequent calculi logical assertions are presented in the form of sequents.

A *sequent* is a pair denoted by $\Gamma \vdash \Delta$, where both the *antecedent* Γ and the *succedent* Δ are sequences of formulae. The intuitive meaning of $\Gamma \vdash \Delta$ is: if all of the formulae in Γ are true, then at least one formula in Δ is true.

Manuscript received December, 11, 2006.

Author is with Department of Applied Mathematics, Faculty of Electrical Engineering, University of Belgrade, Belgrade, Serbia (e-mail: tlutovac@eunet.yu).

A sequent calculus system gives a set of rules for manipulating sequents, in the form of set of sequent rules. Sequent calculus rules generally are written as follows.

$$\frac{sequent_1,\ sequent_2,\ \ldots sequent_k}{sequent}\ \text{rule-name} \qquad k = 0, 1, ...$$

The sequents above the line in a rule are called *premises* of the rule and the sequent below the line is called the *conclusion*.

A *proof* of a sequent $\Gamma \vdash \Delta$ is a tree whose nodes are labelled with sequents such that the root node is labelled with $\Gamma \vdash \Delta$ (which is then called the *end-sequent*), the internal nodes are instances of one of the inference rules, and the leaf nodes are labelled with axioms.

A sequent $\Gamma \vdash \Delta$ is *provable* in the sequent calculus formalization of a logic (or logical fragment) if there is a proof with $\Gamma \vdash \Delta$ as the end- sequent.

Bottom-up ie. *backwards* proof search consists of building the proof from the given end-sequent. The end-sequent is reduced by reversed (conclusion-to-premise) application of a sequent calculus rule to produce subgoal branches.

The theory of the exploration of the search space generated by inference rules used in this way is subject of proof search . A number of computational difficulties arise in designing algorithms for proof search . Prominent among these is the need, in order to minimize complexity, to control and/or eliminate as much redundancy from the search space as possible.

There have been a variety of proof-theoretic techniques used to analyze and design strategies for efficient sequent calculi proof search in theorem proving and logic programming [13, 16, 15]. It is notable that many of the existing strategies are all rather sophisticated and involve complex manipulations of proofs. Many are restricted to particular logic or classes of formulae. Almost all are designed for analysis on paper by a human and many of them are ripe for automation, being formally defined in

precise detail, and yet somewhat overwhelming for humans.

We investigate systematic and automated-oriented techniques for managing some forms of redundancies in proof search in sequent calculi. This paper is a summary of some results, being developed, presented and formally proved in [7][1], on the development of systematic techniques:
- for detection of redundant and eliminable formulae from a given sequent proof and
- for ensuring termination ie. eliminating infinite loops during a backward proof search.

Our solutions are based on the widely used practice of including in the representation of the sequents some additional information about the status of the search.

Our work on detection of redundant parts of sequent proof is motivated by the fact that none of the existing algorithms for efficient implementation of proof search (in linear logic, at least) can distinguish redundant formulae that can be freely, unconditionally eliminated from the proof from those redundant formulae whose elimination lead to an invalid proof, as far as we are aware. Our mechanism makes such a distinction. Furthermore it allows selection in a sense that a redundant (sub)formula can be either eliminated or replaced by an arbitrary formula. This induces a class of equivalent formulae (in terms of provability) and a class of equivalent proofs modulo the redundant parts. This allows, for example, more flexible reuse of previously successful searches and is potentially useful for implementations.

There have been a variety of systems for preventing and detecting infinite loops during proof search. Surprisingly, no loop detection mechanism has been developed or described in the literature (as far as we are aware) for resource sensitive logic (such as, for example, linear and affine logic). We have developed a (terminating) sequent calculi with loop-detection mechanism for intuitionistic, propositional[2] affine logic. We have followed the overall approach of [3] and [4], i.e., to incorporate the loop-detection mechanism into the sequent rules. The proposed conditions are independent of the search strategy used and no explicit loop checking is needed in the interpreter.

[1] Parts of this material are presented in [8, 9].

[2] Quantifier free.

Given the evolution of programming languages towards higher and higher level languages, with a corresponding increase in the computational power of the execution models of these languages, a natural demand is a wide range of expressive logics and more powerful inference facilities in which to write programs. That is why our starting point is propositional linear logic and affine logic. Linear logic is a refinement of classical logic, in that there is a fragment of linear logic which has precisely the same properties as classical logic; at the same time however, linear logic contains features which are not present in classical logic. In essence, these features are due to removing the rules for contraction and weakening and reintroducing them in a controlled manner. It is simpler and more natural to express certain resource management problems in linear logic than in classical logic. For example, the property of having two dollars we may represent by the LL conjunction $\$1 \otimes \$1$. In classical logic, this would be represented by $\$1 \wedge \$1$, which is equivalent to $\$1$, ie. that having two dollars is equivalent to one dollar, which is clearly nonesential. However, in linear logic $\$1 \otimes \$1$ and $\$1$ are not equivalent, which is more appropriate. For this reason linear logic is often described as a logic of resources rather than logic of truth (such as classical logic) in that different amounts of the same thing are considered to be different.

We are interested in affine logic because of its relationship to logic programming. Due to the presence of the weakening rules[3], there are many problems where affine logic is better suited than linear logic. Propositional affine logic [5] is decidable[4] , and so it seems reasonable to expect a complete proof search procedure with a loop detection mechanism.

This paper is organized as follows. In Section 2 we briefly explain our mechanism for detection and elimination of redundant formulae in sequent proof.

[3] While proofs in linear logic must use each linear formula exactly once, proof derivations in affine logic use each affine formula at most once.

[4] Let us recall that a logical system is *decidable* iff there exists an algorithm such that for every well-formed formula in that system there exists a maximum finite number N of steps such that the algorithm is capable of deciding in less than or equal to N algorithmic steps whether the formula is (semantically) valid or not valid.

In Section 3 we illustrate our solution for prevention of infinite loops in propositional affine logic. In Section 4 we present our conclusions.

2. Redundant formulae in sequent proofs

Proof search often involves managing information which later, when the proof is completed, turns out to be redundant. For example, the sequent $p, q \vdash p, r, s$ is provable in classical (and affine) logic. However q, r and s are redundant, and the "core" provable sequent is $p \vdash p$. More complex examples involve choosing between subformulae.

Example 1. Consider the (linear logic) proof Π below:

$$\Pi: \quad \dfrac{\dfrac{}{t \vdash t}\,Ax \quad \dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{r \vdash r}\,Ax}{r \vdash ?p, r}\,?wR}{r \vdash ?q, ?p, r}\,?wR}{r \vdash ?q \wp ?p, r}\,\wp R}{r \vdash (?q \wp ?p) \oplus s, r}\,\oplus R}{r, t \vdash t \otimes ((?q \wp ?p) \oplus s), r}\,\otimes R \qquad \dfrac{?}{r, t \vdash t, r}$$

What is the core of this proof ie. what is the minimal set of formulae which guaranties success of the proof? The (sub)formulae p, q and s are *unused*, but only s can be freely deleted from the proof while formulae p and q cannot be simultaneously eliminated. Elimination of the whole formula $(?q \wp ?p) \oplus s$ will disable proof branching ie. distribution of formulae across the multiplicative branches of the proof. Elimination of the subformula $?q \wp ?p$ will also lead to the unprovable sequent (on the right-hand side above). So, we have that p, q and s are unused and that p and q cannot be simultaneously eliminated from the proof. For each unused atom we have three possibilities: to omit it from the proof; to leave the atom unchanged or to replace it with an arbitrary formula. So proof Π can be thought of as a template for $(3^2-1)\cdot 3$ proofs (ie. some variations of the given proof) which can be generated by alterations of p, q and s. All that proofs do not alter the search strategy used, in that the order of application of the rules is not changed.

This knowledge allows later computations to make use of earlier work. So a proof search strategy can retain the results of a previous successful search and to apply and combine them to a new situation. The knowledge about redundant and eliminable formulae can be potentially useful when composing programs (and hence proofs), for debugging, and for teaching purposes.

2.1 *Our technique for detection of redundant formulae*

In Chapter 4 of [7] we have proposed a mechanism for distinguishing between the necessary and unnecessary formulae in a linear logic proof. We have enriched the standard sequent structure with labels and Boolean constraints as follows:

$\phi_{1,[v_1]}, \phi_{2,[v_2]}, \ldots \phi_{n,[v_n]} \vdash \psi_{1,[w_1]}, \psi_{2,[w_2]}, \ldots \psi_{m,[w_m]} - \mathcal{C}$

where $\mathcal{C}$ is a set of constraints being generated so far on the branch of a proof tree by application of the particular rules and labels $[v_1], \ldots, [w_1], \ldots [w_m]$ trace formulae duplicated by the contraction rules. Labels and constraints allow us to store the necessary information about the usage of formulae in a proof.

We have defined a labelled sequent calculus, called LL^{PRE}, for elimination of redundant formulae in propositional linear logic. LL^{PRE} sequent calculi is defined as shown in Figure 4.1, Chapter 4 of [7]. Examples of some LL^{PRE} rules and the corresponding constraints are given below.

$$\dfrac{- \mathcal{C} \cup \{p > 0\}}{p \vdash p \quad - \mathcal{C}}\,Ax \qquad \dfrac{\Gamma \vdash \psi_{[w]}, \Delta \quad - \mathcal{C} \cup \{\phi_{[w]} \leq \psi_{[w]}\}}{\Gamma \vdash (\phi \oplus \psi)_{[w]}, \Delta \quad - \mathcal{C}}\,\oplus \mathrm{R}$$

$$\dfrac{\Gamma \vdash \phi_{[w]}, \Delta \; - \mathcal{C} \cup \{\phi_{[w]} > 0\} \quad \Gamma' \vdash \psi_{[w]}, \Delta' - \mathcal{C} \cup \{\psi_{[w]} > 0\}}{\Gamma, \Gamma' \vdash (\phi \otimes \psi)_{[w]}, \Delta, \Delta' - \mathcal{C}}\,\otimes R$$

We begin with the empty labels (denoted as $[\,]$) on each formula of the end-sequent and with the empty set $\mathcal{C}$. The labels and constraints generated and accumulated during a proof construction (in the labelled system LL^{PRE}) place some restrictions on the elimination of the corresponding formulae. For example, the intuition underlying the constraint $\phi_{[w]} \leq \psi_{[w]}$ is that elimination of formula $\psi_{[w]}$ is a necessary condition for elimination of formula $\phi_{[w]}$. The intuition underlying the constraint $\phi_{[w]} > 0$ is that formula $\phi_{[w]}$ cannot be entirely eliminated.

A labelled proof tree (generated by the LL^{PRE} sequent rules) is forwarded to the main algorithm, called algorithm PRE, being defined as follows.

Algorithm PRE (input: labelled proof π)

1. *Generate Boolean expressions and constraints on Boolean expressions;*
2. *Calculate possible assignments for Boolean variables;*
3. *If there is an assignment with at least one Boolean variable being assigned the value 0 then:*
   Delete atoms being assigned 0 ie. delete formulae made up of such atoms and the corresponding inferences

   Else EXIT: 'Simplification of proof π is not possible'

Algorithm PRE has to interpret the set of accumulated constraints via the set of Boolean constraints and to find an assignment for Boolean variables. Intuitively, we associate each atom in a proof π with a Boolean variable. Atoms associated with Boolean variables annotated to 0 and the (sub)formulae made up of such atoms can be safely eliminated (ie. deleted) from the proof π.

For example for the proof Π from Example 1, we will have the labelling as follows.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\Pi_1}{r_{[]}, t_{[]} \vdash (t_1 \otimes ((?q \wp ?p) \oplus s))_{[]}, r_{1,[]} \quad - \emptyset}\otimes R}{r \vdash (?q \wp ?p) \oplus s, r_1 \quad - \{(?q \wp ?p) \oplus s > 0\}}\oplus R}{r \vdash ?q \wp ?p, r_1 - \{(?q \wp ?p) \oplus s > 0,\ s \leq ?q \wp ?p\}}\wp R}{r \vdash ?q, ?p, r_1 \ - \{(?q \wp ?p) \oplus s > 0,\ s \leq ?q \wp ?p\}}w?R}{r \vdash ?p, r_1 \ - \{(?q \wp ?p) \oplus s > 0,\ s \leq ?q \wp ?p\}}w?R}{r \vdash r_1 \ - \{(?q \wp ?p) \oplus s > 0,\ s \leq ?q \wp ?p\}}w?R}{- \{(?q \wp ?p) \oplus s > 0,\ s \leq ?q \wp ?p,\ r = 1,\ r_1 = 1\}}Ax$$

where the subproof Π_1 is as follows. $\dfrac{- \{t = 1,\ t_1 = 1, t_1 > 0\}}{t \vdash t_1 \ - \{t_1 > 0\}}\ Ax$.

For the proof Π Algorithm *PRE* will extract the following Boolean constraints: $q + p + s > 0$, $s \leq q + p$, $r_1 = 1$, $r = 1$, $t_1 = 1$, $t = 1$. Possible solutions for the (unassigned) Boolean variables are: $(p, q, s) \in \{ (0,1,0), (1,0,0), (1,1,0), (1,1,0), (1,0,1) \}$. Hence, there are five possible simplifications of proof Π. Below we illustrate one of them:

$$(p,q,s) = (0,1,0) \mapsto \dfrac{\dfrac{}{t \vdash t_1}\ Ax \quad \dfrac{\dfrac{}{r \vdash r_1}\ Ax}{r \vdash ?q, r_1}\ ?wR}{r, t \vdash t_1 \otimes ?q, r_1}\otimes R$$

As we already pointed out none of the existing algorithms for efficient implementation of proof search (in linear logic, at least) is able to make the distinction between unused formulae that can be freely eliminated from the proof from those unused formulae whose elimination will cause a 'crash'. Our labelled system makes such a distinction. Furthermore it allows selection in a sense that a redundant (sub)formula can be either eliminated or replaced by an arbitrary formula. Thus, we can get a class of equivalent formulae (in terms of provability) and a class of equivalent proofs modulo the redundant parts. This allows, for example, more flexible reuse of previously successful searches and is potentially useful for implementations.

Our intention was not to find all different proofs of a given sequent but to generate all the concrete simplifications which are instances of a generated proof. Our solution for detection of redundant, eliminable formulae implies elimination which is independent of the search strategy used; elimination which does not alter the search strategy applied and does not require additional proof search ie. redundant formulae remaining in the resulting proof cannot be eliminated without additional proof search. Soundness and completeness of our solution are proved formally in Chapter 4 of [7].

3. Detection and prevention of infinite loops during proof search

It is well known that for many logics, backward proof search in the usual sequent calculi generally does not terminate in general, in the sense of Dyckhoff[1]: 'By "terminating" we mean just that every sequence of steps, in a backward proof search, is finite.' This makes loop detection, where possible, a critical aspect of systems based on backward proof search. For example, consider the (unprovable) linear sequents $!(p \multimap q), !(q \multimap p) \vdash q$ and $!((r \multimap p) \multimap p), r \vdash p$, and the following attempts for proof construction:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\vdots}{\mathcal{P} \vdash q} \quad \dfrac{}{p \vdash p}\ Ax}{\mathcal{P}, q \multimap p \vdash p}\multimap L}{\mathcal{P}, !(q \multimap p) \vdash p}\ !L}{\underbrace{!(p \multimap q), !(q \multimap p)}_{\mathcal{P}} \vdash q}\ !C \quad \dfrac{}{q \vdash q}\ Ax}{\mathcal{P}, p \multimap q \vdash q}\multimap L}{\mathcal{P}, !(p \multimap q) \vdash q}\ !L}{!(p \multimap q), !(q \multimap p) \vdash q}\ !C$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\vdots}{\mathcal{P}_1, r, r \vdash p}}{\mathcal{P}_1, r \vdash r \multimap p}\multimap R \quad \dfrac{}{p \vdash p}\ Ax}{\mathcal{P}_1, r, (r \multimap p) \multimap p \vdash p}\multimap L}{\underbrace{!((r \multimap p) \multimap p)}_{\mathcal{P}_1}, r \vdash p}\ !C, !L$$

The sequent $\mathcal{P} \vdash q$ may continue to occur in the left-hand proof construction. The (sub)formula r may continue to occur in the antecedents during the right-hand proof construction. Note that this is due to the deterministic nature of the proof search process; as in many logic programming languages, the selection of a formula from the antecedent is determined largely by the formula position, and so an interpreter will always handle variants of a given antecedent and a given succedent in exactly the same way (up to renaming variables). Thus far, there is no satisfactory solution to this problem.

There have been a variety of systems [3], [4], [1] for preventing and detecting loops during proof construction. Due to the lack of non-decreasing number of formulae in antecedents, the existing techniques are not directly applicable for resource sensitive logic (such as, for example, linear and affine logic). No loop detection mechanism (apart from a naive history mechanism) has been developed for resource sensitive logic. Naive history

mechanism implies unintelligent searching through the history ie. through is the list of all sequents that have appeared so far on the branch of a search tree. At every step of a proof construction the latest generated sequent is checked to see whether it is a member of the history list. If so, a loop has been generated and so the search backtracks. If not, the history is extended with the sequent and the proof search continues. Implementation of this scheme is clearly inefficient as it requires a great deal of information to be stored. It is important to note that situation illustrated by the right-hand derivation above cannot be detected and solved by the naive history mechanism.

We have identified two reasons for non-termination in propositional linear and affine logic: *simple loops* (ie. appearance of identical sequents in the same branch of a proof tree, as illustrated by the left-hand derivation above) and *special loops* (ie. an infinite repetition of a particular formula, as illustrated by the right-hand derivation above). Our strategy for detecting and preventing loops is twofold. First, we add a simple history list to sequents to allow detection of simple loops, and then introduce machinery that, essentially, turns infinite special loops into simple loops. The idea behind the treatment of special loops is not to make the rule (or sequence of rules) 'responsible' for the special loop inapplicable. The idea is to identify a special loop, and to continue and complete the proof search bearing in mind that some formulae (originated from that special loop) can be used as many times as needed (such as, for example, the formula r in the right-hand example above).

We have proposed sequent calculi rules with zones and side conditions (denoted PIA_{ff}^{Hist-}) for propositional affine logic, as shown in Section 5.7 of [7]. Example of a PIA_{ff}^{Hist-} rule is given below.

$$\frac{\Gamma\,;\,\Upsilon,\,\mathcal{U};\,\Delta-\mathcal{U}\vdash\chi_1\ \blacktriangleright\ f\,;\,\mathcal{H}_1\,;\,h_1}{\Gamma\,;\,\Upsilon;\,\Delta\vdash\chi\ \blacktriangleright\ f\,;\,\mathcal{H}\,;\,h}\ \oplus R\ \Longleftrightarrow\ \mathbf{TEST}(\,\Upsilon;f;\mathcal{H};h;\Delta;\chi_1;\mathcal{U};\mathcal{H}_1;h_1)$$

We have divided the antecedent into zones to isolate and control the non-decreasing part of antecedent ie. to assemble all formulae originated from special loops into a new zone (called $*$context zone) of antecedents. This allows to cut down the amount of storage and checking in the history list. We have extended the sequent rules with the set of side conditions (denoted as procedure TEST). Procedure TEST has to maintain history list, and detect and control simple and special loops. If the latest generated sequent is a member of the history list, procedure TEST make the rule inapplicable (ie. forces the system to backtrack). In the case of special loop, procedure TEST classify all formulae originated from recognized special loop into $*$context zone. As search proceeds, the unrestricted resources of every identified special loop will be classified into the $*context$. Thus, example of the special loop, shown on the right-hand side at the beginning of this section, would be interpreted in the PIA_{ff}^{Hist-} calculi as shown on the left side below.

$$\dfrac{\dfrac{\overbrace{\mathcal{P}_1\,;\,r\,;\quad\ ;\,r,r}\vdash p}{\mathcal{P}_1\,;\quad\ ;\,r\vdash r\multimap p}\multimap R\qquad\dfrac{}{\mathcal{P}_1\,;\quad\ ;\,p\vdash p}\,Ax}{\dfrac{\mathcal{P}_1\,;\quad\ ;\,r,(r\multimap p)\multimap p\ \vdash\ p}{\mathcal{P}_1\,;\quad\ ;\,r\ \vdash\ p}\,!C,!L}\multimap L \qquad \begin{array}{l}\mathcal{P}_1\,;\,r\,;\quad\vdash p\\ \vdots\\ \mathcal{P}_1\,;\,r\,;\quad\vdash p\\ \vdots\\ \mathcal{P}_1\ ;\ ;\ r\vdash p\end{array}$$

$\underbrace{\qquad}_{*\,context\ zone}$

Formula r (being classified in the $*$context) can be further used as many times as needed. An attempt to reiterate the above derivation will cause detection of simple loop (as shown on the right side above). Detection of a simple loop makes the last applied inference rule inapplicable ie. forces the system to backtrack to the most recent decision point and to try to find alternative solution(s). We get failure ie. the sequent at the root of the proof tree is not provable if at any point no rule instance can be applied.

Soundness and completeness of the PIA_{ff}^{Hist-} sequent calculi system as well as the fact that backward proof search in this calculus will always terminate, are proved formally in Chapter 5 of [7].

Providing a terminating procedure for a propositional affine logic could be very useful in the design of tabling mechanisms for linear logic. Another natural extension of this work is to apply the ideas behind the PIA_{ff}^{Hist-} sequent calculi to (fragments of) linear logic. The well-known problem concerning an implementation of logic programming language Forum [11] is connected with the special loops originating from $\perp-$headed implications. Thus far, there is no satisfactory solution this problem. It is our contention that the ideas behind the PIA_{ff}^{Hist-} loop detection mechanism can be used to establish control over the '$\perp-$headed special loops' in Forum.

4. Conclusion

Managing redundancies in sequent calculus proof search is nontrivial. We have briefly presented some mechanisms, being formally defined, developed and proved in [7], for identification and control over different forms of redundancies in sequent calculi proof search . The proposed solutions may contribute to automation of proof search by, for example, fine tuning the search to find one "good" representative of a class of proofs (such as, for example, a class of equivalent proofs modulo redundant formulae).

Our technique for elimination of redundant formulae is limited to sequent proofs and thereby differs from dead-code elimination in functional languages. Developing more general techniques for program slicing and dead-code elimination in advanced logic programming languages are items of future work.

We have developed first terminating sequent system for (a fragment) of propositional affine logic. The perspective to apply the ideas for managing loops in the logic programming language Forum also emphasizes the interest of the results. A natural extension of this work is to apply the ideas behind the PIA_{ff}^{Hist-} sequent calculi to various (fragments of) other resource-sensitive logics.

Our work is intended as a contribution to a library of automatic support tools for managing redundancies in sequent calculi proof search. The proposed strategies and techniques can be implemented and utilized by means of an automated proof assistant such as Twelf [14], possibly in conjunction with constraint logic programming techniques [10].

References

1. Dyckhoff R, " Contraction-free Sequent Calculi for Intuitionistic Logic", The Journal of Symbolic Logic, Vol. 57, No. 3, 1992, pp. 795-807.
2. Harland J, Pym D., " Resource-distribution via Boolean constraints", ACM Transactions on Computational Logic 4:1, 2003, pp. 56-90.
3. Heuerding A., Seyfried M., Zimmermann H., " Efficient Loop-Check for Backward Proof Search in Some Non-classical Propositional Logics", in P. Miglioli at al. (eds.), Proceedings of the 5th International Workshop on Tableaux, Italy, 1996, pp. 210-225.
4. Howe J.M., " Proof Search Issues in Some non-Classical Logics", PhD thesis, School of Mathematical and Computational Sciences, University of St Andrews, 1998.
5. Kopylov A., " Decidability of Linear Affine Logic", Proceedings of the Tenth Annual IEEE Symposium on Logic in Computer Science 496-504, San Diego, 1995.
6. Lutovac, T., Harland J., " Issues in the Analysis of proof search Strategies in Sequential Presentations of Logics", IJCAR'04 Workshop on Strategies in Automated Deduction, Electronic Notes in Theoretical Computer Science, 125(2), 2005, pp. 115–147.
7. Lutovac T., " Issues in Managing Redundancies in Proof Search", Phd Thesis, School of Computer Science and Information Technology, Science, Engineering and Technology Portfolio, RMIT University, Australia, 2005.
8. Lutovac T., Harland J., " A Redundancy Analysis of Sequent Proofs", International Conference on Automated Reasoning with Analytic Tableaux and Related Methods (TABLEAUX 2005),Germany 2005, $LNAI$ 3702, Springer-Verlag, 2005, pp. 76-90.
9. Lutovac T., Harland J., " Detecting Loops During Proof Search in Propositional Affine Logic", Journal of Logic and Computation, Volume 16, Number 1, 2006, pp. 61-133.
10. Marriot K., Stuckey P., " Programming with Constraints", MIT Press, 1998.
11. Miller D., " Forum: A multiple-conclusion specification-logic", Theoretical Computer Science 165(1), 1996, pp. 201-232.
12. Polakov J., " Linearity Constraints as Bounded Intervals in Linear Logic Programming", in D, Galmiche,(eds.), LICS'04 Workshop on Logic for Resources, Process and Programs (LRPP), 2004, pp. 173-182.
13. Pym D., Harland J., " A Uniform Proof-theoretic Investigation of Linear Logic Programming", Journal of Logic and Computation 4:2, 1994, pp. 175-207.
14. Schürmann C., " Automating the Meta-Theory of Deductive Systems", PhD thesis, Carnegie-Mellon University, 2000.
15. Tammet T., " Proof Search Strategies in Linear Logic", Journal of Automated Reasoning 12, 1994, pp. 273-304.
16. Wallen L., " Automated Proof Search in Non-classical Logic", MIT Press, 1990.

Accelerating Conjugate Gradient Solver: Temporal Versus Spatial Data

Korolija, Nenad G.; Milutinovic, Veljko; and Milosevic, Srdjan

***Abstract*—Simulation of an object in the wind tunnel is a long lasting process, and therefore an ideal candidate for making code run in parallel.**

Simulation complexity is still to high for today's computers. With a growing number of processes computation time is falling, but communication time is rising. Memory can also be the problem.

Existing solutions are based on one process being the master, and, as so, communicating with all other processes. That causes both time consuming communication while other processes wait for the master process and memory problem while one process holds all the data at one moment if no special technique is applied.

In this paper, another approach is described. Using load balancing, all processes became equal during the computation phase. That means that each one the n processes tends to hold approximately 1/n of the data, and works without waiting for communication to finish.

Numerical and computational analyses are done in order to show reader the major advantage of this approach.

As a result, in a real case, the speedup when switching from 64 to 128 processors is almost two.

Index Terms— *Master-slave parallelism – all of the processess working in parallel, non-zero block – block of matrix that has at least one non-zero value*

1. INTRODUCTION

BECAUSE making a car model and simulating its air resistance in the wind tunnel is both time and money expensive, the idea of making a simulation has became a reality. It is expected that the first BMW will be made without making a single model soon. In order to evaluate air resistance, one needs to discretize a volume, set PDE-s, and solve them. Solving PDE-s with huge number of unknowns is not possible without using mathematical algorithm, except in some special cases. By discretizing, a system with n linear equations and n unknowns is obtained. The most popular way of presenting these is using matrices. In this case, the matrix will be sparse. Conjugate Gradient method is a method for solving system of linear equations using induction. In each step, we are supposed to be closer to the exact solution.

This paper deals with optimizing code, for parallel execution of Conjugate Gradient algorithm without preconditioning with a sparse matrix that is a result of discretizing a volume and setting correspondent PDE-s. Code was tested on many multi-processors computer architectures, which were suitable for running MPI programs.

Dipl. Ing. Nenad G. Korolija, Dr. Veljko Milutinovic, Mr. Srdjan Milosevic, Scool of Electrical Engineering, University of Belgrade, Serbia

2. PROBLEM STATEMENT

When talking about simulations, where the volume is divided in small amounts of volume, it is obvious that by dividing on smaller pieces leads to the result with better precision. Of course, that also means longer execution time. Therefore, it is natural to try to run a simulation code in parallel.

While the majority of calculation in CG algorithm is matrix vector multiplication, the execution time is easy to calculate, and the multiplication can easily be divided on many processors.

Anyway, when the result should be spread to all of the processors, sometimes it is faster to run the whole simulation on single processor computer, then to run it on many processors, and then deal with the communications. Even if the communication lines are very fast, with every message passing interface one processor has to form a message head and body and send it, and the receiving part should do inverse operations, and the direct communication is not easy to establish, and usually not useful.

For example, with matrix vector multiplication divided on many processors, each of them needs to send the result to all of the others. In case of modeling the volume, the result matrix is sparse, which will be of great interest in further calculations.

3. PROPOSED SOLUTION AND WHY IT IS EXPECTED TO BE BETTER THEN OTHERS

The main idea is not presented by explaining each part of the algorithm/code, but instead, explaining each idea that lead to the final solution. Within each idea, main characteristics are given, a picture demonstrating what we have achieved by implementing it, and the problem to be solved by implementing next idea. But first, the serial implementation will be discussed.

The main ideas were:

- Dividing calculation onto many processes
- Reduced communication
- Master-slave parallelism

3.1 Serial implementation

The most important thing to notice when talking about making a simulation run on many processors is that there are approximately 200 iterations per one time step, where matrix-vector multiplication is the most processor demanding operation in iteration. Beside it, scalar vector product is calculated in all iterations, and the multiplication of the vector with a constant. The most promising thing to do is to split matrix vector multiplication on many processors. Other operations are not to be split at this stage because more time would be needed to send and receive the result than to calculate it on a single processor.

3.2 Dividing calculation onto many processes

Because of the nature of the problem, the matrix is divided onto non-zero blocks same sizes. The number of row blocks is divided by the number of available processors. Each processor is responsible for multiplication of approximately the same number of row blocks with appropriate vector. Figure 1 depicts a non-zero blocks marked as black circles, while the rest of blocks are zero blocks. The acceleration obtained by using this basic principle is obvious. Still, there is a problem to be solved. After each matrix-vector multiplication, a result should be collected, and then delivered to all processes. This approach requires sending and receiving huge amount of data.

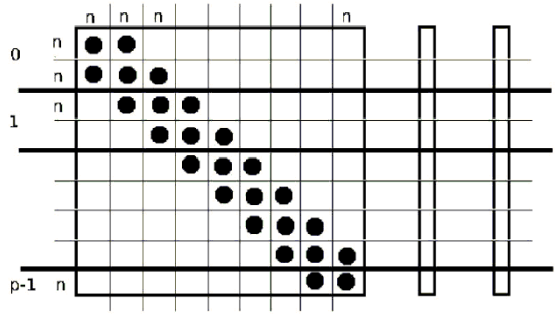


Figure 1 – dividing calculations onto p processors

3.3 Reducing communication

In order to make code more efficient, the structure of the matrix is examined. Figure 2 describes necessary vector data marked as tree dots for multiplication with the corresponding part of the matrix. It is easy to notice that if one processor multiplies any number of row blocks with the vector, it needs to receive only 2*n numbers, and send the same amount of data.

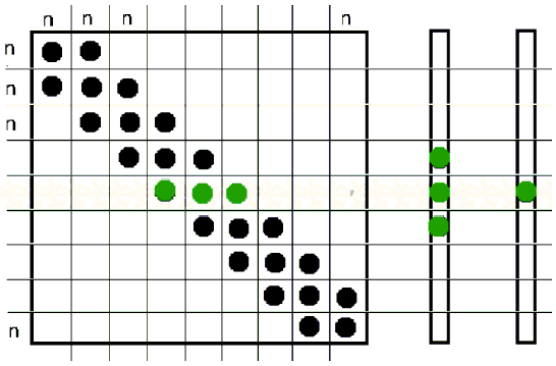


Figure 2 – determining necessary vector parts for multiplying with one row block

Comparing to the previous case, maybe it is not that obvious, but the communication necessary for matrix vector multiplication is reduced almost n/2 times, where n could be even 10000. Anyway, the problem is still a little bit covered, but easier to notice. The whole communication is done by the root process and each other process. In case we have a computer architecture made of equal nodes, each process would have to wait until all of the data has been received by the root process and sent to all other processes. If one process is run on single processor, all processors would have to wait for communication to finish.

3.4 Master-slave parallelism

Now that we know which part of the vector is necessary for which process in order to do the calculation, we can try do determine which process has got the requested data. Even if all of the processes are treated as equal when using MPI, we can mark process zero as the root, and all other processes as slaves. Similarly, we can force the process with any rank to work with corresponding row-blocks, and therefore know rank of the process that is working with any part of the vector. This way, as shown on figure 3, every process needs to send only n real numbers to the upper neighbor and n real numbers to the lower neighbor. Here, upper neighbor is the process whose rank number is less by one than the current process rank number, and lower is the one with the rank number higher by one. Similarly, it needs to receive same amount of data from same processes. While many computer architectures support parallel communication between processes, using master-slave parallelism could lead to almost n times less execution time compared to the one in previous case, where all the communication was done over the root process.

The last, but not least to say is that whole communication could be done in parallel to the calculation, which means that for big problem sizes, the communication time is around zero. This is achieved in tree stages. First is starting sending and receiving operation. Second is multiplying each row block that is independent

from other processes. Third is checking if the communication has been finished. Only in case of having small data sets, processors would have to wait. Otherwise, the remaining thing to do was multiplying the upper and the lower row block with correspondent vector parts.

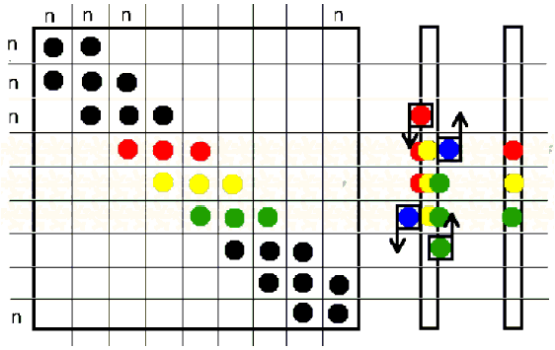


Figure 3 – communication of process with the neighbor processes.

3.5 Further development

For this kind of algorithm as CG is, the limit for making a code parallel as it can be is that in each iteration an scatter/reduce operation must be done. That includes sending a packet of data from each of the processes, than summing data from every process, and then delivering the result to all of the processes. By modifying CG algorithm for parallel implementation, a redundant communication could be obtained. Of course, algorithm should be very different in that case, while prediction of the sum of the data in next step should be made, and the next iteration could depend on the prediction, and not the real sum. Later, the correction would have to be done. Developing such algorithm would show if the result could be reached in a shorter time.

4. CONDITIONS AND ASSUMPTIONS OF THE RESEARCH TO FOLLOW

In this chapter, a brief introduction to the suitable computer architecture for this program is given. Testing was mainly done at cluster Mozart at SGS department on IPVS, in Stuttgart, Germany. It consisted of 64 nodes, each containing two processors with 1GB of RAM. Simulation can be run on any system having MPI installed on it, even single processor. Anyway, in order to obtain considerable less execution time than in case of running a serial version, cluster with big number of processors and good network is needed. Even if the network is not good, with the bigger program size, the process data exchange could be done in parallel to the calculation. Special sparse matrix was produced by Ionel Muntean's code, which was given in 9 vectors for the 2D case and 27 vectors for 3D case. These vectors represented non-zero data in the matrix. It is much more efficient to store only 9 or 27 vectors than to store whole non-zero blocks in the memory while most of the elements of the blocks would be zeros.

5. ANALYTICAL PERFORMANCE ANALYSIS

In this chapter, analysis is done considering memory and time aspects. For each of them, a comparison between serial and parallel version is given. In order to make paragraph more understandable, let n be the dimension of the matrix and the vectors, and p number of processes in parallel version of the program, which will be used in later text.

5.1 Memory usage

Memory usage will be considered with a size needed for representation of a real number defined as a minimum memory usage. For example, size n will represent 8*n bytes if one real number needs 8 bytes.

For serial version the memory size needed is determined in co ordinance with size of vectors and matrix. There are 4 vectors of size n, and matrix 5*n*n for 2D case, and 27*n*n for 3D case. In order to run the serial version of program that executes CG algorithm, all the memory must exist in single computer.

For parallel version, matrix size is 5*n*(n/p) for 2D case, and 27*n*(n/p) for 3D case. Size of each of 4 vectors is n/p. It is obvious that a parallel version is as parallel as it can be considering memory aspects, meaning that all the data that occupies most of the memory is spread over all p processes equally.

5.2 Time needed

Using CG algorithm without prediction, it is calculated that approximately 250 iterations is done in order to have solution enough close to the real system of linear equations result. Calculation time per iteration will be used as a minimum amount of time in order to make analysis more readable. Automatically, same calculation is valid for both 2D and 3D case.

For serial version all the calculation has to be done at single processor unit, meaning that time needed for execution is n.

For parallel version, it is good to define the time needed for data block sending and receiving. Anyway, while the transmitting of data is at least partially done in parallel to calculation of part of matrix and part of vector multiplication, the important thing is not to calculate the time needed for sending, but the difference between calculation time and that time. While this could make the analysis unnecessary complicated and example dependent, only graphics showing algorithm execution time for different problem sizes are given. At this point, it is important to notice that for the big problem sizes, but still real, the data sending/receiving is done completely in parallel to calculation even for considerably high number of processors. Therefore, the time needed to finish execution of parallel version of

code is: n/p plus time needed for reduce/broadcast operations, where every process sends/receives one real number. For big problem sizes, by doubling processor number, the execution time is reduced around twice! Therefore, it is obvious that a parallel version is as parallel as it can be considering processing time also, when the problem size is great enough to be reasonable to run it on more then one processor.

6. SIMULATION ANALYSIS

In this chapter, graphics and tables are given, for both 2D and 3D cases, in order to make possible for reader to realize the benefits of the proposed solution at the first glance. Figures 4 and 5 depict tables showing running time in seconds depending on the number of processors used for calculation in each row, and the problem size in each column. On figures 6 and 7, graphics are given for chosen problem sizes to show the dropping of the execution time with growing number of processors. Note that x-axis is exponential.

	20 x 20	50 x 50	100 x 100	200 x 200	400 x 400
np 1	0.014907	0.05483	0.202765	1.045567	4.306912
np 2	0.011452	0.037967	0.116468	0.653028	3.310686
np 4	0.014715	0.027751	0.067383	0.250178	1.761521
np 8	0.017923	0.024089	0.044449	0.121511	0.83695
np 16	x	0.025532	0.035872	0.074576	0.270791
np 32	x	x	0.03601	0.055549	0.147883
np 64	x	x	x	0.053953	0.112273
np 128	x	x	x	x	0.102522

Figure 4 – Measurements on cluster Mozart for 2D case (number of processes vs. problem size)

	20 x 20 x 100	30 x 30 x 100	40 x 40 x 100	30 x 30 x 600	60 x 60 x 1200
np 1	3.770388	8.015547	14.47042	49.212649	466.394529
np 2	2.786651	6.169195	12.845254	37.632147	555.831802
np 4	1.556983	3.684659	7.64594	20.757366	400.517335
np 8	0.788501	2.069098	4.260279	10.569032	204.146744
np 16	0.334108	1.229958	2.336778	5.592154	102.672306
np 32	0.212199	0.621971	1.647089	3.012153	51.750987
np 64	x	x	x	1.87647	26.920537
np 128	x	x	x	1.072528	14.734521

Figure 5 – Measurements on cluster Mozart for 3D case (number of processes vs. problem size)

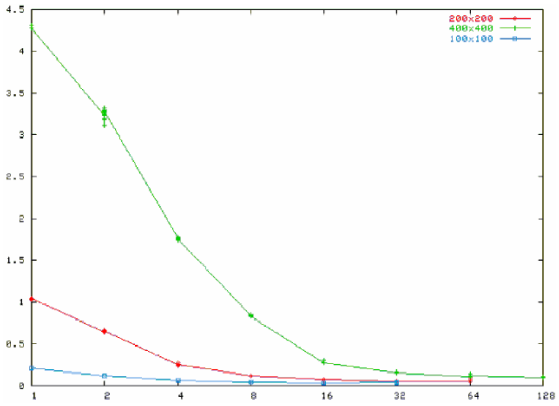


Figure 6 – Tree graphs for tree different problem sizes for 2D case (X-axis - number of processes, Y-axis – time given in seconds)

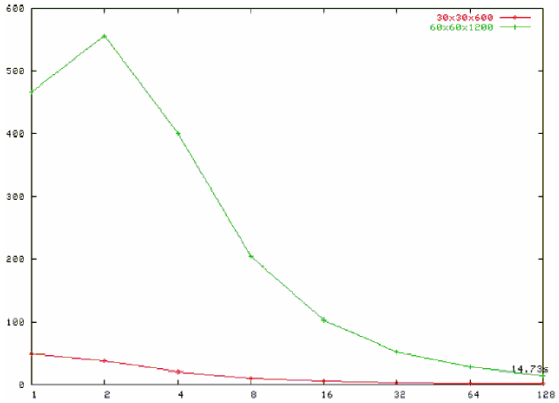


Figure 7 – Two graphs for two different problem sizes for 3D case (X-axis - number of processes, Y-axis – time given in seconds)

7. CONCLUSION

Even though most of today's computers are single processor computers, there are a lot of clusters and special purpose computers, and a lot of them still to come. Therefore, making computer programs for parallel execution on many processors is a very promising activity. Simulations are the best examples of the heavy computing programs, and as such, ideal candidates for making the code parallel. By making parallel version of CG algorithm, the time necessary for making a program was reduced in some cases even by 40 times. For achieving such a result one needs to be able to run a code on 128 processors architecture, like one in SGS department on IPVS in Stuttgart, Germany.

ACKNOWLEDGMENT

This work would not be possible without help of Prof. Joachim Bungartz and Ionel Muntean. Many thanks to them for inviting me on a three months practice in Stuttgart, and providing me access to the cluster Mozart, as well as for the great classes I had taken.

REFERENCES

[1] "A message passing standard for MPP and workstations", Jack J. Dongarra, Steve W. Otto, Marc Snir, Yorktown Heights, David Walker

[2] "Comparing the OpenMP, MPI, and Hybrid Programming Paradigm on an SMP Cluster", Gabriele Jost, Haoqiang Jin, Dieter an Mey, and Ferhat F. Hatay

[3] "MPI: The Complete Reference", M. Snir, S.W. Otto, S. Huss-Lederman, D. W. Walker and J. J. Dongarra. Published by the MIT Press, 1995.

[4] "The Emergence of the MPI Message Passing Standard for Parallel Computing", R. Hempel and D. W. Walker, Computer Standards and Interfaces, Vol. 7, pages 51-62, 1999.

[5] "Redistribution of Block-Cyclic Data Distributions Using MPI", D. W. Walker and S. W. Otto, Concurrency: Practice and Experience, Vol. 8, No. 9, pages 707-728, November 1996.

[6] "MPI: A Standard Message Passing Interface", J. J. Dongarra and D. W. Walker, Supercomputer, Vol. 12, No. 1, pages 56-68, January 1996.

[7] "The Design of a Standard Message-Passing Interface for Distributed Memory Concurrent Computers", D. W. Walker, Parallel Computing, Vol. 20, No. 4, pages 657-673, April 1994.

[8] "Standards for Message Passing in a Distributed Memory Environment", D. W. Walker, Technical Report ORNL/TM-12147, Oak Ridge National Laboratory, August 1992.

[9] "On Characterizing Bandwidth Requirements of Parallel Applications", Anand Sivasubramaniam, Aman Singla, Umakishore Ramachandran, and H. Venkateswaran, College of Computing, Georgia Institute of Technology, Atlanta, GA 30332-0280.

[10] Milutinovic, Veljko, “The Best Method for Presentation of Research Results”
IEEE TCCA NEWSLETTER, September 1996.

[11] Lectures from Prof. Dr. Hans-Joachim Bungartz, http://www5.in.tum.de/persons/bungartz.html

[12] Lectures from Ioan Lucian Muntean http://www.ipvs.uni-stuttgart.de/abteilungen/sgs/abteilung/mitarbeiter/ioan_lucian.muntean/de

[13] “Teaching High-Performance Computing on a High-Performance Cluster”, Ralf-Peter Mundani, and Ioan Lucian Muntean. IPVS, Universitat Stuttgart http://www.springerlink.com/index/YWAG0MWXGQ95HL21.pdf

The Pattern-Oriented Decision-Making Approach

Delibašić, A., Boris; and Suknović, B., Milija

Abstract-This paper introduces a pattern-based approach when solving decision-making problems. We believe that integrated solutions of algorithms and methods in multiattribute decision-making (MADM) and data mining (DM) do not support the decision making process as they could in the process of finding an acceptable solution, or gaining knowledge. Most methods and algorithms in MADM and DM provide the decision maker an acceptable solution. On the other side the analysts have no freedom to adapt the methods or algorithms to the subtle details of the problem; so many new problems can't be handled well. We propose a solution of building pattern solutions for algorithms and methods of MADM and DM, where these patterns have passed experience validation and could be used well as building components in modular development environments. We believe this way that analysts could be able to generate their own algorithms and methods which could better adapt to new problems and generate better solutions. In this paper we present the four big patterns of decision-making and their common realizations. We present also the developed platform for modular MADM.

Index Terms: ***Decision-making patterns, Decision-making process, Business intelligence, Modular multiattribute decision-making application***

1. INTRODUCTION

When designing solutions in different areas there are integrated and there are modular solutions. Integrated solutions provide us with simplicity, but the lack of choice. Modular solutions give us the ability to have greater influence on our solution, but ask for more knowledge and attendance of the decision maker. In reality all solutions are between full modularity and full integrality [10,s.4]. We believe that for solving problems in the decision-making area, it is more appropriate to use modular solution, than integrated one.

We think that MADM methods and DM algorithms could be used more intensive in the decision-making process. Decision-making is sometimes an innovative activity. Decision-making and the process of design are interrelated [14]. Sometimes, it is unacceptable to model a decision or to seek knowledge with a method or algorithm which is not adaptable to the new problem. For the decision-makers the formal decision-making process is a black box. We think the time is right for making a white box of the algorithms and methods in DM and MADM.

With the development of pattern theories in various areas (architecture, IS, tele-communications, organization) it seemed that the problems of adaptability and maintenance of decision-making methods and DM algorithms could be solved using patterns. That is why decision-making patterns in the area of business intelligence and business decision-making have been identified and presented.

In this paper an original software platform for modular MADM is being presented. The decision analyst is now equipped with a tool that allows him generating his own MADM method.

2. PATTERNS ARE BUILDING BLOCKS

The ideologist of the pattern movement, Cristopher Alexander defines pattern as *a three-part rule that expresses the relation between a certain context, a problem and a solution. It is at the same time a thing that happens in the world and a rule that tells us how to create that thing and when to create it. It is at the same time a process, a description of a thing that is alive and a process that generates that thing.* [3] This definition tells that for problems in certain context, there is most often a solution. The solution a pattern provides is a thing (the solution) and a process of achieving the solution. A pattern is a building block that can be used for generating solutions. What is more important is the following statement of Alexander [1]: *The human feeling is mostly the same, mostly the same from person to person, mostly the same in every person. Of course there is that part of human feeling where we are all different. Each of us has our unique individual human character. That is the part people most often concentrate on when they are talking about feelings, and comparing feelings. But that idiosyncratic part is really only about 10% of what we feel. 90% of our feelings is stuff in which we are all the same and we feel the same things. So, from the very beginning when we made the pattern language, we concentrated on that fact, and concentrated on that part of human experience and feeling where our feeling is all the same. That is what the pattern language is – a record of that stuff in us, which belongs to the 90% of our feeling, where our feelings are all the same.* This means that in the Alexandrian way [1][2][3] a pattern is something that belongs to the common value of humans, that is, a pattern

is understandable for most decision-makers, if not all.

So, patterns can be used for building systems and patterns are understandable for most people. If patterns could be identified in methods and algorithms of MADM and DM, then the decision-maker could be provided with a tool of generating their own methods and algorithms, depending on the problem he/she is solving. That way, black boxes from business intelligence could transform into white boxes.

For the formal representation of patterns in this paper the J.O. Coplien pattern formalization form has been used [4] [5, s. 8]. This form consists of the following elements:

- Context,
- Problem,
- Forces,
- Solutions and
- Resulting context.

Pattern starts with a context and result with a resulting context. Patterns change the space in which they are used. A context describes the surrounding that should be able to fit the pattern. Patterns are dependant of the context in which they are used.

Problem describes what produces the uncomfortable feeling in a certain situation, and what forces the decision-maker to seek a solution. The uncomfortable feeling comes from the forces. Forces are keys for pattern understanding. They describe the problem space. They are directed in different ways (example: force 1: excellent book, force2: very expensive), so they are not harmonized and do not produce a solution.

When a problem and its forces are well understood, then a solution is easily recognized. The reason is because patterns are familiar to people. People and the nature alone are made of patterns. The solution, the pattern itself, resolves forces and provides a good solution. On the other hand, a pattern is always a compromise, while it resolves some forces, it adds to the context space new ones.

The resulting context describes what the pattern has done positive and what negative. The five steps or elements can be done many times in order to find better patterns. Patterns change because new forces can make the pattern become inapplicable, so a new pattern can replace previous used pattern. So the process of reaching pattern is continual.

Alexander says [2]:

1. Patterns contain life.
2. Patterns support each other: the life and existence of one pattern influences the life and existence of another pattern.
3. Patterns are built of patterns, this way their composition can be explained.
4. The whole (the space in which patterns are implemented to) gets its life depending on the density and intensity of the patterns inside the whole.

3. THE FOUR GREAT DECISION-MAKING PATTERNS

We have conducted an analysis in MADM, DM [12], case-based reasoning [16], and artificial neural networks [15] [9] and identified four patterns that fully describe and support the decision-making process. Four patterns that were identified are:

1. Divide et impera,
2. Assign a common value,
3. All as one (More information in one), and
4. Analysis (Feedback).

3.1. Divide et Impera

Divide et impera is a common pattern. People have been using this pattern since their creation, this pattern can be found everywhere in nature. Every problem is usually separated into small parts in order to be handled. All our problem solving is dependent upon this pattern. It is referred to as analysis. A symbolical representation of the pattern is shown on Figure 1. The height can not be conquered “at once”, but it is necessary to divide the height into smaller, manageable heights and the big problem can then easily be solved.

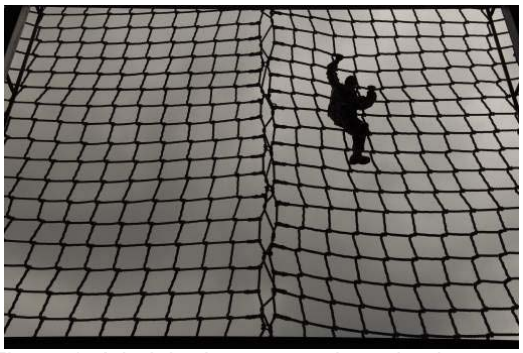


Figure 1. A height that can not be solved at once

In decision-making this pattern is used for structuring problems. It is necessary to define for a problem several attributes, their weights and relations. In the Coplien form it can be written down as follows:

Context: A decision-making problem has aroused.

Problem: There is not a criterion that can describe the problem by itself. There are a lot of attributes that have different intensity (weights), that have relations among them, and that can define the problem. How to know what attribute to choose?

Forces:

- By dividing the problem into sub problems (attributes) it is possible to handle the problem.
- It is not clear how much each attribute influences the problem, its weight is unknown.
- It is possible to choose an attribute that has no influence on the solution of the problem or to

leave out an attribute that has big influence on the solution of the problem.
Solution: The problem should be divided into attributes, so the problem model will be gained. In this model all attributes should be assigned with weights and their relation towards other attributes. It is a crucial task, so it is advisable to consult an expert.
Resulting context: Problem structure is given. One can not be sure if the problem structure will solve the problem correctly. It is known that the problem structure (model) has big influence on the solution.

3.2. Assign a Common Value

Assign a common value is a crucial pattern in team building, organization and all spheres of life where cooperation between objects is necessary. It can be found in all armies on the world and it is one of the army's basic principles. On Figure 2, it is shown what this pattern does.

Figure 2. Uniforms in armies (common values assignment)

All soldiers wear the same uniform. In order to have a good army, every soldier must accept common values on behalf of his/her own individuality. Only this way an army can be successful.

In the decision-making area, this pattern is known as normalization. Normalization provides the data with uniformity, but on the other hand makes a lot of information loss. It is the same principle like in the army. It is a compromise. The Coplien pattern form is:
Context: A problem is structured and the cases (alternatives) have been selected according to the problem structure.
Problem: Not all attributes have the same data type. How to compare and manage the data?
Forces:
- Assigning a common value to all data, the case table becomes easy to manage and analyze.
- A huge amount of information is lost in the data normalization process.
- Expert knowledge about the decision problem is needed.

Solution: Assigning a common value and taking care about minimal loss of information.
Resulting context: The case table is ready for analysis. It is possible that the wrong normalization (assignment of common value) will produce a wrong solution.

3.3 All As One

All as one is a pattern that is known as synthesis. When a team, an organization etc., has reached a common value, it produces excellent results. The team acts as one and solves problems very efficiently. On Figure 3 it is shown how a well prepared team looks like.

Figure 3. The whole team is aiming at the same direction

In decision-making it is necessary to have a summarized value upon which the decision-maker can act. If the data is not well normalized, the summarized value and the solution will not be good. The pattern can be presented as:
Context: The problem has been structured and a common value has been assigned to data.
Problem: The decision maker can not decide upon multiple attributes. How to decide upon multiple attributes? How to sort a case table according to many attributes at the same time?
Forces:
- Having many attributes provides the decision-maker with good information for decision-making.
- The decision-maker needs a summarized attribute upon which he can decide what to do.
- Summarizing attribute values is followed with information loss.

Solution: Choose an aggregate value that can summarize more information in one. One should test more summarizing functions and select the one that gives best information to the decision-maker.
Resulting context: All alternatives (cases) have been assigned with a summarized value upon which the decision-maker knows what to do. All cases can now be sorted. The problem is that the summarized value often does not reveal all aspects of the solution, because a huge amount of data has been lost in data aggregation.

3.4 Analysis

Analysis is the last, but maybe the most important pattern in decision-making. It is a feedback pattern that allows improvement and learning of every system. On Figure 4 a feedback is shown. Though all soldiers are experienced, the trainer explains them their mistakes and explains how they could get even better.

Figure 4. Analysis and preparation for new tasks

In decision-making it is important to see the accuracy of the solution and how it can be improved. It is often measured as the distance of the expected results and the results produced by the model. This pattern checks if the problem has been structured well, if the common value has been assigned properly and if the summarizing function has been well chosen. The Coplien form of the pattern is:

Context: A model has generated a solution for a problem.

Problem: The question is if the solution responds well to the problem? It is not possible to check if all of the previous performed patterns are done well separately.

Forces:

- Analysis can help improve solutions, by changing problem structuring, common value assignment, and aggregate functions.
- When the improvement of a previous mentioned pattern is done, it is assumed that the other patterns are performed well, that is, one can not see at the same time which pattern is good and which not.

Solution: Analysis should be done. A perfect solution can not be reached, because it is always assumed that some patterns are correct in order to improve the other ones. One must always have a leaning point.

Resulting context: Analysis helps in improvement of the system. On the other side, it can not guarantee reaching a perfect solution, just an acceptable one.

3.5 The Decision-making Process

Using the four great patterns of decision-making, it is now possible to define the decision-making process by forming a pattern language. The decision-making process, which is at the same time knowledge retrieval process, is shown on Figure 5.

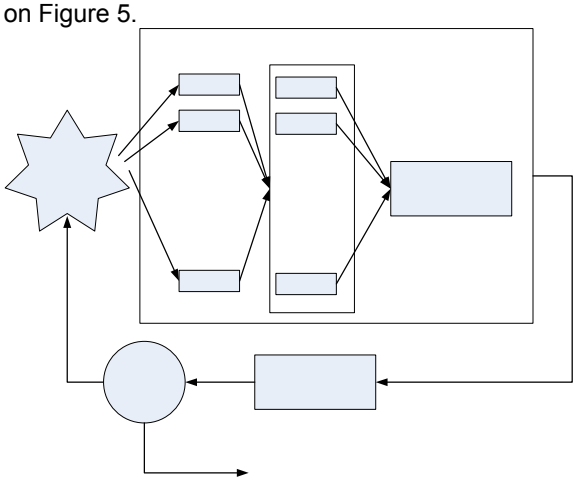


Figure 5. The pattern decision-making process

After a problem arises, it is necessary to divide the problems into sub-problems (attributes) to handle the problem. On the other side, one can not solve problems using multiple attributes. One attribute is needed in order to reach a solution. It is necessary to transform more attributes into one aggregate attribute. One can not get aggregate functions, until all data has been assigned a common value.

At last, one needs to check if the problem is solved well, and if not, to improve the system. So, the process of decision-making and problem solving is eternal and consists of analysis-normalization-synthesis-feedback-analysis etc.

4. THE MODULAR DECISION-MAKING APPLICATION (MADAM)

The ideas in the previous part where used to identify common patterns in MADM. Several MADM methods were analyzed in order to identify patterns of MADM. The analyzed methods are [6]: Simple additive method, Hierarchical additive method, AHP [6], ELECTRE [6], PROMETHEE [6], utility theory etc. In Figure 6 the identified patterns are classified depending on that to which pattern they belong.

In *Divide et impera* there are three patterns: Attribute rank, Attribute weight, Estimation matrix. In A*ssign a common value* there are six patterns: Quantification, Estimation matrix, Min to Max, Preference types, Normalization, Utility functions. In *All as one* there are three patterns: Maxmin, Maxmax and Expected utility. In *Analysis* there is one pattern: What-if analysis. These patterns have been implemented in an application we call MADAM. It is developed in Python program environment and represents a modular MADM system.

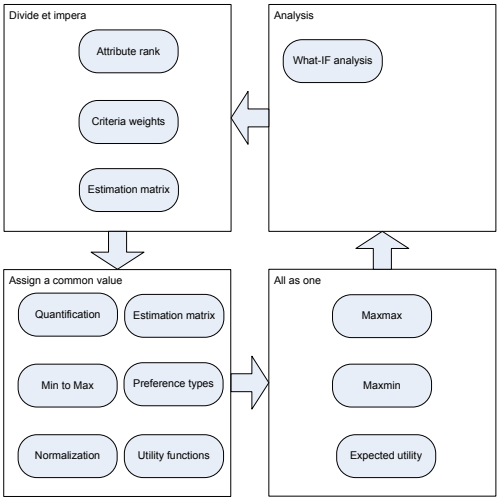


Figure 6. MADM pattern language

4.1 An example of application

We have developed a system we call modular multiattribute decision-making platform (MADAM) [7][8][9]. MADAM has user interface like depicted on Figure 7. It resembles the user interface of SPSS Clementine. The decision maker has to lead the decision-making process in order to reach an acceptable solution. He is building a stream that represents a new generated MADM method.

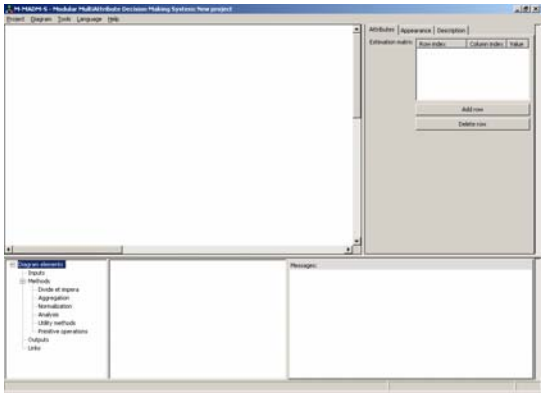


Figure 7. MADAM Interface

Suppose we solve a MADM problem. In general, there are *n* attributes and *m* alternatives. The goal is to find a most acceptable solution, to choose an alternative.

How was this done traditionally? [11] One would choose several MADM methods and see if they generate a common solution. One would, after several iterations of testing and analysis, choose the alternative which came out as the most acceptable for most MADM methods.

Our approach is different. We don't want the decision-maker to work with many MADM methods (many integrated methods) but we want him into the decision-making process. That means that we want the decision-maker, if needed, to model his own MADM method. The decision-maker has to be aware of the compromises he makes in that process. He has to be aware what he gets, and what he looses in every step in decision-making. He/she builds the solution by composing patterns. Every pattern gives something to the decision-maker, but takes something away. We think this is a more natural way of modeling the MADM problem.

Every MADM method has some good and some bad things. But, they are integrated. Often we can not "pull out" what is good from a method and leave out the inappropriate stuff. Traditionally, several MADM methods are being used and there results are being compared in order to look for some consensus between methods solutions.

As an example, suppose there are five attributes when solving a MADM problem. The attributes are k_1 to k_5. Figure 8 presents a stream in MADAM that could solve the MADM problem. After the data from the case table has been structured and read, the decision-maker does the following:

1. Transforms qualitative data to quantitative (Quantification). All data becomes numerical, but a lot of information is lost.
2. The decision maker can decide to use L1 metrics for normalizing attribute k_1, and $L\infty$ metrics to normalize attribute k_2. Every normalization type has its good sides and its bad sides. Choosing metrics depends upon the nature of data and the nature of the problem.
3. Normalizing attribute k_3 using the linear preference type from PROMETHEE (Preference type).
4. Normalizing attribute k_4 using an estimation matrix from AHP, and
5. Normalizing k_5 using utility functions.

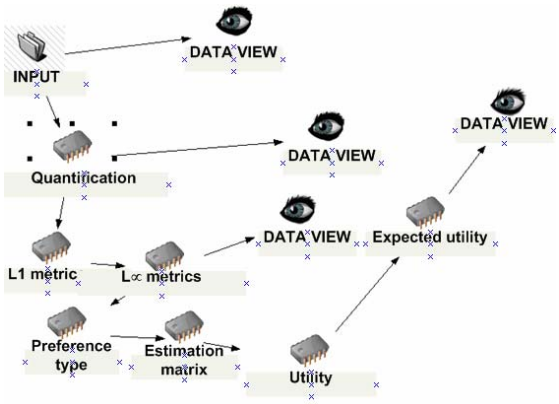


Figure 8. The stream of generating a new method for problem solving in MADAM

All values in the case table are afterwards normalized [0,1] and the decision maker can choose an aggregate function. He selects the Expected utility function and uses estimation matrixes from AHP to define attribute weights (AHP weights module). The decision maker reaches a satisfactory solution and ends the decision-making process.

A method editor is implemented inside MADAM which allows easy implementation of new patterns that could be identified in praxis. The interface of the editor is shown on Figure 9. When implementing the mentioned patterns inside MADAM, the idea was to implement the frequently used patterns in MADM.

MADAM presents a platform that supports the decision process in MADM. The decision maker influences the process and has great influence on the solution. The decision maker has now the ability to make a compromise between using integrated solutions (old streams) or to build its own solution (stream). This decision depends mainly on the problem structure. If the problem is new, it is better to form a stream, or modify an existing one.

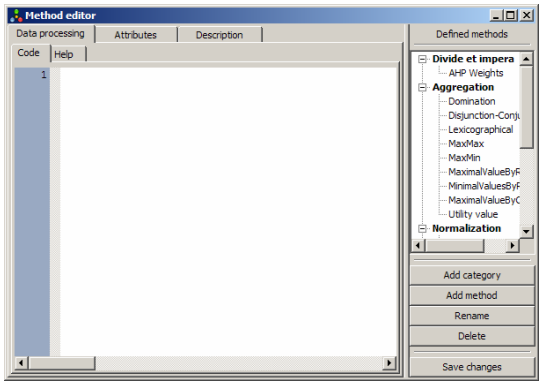


Figure 9. The method editor for modules maintenance and insertion.

5. CONCLUSION

When solving real MADM problems in the praxis, one can notice that the solutions are some kind of modification of existing MADM methods. So, the decision-maker always adapts his knowledge to new problem solving. It is like case-based reasoning [16].

What is important is that for a new problem, often a new method is created. We think that the expert decision maker does not like integrated solutions, but often uses modular approaches (combinations, mutations, etc.) which help him/her generate a solution.

It is also possible to help decision-makers explain the logic that is behind mathematical methods and algorithms. This is possible if the modules of the algorithms and methods are written down as patterns. The decision maker can then see the logic behind the methods and algorithms. He is then enabled to create his/her own method or algorithm that is better adapted to the problem.

There is not an infinite number of patterns, but a limited number. That is why it is possible to discover them in every area of human activity. *A decision maker becomes an expert in his area when he learns to use patterns from this area. A decision-maker is not an expert if he only knows integrated solutions*.

We have shown how this is possible to do in MADM. If the decision maker is looking for knowledge, it is a better approach to let him into the process, to see what he gets on every step, and what he looses. It is a more natural solution and in our opinion one direction of future data mining development.

REFERENCES

[1] Alexander, C., "The Nature of Order Book 1: The Phenomenon of Life", *The Center for Environmental Structure*, Berkeley, California, 2002.

[2] Alexander, C., „The Nature of Order Book 2: The Process of Creating Life", *The Center for Environmental Structure*, Berkeley, California, 2002.

[3] Alexander, C., "The Timeless Way of Building", *Oxford University Press*, 1979.

[4] Coplien, J.O., Zhao, L., "*Toward a General Formal Foundation of Design - Symmetry and Broken Symmetry", Brussels: VUB Press*, 2005.

[5] Coplien, J.O., "Software Patterns", *SIGS Books & Multimedia*, 1996.

[6] Čupić, M., Suknović, M., „Decision-Making: A Formal Approach", *FOS-Belgrade*, 2003.

[7] Delibašić, B., Suknović, M., "Pattern Decision Making", *Management,* No.– 39, 2005.

[8] Delibašić, B., Čupić, M., Suknović, M., Krulj, D., "A pattern based decision support system", *SIM-OP-IS*, Fruška Gora, Serbia, 2004.

[9] Delibasić, B., "Formalization of the business decision making process using patterns", doctor thesis, *FOS-Belgrade*, 2007.

[10] Eckert, C., Clarkson, J., "Design Process Improvement: a review of current practice", *Spriger Verlag London*, 2005.

[11] Holsapple, C., Whinston, W., Andrew, B., "Decision Support Systems - A Knowledge-Based Approach", *West Publishing Company*, 1996.

[12] Larose, D.**,"**Discovering knowledge in data, an introduction to data mining", *John Wiley & Sons Inc.*, December 2004,

[13] Lahdelma, R., Salminen, P., Hokkanen J., "Using multicriteria methods in environmental planning and management", Environmental Management 26(6),2000, 595-605.

[14] Simon, H.A, "The New Science of Management Decision", *Harper & Row*, 1960.

[15] Turban, E., "Decision Support and Expert Systems: Management Support Systems", Fourth Edition, *Prentice-Hall*, 1995.

[16] Watson, I., "Applying Knowledge Management – Techniques for Building Corporate Memories", *Morgan Kaufmann Publishers*, 2003.

Delibašić A. Boris is teaching and research assistant at the University of Belgrade, Faculty for Organizational Sciences, Center for Business Decision-Making.
Website: www.odlucivanje.fon.bg.ac.yu
Email: boris.delibasic@fon.bg.ac.yu

Suknović B. Milija is associate professor and vice-dean for curriculum at the University of Belgrade, Faculty for Organizational Sciences, Center for Business Decision-Making.
Website: www.odlucivanje.fon.bg.ac.yu
Email: milija.suknovic@fon.bg.ac.yu

Development of user-friendly didactic climate models for teaching and learning purposes

Goyette Stéphane, Hervé Platteaux, and François Jimenez

Abstract— ***This study reports on the development and application of two e-learning tools dedicated to climate science: these are Energy Balance Models, or EBMs. Such physically-based models form the ideal framework for studying fundamental energy processes at the basis of global climate and climate changes. The main assumption behind this development is that learning strategy would enhance the student's conceptual understanding from improved pedagogical technologies by allowing a greater interactivity and faster turn around, thus allowing a large number of experiments per unit time where all features are interfaced to appealing graphic displays. Consequently, these tools would contribute to learning efficiency. An analysis of the sort of reception such tools obtained in the student community in terms of their structural design, ergonomy and overall learning performances was carried out. The results show that their understanding of basic climate concepts may improve due to the interactivity and the graphic interfaces, allowing a visual display of the basic climate processes driving the energy balance of the Earth.***

Index Terms—*climate models, computer simulations, higher education, learning tools, Fortran, Java, JSP*

1. INTRODUCTION

The theoretical concepts fundamental to climate and climate change are taught at the bachelor level in a number of science departments (e.g., Geography) around the world. As is often the case, undergraduates do not have a profound knowledge of energy flow in the global climate system, and many of them are still having problems understanding the Earth's greenhouse effect, its anthropogenic disruption and the potential links to climatic change. Moreover, lecturers are expected to deal with a broad spectrum of student ability and background [1]. Courses and teaching methods require constant improvement and must also be adjusted to deal with classes having wider objectives. Recently, a growing interest in computer-based e-learning tools has prompted the development of innovative learning strategies [e.g., 2]. One of them is provided by web-based applications for climate processes. Nowadays, only a few, scattered and more-or-less user-friendly options with graphic interfaces exist to facilitate learning and better understanding of the complexity of the climate system. For example, in Java [3,4,5,6]; others present the many steps needed to achieve a climate model by means of Matlab [7,8] or with the Stella environment [9]. More complex software, allowing students to learn and experience the full climate system are available on the web [EdGCM; 10]; their use is, as yet, restricted to graduate students and people having the necessary scientific background, and form excellent methods for those who need to have a comprehensive knowledge of the climate system. Learning with increasingly innovative pedagogic methods may turn out to be more beneficial for learners than plugging numbers into memorized equations for which no connection to the real world exists, such as is the case in a classical teaching environment. Nowadays, computer simulations and virtual labs are becoming efficient tools for learning [11,12]. Traditional pedagogical supports such as blackboards, textbooks, transparencies and videos have been complemented by computer-based e-learning tools, allowing teaching to take place in a more polyvalent, ordered and appealing educational environment. Such new technologies are not intended to replace lecturers, however. The latter, rather than having to change their roles, may be less focused on teaching theoretical aspects of climate science and should concentrate more on the learning strategies to be adopted by the students, who would feel more involved in their training.

The goal of this study is to develop and apply a number of simple climate model interfaces aiming to improve teaching of climate and climate change concepts. In addition, these would help learning of climate processes by interacting with an easy-to-use interface, thus allowing fast turn-around experiments. One main advantage is that these interfaces can be used remotely, outside the lecture theatre, thus helping to optimise the

Revised manuscript received November 28, 2006. This work was supported in part by the Centre NTE of the University of Fribourg. Dr. Stéphane Goyette: Climate Research Group, University of Geneva, Route de Drize 7, 1227, Carouge, Geneva, Switzerland (http://homeweb2.unifr.ch/goyette/Pub/), Dr. Hervé Platteaux and Francois Jimenez, Centre Nouvelles Technologies & Enseignement, Pérolles 90, 1700, Fribourg, Switzerland.

absorption of climate concepts by students wherever computers connected to external networks are available. In the following study, an e-learning method is described, forming user-friendly didactic climate models for teaching and learning purposes. This is currently done by interfacing Fortran programs, using input files with Java script, thus allowing learners to interact easily by means of sliders and boxes where parameter values can be modified, and thus provide an appreciation of the results by means of an attractive built-in graphic interface. A number of students have been requested to evaluate these model interfaces, and the results have been compiled and analysed in order to give some credit to this method.

2. SIMPLE ZERO- AND ONE DIMENSIONAL ENERGY BALANCE CLIMATE MODELS

Modelling of the Earth's energy balance is founded upon physically-based calculations of the greenhouse effect. Such studies began when Joseph Fourier [13] explained that the atmosphere retains heat radiation. The energy budget then started with bulk calculations for the energy balance of the whole planet, as if it were a rock hanging in front of a fire. Tyndall [14] discovered that certain gases, such as water vapour and carbon dioxide (CO_2) are opaque to infrared rays, thus helping to keep our planet warm by preventing this radiation from escaping into outer space. Arrhenius [15] then developed calculations of the radiation transfer for atmospheres with differing amounts of CO_2, and speculated that changes could have caused Ice Ages and interglacial periods. Later, after continued theoretical and empirical work on the embryonic climate-change theory induced by variations of the concentration of greenhouse gases [e.g., 16, 17, 18, 19], the advent of digital computers helped to perform extended calculations of infrared absorption in the atmosphere [20, 21], revealing that significant climate changes were effectively plausible. It also became obvious that feedback had to be taken into account in order for the calculations to be considered realistic. The structures that scientists tried to build upon these findings may be called "models" and their first application was to explain the world's climates and their variation. Budyko [22] computed the balance of incoming and outgoing radiation energy according to latitude, and found that the heat balance worked very differently in high latitudes compared to those of low latitudes. Soon after, Sellers [23] built on these ideas and computed possible variations of the actual atmosphere separately for each latitude zone. Consequently, the theory that increasing industrial activities may eventually lead to a global climate much warmer than today has been corroborated. These thought-provoking results fostered interest in simple models where they served as a helpful starting point for testing a number of assumptions. Energy Balance Models (EBMs) of this type, although simple, remain interesting tools for studying global climate and climatic change [e.g., 24]. They may thus be considered as valuable virtual laboratories for modern students. These "simple" models that run on desktop computers were comparable to those that had been considered state-of-the-art for the most advanced computations in the 1960s.

A brief overview of the theoretical background used to build these models follows. The equilibrium between absorbed solar and emitted infrared radiation forms the first approximation to a model of the Earth's global climate. The version proposed in this project can be run through a user-friendly web page where a number of sliders can be used to modify parameter values. Step-by-step calculations are not revealed, but final results are displayed on the screen so that users can modify the displayed plots. These can be executed quite rapidly on desktop computers connected to a local network by establishing a connection to a server equipped with Java facilities [JSP; e.g., 25, 26], and where users can interact and visualise two EBMs (driven by Fortran programs) through sophisticated visual graphic displays (provided by Java scripts).

2.1 Description of the EBM 0D

This climate model may be used to numerically simulate and display the mean temperature of the global Earth-Atmosphere system, represented by a single point in space under the influence of absorbed solar and emitted infrared radiation [27; Chap. 10]. The temperature, T, evolves to a constant value after the equilibrium between the absorbed solar energy and the outgoing thermal infrared energy is reached, as follows:

$$D_m C \frac{dT}{dt} = \frac{F_s I_o}{4}\left(1-\alpha_p\right) - \tau_a \varepsilon \sigma T^4 \qquad (1)$$

where I_o represents the solar constant (1370 W m^{-2}) and α_p is the planetary albedo determining the solar absorption. All the parameter values are displayed in Table 1 of the Appendix. The time- and space-averaged energy input rate is therefore I_o / 4 over the whole Earth and the reflection is given by I_o α_p / 4. A rheostat, F_s, is introduced in order to change the solar power input. The surface may be considered as a blackbody in the infrared so that the infrared emissivity ε, may be taken as unity. The atmospheric transmissivity in the infrared is τ_a, and σ is the Stefan-Boltzmann constant (5.67 x 10^{-8} W m^{-2} K^{-4}); consequently the infrared loss is a function of the fourth power of the system temperature. The thermal inertia is controlled by C, the volumetric heat capacity of the system (J m^{-3} K^{-1}), mainly sea water, and D_m

is the ocean mixed layer depth (m). A finite difference equation used to simulate the temperature evolution may be written in the form:

$$T_n = T_{n-1} + \frac{\Delta t}{D_m C}\left[\frac{F_s I_o}{4}(1-\alpha_p) - \tau_a \varepsilon \sigma T_{n-1}^4\right] \qquad (2)$$

where iterations noted by integer values n $\in$ [1, 2, … , N], are achieved with a timestep Δt, here equal to one day, starting at 0 with T_o as the initial temperature followed by T_1 after Δt, etc., ending at T_N after N Δt (50 years). Under normal conditions, the equilibrium temperature is reached before the Nth iteration. In this version of the model, sliders allow users to modify the default parameter values of {T_o, α_p, τ_a, F_s, D_m} equal to {200 K, 30 %, 62 %, 1, 20 m} respectively (Fig. 1), producing an equilibrium temperature of 287.4 K (~14.2°C). If one or more values are modified, the rate of change and the equilibrium system temperature are also modified as shown by the different curves in Fig. 1. The parameters can be modified within reasonable values determined by each slider. One or more parameters may be changed simultaneously in order to visualise and therefore to understand the respective roles played by radiation and the thermal characteristics of the climate system with respect to global average temperature evolution.

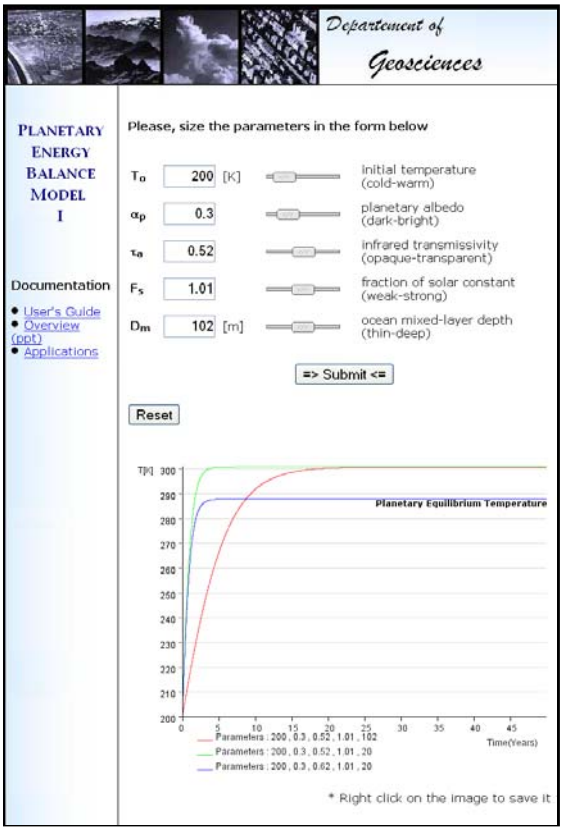


Figure 1. Planetary Energy Balance Model I (EBM 0D) home page. This interface allows the user to choose different parameter values and up to three experimental curves can be displayed on the orthogonal axis {Temperature (in K) vs Time (in years)} shown above. This page also hosts a number of documents available to the users, such as a user's guide, an overview as well as a reference manual for experiments.

2.2 Description of the EBM 1 D

If the surface of the Earth is partitioned into a fixed number of latitudes and the energy balance principle is applied, a one-dimensional climate model is thus developed [e.g., 27; Chap 10]. It computes the zonal average surface air temperature distribution from the Equator northward as a function of the absorbed solar radiation, the emitted infrared and the meridional heat transport at a given latitude according to the ideas of Budyko [22] and Sellers [23]. Assuming a zonal thermal equilibrium, and considering each zone separately, the formulation of this simple steady-state EBM 1D may be written as follows:

$$K_j^{\downarrow}(1-\alpha_j) = L_j^{\uparrow} + Tr_j \qquad (3)$$

In this application, $K_j^{\downarrow}$ = $K{\downarrow}(\varphi_j)$ represents the downwelling incident solar radiation at latitude φ_j, α_j = $\alpha(T_j)$ is albedo computed as a function of zonal temperature in zone φ_j, $L_j^{\uparrow}$ = L↑(T_j) is the zonal average infrared loss at latitude φ_j, and Tr_j = Tr (T_j) is the rate of transport of energy in/out latitude φ_j by the combined atmospheric and oceanic circulations. Here, T_j = $T(\varphi_j)$, represents the steady-state zonal-average temperature at latitude φ_j, j = 1, 9. In order to resolve Eq. (3) analytically, the infrared loss and the rate of transport of energy may be linearized as follows [24; Chap 3]:

$$L_j^{\uparrow} = A + B\,T_j \qquad (4)$$

where A and B are empirical parameters whose values account for the greenhouse gas concentration in the atmosphere, and the heat transport is parameterised as a heat diffusion process as follows:

$$Tr_j = K\left(T_j - \overline{T}\right) \qquad (5)$$

where $\overline{T}$ represents the globally-averaged temperature. The surface albedo can be parameterised by a step function as:

$$\alpha_j = \begin{cases} = 0.6 & \text{when } T_j \le T_{crit} \\ = 0.3 & \text{when } T_j > T_{crit} \end{cases} \qquad (6)$$

which represents the albedo increasing at the snowline where T_{crit}, is a critical temperature whose typical values are between -10°C and 0°C. By combining Eqs. (4), (5) and (6) into Eq. (3) and rearranging it, one can obtain the following solution for the steady-state zonal-average temperature:

$$T_j = \frac{K_j^{\downarrow}(1-\alpha_j) + K\overline{T} - A}{B + K} \qquad (7)$$

In this version of the model, individual sliders can be used to vary one or many parameter values within a range of reasonable values determined by each slider (Fig. 2). The effective albedo is a function of the surface albedo, α_{sfc}, of the cloud, α_{clouds}, and of the ice, α_{ice}. The longwave parameters A and B are set to 204 W m^{-2} and

2.17 W m^{-2} $°C^{-1}$, values first devised by Budyko [22]. After a few seconds of computations using default parameter values, it computes and displays zonal–averaged quantities such as temperature, absorbed solar energy flux, emitted infrared energy flux, etc. One or more parameters may be changed simultaneously in order to understand and visualise on the maps generated the respective roles played by radiation and thermal characteristics of the climate system with respect to global average temperature, where the snowline appears directly on these maps as shown in Fig. 2.

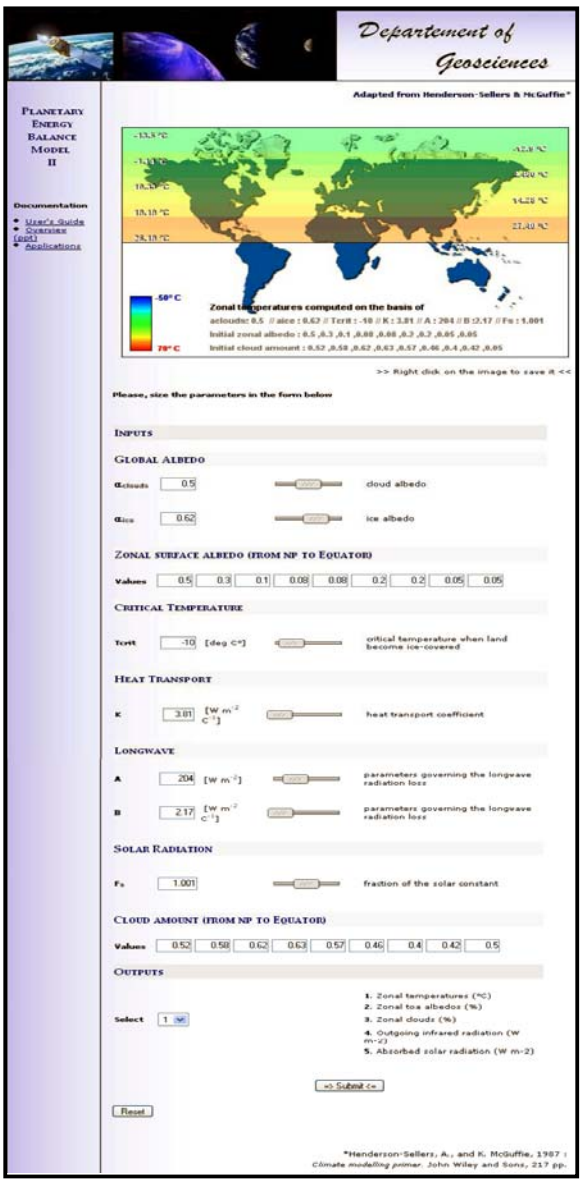


Figure 2. Planetary Energy Balance Model II (EBM 1D) home page. This interface allows the user to choose different parameter values, and different maps can be displayed at the top of this page. The page also hosts a number of documents available to the users, such as a user's guide, an overview and a reference manual for experiments as well.

3. Model Evaluation by a Sample of Learners

To undertake a combined qualitative and quantitative assessment of these model interfaces, a sample of M. Sc. students was prompted to make a survey. They were asked to answer a number of questions regarding the models' formal (visual display, graphic) in relation to their structural (functions, parameters, sliders) aspects; also taken into consideration were their essential values as a learning tool, and ultimately to their usefulness by employing the models to perform simple tests and applications in response to meaningfully prepared assignments. Individual answers were then compiled and analysed in the context of the e-learning environment to verify if these simulators met certain learning requirements and if they were bringing added value to the teaching environment.

3.1 Method

A number of students, studying Physical as well as Human Geography, were gathered in a room and equipped with enough computers for them to participate in this survey individually. The background of these students was very heterogeneous and not all of them had a solid scientific basis. The two model interfaces were directly available on-line. A questionnaire was distributed to all and a period of one and a half hours was allowed to go through it so as to obtain meaningful and detailed answers. In order to facilitate the compilation of all of the replies, students were asked to complete a document that was included in a Learning Management Environment (MOODLE), so that the textual portion of the answers could be illustrated by graphs copied and pasted directly from the simulator web page into this document. A short theoretical overview of the models had been provided prior to the evaluation. The first part of the assessment concerned formal aspects of the simulators regarding their overall visual aspects and their display features such as the use of sliders to change parameter values and graphic outputs; these were set up to test if the climate simulators were visually appealing or not. Other important questions concerned the usage (how often) and the time spent (how long) on each of the climate simulators. The second part of the assessment was aimed at general learning objectives. What were the most significant processes that the students had understood better, updated, or had been revealed to them using these models? Some more targeted questions were asked, for instance. How it showed that the inertia of the Earth's climate system influences the time to reach thermal equilibrium? Does the thermal inertia depend upon the initial temperature? What is the decrease of the solar energy necessary to produce an ice-covered Earth? What is the role of infrared transmissivity (a function of the greenhouse gas concentration in the atmosphere) on the maintenance of the Earth's surface temperature? What happens if the greenhouse gas concentration increases? What

is the role of meridional heat transport in the maintenance of zonal climates? With carefully chosen 1D EBM parameters, are the many results simulated by the 0D EBM reproduced exactly? How? etc.

3.2 Results

Ten questionnaires were returned after the survey. The compilation of the results indicate that all of the students considered that the simulators are visually appealing, and that their setup and objectives were satisfactory; i.e., simulating global-averaged temperature for the first (EBM 0D) and zonal-averaged temperatures for the second (EBM 1D). All of them mentioned that the structural organisation, sliders and the graph position were quite satisfactory. However, many of them proposed that online definitions of all of the parameters and functions should be available when the mouse-pointer encounters appropriate locations. All of them used the simulators several times during their learning phase (e.g., prior to this evaluation), and spent between fifteen minutes and an hour on it; they spent thirty minutes on average each time they logged on. When using these climate simulators, they also learned more about useful concepts, such as the order of magnitude of changes and model sensitivity, thanks to the sliders and to the graphic displays. An example is related to the thermal inertia of the earth, mainly controlled by the product «$D_m\, C$», as a determiner of the time needed to reach the global thermal equilibrium. This temperature is computed on the basis of Eq. (1) in which $dT / dt = 0$, i.e., when the absorbed solar energy is equal to the infrared lost, and it follows that:

$$T_{eq} = \sqrt[4]{\frac{F_s\, I_o \left(1 - \alpha_p\right)}{4\, \tau_a\, \varepsilon\, \sigma}} \qquad (8)$$

This temperature is independent of the thermal inertia. Given the same initial temperature (cold start at 200 K) the differences are greatest after two years, and it takes forty years for both experiments to converge to within 0.1 K when the inertia is multiplied by ten, as shown in Fig. 3. They found that the global average temperature is rather sensitive to infrared transmissivity τ_a. An important feature, related to this value, is that without an atmosphere i.e., $\tau_a = 1$, $T_{eq} = 255$ K (-18°C). The presence of greenhouse gases, such as water vapour, CO_2, CH_4 and N_2O in an atmosphere will decrease the infrared transmissivity and consequently trap heat in the lower atmosphere.

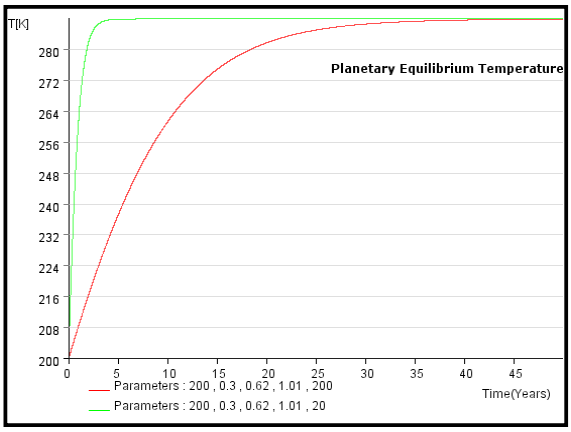


Figure 3. By using EMB 0D many users appreciated how long it took to reach thermal equilibrium and finally understood the meaning of the value of the steady state temperature.

They found that $\Delta\tau_a$ = -3.2 %, with τ_a decreasing from .62 to 0.60, which is roughly what is needed to simulate the effect of enhanced infrared trapping by doubling the CO_2 concentration in the air, produced a ΔT_{eq} = +2.4°C increase in global average temperature. This is a realistic prognosis for future climatic change that will occur during the course of the XXI^e Century. Also, an important feature of the behaviour of these simple climate models that the learners understood well is that an Ice Age can be easily simulated by decreasing the value of the solar constant. The global average temperature reached the equilibrium value of 273 K by changing the value of the solar constant by roughly ΔI_o = -19%. During theoretical lectures, they learned that meridional heat transport is induced by the energy deficit that exists between the Equator and the Poles. With the EBM 1D, it was obvious that reducing or increasing the heat diffusion coefficient *K*, by 10% produced a larger or a smaller North–South temperature gradient - as large as 3.7°C at the Poles and as small as 1.5°C at the Equator - but at the same time leaving the global average temperature unchanged. This is not so obvious, since the heat transport coefficient appears in both the numerator and the denominator in Eq. (7). There was a better understanding that reduction of *K* further acts as a decoupler for these two zones, i.e., the former getting hotter and the latter colder respectively, as shown in Fig. 4 for a 5% decrease of *K*. The students had been asked to emulate the increase in the greenhouse gas in the atmosphere by properly "tuning" the infrared parameters *A* and *B*. They soon found that the zonal temperatures are very sensitive to these. An increase of zonal temperatures can be simulated by decreasing the parameter *B*, that is decreasing further the infrared loss to space when temperature is increasing. In order to simulate an Ice Age, a similar decrease of the solar constant is needed, such as ΔI_o = -19%, where the former is defined in this context as an Earth and where the ice margin reaches the Equator.

3.3 Discussion

All the participants agreed that the conceptual aspects at the basis of a particular mathematical

model complemented by visual support aided a better understanding of climate and climate change. They also agreed that the theoretical lectures are essential prior to using these e-learning tools. The participants were also in agreement when they pointed out that one drawback following the use of such a tool is that it may not be conducive to further study, and to a feeling by learners that climate science is trivial and straightforward. This is one reason why these tools must be used as a complement to the theory, not as a surrogate for climate science.

To provide an example of this word of caution, we might mention that these models are generating a number of "raw" results which need to be analysed further. In that respect, learners were asked to appreciate the so-called "backward" consistency of the results. In particular, the EBM 1D results, when properly diagnosed, may be compared directly with those of the EBM 0D when the model input parameters are "similar". A comparison is made between the zonal averaged temperatures and the global average, where the former are averaged with weights proportional to the cosine of the latitude $\cos\varphi\,\Delta\varphi$. For example, when default parameters are used, the global average temperature simulated with the EBM 0D, gives 14.2°C compared with the 15.0°C of the EBM 1D. This difference, in simulated temperatures attributed to the different model parameterisations, is not deemed significant. The same diagnostics apply for the absorbed solar and the infrared lost to space, and also give comparable results with both models in the order of 235 ± 5 W m^{-2}.

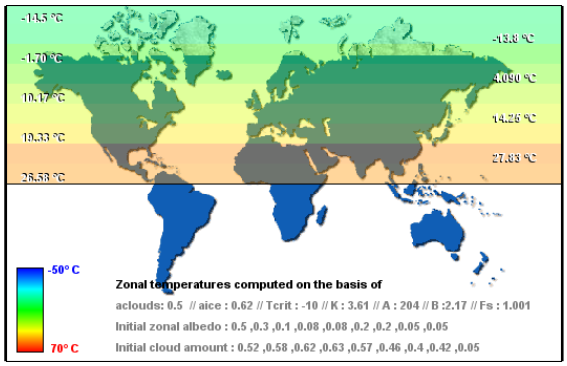


Figure 4. By using EBM 1D users found that decreasing the meridional heat transport coefficient by 5% decouples the Equator and the Poles, i.e., the Equator warms and the Poles cool by 0.5°C and -1°C respectively, but the global average remains unchanged.

A number of climatic aspects have not been considered here, such as a more complete sensitivity study of the Earth's global temperature relating to the variation of different parameters of the model: these include the amount of cloud, the surface albedo, etc., but this is well beyond the scope of the present study. Nevertheless, one important aspect they found is that, despite the relative success of these two EBMs to simulate general aspects of the thermal structure of the Earth, the models have limitations. Indeed, the geographical distribution of the temperatures is not represented, and neither is the vertical thermal structure. Consequently, they also considered that, to obtain a deeper understanding and analysis of the Earth's climate, one would have to use more complex modelling systems, such as Radiative–Convective models (RC), Global Climate Models (GCM) or the more detailed Regional Climate Models (RCM).

4. Concluding remarks

The development of computer-based e-learning climate model interfaces has been outlined in this paper. They are mostly dedicated to learning strategies for undergraduates, but may also be used by graduate students and others having to deal with the basics of climate science. An evaluation of two of these tools has been performed where a survey put emphasis on their usefulness, ease of use (ergonomy), added learning values, and on the quality of the graphic displays, thus helping to better absorb and understand the theory. The analysis of the survey indicates that the students are keen to use new technologies, and the added value of such pedagogic tools proved significant. With respect to the learning problems mentioned in the introduction, the analysis of the student comments concerning the quality of these climate model interfaces leads us to the conclusion that, by complementing the learning context with appropriate tools, students may absorb information more easily. We believe this justifies the development of such e-learning technologies where a more complete toolbox using similar "models" is definitely needed. This study is the first phase of a larger project to develop a much wider climate-oriented toolbox. A computer-based virtual learning laboratory similar to the enquiry-based facility in Physics [28] is also envisaged, which will include a set of virtual laboratory-based modules (e.g., models) that will provide a step-by-step introduction to climate science. Through the study of simple climate systems and their interaction, it is presumed that students would gain a better understanding of the fundamental processes controlling the climatic system. Along the lines described above, a global radiative-convective model (RC), based on the same system (i.e., Fortran + Javascript application), is currently under development according to the model equations set out in Chap 10 of [27] and those in Chap. 4 of [24]. This model will allow a deeper understanding of the role played by atmospheric greenhouse gas on the atmospheric temperature profile and thus on the thermal equilibrium of the Earth using an interface similar to that developed so far. One of the big challenges we are facing is to maintain and improve the quality of education on one hand and of the learning on the other in order to draw a large number of students keen on learning

climate science in an appropriate manner so that they gain a useful understanding needed later for research in other activities. E-learning resources provide a quantity of innovative methods, and in this study we have demonstrated to some extent their potential to develop efficient and useful tools having an indisputable added pedagogical value.

APPENDIX

Table 1. Definitions and ranges of the values of the adjustable parameters of EBM 0D and EBM 1D. Below, NP stands for North Pole and EQ for Equator.

0D EBM Parameters	Definitions	Ranges	Default values
T_o	Initial temperature	150 – 350 K	200 K
α_p	Planetary albedo	0 - 1	0.3
τ_a	Atmospheric transmissivity	0 - 1	0.62
F_s	Rheostat	0.5 – 1	1.0
D_m	Mixed/layer	1 – 200 m	20
1D EBM			
α_{clouds}	Cloud albedo	0 - 1	0.5
α_{ice}	Ice albedo	0 - 1	0.62
α_{sfc}	Surface albedo	0 - 1	.5 (NP), .3, .1, .08, .08, .2, .2, .05, .05 (EQ)
T_{crit}	Critical temperature	-15 – +5°C	-10.0°C
K	Heat transport coefficient	0 – 50 W m^{-2} $°C^{-1}$	3.81 W m^{-2} $°C^{-1}$
A	Infrared parameter	150 – 310 W m^{-2}	204 W m^{-2}
B	Infrared parameter	0 – 20 W $m^{-2}°C^{-1}$	2.17 W $m^{-2}°C^{-1}$
F_s	Rheostat	0.5 – 1	1.0
Ci	Cloud amount	0 - 1	.52 (NP), .58, .62, .63, .57, .46, .4, .42, .5 (EQ)

ACKNOWLEDGMENT

The authors are thankful to the anonymous reviewers that have helped to finalise this manuscript. Also, they are grateful to the Centre NTE, to the Department of Geosciences of the University of Fribourg, Switzerland. Lastly, the authors wish to thank the "Service informatique de l'Université de Fribourg" for hosting these model interfaces - whose URL is the following: `http://elearning.unifr.ch/ebm`

REFERENCES

[1] Ramsden, P., "Learning to teach in higher education," *Routledge*, London, UK, 1992.

[2] Hantsaridou, A. P., Theodorakakos, A. Th., Polatoglou, H. M., "A didactic module for undertaking climate simulation experiments," *Inst. Phys. Publ.*, Eur. J. of Phys., Vol. 26, 2005, pp. 727-735.

[3] Global Energy Balance Model, Clark College, Computer Science Lab., http://cs.clark.edu/~mac/physlets/GEBM/ebm.htm (October, 2006)

[4] Energy Balance Model, The Shodor Education Foundation, Inc., http://www.shodor.org/master/environmental/general/energy/energy.html (October, 2006)

[5] Latitudinal Temperature Computations, Rowland's Learning and Teaching Home Page, http://www.worc.ac.uk/LTMain/Rowland/mec/MODELS/1dmodel.htm?selectnav=1DModelOutline.html (October, 2006)

[6] Energy Balance Climate Model 1, Center for Climatic Research, University of Wisconsin-Madison, http://ccr.aos.wisc.edu/model/ebcm/EBCM1.html

[7] A Matlab implementation of a Zero Dimensional Energy Balance Model, http://www.noc.soton.ac.uk/soes/research/groups/ocean_climate/demos/ebm/mocha_install.html

[8] One-Dimensional Energy Balance Model, Department of Physics Gustavus Adolphus College, http://physics.gac.edu/~huber/envision/instruct/ebm2doc.htm (October, 2006)

[9] Modeling Earth's Climate System with STELLA, Carleton College, Minnesota, http://www.carleton.edu/departments/geol/DaveSTELLA/climate/climate_modeling_1.htm (October, 2006)

[10] EdGCM: The Project, http://edgcm.columbia.edu/ (January, 2007)

[11] De Jong, T., Sarti, L. E., "Design and production of multimedia and simulation-based learning material," *Kluwer Academic Publisher*, Dordrecht, 1994.

[12] Huang, C., "Changing learning with new interactive and media-rich instruction environments: virtual labs case study report," *Elsevier Science*, Comput. Med. Imag. Graph., Vol. 27, 2003, pp. 157-164.

[13] Fourier, J., "Remarques générales sur les températures du globe terrestre et des espaces planétaires," Annales de Chimie et de Physique, Vol. 27, France, 1824, pp. 136-167.

[14] Tyndall, J., "On Radiation through the Earth's Atmosphere," Philos. Mag., Vol. 25, UK, 1863, pp. 200-206.

[15] Arrhenius, S., "On the influence of carbonic acid in the air upon the temperature of the ground," Philos. Mag., Vol. 41, UK, 1896, pp. 237-276.

[16] Simpson, G. C., "The distribution of terrestrial radiation," Memoirs of the Royal Meteorological Society, Vol. 23, UK, 1929, pp. 53-78.

[17] Hulburt, E. O., "The temperature of the lower atmosphere of the Earth," *Nat. Acad. Sci.*, Phys. Rev., Vol. 38, USA, 1931, pp. 1876-1890.

[18] Callendar, G. S., "Infra-red absorption by carbon dioxide, with special reference to atmospheric radiation," Q. J. Roy. Met. Soc., Vol. 67, UK, 1941, pp. 263-275.

[19] Chandrasekhar, S., "Radiative Transfer," *Clarendon Press,* Oxford, 1950, reprinted by Dover Publications, New York, 393 pp.

[20] Plass, G. N., "Effect of carbon dioxide variations on climate," *Am. Assoc. Phys. Teach., Publ.*, American J. Physics, Vol. 24, USA, 1956, pp. 376-387.

[21] Möller, F., "On the influence of changes in the CO_2 concentration in air on the radiation balance of the Earth's surface and on the climate," *AGU*, J. Geophys. Res., Vol. 68, USA, 1963, pp. 3877-3886.

[22] Budyko, M. I., "The effect of solar radiation variations on the climate of the Earth," *Blackwell Publ.,* Tellus, Vol. 21, Sweden, 1969, pp. 611-619.

[23] Sellers, W. D., "A global climatic model based on the energy balance of the Earth-Atmosphere system," *AMS*, J. Appl. Meteorol., Vol. 8, USA, 1969, pp. 392-400.

[24] Henderson-Sellers, A., McGuffie, K., "Climate modelling primer," *John Wiley and Sons*, New York, 1997.

[25] Negrino, T, Smith, D., "JavaScript & Ajax for the Web,", *Peachpit Press*, 6th ed., Berkeley, CA, USA, 2006.

[26] Hall, M., "Servlets & JavaServer Pages," *Campus Press*, Boulder, Co., USA, 2000.

[27] Trenberth, K. (Ed.), "Climate system modelling," *Cambridge University Press*, UK, 1992.

[28] McDermott, L. C. "Physics by Inquiry Volumes I & II," *John Wiley and Sons*, New York, 1996.

Knowledge Processing and Computer Architecture

Omerovic, S., Tomazic,S., Milovanovic, M., and Torrents, D.

Abstract— ***This position paper argues that the most suitable computer architecture for knowledge processing in bioinformatics is TM (transactional memory), ported into the DSM (distributed shared memory) environment, and expanded with elements of SMT (simultaneous multithreading. Current implementations of TM are in the SMP (shared memory multiprocessor) environment and without extensive support for SMT. In order to justify this position, the paper treats the field of decision making (DM) applied to knowledge processing for the need of bioinformatics. The basic idea is to have an automated reasoning mechanism Decision Making System (DMS) able to make a decision (related to the corresponding question), if the input data are in a text form (like it is the case in genomic processing). An illustration of Data modelling and Analysis layer, as a part of DMS and for the purpose of genomic processing, is given next. Bioinformatics experts mostly use BLAST software output in order to make decisions concerning genomic data. This DM process is mostly done manually, making it dependent on the expert knowledge and talent, in a way which is (for the most part) not automated and therefore not uniform and not eligible for global data exchange and comparing. We have proposed an approach that may lead to automatization of the DM process, based on the theory of concept processing (upgrade of Data Mining and Semantic Web). Analysing the typical processing needs in the set of genomic data (atomic access and high levels of parallelism), conclusion is that TM Systems (TMS) offer the processing capabilities demanded by both the Data Modelling and Analysis (layer 2) and Concept Processing (layer 3), but first have to be ported from SMP to DSM and enhanced with SMT (to support the higher levels of parallelism), before they can be successfully applied in genomic processing and other areas of science where huge-volume knowledge processing through DM is required.***

Index Terms— ***genomic, knowledge, memory, modelling***

Manuscript received January 30, 2007. This work was supported in part by the Supercomputing Centre, Barcelona, Spain.

Sanida Omerovic and Saso Tomazic are with the Faculty of Electrical Engineering, University of Ljubljana, Slovenia (e-mail: sanida.omerovic@lkn1.fe.uni-lj.si, saso.tomazic@fe.uni-lj.si). Milos Milovanovic is with Supercomputing Centre, Barcelona, Spain (e-mail: milos.milovanovic@bsc.es). David Torrents is with the ICREA-Supercomputnig Centre, Barcelona, Spain (e-mail: david.torrents@bsc.es).

1. *Introduction to Decision Making Systems*

Every day surrounds us a vast amount of data: newspapers, radio, TV, Internet, etc. And every day a person makes hundreds of decisions: what to eat, what to wear, where to go, who to contact, etc. The whole DM process is happening person's head, so the most general and oldest DMS is human brain. Decisions are made by input data (based on person's everyday perception), person's DM criteria's (already stored in person's brains based on life experience) and predefined knowledge (everything what person have learned from birth up to the present moment).

Every system that has unstructured data and question/s as input and decision/s as output, can be observed as a DMS consisting of the following four layers: Data retrieval layer + Data modeling and Analysis layer + Concept processing layer + Decision making layer. For each one of these layers, have its organization presented, and discussed its processing needs (implementation requirements). Observed DMS for the purpose of DM in the machine world has the analogy of the human DM process mentioned above. DMS layers are presented in Figure 1 and one can conclude that these layers are involved iteratively.

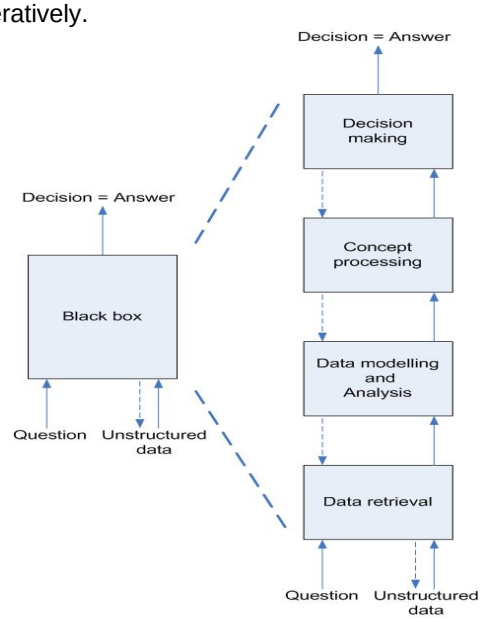


Figure 1. DMS – from general to layered view. Inputs are question and unstructured data, and output is decision (usually in a form of answer to the specific question). DMS core consists of the following four layers: Data retrieval, Data modelling and Analysis, Concept processing, and Decision making.

In the case of the Data retrieval (DR) layer, the essence is (as the name itself says) retrieval of unstructured data. So, in this layer one is gathering all types of data (text, audio, video, and pictures) into DMS. Here, one is dealing with a different data sources, and filtering helps extract only the data needed for Decision.

In the case of the Data modelling and analysis (DMA) layer, the essence is modelling and analysis of the unstructured data. So, after gathering as much as possible data related only to output Decision, data are modelled in a uniform manner, so that they can be compatible for the further analysis. At this point, one tries to eliminate noisy data (data that may be invalid), so that only valid data are used to make concepts from, which is next step.

In the case of the Concept processing (CP) layer, the essence is that it includes two sub-layers: Concept Modelling and Concept Search. This is the core of proposed DMS, and the idea is that reasoning mechanisms for Concept definition, Concept population, and Concept replacement are embedded into this level. In this way, knowledge is presented by concepts, and DMS operates on Conceptual-level, instead of Semantic-level like search engines today (Google, Yahoo, etc). This can be done by using Neural Networks, Fuzzy logic, Space Vector Model [1], or similar statistical methods.A detailed idea of CP layer is presented in Section 3.

In the case of the DM layer, the essence is that this layer contains a reasoning mechanism that combines concepts (from below layer) and DM criteria's that are stored in this layer and directly related only to output Decision. DM criteria are defined by the party that is using DMS (it can be a person or a company).

2. Data Modeling and Analysis for Genomic Processing

This section gives practical example of the first two DMS layers, namely DR and DMA, in the case of genomic processing. As mentioned before, in a DMS, input can be any kind of unstructured data. For the purpose of this section, analysis is limited on text only, because genomic processing uses only text as input. That fact serves to us as a justification to apply DMS in genomics.

Genomic researchers mostly deal with similarity issues between genomic sequences. Genomic sequences are treated as long sequences of letters A (Adenine), G (Guanine), C (Cytosine), and T (Thymine) which represents nitrogenous bases in protein structure.

As a illustration for previous sentence, DNA nitrogenous base pairs and a DNA sequence example is presented in Figure 2 and Figure 3, respectively

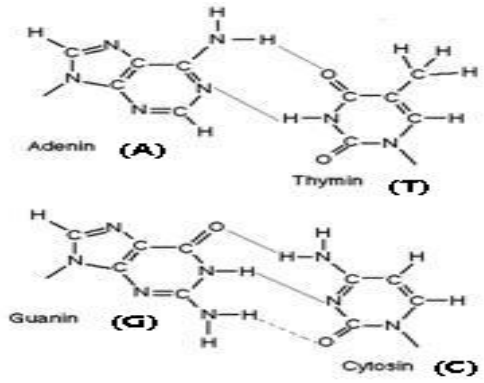


Figure 2. DNA nitrogenous base pairs

```
CCTTGCTCTA CCGGCAGCAC CAATTTCCAA

AGACTTTGGC ACAACCTTTT TATTGTAATG

CTTTTTTAAC TTTTTCAAGA GCTCGCAATC

GATTCATCGA CCATCAAATT ATTGATAATC
```

Figure 3. A DNA sequence presented as an array of letters which are mapping the nucleotides in DNA (consisted of one of four types of nitrogenous bases A/G/C/T, a five carbon sugar, and molecule of phosphoric acid).

Observing through the DMS model described above, genomic processing moves automatically to the DMA layer. Obtaining the available genomic sequences (basically DR layer) can be done by Internet for no cost in academic purposes. One can download the genomic sequences (of a human, chimpanzee, mouse, etc) from web pages like: www.ensembl.org, www.ncbi.nlm.nih.gov, genome.ucsc.edu, etc.

For the purpose of the DMA layer, genomic experts use software that is able to find similar patterns (words) within the long genomic segment.

Examples of this software are BLAST [2] (the mostly frequently used software), Smith Waterman [3], FastA [4], and others. These programs find similarity between a query sequence and the sequences within the database.

In the example shown at Figure 4 and Figure 5, one can see a fraction of the results obtained from a BLAST comparison of protein SLC7A7 (human) against a SwissProt (http://www.isb-sib.ch) database of proteins. Two illustrative examples that show from a perfect (word) mach to a similar mach are presented.

```
>gi|12643348|sp|Q9UHI5|LAT2_HUMAN
<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=Protein&list_uids=12643348&dopt=GenPept> Gene info
< http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=search&term=12643348%5BPUID%5D> Large neutral amino
acids transporter small subunit 2 (L-type  amino acid transporter 2) (hLAT2)
Length=535
Score =  665 bits (1717),  Expect = 0.0, Method: Composition-based stats.
Identities = 332/332 (100%), Positives = 332/332 (100%), Gaps = 0/332 (0%)

Query 1     MGIVQICKGEYFWLEPKNAFENFQEPDIGLVALAFLQGSFAYGGWNFLNYVTEELVDPYK  60
MGIVQICKGEYFWLEPKNAFENFQEPDIGLVALAFLQGSFAYGGWNFLNYVTEELVDPYK
Sbjct  204   MGIVQICKGEYFWLEPKNAFENFQEPDIGLVALAFLQGSFAYGGWNFLNYVTEELVDPYK  263

Query 61    NLPRAIFISIPLVTFVYVFANVAYVTAMSPQELLASNAVAVTFGEKLLGVMAWIMPISVA  120
NLPRAIFISIPLVTFVYVFANVAYVTAMSPQELLASNAVAVTFGEKLLGVMAWIMPISVA
Sbjct  264   NLPRAIFISIPLVTFVYVFANVAYVTAMSPQELLASNAVAVTFGEKLLGVMAWIMPISVA  323

Query 121  LSTFGGVNGSLFTSSRLFFAGAREGHLPSVLAMIHVKRCTPIPALLFTCISTLLMLVTSD  180
LSTFGGVNGSLFTSSRLFFAGAREGHLPSVLAMIHVKRCTPIPALLFTCISTLLMLVTSD
Sbjct  324   LSTFGGVNGSLFTSSRLFFAGAREGHLPSVLAMIHVKRCTPIPALLFTCISTLLMLVTSD  383

Query 181  MYTLINYVGFINYLFYGVTVAGQIVLRWKKPDIPRPIKINLLFPIIYLLFWAFLLVFSLW  240
MYTLINYVGFINYLFYGVTVAGQIVLRWKKPDIPRPIKINLLFPIIYLLFWAFLLVFSLW
Sbjct  384   MYTLINYVGFINYLFYGVTVAGQIVLRWKKPDIPRPIKINLLFPIIYLLFWAFLLVFSLW  443

Query 241  SEPVVCGIGLAIMLTGVPVYFLGVYWQHKPKCFSDFIELLTLVSQKMCVVVYPEVERGSG  300
SEPVVCGIGLAIMLTGVPVYFLGVYWQHKPKCFSDFIELLTLVSQKMCVVVYPEVERGSG
Sbjct  444   SEPVVCGIGLAIMLTGVPVYFLGVYWQHKPKCFSDFIELLTLVSQKMCVVVYPEVERGSG  503

Query 301  TEEANEDMEEQQQPMYQPTPTKDKDVAGQPQP  332
TEEANEDMEEQQQPMYQPTPTKDKDVAGQPQP
Sbjct  504   TEEANEDMEEQQQPMYQPTPTKDKDVAGQPQP  535
```

Figure 4. BLAST Sample session, perfect match. Comparison of protein SLC7A7 (human) against the same protein.

```
>gi|12643378|sp|Q9UM01|YLA1_HUMAN
   <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=Protein&list_uids=12643378&dopt=GenPept> Gene info
   <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=search&term=12643378%5BPUID%5D> Y+L amino acid
    transporter 1 (y(+)L-type amino acid transporter
    1) (y+LAT-1) (Y+LAT1) (Monocyte amino acid permease 2) (MOP-2)
    Length=511
   Score =  257 bits (656),  Expect = 4e-68, Method: Composition-based stats.
   Identities = 138/315 (43%), Positives = 203/315 (64%), Gaps = 10/315 (3%)

    Query  2      GIVQICKGEYFWLEPKNAFENFQEPDIGLVALAFLQGSFAYGGWNFLNYVTEELVDPYKN  61
                  GIV++ +G    E  N+FE      +G +ALA     F+Y GW+ LNYVTEE+ +P +N
   Sbjct  202    GIVRLGQGASTHFE--NSFEG-SSFAVGDIALALYSALFSYSGWDTLNYVTEEIKNPERN  258

    Query  62     LPRAIFISIPLVTFVYVFANVAYVTAMSPQELLASNAVAVTFGEKLLGVMAWIMPISVAL  121
                  LP +I IS+P+VT +Y+  NVAY T +  +++LAS+AVAVTF +++ G+  WI+P+SVAL
   Sbjct  259    LPLSIGISMPIVTIIYILTNVAYYTVLDMRDILASDAVAVTFADQIFGIFNWIIPLSVAL  318

    Query  122   STFGGVNGSLFTSSRLFFAGAREGHLPSVLAMIHVKRCTPIPALLFTCISTLLMLVTSDM  181
                  S FGG+N S+  +SRLFF G+REGHLP  + MIHV+R TP+P+LLF  I  L+ L   D+
   Sbjct  319    SCFGGLNASIVAASRLFFVGSREGHLPDAICMIHVERFTPVPSLLFNGIMALIYLCVEDI  378

    Query  182   YTLINYVGFINYLFYGVTVAGQIVLRWKKPDIPRPIKINLLFPIIYLLFWAFLLVFSLWS  241
                  + LINY  F  + F G+++ GQ+ LRWK+PD PRP+K+++ FPI++ L   FL+   L+S
   Sbjct  379    FQLINYYSFSYWFFVGLSIVGQLYLRWKEPDRPRPLKLSVFFPIVFCLCTIFLVAVPLYS  438

    Query  242   EPVVCGIGLAIMLTGVPVYFL--GVYWQHKPKCFSDFIELLTLVSQKMCVVVYPEVERGS  299
                    + +  IG+AI L+G+P YFL   V    +P     +  T   Q +C+ V  E++
   Sbjct  439    DTINSLIGIAIALSGLPFYFLIIRVPEHKRPLYLRRIVGSATRYLQVLCMSVAAEMDLED  498

   Query  300   GTEEANEDMEEQQQP  314
                 G E    M +Q+ P
   Sbjct  499   GGE-----MPKQRDP  508
```

Figure 5. BLAST Sample session, similar match. Comparison of protein SLC7A7 (human) against a SwissProt database of proteins.

As it is shown in the above two figures, BLAST expresses the level of similarity between query sequence and database sequence in terms of: score, expectations, method, identities, positives, and gaps. Here is where proposed DMA layer is finishing, and from this point inferring needs to be done by genomic experts on the bases of software (ex. BLAST) output, and knowledge gathered elsewhere (brains, book, computers, etc).

The complete philosophy of sequence comparison in biological context relies on the fact that similar sequences have similar functions. Therefore, comparative analysis of sequences help researchers infers possible functions, which guides them for further molecular analysis.

The possible interpretations of this type of comparative analysis are very wide and depend very much on initial question. Following authors initial scheme (Figure 1), one could see this step as the core of building the CP layer. This topic is further discussed in Section 3.

Also, a forthcoming challenge in the field of comparative genomic analysis is to compare large amounts of genomic data (letters). Current databases are already reaching size limits that

make simple comparisons not possible. These limitations are probably due to lack of memory that could be eventually solved at hardware level or by modifying the structure of data to make them more efficient for processing. For example, if one wants to compare one mammalian genomic sequence against all existing mammalian sequences, one would need a database with memory storage of 60 GB. Every day, researchers are producing more and more genomic sequences. Scientific community expects a large amount of genomic data coming from meta-genomic (Environmental genomic) projects like Sargasso Sea Project [5]. In general, if one wants to have complete and accurate results in the domain of knowledge extraction, big database is an advantage. That is the problem that can be solved with TM, which is explained more in Section 4

3. Concept Processing

State of the art Knowledge-retrieval systems are based on Semantic retrieval, but Knowledge-retrieval systems will become much more efficient once they start using CP. If one says “I am married” and “I have a husband” – these two statements are semantically different, but they both refer to the same concept. If the retrieval is based on semantics, only a subset of knowledge will be retrieved from the database or the Internet. If the retrieval is based on the concepts, all the relevant knowledge that points to the same concept will be retrieved.

3.1. Internal structure - Core Concept and Onion-layers

One important research problem is how to represent concepts. A trivial solution (which does not make sense for large sets of data) is to come up with a huge Case Statement that will include all semantic structures that lead to the same concept. This is a brute force approach, and can be applied only to limited vocabulary problems, like the processing of patent data [6], or similar. The only realistic solution is to employ a CP software architecture, based on Concept Networks, expanded into Concept Web, using a modular Onion-layered type structure, as indicated in Figure 6.

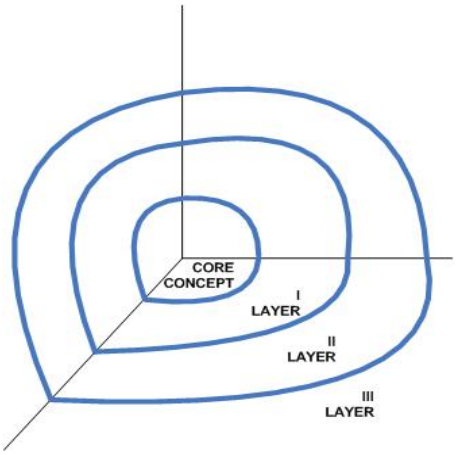


Figure 6. Concept internal organization - Onion-layered structure. The main idea is to have core of the concept in which minimum data is stored for basic concept understanding. Here, in the core, should be stored the essence of a concept, so that when a person not familiar with the topic processes it, it is able to understand it. Like, when learning completely new things, first one has to be aware of the essence in order to understand anything related to the topic/concept involved. Next layer (I LAYER) gathers a set of concepts related to the observed CORE CONCEPT. So, if one is not able to understand the core itself, then one moves to LAYER I, which contains other concepts related to the observed concept, so one have more knowledge and therefore more ability to understand it. If the amount of knowledge stored in the LAYER I is still not enough for understanding the CORE CONCEPT then one move to LAYER II having on disposal even more concepts related to the observed CORE CONCEPT. And so on for LAYER III…

Onion-layered type structure is similar to the process of learning. When one is not able to learn from the essence of the matter presented, one searches for more data (like the examples or relations to the other topics) in order to understand it. CORE CONCEPT represents the essence of the matter presented and examples and relatedness to other topics is what LAYER I / II / III / etc are presenting.

Basically, if one observes CORE CONCEPT as a sphere, I LAYER is a sphere with a bigger radius containing CORE CONCEPT; LAYER II is a sphere containing both LAYER I and CORE CONCEPT; LAYER III is a sphere containing LAYER II, LAYER I, and CORE CONCEPT, and so on….

The CP software architecture, with an indication that its efficient execution implies the existence of the underlying computer architecture that represents a perfect concept match for the processing needs, is presented in Figure 7. The CORE CONCEPT includes the essence of the concept definition, and outer layers represent concept refinements.

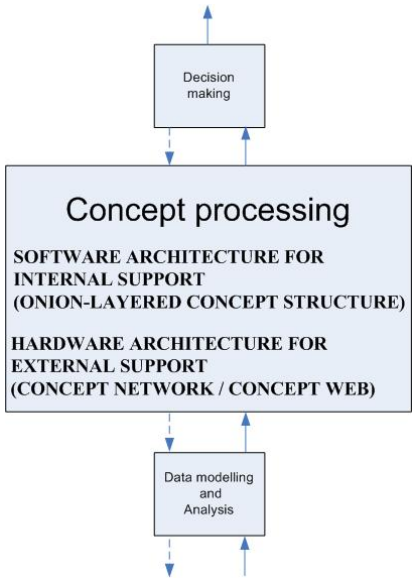


Figure 7. Software and Hardware demands for CP layer. Software architecture in CP layer should support Onion-layered concept structure, and hardware architecture should support Concept Network/Concept Web structure.

3.2. External structure – Concept Network and Concept web

Now, the question arises how to organize CORE CONCEPTS (or just short CONCEPTS) among each other. Two solutions are proposed: CONCEPT NETWORK where concepts are related with one directional arc that has verb attached to it (original idea taken from RDF ontology language - www.w3.org/**RDF**/), and CONCEPT WEB, which is the extension of CONCEPT NETWORK, where relations between CONCEPTS are also CONCEPTS.

Figure 8 and Figure 9 show examples of CONCEPT NETWORK and the related CONCEPT WEB in it's the simplest form. The former includes the nodes that refer to subject and object matter, while the predicate matter is referred to as arcs. The later is derived from the former by promoting the predicate arcs into nodes, using a generalization approach (verbs of the arcs are converted into the nouns of the nodes). Subjects and objects are observed as core concepts, and verbs are observed as relations/core concepts respectively.

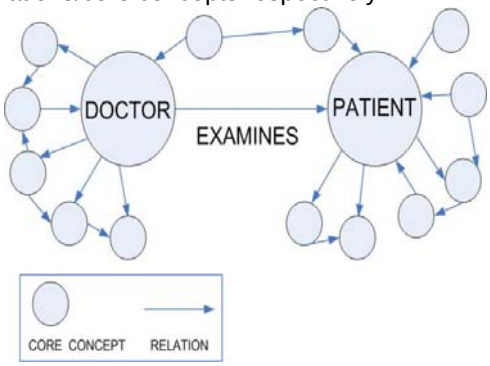


Figure 9. Concept network. The core concepts DOCTOR and PATIENT contain subject matter related to those two concepts.

For example: DOCTOR = a person trained in the healing arts and licensed to practice. PATIENT = one who receives medical attention, care, or treatment.

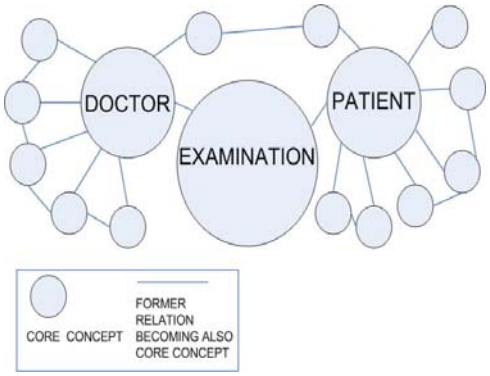


Figure 10. Concept web. The relations are also core concepts. So, besides core concepts DOCTOR and PATIEN, core concept EXAMINATION is added = a medical inquiry into a patient's state of health.

Programs already do exist (e.g., OntoLearn [6]) that build concept network (sometimes called semantic nets) using the lexicosemantic relations like: Hyponym (a word or phrase whose semantic range is included within that of another word), Hypernom (opposite of a hyponym), Gloss (concept appears in the definition of another concept), Topic (concept often co-occurs with another concept), etc.

CP consisted of concepts and their relations which are built on semantic bases is avoided in proposed DMS, because that brings CP which is language and grammar independent. Idea is to have processing of the conceptual level, with the words that have unique meaning (for that purpose genomic processing is ideal because letters A, C, G, T in genomic do have unique meaning).

The move from semantics to concepts implies the activities that (in their simplest form) convert the above mentioned semantic web into a related concept web.

4. Computer Architecture Considerations

Knowledge is often ambiguous and therefore not scalable and not suitable for further processing. In most of the cases, user is provided with a huge amount of data, but without any possibility for automatic logical reasoning on the top of those data. Concept processing based on transactional memory (extended into DSM and expanded with SMT) as a possible solution to overcome this problem. Actually, TM [7] is the solution for a wide variety scientific of supercomputing problems not discussed in this paper.

All problems discussed so far use concepts as atoms of knowledge (which fits into the atomic transaction structure of TM). Concept definitions in their atomic form directly map on the TM constructs, and concept organization presented in this paper goes in two directions: one related

to application issues, and the other related to technology issues.

4.1. Application Issues

The relationship between application demands and constructs offered by the underlying architecture is essential for the required “perfect match” of the application and the architecture. No matter if one talk about CP for Internet oriented DM in business (to detect potential profit strategies), or about genomic processing in distributed database (to detect potential genetic developed diseases), application requirements can be described as follow:

```
start atomic transaction
  access a data structure
  perform the related processing
  detect potential hazards in the
  system
  commit or rollback
end atomic transaction
```

This very same computational structure is built into a typical TMS, and can be directly translated into the code for a TMS, as indicated in the following example:

This computational structure is built into a typical TM like system and can be directly transformed into the proper TM form. This transformation will be presented using AMMP application [9]. AMMP is a modern full-featured molecular mechanics, dynamics and modelling program. It can manipulate both small molecules and macromolecules including proteins, nucleic acids and other polymers. This application is also part of Spec OMP 2001 benchmark [9] for testing platforms for execution of the parallel application. The idea is to make TM completely transparent from the researcher/programmer because in that way all difficulties of writing parallel applications are hidden and researcher productivity is much better. Idea is that “useful” code should be just surrounded into atomic block and everything else should be as it was before. After that, specialized compiler should transform the code. Original code and the transformed code are presented in Figure 11 and Figure 12 respectively.

```
atomic {
  ux = (a2->dx -a1->dx)*lambda
        +(a2->x -a1->x);
  uy = (a2->dy -a1->dy)*lambda
        +(a2->y -a1->y);
  uz = (a2->dz -a1->dz)*lambda
        +(a2->z -a1->z);
  r = one/( ux*ux + uy*uy +
      uz*uz);
  r0 = sqrt(r);
  ux = ux*r0;
  uy = uy*r0;
  uz = uz*r0;
  k = -dielectric*a1->q*a2->q*r;
  r = r*r*r;
  k = k + a1->a*a2->a*r*r0*six;
  k = k - a1->b*a2->b*
      r*r*r0*twelve;
  a1fx = a1fx + ux*k;
  a1fy = a1fy + uy*k;
  a1fz = a1fz + uz*k;
  a2->fx = a2->fx - ux*k;
  a2->fy = a2->fy - uy*k;
  a2->fz = a2->fz - uz*k;
}
```

Figure 11. Critical part of the AMMP application which should be executed atomically. This code presents standard C/C++ code and with the simple surrounding that code into atomic block one have the TM application. This way TM mechanism is completely transparent for the researcher, it hides the difficulties of writing the parallel applications and extends the researchers productivity.

```
{ startTransaction(); {
   write(t, &ux, (*read(t,&((
    *read(t, &a2))->dx)) -
    *read(t, &((*read(t, &a1))->dx))
)* *read(t, &lambda) +
    (*read(t, &((*read(t, &a2))-
>x))-
    *read(t, &((*read(t, &a1))-
>x))));
   write(t, &uy, ( *read(t, &( (
*read(t, &a2)  ) ->dy) ) -
    *read(t, &( ( *read(t, &a1)  ) -
>dy) ) )* *read(t, &lambda) +
    (*read(t, &( ( *read(t, &a2)  )
->y) ) -
    *read(t, &( ( *read(t, &a1)  ) -
>y) ) ));
   write(t, &uz, ( *read(t, &( (
*read(t, &a2)  ) ->dz) ) -
    *read(t, &( ( *read(t, &a1)  ) -
>dz) ) )* *read(t, &lambda) +
    ( *read(t, &( ( *read(t, &a2)  )
->z) ) -
    *read(t, &( ( *read(t, &a1)  ) -
>z) ) ));
   write(t, &r,  *read(t, &one) /(
*read(t, &ux) * *read(t, &ux) +
    *read(t, &uy) * *read(t, &uy) +
*read(t, &uz) * *read(t, &uz) ));
   write(t, &r0, sqrt( *read(t, &r)
));
   write(t, &ux,  *read(t, &ux) *
*read(t, &r0) );
   write(t, &uy,  *read(t, &uy) *
*read(t, &r0) );
   write(t, &uz,  *read(t, &uz) *
*read(t, &r0) );
   write(t, &k, - *read(t,
&dielectric) *
    *read(t, &( ( *read(t, &a1)  ) -
>q) ) *
    *read(t, &( ( *read(t, &a2)  ) -
>q) ) * *read(t, &r) );
   write(t, &r,  *read(t, &r) *
*read(t, &r) * *read(t, &r) );
   write(t, &k,  *read(t, &k) +
*read(t, &( ( *read(t, &a1)  ) ->a))
*
    *read(t, &( ( *read(t, &a2)  ) -
>a) ) *
    *read(t, &r) * *read(t, &r0) *
*read(t, &six) );
   write(t, &k,  *read(t, &k) -
*read(t, &( ( *read(t, &a1)) ->b) )
```

```
*
    *read(t, &( ( *read(t, &a2)  ) -
>b) ) *
    *read(t, &r) * *read(t, &r) *
*read(t, &r0) * *read(t, &twelve) );
   write(t, &a1fx,  *read(t, &a1fx)
+ *read(t, &ux) * *read(t, &k) );
   write(t, &a1fy,  *read(t, &a1fy)
+ *read(t, &uy) * *read(t, &k) );
   write(t, &a1fz,  *read(t, &a1fz)
+ *read(t, &uz) * *read(t, &k) );
   write(t, &((*read(t, &a2)) ->fx),
*read(t,&((*read(t, &a2)) ->fx))-
    *read(t, &ux) * *read(t, &k) );
   write(t, &( ( *read(t, &a2)  ) -
>fy ) ,
    *read(t, &((*read(t, &a2))->fy))
- *read(t, &uy) * *read(t, &k) );
   write(t, &( ( *read(t, &a2)  ) -
>fz ) ,
    *read(t, &( ( *read(t, &a2)  ) -
>fz) ) - *read(t, &uz) *
    *read(t, &k) );
  } endTransaction();
}
```

Figure 12. Generated code for the part of the AMMP application presented in Figure 11). This is generated code suitable for STM (Software transactional memory). It presents how difficult it would be to write program without support of the external tools. The idea is to make TM completely transparent for the researcher and to create external tools which will provide support for the transactional memory. Basic support for the STM as a Proof of Concept is the next step together with research in Hardware TM and Hybrid TM in order to produce high-performance applications.

In this way one can have a very user friendly system, because programmer should just surround critical code into atomic locks. Also, system will have high performances because it can be ported for HTM (Hardware Transactional memory) of HyTM (Hybrid Transactional Memory), which will speedup the execution of the applications. Next step is to research and develop specialized hardware that will accelerate our TM applications. In that way one can have a high-performance system for extraction of knowledge.

However, since TM is defined in the SMP environment, and researchers in Genomic processing report the need for huge and distributed memory and databases, it has to be ported to DSM and expanded with SMT, before it can be used in applications of interest for this work.

A good source of information about DSM is the survey paper [11] For the DSM and SMT extensions to be applicable both to the second and the third DM layers presented here, the development has to take into consideration the needs of both, applications like BLAST and algorithms like those suggested in CP layer.

4.2. Technology Issues

The TMS architecture has been designed for SMP paradigm, and its potentials for parallel processing are limited. Typical number of nodes on a SMP bus is equal to 16, 32, or 64, and further expansions are not possible, due to current technology limitations (current bus speed is not large enough to enable a larger number of processors). On the other hand, amount of parallelism involved in the above discussed applications is enormous, and may be of the order of 1M, 2M, or even 4M.

Fortunately, with introduction of optics into the domain of system communications, in the future one can expect that the size of SMP systems can grow beyond 64, 128, or 256. Also, if the TMS concept is ported from SMP to DSM (where the current levels of parallelism are 1K, 2K, or 4K), then parallelisms above 1M become readily available, and the transactional memory paradigm can find its usage in complex systems oriented to concept modeling for bio-informatics.

5. Generalized Structure of a System for Concept Processing

With all above in mind, proposed structure of a future Knowledge Processing System is as indicated in Figure 13. The bottom, Computer Architecture, layer is consisted of two sub layers: TMS/DMS environment layer and appropriate Operating System (OS) and Optimizing Compilers layer. On the top of the bottom layer (including both sub layers) works Knowledge Understanding layer that helps build the semantics, and even more importantly, the concept models, used to extract knowledge for the applications of interest. Finally, the top layer, called Application layer, is related to applications like Internet search, genomic processing, etc (all based on text as the input data).

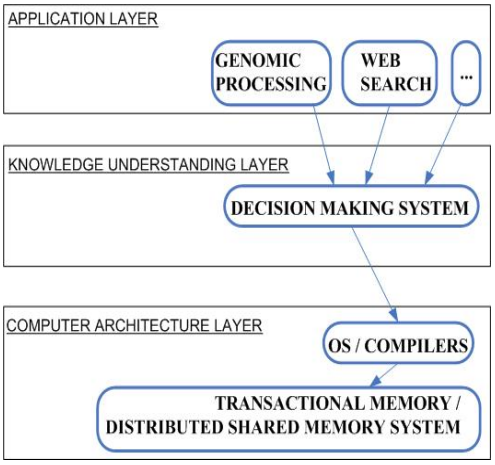


Figure 13. Knowledge Processing System: Computer Architecture layer, DMS layer, Application layer

A possible strategy leading to design and implementation of a system from Figure 10 implies the following steps:

- Prior to application, experts in fields like Knowledge processing (in general) or Genomic processing (specific), are consulted to see what are their typical problems and what are the typical computational patterns involved.
- Appropriate algorithms are developed, which are oriented to fast execution on TMS architectures. If needed, the architecture of the underlying machine can be modified.
- Software is developed that converts the existing tools into forms that can run efficiently on the TMS architecture.
- Performance is measured, and possibly some of he software constructs are ported into hardware.

Of course, once the system is designed, and lessons are learned from the deployment of the system, after the incubation period is lived through, ideas will be generated on how one can further improve the speed and other important aspects of the processing system involved.

6. CONCLUSION

In this paper, we have presented an architectural and algorithmic support for a Knowledge Processing System in selected high-demand applications. One possible scenario implies a three layer system:

- Top: Application (like Genomic Processing).
- Medium: Knowledge Understanding, based on Data Mining, Semantic Web, and CP (as the most sophisticated approach).
- Bottom: Computer Architecture, along the concepts of SMP and DSM, with a special emphasis on TM), which uses a computational paradigm compatible with the needs of CP.

ACKNOWLEDGEMENT

The authors would like to thank Prof. Pavle Andjus, Institute for Physiology and Biochemistry School of Biology; University of Belgrade; Serbia, for his useful comments.

REFERENCES

[1] Salton, G., Wong, A., "A Vector Space Model for Automatic Indexing," Communications of the ACM, 1975, pp. 613 - 620 Vol. 18, Issue 11.
[2] BLAST, December 2006. Available online: www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=9606
[3] "Identification of Common Molecular Subsequences," Temple F. Smith and Michael S. Waterman, Journal of Molecular Biology, 1981, pp. 195-197.
[4] "Rapid and sensitive protein similarity searches," D. J. Lipman, W. R. Pearson, Science 22 March 1985, Vol. 227. no. 4693, pp. 1435 – 1441.
[5] Sargasso Sea project, December 2006. Available online: http://www.genomenewsnetwork.org/articles/2004/03/04/sargasso.php.
[6] "A Proposed Hybrid Approach for Patent Modeling," Ognjen Scekic, Djordje Popovic, Veljko Milutinovic, Transactions on Internet Research, July 2006, Vol. 2, Number 2.
[7] "Ontology Learning and Its Application to Automated Terminology Translation," Roberto Navigli, Paola Velardi, Aldo Gangemi, IEEE, Intelligent Systems, 2003, pp. 22-31
[8] "Transactional Memory: Architectural Support for Lock-free Data Structures," M. Herlihy, J. Eliot, B. Moss, Proceedings of the 20th Annual International Symposium on Computer Architecture, 16-19 May 1993, pp. 289-300.
[9] AMMP Home Page, December 2006. Available: http://www.cs.gsu.edu/~cscrwh/ammp/ammp.html
[10] SPEC OMP 2001 Benchmark, December 2006. Available: http://www.spec.org/omp/
[11] "A Survey of Distributed Shared Memory," Jelica Protic, Milo Tomasevic, Veljko Milutinovic, Proceedings of the 28th Annual Hawaii International Conference on System Sciences, 1995.

Development of a Biomechanical Knowledge System to Identify Brain Injuries in Emergency Department

Kou, Zhifeng and Ziejewski, Mariusz

Abstract—*Traumatic brain injury (TBI) has an annual incidence rate of over one million emergency department (ED) visits in the United States (U.S.). A patient subjected to trauma-induced alternation of mental status could have an TBI that may, or may not, involve any loss of consciousness. In clinical practice, diagnosis of TBI is very difficult because the presence of a head injury may be masked by a serious injury to another body part, subtle and changeable symptoms, or the delayed onset of symptoms. Many people with TBI do not receive medical care at the time of the injury and may complain to their physicians of their persistent symptoms for days, weeks, months, or even years after the injury. Currently, there is no reliable diagnostic tool to assist the ED physician when he, or she, sees a patient with TBI, especially mild traumatic brain injury (MTBI). Meanwhile, over forty years of injury mechanism study in the area of impact biomechanics proved to be effective in predicting brain injuries. To date, there is no diagnostic tool using impact biomechanics to quantify the risk factors of motor vehicle crash (MVC) occupants for MTBI in EDs. To the best of our knowledge, no one has explored how to prepare and model the knowledge of impact biomechanics into an information system for EDs. Our overall hypothesis was that an MVC scenario in association with the injury mechanism are important risk factors for MTBI. As part of our study series, this paper reports the development of a Web-based application system using the knowledge of impact biomechanics, based on MVC scenarios, in order to identify the patients, in EDs, at risk for MTBI and to stratify their risk levels. The system has been able to capture 94% of hypothetical MTBI patients at risk. The system could potentially assist the ED physicians in decision making for a proper referral pattern and clinical diagnosis of MTBI. The study also provides a novel approach to modeling the knowledge in impact biomechanics into a database, a shell for managing the knowledge rules, and a generic interface for editing the rules. The system shell could be easily adapted to other knowledge based systems to provide domain expertise from other fields for biomedical applications.*

Index Terms—*Traumatic Brain Injury, Mild Traumatic Brain Injury, Emergency Medicine, Knowledge System, Biomedical Information System, Expert System, Telemedicine*

1. INTRODUCTION

Over the past few years, a large number of clinically useful tools and reference resources for ED physicians have become available, either as stand-alone software, or as resources available through the Internet. To the best of our knowledge, however, there is no such system providing expertise in impact biomechanics in EDs to help identify brain injury, which is a very prevalent disease in the U.S. Furthermore, how to model the knowledge in impact biomechanics has not yet been explored.

There are approximately 1 million ED visits annually for TBI in the U.S. [1]. The majority of them are MTBIs primarily resulting from MVCs and falls [1]. However, the consequences of MTBI are often not mild [2]. According to the definition of MTBI by the American Congress of Rehabilitation Medicine (ACRM) [3], "A patient with MTBI is a person who has had a traumatically induced physiological disruption of brain function, as manifested by at least one the following: 1) any period of loss of consciousness, 2) any loss of memory for events immediately before, or after, the accident, 3) any alternation in mental state at the time of the accident, and 4) focal neurological deficit(s).

Clinically, however, there is no reliable diagnostic tool to assist an ED physician when he, or she, sees a patient with TBI, especially MTBI. Reliance on patients to report risk factors can be highly unreliable due to the lack of the patient's knowledge about the importance of certain risk factors, or the patient may have amnesia from the event. It is very hard for healthcare providers to record, or for a patient to report, a brief loss of

Manuscript received Feb 24, 2007. This work was supported in part by the MeritCare Medical Foundation at Fargo, North Dakota.

Z. K., Ph.D., is with the Department of Radiology, Wayne State University School of Medicine, Detroit, Michigan 48201, USA. (e-mail: zhifeng_kou@yahoo.com). The author was affiliated with the North Dakota State University by the time performing this research.

M. Z. (corresponding author), Ph.D., is with the Mechanical Engineering Department, North Dakota State University, Fargo, North Dakota 58105, USA. (email: mariusz.ziejewski@ndsu.edu).

consciousness, or memory loss, caused by a blow to the head. These patients are often given a nebulous diagnosis of “brain concussion” and discharged home without any concrete follow-up plan or neuropsychological assessment. The mis-diagnosed and un-diagnosed rates range between 20%-50%.

Recently, in emergency care, the consideration of an injury scenario began to be appreciated in association with the risk factors for TBI. The EFNS (European Federation of Neurological Societies) Guideline on MTBI considers a high-energy MVC as an important risk factor associated with an increased risk of intracranial injuries [4]. Current advanced trauma life support system suggests the consideration of the injury mechanisms and vehicle occupant kinematics for identification of risk factors [5-7]. However, the identification and quantification of these risk factors is a unique piece of science called impact biomechanics which is the “basic science of injury causation” [8]. Without the analysis of an impact scenario and the mechanical load applied on a patient’s head/brain, the fast and accurate diagnosis of MTBI is a great challenge to the currently available clinical diagnosis and detection techniques.

Meanwhile, the lack of 24/7 available expertise in impact biomechanics is a considerable barrier for the biomechanical identification of brain injury in an ED. To the best of our knowledge, there is no diagnostic tool using impact biomechanics to quantify risk factors of MVC in ED settings for MTBI, nor has anyone explored how to prepare and model the knowledge of impact biomechanics into an information system for emergency medicine. Our overall hypothesis was that MVC scenarios in association with the injury mechanism are important risk factors for MTBI. This is also well supported by the current practice of EFNS guidelines on MTBI and ACSTC field triage criteria of trauma patients. As part of our systemic effort, our specific aims in this study were 1) to develop a Web-based system using the knowledge of impact biomechanics to identify the patients at risk for MTBI and to stratify their risk levels in EDs and 2) to investigate an approach to modeling the knowledge in impact biomechanics into an information system.

2. MATERIALS AND METHODS

2.1 Development of Data Collection Instrument

To collect necessary MVC parameters, on scene, that contribute to the risk for MTBI, a data collection instrument (Figure 1) was developed, in a separate study, for EMTs to use. From Jan 2002 to May 2003, we collected 317 crash cases with an overall completion rate of 82%. Our results demonstrated that EMTs could collect the MVC data, on scene, without affecting their primary duties. The description of the instrument and justification of MVC parameters were narrated in the separate study. We include a brief summary here, however, for completeness.

The instrument included impact data forms and a digital camera (Figure 1). There were four sections to complete on the impact data form. The first section was general information. This was where information, such as the date of the accident, the type of information source, and patient’s information, such as gender, height, and weight was recorded. The second part dealt with the information about the vehicle involved in the accident, such as the year, make, model, seatbelt use, and airbag deployment. Next was the information involving the seat location and body position of the patient in the vehicle prior to impact. The last of the overall four sections was three photo entries that were taken by the data collector using the digital camera to identify impact depth and impact direction and location. When the impact data form was completed, it was placed into an envelope along with the memory card from the camera and given to the emergency medical technician (EMT) dispatcher.

There was no patient identification information, such as patient’s name, social security number, telephone number, home address, etc., on the data form. The only data collected were injury mechanism-related MVC information.

2.2 Justification of Crash Parameters

Among many MVC factors, the seatbelt use, airbag deployment, and vehicle make, year, and model are crucial factors to record in our data collection kit. It has been well recognized by the public and documented by the U.S. National Highway Traffic Safety Administration (NHTSA) that the occupant’s seatbelt use and the availability of an airbag can significantly reduce the death and disability during vehicle crash accidents [9]. Furthermore, considering the size, body weight, and design of vehicles, different vehicles also have different safety parameters (http://www.nhtsa.gov).

In addition, the vehicle occupant’s body height and weight signifies the possible interaction of the occupant’s head with vehicle interiors [10]. Furthermore, the occupant’s gender and age are also reported to be factors contributing to his or her vulnerability [11].

Another critical vehicle crash factor is the patient’s body posture for being out-of-position (OOP). OOP has been widely recognized by the

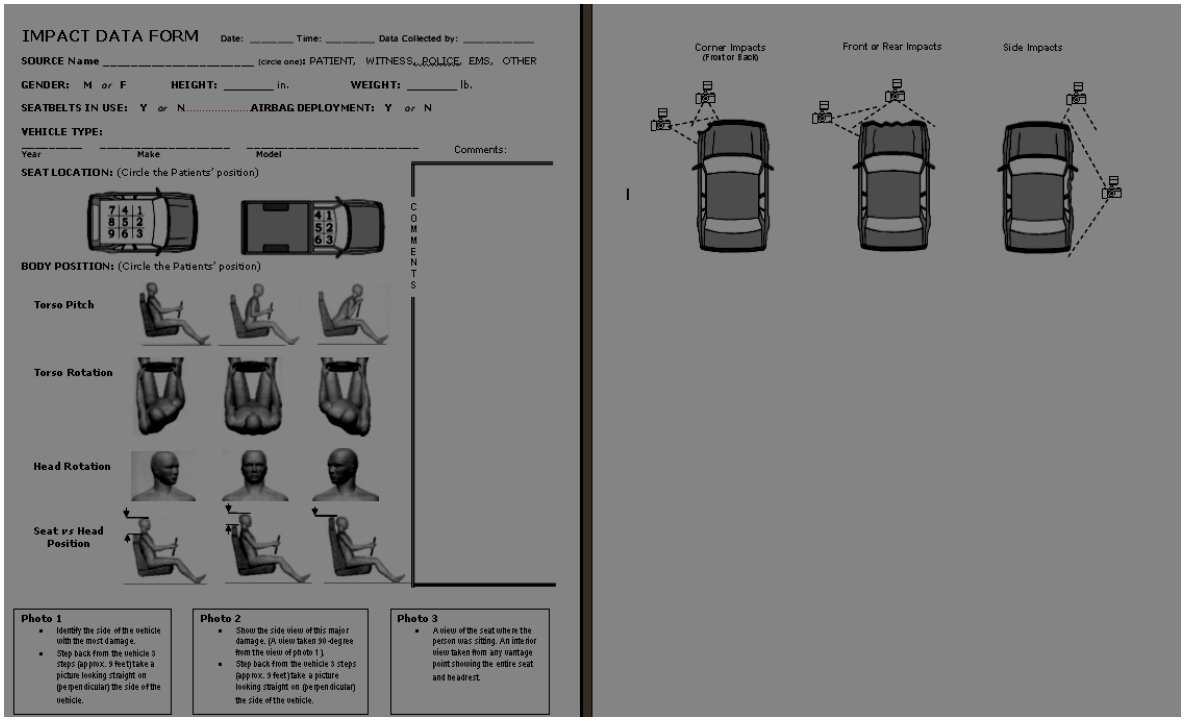


Figure 1. Data Collection Instrument.

U.S. NHTSA, the automotive industry, and the aerospace industry in injury causation analysis [12-16]. The term OOP refers to any patient who is not in the "normal seated position" (NSP) prior to impact. The NSP is defined as follows: the occupant's shoulder blades are pressed firmly into the seat back, and their head is close to the head restraint [17]. Full scale experimental tests on crash dummies and mathematical models of rear-end collisions have shown that only a small variation in occupant position can result in a large increase of impact forces [18-22]. In the 1990s, research on airbags recognized the need to address the problem of occupants being OOP with regard to airbag deployment in frontal collisions [12-16].

In summary, depending on the collision direction (frontal, rear, or side), the use of the seatbelt, the airbag deployment, the stiffness of the seat and the height of head restraint, OOP patients could develop risks for the injuries on their head, neck, and other parts of their body. An occupant not wearing a seatbelt could easily be OOP. In a frontal collision, especially with airbag deployment, an OOP occupant would be at high risk for developing head, neck, or chest injuries. In a lateral collision [23], especially an impact on the occupant's side, an OOP occupant will be more vulnerable than a normal seated patient for developing head/neck injuries. This is due to the greater inertial forces, or possible direct impact forces, regardless of the airbag deployment. In rear-end collisions, if the head restraint is lower than an occupant's head or even neck, the occupant would easily develop whiplash, or possible head injury. Moreover, the deeper the intrusion into the occupant's vehicle, the higher likelihood of injury by considering relative direction of impact versus the occupant's seating position.

2.3 Modeling and Stratifying of the Risk Factors

Based on our research experience in forensic investigation, experimental reconstruction and computer simulation of MVCs, and our search of published literature, we developed a knowledge base to quantify the risk factors and at-risk scenarios of MVCs. A risk factor for MTBI could be a significant MVC parameter, e.g. non-use of seatbelt, or a combination of several MVC parameters to form a crash scenario. Each risk factor is empirically assigned a level of risk for MTBI from low, moderate, to high. Three low level risks are equivalent to a moderate level risk; three moderate risks are equivalent to a high level risk; and the overall risk level of a patient is the summation of all risk factors.

By structuralizing the knowledge information and considering the potential interactions of knowledge, we developed a set of rules to represent this knowledge base. For the ease of modeling into a database, we further structuralized this rule set. Each rule has six parts:

1) The rule name.
2) The IF conditions to make the rule fire, which is the logical combinations of crash variables.
3) The weight of each rule in contribution to a patient at risk.
4) The rule status: valid or invalid. (This will be discussed in the Rule Management section.)
5) Rule description, which is to be displayed on the result page as explanatory information about the risk factors of this rule if the rule is fired.

6) Reference information, which is the published biomechanical data in support of this rule definition.

By structuralizing the rules, we could easily model the rules into a database by storing each part of a rule as a data entry. One rule was a data record. In the evaluation of a case, the software queries rule set from the database, evaluated the true, or false of an IF conditions. Once the evaluation was true, this rule's specific weight for a risk factor was counted; and the description regarding its risk factors and the reference information about this rule was also compiled into an evaluation result page to display for end users.

2.4 System Design

We designed the system in a three-tier architecture model using the currently most popular LAMP technology (Figure 2): Linux operating system, Apache Web server software, MySQL relational database, and PHP script language for server programming. All of these technologies were open source software without a license charge.

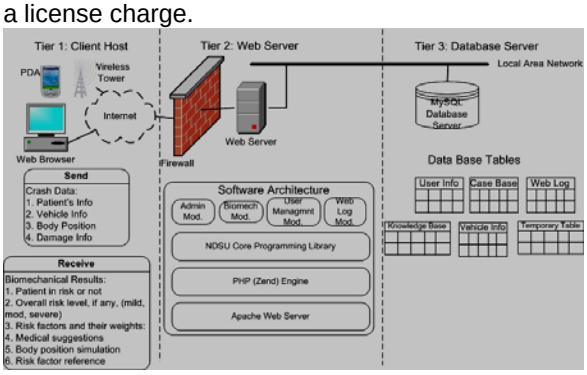


Figure 2. System Architecture.

At tier 1, the ER users could use the Web browser either in desktop/laptop personal computer (PC), or in personal digital assistance (PDA) to send MVC data in an HTTP request to the web server at tier 2. The PHP script program on the Web server processed the hypothetical case information, stored the case data, queried the vehicle information from the database located at tier 3, evaluated the risk factors of the hypothetical patient, and stratified the risk level of that patient. Both the Web server and database server were located within the North Dakota State University (NDSU) campus firewall to protect against hacking.

Two versions of the system were developed: one was for the desktop PC using a regular Web browser and the other was for a PDA. The PDA version could only evaluate the risk factors of the hypothetical patient and retrieve previous cases. There was no administrative function for PDA version.

A typical process of a hypothetical patient case evaluation consisted of three steps:

a.) An emergency staff user logged into their account in our system (URL: http://www.ndsu.edu/biomech).

b.) The user entered the impact MVC data on the Web, which included the general information of the hypothetical patient, vehicle information, patient's body position, and vehicle damage information.

c.) The user waited for 2-5 seconds to allow the system to evaluate the case and return the results. The result page consisted of the following information shown in Figure 3 as a sample page:

1) Patient at risk, or not.
2) Overall risk level, if any.
3) Detailed risk factors, if any.
4) Medical suggestions based on risk level, if patient was at risk.
5) Illustration of patient's body position prior to moment of crash.
6) Published reference data in support of risk factors involved for patient.

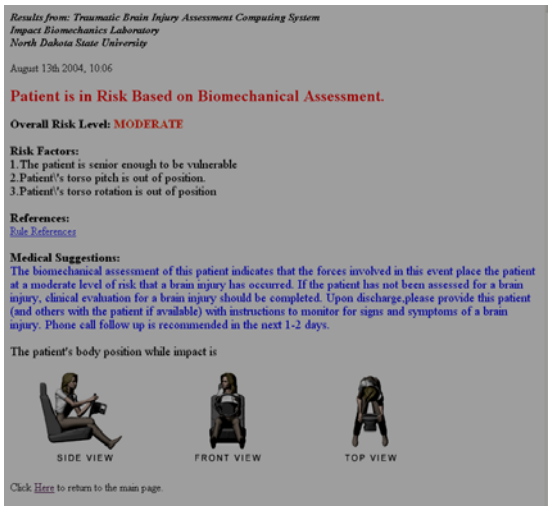


Figure 3. A sample result page.

2.5 System Development

The system development process was performed in a software engineering paradigm using the Object-Oriented approach. The system was designed in a layered architecture in order to guarantee reusability, portability, and easy maintenance. There were three layers: the interface layer, the business layer, and the data access layer.

The interface layer was responsible for the user interaction with the system. The business layer was responsible for the system's logic and was comprised of the objects inherent to the application domain. The data access layer was responsible for accessing the data storage medium. It contained classes that actually implemented the interface for a specific storage medium and database table. We used the NDSU Core Programming Library to implement the data access layer, as shown in Figure 2. The library is a repository of class objects developed, NDSU's Information Technology Services (ITS), in an

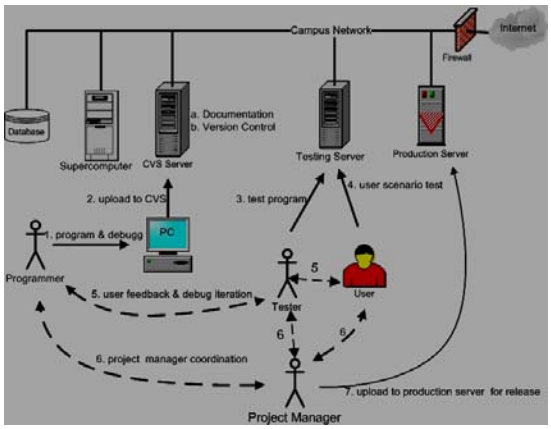
Object Oriented Programming paradigm. It provided commonly used components for the development of web database systems, and was tailored to the specific configuration of the information infrastructure at NDSU.

The implementation process, as shown in figure 4, mainly consisted of the following interweaving steps:

1. Programmers wrote and debugged programs on their local machines.
2. Programmers submitted their code to the Concurrent Versions System (CVS) server, which kept track of the version evolution and documentation; and submitted the inline and implementation documentation when the debugging work was done.
3. The testers copied the code to a testing server for software; and provided feedback to the programmers.
4. The users worked with the system testers to evaluate the user interface, use cases, capacity requirements, etc.
5. Based on the testing results and users' comments, the testers interacted with programmers to provide feedback for debugging iteration. All of the testing reports were documented in the CVS server associated with each version.
6. The project manager coordinated with programmers, testers, and users for the software development iteration during the entire process.
7. Once the software passed the testing, the project manager released the software system by uploading it onto the production server which was NDSU's campus web server for end user field evaluation.

Based on our experience on project development, we set up a NetOffice environment for project management. It was an open source web-based virtual working environment for project management, where the project team members had different access privileges: the manager generated a work task, the testers reported bugs, programmers generated documents, users initiated new requirement and any member could initiate a discussion. All of the project activities were recorded and generated as report files. The project team members received email as a reminder of any activity.

Figure 4. System Development Process.



2.6 A Shell for Knowledge Rule Management

In the design and maintenance of a rule-based knowledge system, one of the difficulties is to manage the knowledge rules. Potential problems such as adding new rules, deleting old rules, testing of the rules and checking conflict of the rules must be addressed [24]. An administrative module was developed to provide an interface for rule management by an administrative user. By querying the rules stored into database tables, the module allowed the administrative user to perform the following functions:

a) Adding a new rule(s): A generic rule editing interface was designed to allow the expert in biomechanics add a new rule even without any knowledge of information systems. The interface will be discussed in the next section.
b) Testing rule(s): The administrative user could select any subset of the rules and input an MVC case to evaluate the functions of this rule set. During the testing process, the system could trace which specific rule had been fired under what condition, which was extremely useful to test the compatibility of the rule set and find potential conflict of rule(s).
c) Deleting rule(s): Any subset of the rules could be deleted by the administrative user.
d) Updating rule(s): In the current rule base, there were two possible status of each rule: Valid or Invalid. A valid rule meant it was active in the production machine and would be fired if the condition for this rule was met, and an invalid rule meant the rule was hibernating and was used only for testing purposes by the system administrator. The rule status could be changed by our administrative user. The reason for this feature was that during the long-time maintenance of the rule set, a domain expert could be unsure of some rule(s) and want to test it further rather than adding it directly into, or deleting it from, the rule base. It was determined that it would be better to let the unsure rule(s) hibernate.

2.7 A Generic Interface for Rule Editing

Compatible with the structuralized rule set, a generic user-friendly interface was developed for rule editing. A domain expert, who is even a novice in information system, could easily edit new rule(s) into the knowledge base. As shown in figure 5, all of the potential MVC variables to be considered in the rule base, e.g. gender, passenger seating position, crash position and depth, were incorporated into pull down menus in the interface. All of the potential relationships of these variables, e.g. =, >=, <=, were also provided in pull down menus as well. The parentheses potentially used in the logic combinations of these variables were also provided. The IF condition for a new rule could be designed by taking any combinations of these variables. For example, if a rule said "if the vehicle occupant is a female no less than 65 years old, and the damage depth is over 6

inches, of the female had direct head impact onto the vehicle interior, she could be at moderate risk," the logic expression for the IF condition would be:

((Patient Age >=65)AND(Patient Gender==F)AND((Impact Depth>=6inches)OR(Head Strike==Y)))

as shown in Figure 5.

The MVC variables were also defined as programming variables in the server script program. To determine if a specific rule's condition was met, the system retrieved this rule from the database and evaluated the true, or false, of the logic expressions stored in the IF condition entry.

The risk factors, rule name, rule status, brief description of the rule, reference source of the rule and a brief digest from the reference regarding this rule were also provided in the interface along with the IF conditions (Figure 5).

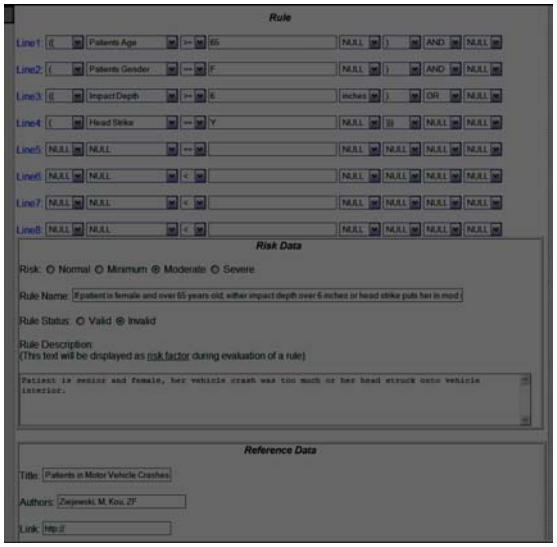


Figure 5. An Interface for Rule Editing.

3. RESULTS

The validation results of the knowledge base were encouraging. We used 29 hypothetical cases (n=29) to validate the system. Based on detailed biomechanical analysis, in conjunction with human brain tolerances, 16 were affirmative MTBI cases (n1=16) and 13 were non-MTBI cases (n2=13). The system identified 22 of the total 29 cases at risk. Among these 22 at-risk hypothetical cases, 15 were affirmative MTBIs, which takes 94% of the 16 affirmative MTBI cases and results in the system sensitivity of 94%. By stratifying the at-risk cases into different risk levels, the results show that the hypothetical patients' risk levels signified their frequency of MTBI. Two out of 6 cases (33%) at low risk level were affirmative MTBI, 7 out of 10 cases (70%) at moderate risk level were affirmative MTBIs and all 16 cases (100%) at high risk level were affirmative MTBIs. A logistic regression analysis also confirmed that patient risk level was significant in predicting their injury probabilities (Wald Chi-Square test value 8.073, p=0.0045).

4. DISCUSSION

By stratifying patients into different levels of risk, different probabilities of brain injury could be determined, and appropriate measures and referral patterns could be taken by emergency physicians to confirm the presence of MTBI and to manage the patients timely. Our system demonstrated the low risk level a 37% probability of brain injury in the hypothetical patients; moderate risk level at a 75% probability; and high risk level at a 94% probability. This system could potentially assist ED physicians in identifying a patient at risk for MTBI in order for a proper referral pattern to occur and for clinical decision making for further neuroclinical examinations. Given the different MTBI probabilities at different risk levels of patients, measures to be considered and referral patterns could be, but no limited to, clinical evaluation for brain injury, education of patients and patients' significant others, neuropsychological consultation, and follow-up phone calls within the next 1-2 days.

One-third of the nation's population lives in "rural" America and a disproportionate number of deaths due to MVCs (56.9%) occur in rural areas [25]. Telemedicine could bridge the gap between biomechanics research of head injury and its day-to-day clinical applications. Our system proved effective in identifying MTBIs in hypothetical patients. The system could be used in rural areas, especially in mid west states like North Dakota to help ED physicians identify biomechanical risk factors and to stratify patients into different risk levels for the appropriate referral pattern and management in a timely manner.

Currently no generally accepted standards exist for the treatment and management of MTBI, appropriate diagnosis, referral, and patient and family education are critical for helping MTBI patients achieve optimal recovery and to reduce, or avoid, significant sequelae [26]. Diagnosing MTBI, however, can be challenging because symptoms often are common to other medical problems, and onset of symptoms may occur days to even months after the initial injury [26,27]. The primary goal of initial management in MTBI is to identify the patients at risk of intracranial abnormalities. In view of the difficulties for evidence-based medicine to provide definitive strategies, additional knowledge of the injury risk factors can be beneficial in an evaluation process.

The American College of Surgeons [28] includes injury mechanism in its life trauma support protocol. Over 40 years of extensive research in the aerospace and automotive industries in the area of impact biomechanics has accumulated a large amount of data and corresponding theories in national databases, as

well as biomedical and biomechanical publications [29-33]. Due to non-availability of this knowledge in ED settings, however, ED physicians have rarely benefited before now from the achievements in this field. Due to the urgent nature of emergency medicine, computerized systems could ideally provide instant assistance to ER physicians on the basis of 24 hours a day and 7 days a week.

The system shell proved to be very effective in managing the rules. The function of testing rules can be especially effective in identifying the potential conflict of the rules by tracing which rules fire under what specific conditions. The generic interface for adding new rules proved to be easy to use by experts in biomechanics who had no training. The system could handle multi-users concurrently. By placing the system inside the firewall of the NDSU campus server and reusing the NDSU Core Programming Library, we saved up to 60% of the maintenance effort and 30% of developing time.

5. *CONCLUSIONS*

We developed a Web-based application system to help identify the potential brain injury patient, in EDs. who have been involved in MVCs. The system proved to be effective in identifying 94% hypothetical brain injury patients by stratifying the patients in different risk levels for MTBI. This system could potentially assist the ED physicians in identifying a patient at risk for MTB, that could lead to a proper referral pattern and clinical decision making for further neuroclinical examinations. This would be particularly useful in rural areas that lack an advanced trauma care system. Furthermore, by modeling the knowledge in impact biomechanics into a database, the system provides a novel approach to handling the structural knowledge and checking the potential conflicts of the knowledge base. The system shell could be easily adapted to other knowledge based systems to provide domain expertise from other fields for biomedical applications.

ACKNOWLEDGEMENTS

We thank Dr. James Ross and his Web Development Team from the ITS at NDSU for their help in programming, especially the hard work from MD Rahaman, Sunil Nanam, and Sampath.Velupula. We also thank Chris Forseth for his help in the early-stage setup of the system.

REFERENCES

[1] Jager T, Weiss H, Coben J, et al. “Traumatic brain injuries evaluated in US emergency departments, 1992-1994.” Acad Emerg Med. 2000;7: pp134-140.

[2] CDC. “Heads Up: Facts for Physicians About Mild Traumatic Brain Injury (MTBI).” Centers for Disease Control and Prevention, National Center for Injury Prevention and Control; Atlanta, GA, September 05, 2003.

[3] Kay T, Harrington DE, Adams R, Anderson T, Berrol S, et al., “Definition of mild traumatic brain injury.” J Head Trauma Rehabil, 1993. 8: pp. 86-87.

[4] P. E. Vos, L. Battistin, G. Birbamer, et al. “EFNS guideline on mild traumatic brain injury: report of an EFNS task force.” European Journal of Neurology 2002, 9: pp. 207–219.

[5] “Advanced Trauma Life Support for Doctors 6th ed.” American College of Surgeons Committee on Trauma, 1997.

[6] Bartlett J, Kett-White R, Mendelow AD, et al. “Recommendations from the Society of British Neurological Surgeons.” Br J Neurosurg, 1998, 12: pp. 349–352.

[7] Minczak BM “Principles of Advanced Trauma Care v9.” Academic Emergency Medicine, 2002, pp. 871-872.

[8] Ommaya AK, “Head injury mechanisms and the concept of preventive management: a review and critical synthesis.” J Neurotrauma, 1995, 12: pp. 527-546.

[9] Digges, KH FHWA/NHTSA National Crash Analysis Center, “Summary Report of Rollover Crashes,” June 2002.

[10] Augenstein J, .Bowen J, Perdeck E, Singer M, Stratton J, Horton T, Rao A , Digges K, Malliaris A, Steps J SAE International 2000 World Congress, “Injury Patterns in Near-Side Collisions, 2000-01-0634, March, 2000.

[11] Quinlan K, et al., “Neck strain and sprains among motor vehicle occupants-United States, 2000,” Accident Analysis & Prevention 36 (2004) 21-27.

[12] Department of Transportation National Highway Traffic Safety Administration, Federal Motor Vehicle Standard (FMVSS) 208, Washington, DC. 2001.

[13] Society of Automotive Engineers, “Guidelines for evaluating out-of-position vehicle occupant interactions with deploying frontal airbags,” SAE Information Report. SAE J1980 Dec 2001, Society of Automotive Engineers, Inc.: Warrendale, PA.

[14] Society of Automotive Engineers, “Human mechanical impact response characteristics: response of the human neck to inertial loading by the head for automotive seated postures.” SAE Information Report. SAE J1460-2, JAN 1998, Society of Automotive Engineers, Inc.: Warrendale, PA.

[15] Robbins DH, Schneider LW, et al., “Seated Posture of Vehicle Occupants,” SAE Technical Paper No. 831617, Society of Automotive Engineers, Inc.: Warrendale, PA. 1983,

[16] Reed MP, Manary MA, and Schneider LW, “Methods for measuring and reporting automobile occupant posture.” SAE Technical Paper No.1999-01-0959. Society of Automotive Engineers, Inc.: Warrendale, PA. 1999,

[17] Strother CE, James MB, Gordon JJ, “Response of Out-of-Position Dummies in Rear Impact.” SAE Technical Paper No. 941055. Society of Automotive Engineers, Inc: Warrendale, PA, 1994,

[18] Berton RJ, “Whiplash: Tests of the influential variables.” SAE Technical Paper No. 680080. Automotive Engineering Congress: Detroit, MI, 1968.

[19] Foret-Bruno JY, Tarriere C, LeCoz JY, Got C, and Guillon F, “Risk of cervicallesions in real-world and simulated collisions,” in 34th AAAM Conference Proceedings. 1990: Scottsdale, Arizona.

[20] Hu AA, Bean SP, Zimmerman RM. “Response of belted dummy and cadaver to rear impact.” SAE 770929. Proceedings of the Twenty-First Stapp Car Crash Conference. Warren, MI: Society of Automobile Engineers, 1977.

[21] Olsson I, Bunketorp O, Gustafsson C, Planath I, Norin H, Ysander L. “An in-depth study of neck injuries in rear end collisions.” Proceedings of International Research Council on Biomechanics of Impact, 1990, Rotterdam, Netherlands.

[22] Romilly DP, Thomson RW, Navin FPD, Macnabb MJ, “Low speed rear impacts and the elastic properties of automobiles,” Proceedings of the 12th ESV conference 1989: Gothenburg, Sweden.

[23] Bazarian JJ, Fisher S, Flesher W, Knox K, Lillis R, Pearson T. “Lateral automobile impacts and the risk of traumatic brain injury.” Annals of Emergency Medicine. 2004;44:pp.142-152.

[24] Luger, George F. “Artificial Intelligence: Structures and Strategies for Complex Problem Solving.” 4th edition. Reading, MA: Addison-Wesley, 2002.
[25] “Rural Emergency Medical Services: Special Report,” Publication OTA-H-445, Congress Office of Technology Assessment, Washington, DC, 1989.
[26] De Kruijk JR, Twijnstra A, Leffers P, “Diagnostic criteria and differential diagnosis of mild traumatic brain injury.” Brain Injury, 2001. 15: pp. 99–106.
[27] Ingebrigtsen T, Romner B, Marup-Jensen S, Dons M, Lundqvist C, Bellner J, et al., “The clinical value of serum S-100 protein measurements in minor head injury: a Scandinavian multicentre study.” Brain Injury, 2000. 14: pp. 1047–1055.
[28] Chambers J, Cohen SS, Hemminger L, Prall A, Nichols J, “Mild traumatic brain injuries in low-risk trauma patients.” The Journal of Trauma: Injury, Infection and Critical Care, 1996. 41(6): pp. 976-980.
[29] Alexander MP, “Mild traumatic brain injury: pathophysiology, natural history, and clinical management.” Neurology, 1995. 45: pp. 1253–1260.
[30] EFNS Talk Force, “EFNS Guideline on Mild Traumatic Brain Injury.” European Journal of Neurology, 2002. 9: pp. 207-219.
[31] Haydel MJ, Preston CA, Mills TJ, et al., “Indications for computed tomography in patients with minor head injury.” New England Journal of Medicine, 2000. 343: pp. 100-105.
[32] Stiell IG, Wells GA, Vandemheen K, et al., “The Canadian CT head rule for patients with minor head injury.” Lancet, 2001. 357: pp. 1391-1396.
[33] Goldstein M, “Traumatic brain injury: a silent epidemic [editorial].” Ann Neurol, 1990. 27: pp. 327.

Zhifeng Kou obtained his Ph.D. in Mechanical Engineering, specialized in head injury biomechanics, and M.S. in Computer Science, in 2005, from North Dakota State University in Fargo, North Dakota, USA. His research interest is traumatic brain injury with publications covering from injury mechanism, clinical decision making system, to state of the art MR imaging of brain injury. His honors include serving as a panelist in the 2006 Sigma Xi International Conference, several travel awards from the International Society for Magnetic Resonance in Medicine and the National Neurotrauma Society, and a Merit Certificate from the United States National Committee on Biomechanics.

Mariusz Ziejewski, Ph.D., is a tenured Associate Professor, Director of Impact Biomechanics Laboratory, College of Engineering and Director of the Automotive Systems Laboratory, Department of Mechanical Engineering, North Dakota State University. He is also an adjunct Associate Professor in the Department of Neuroscience, University of North Dakota School of Medicine. For many years, he has been involved in human body dynamics research with the U.S. Air Force. He was a member of the NHTSA Collaboration Group on Human Brain Modeling and has authored articles and book chapters on neck and brain injury.

Literature Review of Water Demand

Milutinovic, Milan

Abstract— *The field of water demand analysis is becoming increasingly important, due to the problems that water utilities are faced with, when supplying the constantly increasing water quantities. This review paper starts with an introduction to water demand modeling and continues with the specification of the demand models and variables used. Also, effects of non-price policies and technology changes are reviewed.*

1. Introduction: Demand Modeling

With the increase in worldwide water demand over the last few decades, water utilities face problems of supplying the quantity of demanded water. Water pricing, together with other options, showed to be an efficient tool in controlling water consumption. Many studies have researched the influence of pricing. The journals "Land Economics" and "Water Resources Research" have dedicated much space to this study.

A number of the studies were influenced by or used previous research developed in the study of electricity demand (i.e. Taylor 1975, Nordin 1976). Most of the studies are regression models based on data collected during various surveys, in regions where water prices increased.

In a large number of water demand studies, there are many different approaches. There is no consensus on the correct method to predict the demand for water. This is in part influenced by the fact that every region has its own characteristics regarding water use and socio-economic influences. Most studies find that household characteristics, water prices, climate and seasonal changes and conservation campaigns influence price elasticity.

Water demand studies started in the 1960 and 70s mainly in the USA. In the 1980s, the number of studies increased significantly, mostly encountering regression models based on various data sets in water scarce areas of the US. In the 1990s, conservation methods and water efficient technologies received more attention. Also, a number of studies were done in European and other countries. In addition, some new methods were investigated in order to predict the water demand.

This literature review presents specifications of the models, variables used, technology changes, non-price policies, and some new studies in this field that differ from earlier research.

Manuscript received May, 2006. Milutinovic Milan is a student of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 03126t.

2. Models Specification

2.1 Form

Most of the demand models are regression models. They typically use the form Q = f (P, Z) where P are the price variables and Z are factors such as income, household characteristics, weather, etc (Arbues et al. 2003). The most common forms are linear and logarithmic. There is no agreement about which functional form gives better results. Some researchers specify the form by seeing which model better fits their data set. Billing and Agthe (1980) cite that the elasticity in the log model is more useful if the demand is a rectangular parabola, while the elasticity in the linear form is more useful if water demand is linear over a relevant range.

The main flaw that researchers attribute to the linear model is that at some price, the demand for water will be zero, which is not logical as a minimum level of water consumption is needed to survive (Arbues et al, 2003).

2.2 Estimation methods

Different estimation methods are used in the studies. The most common are Ordinary Least Squares (OLS), Two and Three -Stage Least Squares (2SLS, 3SLS), and Maximum Likelihood. The choice of the method is somewhat influenced by the data set that the researcher possesses.

2.3 Data sets

A number of different datasets have been used, ranging from individual household data to aggregate data. A number of the studies used surveys conducted on a sample of households (Rizaiza 1991, Dandy et al. 1997, Renwick and Archibald 1998), other researchers used surveys conducted by the American Waterworks Association (Nieswiadomy - 1984 survey, Foster and Beattie – 1960 survey).

Researches used cross-sectional data (Foster and Beattie 1979, Chicione and Ramammurthy 1986, etc.), times- series data (Billings 1982), and most commonly cross-sectional-times series data (Nieswiadomy and Molina 1989, Renwick and Archibald 1998, Chicione and Ramamurthy 1986, etc.). Some models include lagged consumption in their models (Dandy et al. 1997, Nieswiadomy and Molina 1991). The Dynamic model, with an included lagged consumption, is

used because water use tends to respond slowly to changes in price and other variables, because water-using durables, like washing machines, swimming pools, etc. tend to change only steadily (Dandy et al. 1997).

3. VARIABLES

3.1. Household characteristics

Household characteristics are an important factor influencing water demand. All studies include monthly household income as a significant variable that increases water demand. In the deficiency of income data, some demand models use property value as an alternative (Dandy et al 1997).

A number of researchers include lot size as a significant variable (Renwick et al 1998; Dandy et al 1997; Lyman 1992, etc.). Houses with larger lot sizes are expected to have larger outdoor water use (Renwick and Green 1998). Also, household size was frequently used in the demand equation (Nieswiadomy 1992, Renwick et al 1998, Dandy 1997 etc.) as having significant influence on demand. Density of households (Foster, Renwick 1998; Nauges 2000), the number of faucets and age distribution of household members (Lyman1992) are used in some studies too. Table 1 presents income and household size elasticities found in some water demand studies.

3.2. Price Variables

The most common question in the water demand literature is whether the average price or the marginal price combined with the difference variable should be used as the price variable in the demand equation. Although it has been the subject of a thorough debate in the literature, a consensus has not been reached yet.

The Debate: Howe and Linawever (1967) cited that using the marginal price alone will have invalid results in the presence of block tariffs. Taylor (1975) suggested an alternative method by including two price-related variables in the estimating model, when block rates are applied. Nordin (1976) modified it, citing that the second price variable should be the difference between the consumers actual bill and what would be paid if all units were purchased at the marginal price (in the case of a declining block tariffs for electricity). Billings and Agthe (1980) implement the difference variable under increasing block tariffs for water demand, showing that it is correct and statistically significant. Economic theory suggests that the coefficients in front of the difference variable and income variables should be the same magnitude, but with opposite signs. However, empirical evidence shows that the coefficient on income and difference should have different signs, but with a bigger coefficient in front of the income variable.

Billings and Agthe (1980, 1982) argue that the use of the average price will generate bigger elasticities when a block pricing schedule is implied, especially when the marginal price increases, while the intra-marginal rates remain the same. In this case the change in marginal price is greater than the change in average price. A possible situation is that with an increase in MP, the AP remains constant or even decreases. Billings and Agthe (1980, 1982) also cite that the effect of a change in rates may have different effects on water use; the use of average price alone ignores this, and produces less accurate results

In many recent studies on water demand, the MP combined with the difference variable is used to show price elasticities (Renwick and Archibald (1998); Renwick, Green, and McCorkle (1998); Dandy, Nguyen, and Davies (1997); Nieswiadomy and Molina (1989)).

However, many earlier studies use the average price (Wong 1972, Young 1973, Foster and Beattie 1979). In their studies Foster and Beattie (1981) recognize that the Nordin specification (the use of MP and difference variable) was not significantly different that the average price specification. They also emphasize questions regarding the knowledge that consumers have on their MP and the way of block pricing and if their reaction is actually set according to the average price.

Shin (1985) constructed a price perception model for electricity demand that describes the response of consumers to MP or AP. He cited that the average consumer does not know the actual rate schedule. Nieswiadomy (1992) gives reasons supporting the average price variable because of the difficulty of determining the actual water usage during the month, as water meters are difficult to read. In addition he cites the difficulty of knowing when blocks have been switched and the fact sewer charges can confuse the consumer.

Shin (1985) defines the price perception parameter as $P^* = MP\ (AP/MP)\ k$, where k is the price perception parameter. Thus, if $k = 0$ the consumer responds only to the MP, if $k = 1$ then the consumer responds only to the average price. If $0 < k < 1$ then the price perceived is between AP and MP. Shin finds that electricity consumers react to average prices in his study. Nieswiadomy (1992) tests the Shin model for water demand. His results indicate that consumers react more to average prices than to marginal prices; k is approximately equal to 1 (although in his 1991 study he found that consumers react to marginal prices)

Opaluch (1982) also suggests a test concerning the measure of the price to which consumers respond, for a two block tariff schedule. The hypothesis was conducted through a thorough utility theoretical framework by Opaluch (1981). He suggests a demand

equation:

$$Q = B_3 + B_1 \cdot P_x + B_2 \cdot P_2 + B_3 \cdot \left(\frac{(P_1 - P_2) \cdot Q_1}{Q} \right) + B_4 \cdot (Y - (P_1 - P_2) \cdot Q_1)$$

where:

Q – total purchases of the goods subject to block pricing

Px – price index for other relevant goods

P1 – price of Q in the first block

Q1 – quantity of the good which is subject to the initial block pricing (P1)

Y – total income of the consumer

The average price is

$$AP = \frac{P_1 \cdot Q_1 + P_2 \cdot (Q - Q_1)}{Q} = P_2 + \frac{(P_1 - P_2) \cdot Q_1}{Q}$$

If the consumers react to the block tariff schedule, then B3 = 0, and the demand equation reduces to Nordin's specification. If the consumers react to the average price, B2 = B3 the equation uses average price as a variable.

The Conclusion: A number of studies accept the idea that the preferences between different price specifications are influenced by empirical rather then theoretical factors. Foster and Beattie (1979, 1981) state that the price schedule that consumers react to should be a subject for testing with available data. Basically, if the consumers think the water bill is significant, they will put in the effort to learn about the pricing schedule and their exact consumption and marginal price. Otherwise, where the water bill represents a small percentage of income, the consumer will react to the average price (Nieswiadomy 1992, Shin 1985)

A review of accounted price elasticities and price variables used in various studies are presented on Table 2.

Most researchers found that seasonal changes and climates influence water consumption. However, they used different variables. Billings et al. (1980,1982) use evapotranspiration from Bermuda grass minus rainfall, Dandy et al. (1997) use moisture deficit (MD = PE-0.6R, where 0.6R = effective rainfall, MD = moisture deficit, but only for the summer demand), Foster and Beattie (1981) use precipitation during growing season, Ajadi et al (2003) used rainfall, while Nieswiadomy and Molina (1991) used weather as a variable.

A Number of studies also use temperature in their models (Nieswiadomy, Renwick et al., Riaza, etc.). Renwick et al. (1998) included the influence of temperature and rainfall in their water demand model. Following Chesnutt and Mcspadden, they present two equations for influences that temperature and climate have on demand. To include the influence of seasonality these equations used sine and cosine Fourier series for the maximum daily air temp (eq. 1) and cumulative monthly precipitation (eq. 2). These values are then included into the demand equation.

(1)

$$\ln(DTEMP) = \gamma_0 + \sum_1^6 \left\{ \gamma_{1,j}^{tp} \cdot \sin\left(\frac{2\pi jt}{12} \right) + \gamma_{2,j} \cdot \cos\left(\frac{2\pi jt}{12} \right) \right\} + e_{it}^{tp}$$

(2)

$$\ln(DPREC) = \gamma_0 + \sum_1^6 \left\{ \gamma_{1,j}^{pr} \cdot \sin\left(\frac{2\pi jt}{12} \right) + \gamma_{2,j} \cdot \cos\left(\frac{2\pi jt}{12} \right) \right\} + e_{it}^{pr}$$

A number of studies found that summer demand is more elastic to price increase than is winter demand (Lyman 1992, Dandy et al. 1997, Griffin and Chang, etc.). Dandy used seasonal models (winter and summer) in his studies. Also studies have found that outdoor water use is more elastic than indoor.

Nieswiadomy cites that in a log-log model temperature has a nonlinear relationship with demand; the marginal impact of temperature goes up with increases of temperature; he also cites that variations of temperature below 18C have no impact on water demand.

4. *Effects of Non-Price Policy on Household Demand*

Previous studies have shown that non-price policies reduce demand. Renwick and Green (1998) showed that non-price Demand side Management (DSM) policy instrumnets have influence on demand. In their demand equation they included six variables: Public information campaigns (INFO), distribution of free retrofit kits (RETRO), low-flow toilet rebate programs (REBATE), water rationing policies (RATION), water use restrictions (RESTRICT), compliance affirmation policy (COMPLY).

In their study of California Water agencies they find that policies reduce water demand by the percentage presented in table 3.

Table 1: Influence of non-price policies

Variable	% or reduction
INFO	8%
RETRO	9%
RATION	19%
RESTRICT	29%
COMPLY	Not significant
REBATE	Not significant

Logically, more obligatory policies reduce demand for water more than voluntary policies. As the authors conclude, the outcome is influenced by the quality of the implementation of these policies.

Nieswiadomy (1992), using experience in Tucson cites that a campaign is successful in decreasing demand only for a few years. Yet, after a few years use increases back to its previous level. He cites that only a major public campaign accompanied with a price increase will have success in the long run. Nieswiadomy also suggests that education programs will probably

have more effect in water scarce regions, because of the awareness of water scarcity.

5. Influence of Technological change on the Demand for Water

Influence of technology changes only recently became evident. Renwick and Archibald (1998) found that increasing the number of low flow toilets in a household by one would decreases household demand by 10%, while Chesnutt et al. (1992) found that it would decrease the demand by 11%.

Regarding the efficiency of low flow showerheads, the next elasticities were perceived:

Table 1: Elasticities of low-flow showerheads

Renwick and Archibald	Whitcomb	Chesnutt and McSpadden
8%	6.4 -9.7 %	2%

Low flow toilets and showerheads reduce water by having more efficient technologies and insure significant long term demand reduction with no required changes in the behavior of consumers (Renwick and Archibald 1998). In the same study, they perceive that the elasticity for adoptions of water efficient irrigation technologies for low and high density households is 31 and 10 percent, respectively.

Nieswiadomy warns that even if a water efficient device is installed, the consumer may react by using more water knowing about the conservation effect of the device, therefore offsetting the conservation impact of the device.

Agthe and Billings studied effects that would make consumers install water efficient technologies in individual households and apartments. They found that obligations to save money, income, household size and summer marginal prices effected the decision.

6. Recent studies

6.1. Maximum-Likelihood Models

Recently, maximum-likelihood models were used to predict price elasticity (Hewitt and Hanemann (1995), Pint (1999), etc.). Maximum-likehood models were previously applied in the labor supply literature. These two models are specified in a two-stage framework, they are based on likelihood functions that show the probability that a household will choose a particular block, in a discrete way, combined with the probability of its particular level of use in the chosen block, in a continuous way. Hewitt (1993) presented three different maximum-likelihood models: the heterogeneous-preference model, the error perception model and the two-error model. The models are structured based on the assumed source of error in estimating household demand. These errors can be errors in data, missing variables or errors in the household's actual consumption relative to its intended consumption. The models directly allow both economic and non-economic influences, they cite that variation in behavior is due to both price and income and influences represented by various socio-demographic variables (Pint 1999).

However, Hewitt and Hanemann using the two-error model got higher elasticities than in previous studies (-1.6), while Pint pointed out that elasticity is bigger in the two-error model (-0.2 to -1.24) than in the heterogeneous - preferences model (-0.04 to -0.29), concluding that the two models might be upper and lower bounds on the estimates for elasticity of demand for water. Also, they mention that these models are very costly to estimate, since they require a large number of socio-demographic observations and have complex non-linear functions.

6.2. Stone–Geary Form

A few authors used the Stone-Geary form to predict water demand and price elasticity (Matinez-Espineira and Nauges 2004, Gaudin et al 2001, Al-Qunaibet et al 1985). The function has already been used for food products, durable goods, transportation, and energy. Gaudin et al.(2001) propose this form because it includes a quantity of water that does not respond to price, allows elasticity to decrease as the price increases, and uses only two parameters (and) for each good. is defined as a threshold below which water consumption is not affected by prices, while is the preference variable. Basically, "The consumer is faced with a given level of income and set of prices. The consumers first purchases a minimum acceptable level of each good (the $_i$'s) and then portions of each good, for their leftover income, according to their preference parameter (the $_i$'s)" (Gaudin et al. 2001) Gaudin present the next form:

$$Q_w = \gamma_w + \beta \cdot \frac{I^* - P_w \cdot \gamma_w - \gamma_z}{P_w}$$

where I and P are income and price. SGE (γ,β) are linear combinations of exogenous variables. So, the equation for non-constant γ and β in the Gaudin et al (2001) study is:

$\beta_w = (\beta_0+\beta_1 C+\beta_2 SP+\beta_3 APP)$ and

$\gamma_w = \alpha o+\alpha_1 C+\alpha_2 SP+\alpha_3 AAP$

(γ_z was excluded from the model, as insignificant to the study)

Where C – days with rainfall; SP – Spanish population; AAP –average annual precipitation)

Gaudin et al (2001) found summer elasticities bigger than winter elasticities, and that more than half of the water demand does not respond to price increase.

Table 2: Elasticities in studies that use the Stone-Geary Form

Author	Study area	Price elasticity	Income Elasticity
Gaudin et al. (2001)	Texas	0.19-0.28	
Martinez-Espineira and Nauges (2004)	Spain	-0.1	0.1
Al-Quinaibet and Johnston (1985)	Kuwait	-0.77	0.211

6.3. Meta-analysis

Meta-analysis is the use of statistical techniques in a systematic review with a purpose of integrating the results of the included studies. Espey et al. (1997) using meta-analysis studied the factors that affect price elasticity estimates in recent studies in the USA. They tried to explain differences in elasticity using differences in inclusion of variables in the regression models. They found that long-run estimates are more elastic to than short-run estimates; that the inclusion of income, population density, household size, temperature, and seasonable variable do not influence the price elasticity even though they influence the demand; also that evapotranspiration rates, pricing structure (increasing block rates were found to be much more elastic), rainfall and the season influence the elasticity. Also, summer elasticity was found to be bigger than winter elasticity.

Dalhuisen et al. (2003) in their meta-analysis study found that moderately high price elasticities and reasonably low income elasticities are found in studies with increasing block rates. Also, they find that the absolute magnitudes of price and income elasticities are greater for areas with high income, that price elasticities in Europe are bigger than in the US and that elasticities do not change with the date of the study, in other words they did not find differences in elasticities of earlier and more recent studies.

CONCLUSION

Studies in water demand prediction and elasticity have come up with a wide range of results. These studies have been conducted using different datasets, regression methods, price increases and variables, that alter the results. Consequently, some correlation parameters have been empirically proven. However, water demand and price elasticity are, no doubt, influenced by local conditions and socio- economic variables. A consensus has not been reached regarding the best methods to predict demand and elasticity. Most researches conclude that more studies have to be done in water demand.

REFERENCES

[1] Agthe and Billings, 1996 “Water-Price Effect on Residential and Apartment Low-Flow Fixtures.” Journal of Water Resources Planning and Management / January/February

[2] Agthe and Billings,1996 “Water-Price Influence on Apartment Complex Water Use.” Journal of Water Resources Planning and Management / January/February

[3] Al-Qunabeit, Johnston, 1985 “Municipal Demand for Water in Kuwait, Methodological Issues and empirical results.” Water Resources Research 2 (4)

[4] Arbues, Garcia-Valinas and Martinez-Espinera, 2003 “Estimation of residential water demand: a state-of-the-art- review.” Journal of Socio-Economics 32(1)

[5] Ayadi, Krishnakura, and Matoussi; Tunis, 2003 “A panel data analysis of Water Demand in Presence of Nonlinear Progressive Tariffs.”

[6] Billings and Agthe, 1980 “Price Elasticities for Water: A case of Increasing Block Tariffs.” Land Economics 56 (1)

[7] Billings, 1982 “Specification of Price Rate Variables in Demand Models.” Land Economics 58(3)

[8] Chambouleyron, 2004 “Optimal Water Metering and Pricing”, Water resource Management 18

[9] Charney and Woodard, 1984 A test of Consumer Demand Response to Water Prices: Comment Land Economics 60(4)

[10] Chicoine and Ramamurthy, 1986 Evidence on the Specification of Price in the Study of Domestic Water Demand Land Economics 62(1)

[11] Dandy, Nguyen, and Davies, Australia, 1997 “Estimating Residential Water Demand in the Presence of Free Allowances.” Land Economics 73(1)

[12] Dalhuisen, Florax, de Groot, and Nijkamp, 2003 “Price and Income Elasticities of Residential Water Demand: A Meta-Analysis.” Land Economics 79(2)

[13] Espey M., Espey J., and Shaw, 1997 “Price Elasticity of Residential Demand for Water: A Meta-Analysis.” Water Resources Research 33(6)

[14] Foster and Beattie, 1981 “On the specification of price in Studies of consumer Demand under Block price Schedules.” Land Economics 57(4)

[15] Foster and Beattie, 1979 “Urban Residential Demand for Water in the United States.” Land Economics 55(1)

[16] Gaudin, Griffin, and Sickles, 2001 “Demand Specification for Municipal Water Management: Evaluation of the Stone- Geary Form.” Land Economics 77(3)

[17] Griffin and Chang, 1990 “Pretest Analysis of Water Demand in Thirty Communities” Water Resources Research 26 (10)

[18] Hewitt and Hanemann, 1995 “A Discrete/Continuous Choice Approach to Residential Water Demand under block rate pricing” Land Economics 71 (May)

[19] Lyman, 1992 “Peak and Off-Peak Residential Water demand”; Water Resources Research 28(9)

[20] Martinez-Espineira and Nauges 2004 “Is all domestic water consumption sensitive to price control?” Applied Economics 36, 1697-1703

[21] Nieswiadomy, 1992 “Estimating Urban Residential Water Demand: Effects of price structure, Conservation, and Education.” Water Resources Research 28 (3)

[22] Nieswiadomy and Molina, 1991 “A note on Price Perception in Water demand Models”, Land Economics 67(3)

[23] Nieswiadomy and Molina, 1991 “Comparing Residential Water Demand Estimates under Decreasing and Increasing Block Rates Using Household Data.”

[24] Nauges and Thomas, 2000 Privately Operated Water Utilities, Municipal Price Negotiation, and Estimation of Residential Water Demand: The Case of France.” Land Economics 76(1)

[25] Opaluch, 1982 “Urban Residential Demand for Water in the United States: Further discussion.” Land Economics 58(2)

[26] Pint, 1999 “Household responses to Increased Water Rates During the California Drought.” Land Economics 75 (2)

[27] Riazaiza, 1991 "A case of the major Cities of the Western Region of Saudi Arabia." Land Economics 27(5)
[28] Renwick and Archibald, 1998 "Demand Side Management Policies for Residential Water Use: Who Bears the conservation Burden." Land Economics 74(3)
[29] Renwick, Green, and McCorkle, 1998 "Measuring the price responsiveness of Residential Water Demand in California's Urban Areas." A report prepared for the California department of Water resources
[30] Renwick and Green, 1999 Do Residential Water Demand Side Management Policies Measure Up? An Analysis of Eight California Water Agencies" Journal of Environmental Economics and Management
[31] Schefter and David, 1985 Estimating Residential Water Demand under Multi-Pat Tariffs Using Aggregate data Land Economics 61 (3)
[32] Qdais and Nassay, 2001 "Effect of pricing policy on water conservation: a case study" Water policy 3(3)
[33] Woo, 1992 Drought Management, Service Interruption, and Water Pricing: Evidence from Hong Kong Water Resources Research 28 (10)

Author	Study area	Model	Income elasticity	Household size
Hewitt and Hanemann	Texas	D/C	0.15	
Renwick, Archibald 1998	California	Linear	0.36	
Dandy et al 1997*	Australia	Linear	SR: 0.14 | LR: 0.32-0.38	SR 0.04; LR 0.19
Griffin et al(1990)	Texas	Linear	0.3-0.48	

*Dandy et al. in his annual model used property value as an indicator of income (SR –short range; LR – Long range)

Table 3: Income and household size elasticity from various studies

Authors		**Study area**	**Price variable**	**Price Elasticity**
Howe and Linaweaver (1967)		USA	AP	-0.23
Gibbs (1978)		Miami, Florida		-0.51
Foster and Beattie (1980)	Exponential	USA	AP	-0.35 to -0.76
Billings (1982)	Lin/Log	Tucson, Arizona	MP & D	-0.66/-0.56
Schefter and David (1985)		Wisconsin		-0.12
Chicoine et al. (1986)		Illinois		-0.71
Chicoine and Ramamurthy (1986)	Linear	Illinois	MP (AP)	-0.6 on MP
Nieswiadomy and Molina (1989)	Linear	Denton, Texas	MP & D	-0.86
Griffin and Chang (1990)	Linear	USA	AP	-0.16 to -0.37
Riazaiza (1991)	Logarithmic	Saudi Arabia	AP	-0.4 to -0.78
Hansen (1996)		Copenhagen, Denmark		-0.10
Renwick and Archibald (1997)	Linear	California	MP & D	-0.33
Hoglund (1997)	Linear	Sweden	MP & AP	-0.20 on AP
Dandi et al. (1997)	Linear	Australia	MP & D	-0.63 to -0.77
Renwick, Green, McCorkle (1998)	Logarithmic	California	MP & D	-0.16 to -0.21
Nauges and Thomas (2000)	Linear	France	AP (&MP)	-0.22
Ayadi et al.(2003)	Logarithmic	Tunisia	AP	-0.17

Table 4: Summary of price elasticities in some studies of residential water demand

Reviewers by Countries

Argentina
Ovando, Gabriela P.; Universidad Nacional de Rosario
Rossi, Gustavo; Universidad Nacional de La Plata

Canada
Feng, Jing; University of Ottawa
Ollivier-Gooch, Carl; The University of British Columbia
Shewchenko, Nicholas; Biokinetics and Associates
Steffan, Gregory; University of Toronto

Czech Republic
Kala, Zdenek; Brno University of Technology

Finland
Lahdelma, Risto; University of Turku
Salminen, Pekka; University of Jyvaskyla

Germany
Accorsi, Rafael; University of Freiburg
Glatzer, Wolfgang; Goethe-University
Gradmann, Stefan; Universitat Hamburg
Klamma, Ralf; RWTH Aachen University
Wurtz, Rolf P.; Ruhr-Universitat Bochum

Greece
Katzourakis, Nikolaos; Technical University of Athens
Bouras, Christos J.; University of Patras and RACTI

Italy
Badia, Leonardo; IMT Institute for Advanced Studies
Carpaneto, Enrico; Politecnico di Torino

Japan
Hattori, Yasunao; Shimane University

Netherlands
Mills, Melinda C.; University of Groningen
Pires, Luís Ferreira; University of Twente

New Zealand
Anderson, Tim; Van Der Veer Institute

Portugal
Cardoso, Jorge; University of Madeira
Natividade, Eduardo; Polytechnic Institute of Coimbra
Oliveira, Eugenio; University of Porto

Republic of Korea
Ahn, Sung-Hoon; Seoul National University

Singapore
Tan, Fock-Lai; Nanyang Technological University

Spain
Barrera, Juan Pablo Soto; University of Castilla
Gonzalez, Evelio J.; University of La Laguna
Perez, Juan Mendez; Universidad de La Laguna
Royuela, Vicente; Universidad de Barcelona
Vizcaino, Aurora; University of Castilla-La Mancha

Sweden
Johansson, Mats; Royal Institute of Technology

Switzerland
Pletka, Roman; AdNovum Informatik AG
Rizzotti, Sven; University of Basel
Specht, Matthias; University of Zurich

Taiwan
Lin, Hsiung Cheng; Chienkuo Technology University

Turkey
Ozalp, A. Alper; Uludag University

United Kingdom
Ariwa, Ezendu; London Metropolitan University
Biggam, John; Glasgow Caledonian University
Dorfler, Viktor; Strathclyde University
Kolovos, Dimitrios S.; The University of York
Vetta, Atam; Oxford Brookes University

USA
Bach, Eric; University of Wisconsin
Bazarian, Jeffrey J.; University of Rochester School
Bolzendahl, Catherine; University of California
Bussler, Christoph; Cisco Systems, Inc.
Charpentier, Michel; University of New Hampshire
Chester, Daniel; Computer and Information Sciences
DeWeaver, Eric; University of Wisconsin - Madison
Ellard, Daniel; Network Appliance, Inc
Gaede, Steve; Lone Eagle Systems Inc.
Gill, Sam; San Francisco State University
Gustafson, John L..; ClearSpeed Technology
Hunter, Lynette; University of California Davis
Iceland, John; University of Maryland
Kaplan, Samantha W.; University of Wisconsin
Koua, Etien L.; The Pennsylvania State University
Lagunas-Solar, Manuel; University of California Davis
Langou, Julien; The University of Tennessee
Liu, Yuliang; Southern Illinois University Edwardsville
Lok, Benjamin; University of Florida
Minh, Chi Cao; Stanford University
Morrissey, Robert; The University of Chicago
Mui, Lik; Google, Inc
Rizzo, Albert ; University of Southern California
Rosenberg, Jonathan M. ; University of Maryland
Shaffer, Cliff ; Virginia Tech
Sherman, Elaine; Hofstra University
Snyder, David F.; Texas State University
Song, Zhe; University of Iowa
Yu, Zhiyi; University of California

Venezuela
Candal, Maria Virginia; Universidad Simon Bolívar

IPSI Team

Advisor for Mathematics-Related Problems:
Marija Radovic and Visnja Tarbuk

Advisors for IPSI Tutorials:
Tanja Kovacevic and Tanja Petrovic

Advisors for IPSI Research:
Milos Kovacevic and Milos Milovanovic

Advisors for IPSI Developments:
Zoran Babovic and Darko Jovic

Welcome to IPSI BgD Conferences and Journals!

http://www.internetconferences.net
http://www.internetjournals.net

VIPSI-2007 VENICE SPRING
Venice, Italy
March 19 to 22, 2007

VIPSI-2007 AMALFI
Hotel Santa Caterina,
Amalfi, Italy
March 22 to 25, 2007

VIPSI-2007 TOKYO
M.I.T., Tokyo, Japan
May 31 to June 3, 2007

VIPSI-2007 CROATIA - OPATIJA
Opatija/Abbazia
Villa Ariston
June 7 to 10, 2007

VIPSI-2007 CROATIA - ROVINJ
Rovinj/Rovigo
Villa Angelo d'Oro
June 10 to 13, 2007

VIPSI-2007 MONTREAL
Campus of UQAM,
Montreal, Quebec, Canada
June 29 to July 02, 2007

VIPSI-2007 ITALY
Rome and Hotel Castello Chiola,
Loreto Aprutino in Abruzzo,
relatively near Rome, Italy
August 20 to 23, 2007

VIPSI-2007 FLORENCE
Grand Hotel & La Pace,
in Montecatini Terme,
near Florence, Piza, Siena, and Lucca, Italy
Aug 23 to Aug 26, 2007

VIPSI-2007 BELGRADE
Belgrade, Serbia
September 3 to 6, 2007

VIPSI-2007 MONTENEGRO MOUNTAIN SAFARI
Villa Bianca,
Kolasin, Montenegro
September 6 to 9, 2007

VIPSI-2007 MONTENEGRO SVETI STEFAN
Hotel Sveti Stefan
September 9 to 15, 2007

VIPSI-2007 SLOVENIA
Lake Bled in Alps,
Slovenia,
October 8 to 11, 2007

VIPSI-2007 VENICE FALL
Venice, Italy
October 11 to 14, 2007

VIPSI-2007 PORTOFINO
Grand Hotel Miramare,
Santa Margherita Ligure,
Italy (near Genova),
October 14 to 17, 2007

The publication of this journal
is supported in part by
the Ministry of Science of Serbia

CIP – Katalogizacija u publikaciji
Narodna biblioteka Srbije, Beograd

ISSN 1820 – 4511 =
The IPSI BGD Transactions on Advanced Research
COBISS.SR - ID 119128844