Computer Science & Information Technology   **42**

Jan Zizka
Dhinaharan Nagamalai (Eds)

Computer Science & Information Technology

Fifth International Conference on Computational Science, Engineering
and Information Technology (CCSEIT 2015)
Vienna, Austria, May 23 ~ 24 - 2015

AIRCC

Volume Editors

Jan Zizka,
Mendel University in Brno, Czech Republic
E-mail: zizka.jan@gmail.com

Dhinaharan Nagamalai,
Wireilla Net Solutions PTY LTD,
Sydney, Australia
E-mail: dhinthia@yahoo.com

Preface

The Fifth International Conference on Computational Science, Engineering and Information Technology (CCSEIT-2015) was held in Vienna, Austria, during May 23~24, 2015. The Fourth International Conference on Mobile & Wireless Networks (MoWiN-2015), The Fourth International Conference on Information Theory (IT-2015), The Second International Conference on Artificial Intelligence and Applications (AIAP-2015) and The Second International Conference on Bioinformatics and Bioscience (ICBB-2015) were collocated with the CCSEIT-2015. The conferences attracted many local and international delegates, presenting a balanced mixture of intellect from the East and from the West.

The goal of this conference series is to bring together researchers and practitioners from academia and industry to focus on understanding computer science and information technology and to establish new collaborations in these areas. Authors are invited to contribute to the conference by submitting articles that illustrate research results, projects, survey work and industrial experiences describing significant advances in all areas of computer science and information technology.

The CCSEIT-2015, MoWiN-2015, IT-2015, AIAP-2015, ICBB-2015 Committees rigorously invited submissions for many months from researchers, scientists, engineers, students and practitioners related to the relevant themes and tracks of the workshop. This effort guaranteed submissions from an unparalleled number of internationally recognized top-level researchers. All the submissions underwent a strenuous peer review process which comprised expert reviewers. These reviewers were selected from a talented pool of Technical Committee members and external reviewers on the basis of their expertise. The papers were then reviewed based on their contributions, technical content, originality and clarity. The entire process, which includes the submission, review and acceptance processes, was done electronically. All these efforts undertaken by the Organizing and Technical Committees led to an exciting, rich and a high quality technical conference program, which featured high-impact presentations for all attendees to enjoy, appreciate and expand their expertise in the latest developments in computer network and communications research.

In closing, CCSEIT-2015, MoWiN-2015, IT-2015, AIAP-2015, ICBB-2015 brought together researchers, scientists, engineers, students and practitioners to exchange and share their experiences, new ideas and research results in all aspects of the main workshop themes and tracks, and to discuss the practical challenges encountered and the solutions adopted. The book is organized as a collection of papers from the CCSEIT-2015, MoWiN-2015, IT-2015, AIAP-2015, ICBB-2015

We would like to thank the General and Program Chairs, organization staff, the members of the Technical Program Committees and external reviewers for their excellent and tireless work. We sincerely wish that all attendees benefited scientifically from the conference and wish them every success in their research. It is the humble wish of the conference organizers that the professional dialogue among the researchers, scientists, engineers, students and educators continues beyond the event and that the friendships and collaborations forged will linger and prosper for many years to come.

Jan Zizka
Dhinaharan Nagamalai

Organization

General Chair

Natarajan Meghanathan   Jackson State University, USA
Dhinaharan Nagamalai   Wireilla Net Solutions PTY LTD, Australia

Program Committee Members

Abdelkrim   Université de Mascara, Algerie
Abdoulaye Kindy Diallo   Atlassian/ServiceRocket, Malaysia
Abdul Kadir Ozcan   The American university, Cyprus
Abdul Mateen Ansari   Bisha University, Kingdom of Saudi Arabia
Abdullah Kamman   Taif University, Saudi Arabia
Ahmad Al-Khasawneh   Hashemite University, Jordan
Ali Mansoori G   University of Illinois at Chicago, USA
Alireza Nemaney Pour   Islamic Azad University of Abhar, Iran
Amol D Mali   University of Wisconsin-Milwaukee, USA
Anas Abou El Kalam   UCA/ENSA of Marrakesh, Morocco
Ankit Chaudhary   Truman State University, USA
Arockiasamy S   University of Nizwa, Sultanate of Oman
Asadollah Shahbahrami   University of Guilan, Iran
Athirai   Nanyang Technological Universiy, Singapore
Belhamri Azeddine   Universite Mentori, Algerie
Boo-Hyung Lee   KongJu National University, South Korea
Ching-Hsien Hsu   Chung Hua University, Taiwan
Dac-Nhuong Le   Haiphong University,Vietnam
Denivaldo Lopes   Federal University of Maranhao, Brazil
Dhari Ali   University of Technology, Iraq
Farshchi   Allame Tabatabaye University,Iran
Fernando Bobillo   University of Zaragoza, Spain
Ghaznavi-Ghoushchi M. B   Shahed University, Iran
Grienggrai Rajchakit   Maejo University, Thailand
Hacene Belhadef   University of Constantine, Algeria
Halla Noureddine   Tlemcen University, Algeria
Hamdi Hassen   MIRACL Laboratory, Tunisia
Hazem Al-Najjar   Msiurata University, Libya
Hosny M. Ibrahim   Assiut University, Egypt
Hossain Shahriar   Kennesaw State University, USA
Ian Tan   Multimedia University, Malaysia
I-Cheng Chang   National Dong Hwa University, Taiwan
Ihab A. Ali   Helwan University, Egypt
Isa Maleki   Islamic Azad University, Iran
Isaac Agudo   University of Malaga, Spain

Technically Sponsored by

Networks & Communications Community (NCC)

Computer Science & Information Technology Community (CSITC)

Digital Signal & Image Processing Community (DSIPC)

Organized By

Academy & Industry Research Collaboration Center (AIRCC)

TABLE OF CONTENTS

Fifth International Conference on Computational Science, Engineering and Information Technology (CCSEIT 2015)

Fourth International Conference on Mobile & Wireless Networks (MoWiN 2015)

Fourth International Conference on Information Theory (IT 2015)

Second International Conference on Artificial Intelligence and Applications (AIAP 2015)

Second International Conference on Bioinformatics and Bioscience (ICBB 2015)

TWO-LAYER CRYSTAL SQUARE FOR TRANSISTORS ON A SINGLE CHIP

Haissam El-Aawar

Associate Professor, Computer Science/Information Technology Departments
Lebanese International University – LIU
Bekaa-Lebanon
haisso@yahoo.com, haissam.aawar@liu.edu.lb

ABSTRACT

The number of transistors on a chip plays the main role in increasing the speed and performance of a microprocessor; more transistors, more speed. Increasing the number of transistors will be limited due to the design complexity and density of transistors. This article aims to introduce a new approach to increasing the number of transistors on a chip. The basic idea is to construct two-layer crystal square for transistors; this allows to increase the number of transistors two additional times (four times as many) if the number of transistors incorporated in a one layer of crystal square will approximately double every 24 months according to Moore's Law without changing rapidly the design complexity and density in a crystal square and without changing the size of a chip (length and width), in this case the height of a chip must be changed for the two layers.

KEYWORDS

Crystal square, Number of transistors, Feature size, Die size, Design complexity, Density of transistors.

1. INTRODUCTION

Different types of a microprocessor, is also known as a central processing unit (CPU), such as i80x86, Pentium (Intel), PowerPC (IBM), SPARC (Sun), Motorola and many other types perform the same thing in approximately the same way [1].

Moore's Law states that the number of transistors on a chip would double every two years [2, 3]. The designers of CPU are interested to increase the speed and performance of CPU with each new CPU-generation according to the Moore's law by doubling the number of transistors according to the design rule by reducing the size of the transistors, which allows to fit more transistors on a chip without changing rapidly the size of the chip; in general the size of the chip increases slowly over time. Increasing the number of transistors is one of the main factors used to increase the speed and performance of the CPU; more transistors, more clock cycles [4, 5].

2. PHYSICAL LIMITATION OF INCREASING THE NUMBER OF THRANSISTORS

Increasing the number of transistors will be limited due to the following limitation [6, 7, 8, 9]:

2.1 DESIGN RULE, DENSITY AND DESIGN COMPLEXITY [10,11,12,13]

Since the wires and transistors are drawn photographically on the chips, the size of the transistors and the width of the wires are determined by the pixel size of the imaging process. The number of transistors that can be fitted on a given chip size is determined by the size of the transistors. As the feature size on the chip goes down, the number of transistors rises, which means the smaller the transistors, the more transistors can be fitted on the chip. Feature sizes have decreased from 10 micron in 1971 to 0.022 micron in 2012; the semiconductor manufacturing processes have decreased from 10 µm in 1971 to 22 nm in 2012 [4, 13, 14] (see Table 1).

Since the surface area of a transistor determines the transistor count per square millimeter of silicon, the transistors density increases quadratically with a linear decrease in feature size [1]. The increase in transistor performance is more complicated. As the feature sizes shrink, devices shrink quadratically in the horizontal and vertical dimensions. A reduction in operating voltage to maintain correct operation and reliability of the transistor is required in the vertical dimension shrink. This combination of scaling factors leads to a complex interrelationship between the transistor performance and the process feature size.

Due to the shrinking of the pixel size and the increasing of the density, the design complexity increases. If the pixel size shrinks double and the density increases double every two years according to Moore's Law, the physical limitation will definitely appear in few years, which means that it will be very difficult to apply Moore's Law in future. Some studies have shown that physical limitations could be reached by 2018 [6] or 2020-2022 [15, 16, 17, 18].

2.2 NUMBER OF TRANSISTORS AND DIE SIZE [5, 11, 19]

The die size is a size of a chip; the chip size depends on the chip manufacturing processes. As mentioned above, the smaller feature size, the larger number of transistors on a chip, this allows increasing the die size slowly over time. Owing to increasing of the number of transistors the chip size should be increased. The number of transistors of Intel microprocessors has increased from 2300 in 1971 to 3.1 billion in 2012 (1.35 million times during 40 years) and the die size has increased from 12 mm2 in 1971 to 544 mm2 in 2012 (45.34 times during 40 years). The number of transistors of MIPS microprocessors has increased from 110,000 (R2000) in 1986 to 7.15 million (R12000) in 1998 (65 times during 12 years) and the die size has increased from 80 mm2 in 1986 to 229.22 mm2 in 1998 (2.86 times during 12 years). The number of transistors of IBM microprocessors has increased from 1.5 million (Power3) in 1998 to 2.1 billion (Power7+) in 2012 (1400 times during 14 years) and the die size has increased from 270 mm2 in 1998 to 567 mm2 in 2012 (2.1 times during 14 years) (see Table 1).

2.3 PIXEL SIZE AND POWER CONSUMPTION [1,20]

The transistor's switching speed is determined by the pixel size, the smaller transistors switch faster, so they require less power to be switched. Small pixel size allows the battery life of portable devices to be saved for a long time. The energy consumption required for switching transistors is known as dynamic power consumption, which can be increased due to charging and discharging the capacitive output load. Increasing the amount of power requires more hot generation, which can cause an interference of transistors with each other.

Table 1. Evolution of Microprocessors: 1991 to 2012

Manufacturer	Processor	Date of introduction	Number of transistors	Process	Area [mm²]
Intel	Intel4004	1971	2,300	10 µm	12
	Intel8008	1972	3,500	10 µm	14
	Intel8080	1974	4,400	6 µm	20
	Intel8085	1976	6,500	3 µm	20
	Intel8086	1978	29,000	3 µm	33
	Intel80286	1982	134,000	1.5 µm	44
	Intel80386	1985	275,000	1.5 µm	104
	Intel80486	1989	1,180,235	1 µm	173
	Pentium	1993	3,100,000	0.8 µm	294
	Pentium Pro	1995	5,500,000	0.5 µm	307
	Pentium II	1997	7,500,000	0.35 µm	195
	Pentium III	1999	9,500,000	0.25 µm	128
	Pentium 4	2000	42,00,000	180 nm	217
	Itanium 2 McKinely	2002	220,000,000	180 nm	421
	Core 2 Duo	2006	291,000,000	65 nm	143
	Core i7 (Quad)	2008	731,000,000	45 nm	263
	Six-Core Core i7	2010	1,170,000,000	32 nm	240
	Six-Core Core i7/8-Core Xeon E5	2011	2,270,000,000	32 nm	434
	8-Core Itanium Poulson	2012	3.100,000,000	32 nm	544
MIPS	R2000	1986	110,000	2.0 µm	80
	R3000	1988	150,000	1.2 µm	56
	R4000	1991	1,200,000	0.8 µm	213
	R10000	1994	2,600,000	0.5 µm	299
	R10000	1996	6,800,000	0.35 µm	299
	R12000	1998	7,1500,000	0.25 µm	229
IBM	POWER3	1998	15,000,000	0.35 µm	270
	POWER4	2001	174,000,000	180 nm	412
	POWER4+	2002	184,000,000	130 nm	267
	POWER5	2004	276,000,000	130 nm	389
	POWER5+	2005	276,000,000	90 nm	243
	POWER6+	2009	790,000,000	65 nm	341
	POWER7	2010	1.200,000,000	45 nm	567
	POWER7+	2012	2.100,000,000	32 nm	567

3. TWO-LAYER CRYSTAL SQUARE

This article suggests as a solution for avoiding the physical limitations mentioned above and for applying Moore's Law more years a new approach of constructing a crystal square contained transistors on a chip. This approach is called two-layer crystal square of transistors.

Assume a microprocessor (let's say X) has the following specifications: date of introduction – 2015, one-layer crystal square of transistors, transistor count (number of transistors) – 3 billion, pixel size (feature size) – 0.038 micron, die size (area) – 600 mm2 (see Figure 1).
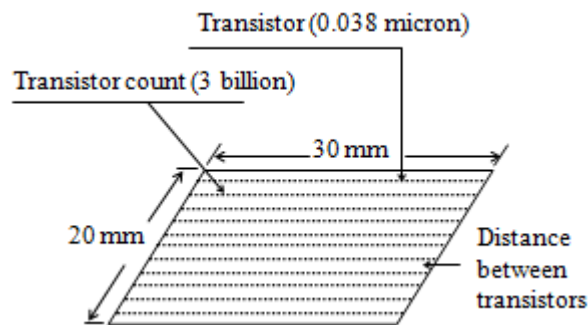
Figure 1. Crystal Square of Transistors

A crystal square of transistors can be constructed using two layers that are equal in size and transistor count, this allows to increase the number of transistors twice (Moore's Law) without touching the feature size and die size (see Figure 2).
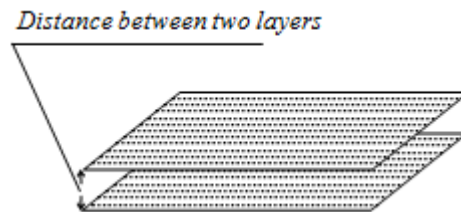
Figure 2. Two-Layer Crystal Square of Transistors

The two-layer crystal square of transistors shown in Figure 2 contains double transistor count without changing feature size and die size. Double number of transistors allows increasing a microprocessor's speed approximately twice. In this case the voltage of transistors must be decreased in order to reduce the dynamic power and energy.

Assume the second generation of the microprocessor (X1) has the following specifications: date of introduction – 2017, Two-layer crystal square, and each layer with the following characteristics: number of transistors – 3 billion (6 billion both), pixel size – 0.038 micron, die size – 600 mm2 (see Figure 3).
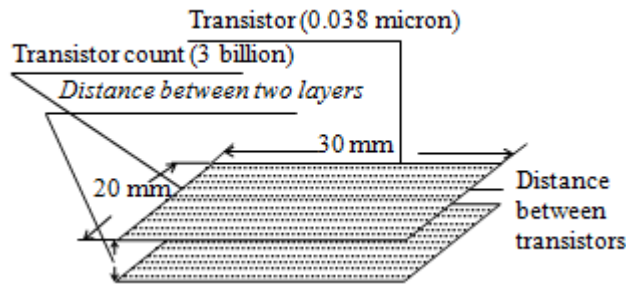
Figure 3. Second Generation

Applying this approach would double transistor count and speed of the microprocessors every two years with slowly increasing the density and area, and decreasing feature size and voltage, so the physical limitations will be reduced and Moore's Law will be continued for many years (may be till 2037 or more).

Assume the third generation of the microprocessor (X2) has the following specifications: date of introduction – 2019, Two-layer crystal square, and each layer with the following characteristics: transistor count – 6 billion (12 billion both), pixel size – 0.035 micron, die size – 688 mm2 (see Figure 4).
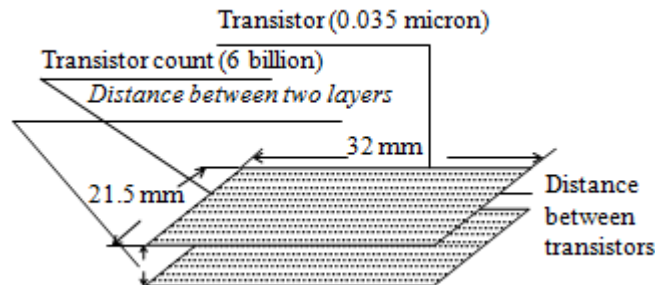
Figure 4. Third Generation

The third generation doubles the transistor count with small decreasing of the feature size (from 0.038 micron to 0.035 micron) and small increasing of the area (from 600 mm2 to 688 mm2), of course, with more density, which allows to fit more transistors in small area.

All the assuming characteristics of the microprocessors (X, X1, X2) and others (XN-1 and XN) are shown in Table 2. This table shows some of the assuming microprocessors from 2015 till the year 2037. The number of transistors will have doubled every two years. The feature size will have increased 3 nm (from 38 to 35) every two years (from the year 2019 – after applying a two-layer crystal technology in the year 2017), which means that the feature size would be reached 0.8 nm by the year 2037, which increases the design complexity and causes reaching the physical limitation. The area will also have increased about 12-15% every two years (from the year 2019); the small increasing of area can be applied if and only if the density of transistors increases with each generation. The increasing of the density would also be reached the physical limitation by the year 2037, as more density needs more power, which requires more hot generation.

Table 2. Assuming Evolution of Microprocessors: 2015 to 2037

Processor	Date of introduction	Number of transistors (billion)	Feature size (nm)	Area [mm²]
X	2015	3	38	600
X1 (Two-layer crystal)	2017	6	38	600 (each layer)
X2 (Two-layer crystal)	2019	12	35	688 (each layer)
X3 (Two-layer crystal)	2021	24	32	770 (each layer)
X4 (Two-layer crystal)	2023	48	29	882 (each layer)
--	--	--	--	---
XN-1 (Two-layer crystal)	2035	3072	11	---
XN (Two-layer crystal)	2037	6144	0.8	---

As shown in the table below, several measures of microprocessors technology, such as transistor count, is improving at exponential rate based on Moore's Law (see Figure 5).
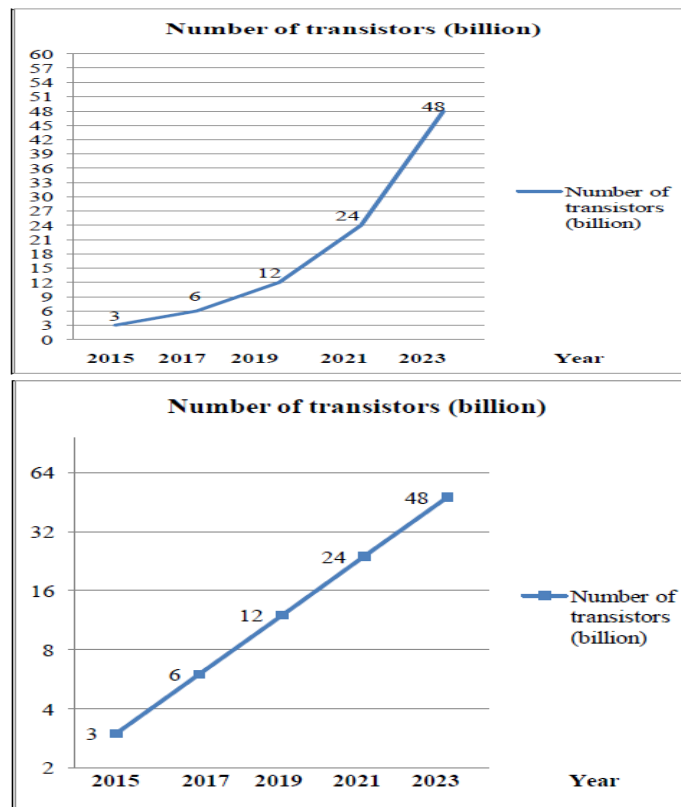
Figure 5. Number of transistors related to Moore's Law

4. CONCLUSION

The problem of applying Moore's law in microprocessor technology as much as possible is still topical research field although it has been studied by the research community for many decades. The main objective of this article is to find as much as possible a complete and logical solution for avoiding physical limitation in manufacturing of microprocessors technology and applying Moore's Law for a long time.

As mentioned above, some studies have shown that physical limitations could be reached by 2018 or 2022. Applying the new approach in microprocessor technology will delay the physical limitation for few more years, because it doubles the transistor count every two years based on Moore's Law, which approximately double the microprocessor's speed, with small decreasing of the feature size and small increasing of the area.

ACKNOWLEDGMENT

REFERENCES

[1]   John L. Hennessy and David A. Patterson, "Computer Architecture, A Quantitative Approach", 5th ed., pp. 17-26, 2011.
[2]   Gorden E. Moore, "cramming more Components onto Integrated Circuits", Electronics, Vol. 38, No. 8, April 19, 1965.
[3]   Jane Laudon, Kenneth Laudon, "Essentials of Management Information Systems", Chapter 4: IT Infrastructure: Hardware and Software, 10th ed., 2012.
[4]   Steve Gilheany, "Evolution of Intel Microprocessors: 1971 to 2007".
[5]   Wolfgang Arden, "Future roadblocks and solutions in silicon technology as outlined by the ITRS roadmap" in Mterials Science in Semiconductor Processing, vol. 5 issue 4-5 August – October, 2002, pp. 313-319.
[6]   Jan M. Rabaey, "Design at the end of Silicon Roadmap", Keynotes Address III, University of California, Berkelev, IEEE, ASP-DAC 2005.
[7]   Damon Poeter, "Intel's Gelsinger Sees Clear Path to 10nm Chips", June 30, 2008.
[8]   Hasan S., Humaria, Asghar M., "Limitation of Silicon Based Computation abd Future Prospects" in Proceedings of Second International Conference on Communication Software and Networks, 2010. ICCSN'10, pp. 599-561.
[9]   Robert W. Keyes, "Physical limits of silicon transistors and circuits", September 2005.
[10] F. Morals, L. Torres, M. Robert, D. Auvergne, "Estimation of layout densities for CMOS digital circuits", Proceeding International Workshop on Power and Timing Modeling Optimization Simulation (PATMOS'98), pp. 61-70, November 1998, Lyngby, Danemark.
[11] Ulrich Sigmund, Marc Steinhaus, and Theo Ungerer, "On Performance, Transistor Count and Chip Space Assessment of Multimedia-enhanced Simultaneous Multithreaded Processors", Workshop on Multi-Threaded Execution, Architecture and Compilation (MTEAC-4), Monterrey, Ca., Dec., 10, 2000.
[12] Singh B.P., Singh Renu, "Advanced Microprocessors and Microcontrollers", New Age International, Jan 1, 2008, 592 pages.
[13] Intel, "Transistors to Transformations, Form Sand to Circuits-How Intel Makes Chips".
[14] "A History of Microprocessors Transistor Count 1971 to 2013".

[15] Ahmad, Khaled; Schuegraf, Klaus, "Transistor Wars: Rival architecture face off in a bid to keep Moore's Law alive", IEEE Spectrum: 50, November 2011.

[16] Brooke Crothers, "End of Moore's Law: it's not just about physics", August 28, 2013.

[17] Robert Colwell, "The Chip Design Game at the End of Moore's Law", Hot Chips, August 2013.

[18] Joel Hruska, "Intel's former chief architect: Moore's law will be dead within a decade", August 30, 2013.

[19] Yale N. Patt, Sanjay J. Patel, Marius Evers, Daniel H. Friendly, Jared Stark, "One Billion Transistors, One Uniprocessor, One Chip", IEEE Computer Vol. 30 Issue 9, pp. 51-57, 1997.

[20] P. Zdebel, "Law Power/Low Voltage CMOS Technologies, A Comparative Analysis", Microelectronics Engineering, Vol. 39, Elsevier, Dec. 1997, pp. 123-137.

AUTHOR

Haissam El-Aawar is an Associate Professor in the Department of Computer Science and Information Technology at the Lebanese International University where he has been a faculty member since 2009. Haissam completed his Ph.D. and M.Sc. degrees at the State University "Lviv Polytechnic" in Ukraine. His research interests lie in the area of Artificial Intelligence, theory of complexity, microprocessors evaluation, CISC- and RISC-architectures, robotics control, mobility control and wireless communication.

Boolean Orthogonalizing Combination Methods

Yavuz Can[1]and Georg Fischer[2]

[1,2]Institute for Electronics Engineering (LTE),
Friedrich-Alexander University, Erlangen-Nuremberg, Germany
`yavuz.can@fau.de`
`georg.fischer@fau.de`

ABSTRACT

In this paper a new logical operation method called "orthogonalizing difference-building" is presented. It is used to calculate the difference, but also the complement of a function as well as the EXOR and EXNOR of two minterms respectively two ternary-vectors or two functions respectively two ternary-vector-lists is presented. On the basis of this new method a further logical operation method called "orthogonal OR-ing" is going to be introduced. The advantages of both methods are their results, which are already available in an orthogonal form that has an essential advantage for continuing calculations. Since it applies, an orthogonal disjunctive normal form is equal to orthogonal antivalence normal form, subsequent Boolean differential calculus will be simplified.

KEYWORDS

Difference-Building, Orthogonality, TVL, Data Memory Request, Computing Time

1. INTRODUCTION

Orthogonality is a special property of Boolean functions because an orthogonal function can be transformed in another form and so it simplifies the handling for further calculations in applications of electrical engineering. For example, the orthogonal form of a disjunctive form is equal to the orthogonal form of an antivalent form, $DNF^{orth} = ANF^{orth}$. This work shows new orthogonalizing methods, which means a new method of OR-ing as well as a method for building the difference of minterms respectively ternary-vectors (TV) or functions respectively ternary-vector-lists (TVL) of disjunctive normal form which result in orthogonal form. On the one hand, the method "othogonalizing difference-building $\ominus$" is similar to the usual difference-building derived from the set theory, but, on the other hand, it provides orthogonal results. It also enables the replacement of EXOR and EXNOR of two minterms (also TVs) or two functions (also TVLs). Thus, orthogonal results can be achieved which has essential advantage for further calculation in particular in the area of TVL-arithmetic. The other method "orthogonalizing OR-ing $\underline{\vee}$" offers the logical disjunction of two minterms respectively two TVs with the characteristic of the resulting in orthogonal form.

2. THEORETICAL FOUNDATIONS

2.1. Isomorphism: Set Theory as Basis for Switching Algebra

In Figure1a), the set difference from the set theory is shown. If the set difference is called $M\backslash S$, the remaining set of these two sets is the set of all elements belonging to set M, but not to set S (M without S). In Figure1b), the union of set S_1 and S_2, that is $S_1 \cup S_2$, is illustrated [2][3].
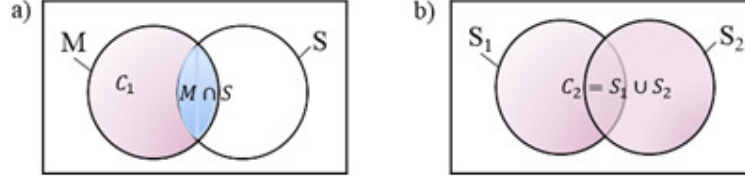
Figure 1: a) Set Difference $M\backslash S$; b) Set Union $S_1 \cup S_2$

Due to the isomorphism the set difference and the set union out of the set theory are expressed in the propositional logic as follows:

$$C_1 = M\backslash S = M \cap \bar{S} \quad \rightarrowtail \quad M \wedge \bar{S} \tag{1}$$

$$C_2 = S_1 \cup S_2 \quad\quad\quad \rightarrowtail \quad S_1 \vee S_2 \tag{2}$$

Thus, an Equation (3) to build the difference of two minterms in the switching algebra is defined out of (1). The minuend minterm is indentified with $m_M := \bigwedge_{m=0}^{n-1} x_{n-m}$, the subtrahend minterm is indentified with $m_S := \bigwedge_{s=0}^{n-1} x_{n-s}$ and with $n \in \mathbb{N}$ follows:

$$m_M \backslash m_S = \left(\bigwedge_{m=0}^{n-1} x_{n-m}\right) \backslash \left(\bigwedge_{s=0}^{n-1} x_{n-s}\right) := \left(\bigwedge_{m=0}^{n-1} x_{n-m}\right) \wedge \overline{\left(\bigwedge_{s=0}^{n-1} x_{n-s}\right)} =$$

$$= \left(\bigwedge_{m=0}^{n-1} x_{n-m}\right) \wedge \left(\bigvee_{s=0}^{n-1} \bar{x}_{n-m}\right) = (x_n \cdot x_{n-1} \cdots x_1)_M \wedge (x_n \cdot x_{n-1} \cdots x_1)_S \tag{3}$$

An Equation (4) building the disjunction of two minterms where the first summand minterm is identified with $m_{S_1} := \bigwedge_{s_1=0}^{n-1} x_{n-s_1}$ and the second summand minterm is identified with $m_{S_2} := \bigwedge_{s_2=0}^{n-1} x_{n-s_2}$ is defined out of (2). And with $n \in \mathbb{N}$ follows:

$$m_{S_1} \vee m_{S_2} = \left(\bigwedge_{s_1=0}^{n-1} x_{n-m}\right) \vee \left(\bigwedge_{s_2=0}^{n-1} x_{n-s}\right) = (x_n \cdot x_{n-1} \cdots x_1)_{S_1} \vee (x_n \cdot x_{n-1} \cdots x_1)_{S_2} \tag{4}$$

Equation (3) and (4) only indicate the value of the variable at the corresponding position in terms of its index. That means $\bar{x}_i$ is the complement value of variable at the i-th place, whereas x_i hows the value of the variable at the i-th place. The general validity is proved by the method of mathematical induction.

2.2. Ternary-Vector-List (TVL)

Ternary-Vector-Lists are representations of Boolean functions which can take one of three possible values for each point. A non-negated variable is characterized by '1', a negated by '0' and not included by '-'. Boolean equations represented by TVL can be treated computationally easier [5][10]. A TVL consists of m-rows (number of conjunctions or disjunctions contained in the function) and n-columns (number of independent variables) and is characterized with $n \in \mathbb{N}$ and $t \in \{0, 1, -\}$. Any form of a function, that means disjunctive form (DF), conjunctiveform (KF), equivalent form (EF) and antivalence form (AF), can be represented as a TVL. As no operators ($\lor$, $\land$, $\odot$, $\oplus$) in the TVL presentation are given, the designations of the matrix by $D(f)$, $K(f)$, $E(f)$ and $A(f)$ show the type of the TVL [5][10][12].

$$TVL = \begin{bmatrix} TV_1 \\ TV_2 \\ \vdots \\ TV_m \end{bmatrix} := \begin{matrix} 1.\,row \\ 2.\,row \\ \vdots \\ m.\,row \end{matrix} \begin{bmatrix} \overset{x_n \ \cdots \ x_2 x_1}{t_{1n} \cdots \ t_{12} \ t_{11}} \\ t_{2n} \cdots \ t_{22} \ t_{12} \\ \vdots \ \vdots \ \vdots \ \vdots \\ t_{mn} \cdots \ t_{m2} t_{m1} \end{bmatrix} = [t_{mn}] \tag{5}$$

The rule for AND-ing ($\land$) of two TVs ($TV_{i,j}$) which represent an average TV (TV_k), is defined by Table 1:

Table 1. Rule for AND-ing of two TVs

$\land$	0	1	-		if $\times \geq 1$	$\times$: empty set
0	0	$\times$	0			$TV_k = \times$
1	$\times$	1	1		else	
-	0	1	-			$TV_k = TV_i \land TV_j$

The rule for OR-ing ($\lor$) of two TVs ($TV_{i,j}$), which represent the union of both TVs, is defined as follows:

$$TV_i \lor TV_j = \begin{bmatrix} TV_i \\ TV_j \end{bmatrix} \tag{6}$$

The difference-building of two TVs with $TV_M = [t_n, t_{n-1}, .., t_1]_M$ and $TV_S = [t_n, t_{n-1}, .., t_1]_S$ is distinguished out of Equation (3):

$$TV_M \backslash TV_S = [t_n, t_{n-1}, .., t_1]_M \backslash [t_n, t_{n-1}, .., t_1]_S := [t_n, t_{n-1}, .., t_1]_M \land \overline{[t_n, t_{n-1}, .., t_1]_S} =$$

$$= [t_n, t_{n-1}, .., t_1]_M \land \begin{bmatrix} \bar{t}_n & - & \cdots & - \\ - & \overline{t_{n-1}} & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ - & - & \cdots & \bar{t}_1 \end{bmatrix}_S \tag{7}$$

2.3. Orthogonality

A function or TVL is orthogonal if its minterms respectively its TVs are disjoint to one another in pairs at least in one column. Consequently, these minterms ($m_{i,j}$) or these TVs ($TV_{i,j}$) then have no common covering after their logical conjunction. An orthogonal function has no redundant minterm. Thus, the following Equations in (8) can be formulated for the proof of orthogonality [17]:

$$m_i \wedge m_j = 0 \qquad \text{or} \qquad TV_i \wedge TV_j = \times \qquad\qquad (8)$$

3. ORTHOGONALIZING DIFFERENCE-BUILDING ⊖

3.1. Method

The method of orthogonalizing difference-building ⊖ illustrated in the Karnaugh map (Fig. 2) corresponds to the removal of the intersection which is formed between the minuend m_M and the subtrahend m_S, from the minuend m_M, which means $m_M \setminus (m_M \wedge m_S)$. The result consists of several disjoint minterms, which cover all of the remaining 1s and are pairwise orthogonal to each other. For this purpose, an example is shown in a K-map with 4 variables (Fig. 2),in which a group of 2 (subtrahend: $x_3 x_2 x_1$) is subtracted from group of 8 (minuend: x_4). The result consists of several blocks (1st Block, 2nd Block, 3rd Block) which are pairwise orthogonal to each other.
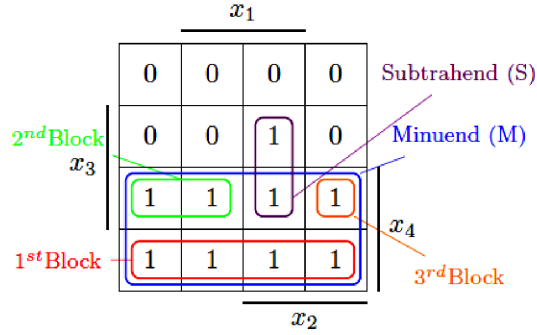
Figure 2: Example of ⊖ in a K-map

$$\underbrace{x_4}_{\text{Minuend}} \quad \ominus \quad \underbrace{x_3 x_2 x_1}_{\text{Subtrahend}} \quad = \quad \underbrace{x_4 \overline{x_3}}_{\text{1st Block}} \ \vee \ \underbrace{x_4 x_3 \overline{x_2}}_{\text{2nd Block}} \ \vee \ \underbrace{x_4 x_3 x_2 \overline{x_1}}_{\text{3rd Block}}$$
$$\underbrace{\phantom{x_4 \overline{x_3} \ \vee \ x_4 x_3 \overline{x_2} \ \vee \ x_4 x_3 x_2 \overline{x_1}}}_{\text{Difference}}$$

- The first literal of the subtrahend, here x_3, is taken complement and AND-ing to the minterm of the minuend, here x_4. Consequently, the first block of the difference is $x_4 \overline{x_3}$.
- Then the second literal, here x_2, is taken complement and AND-ing to the minterm of the minuend and to the first literal x_3 of the subtrahend. Therefore, the second block is $x_4 x_3 \overline{x_2}$.
- Following the next literal, hier x_1, is taken complement and AND-ing to the minterm of the minuend and to the first literal x_3 and second literal x_2 of the subtrahend. Thus, the third block of the difference is $x_4 x_3 x_2 \overline{x_1}$.
- This process is continued until all literals of the subtrahend are singly complemented and linked by AND-ing to the minuend in a separate minterm.

The Equation (9) is applied to calculate the orthogonalizing difference-building of two minterms. In this case, the formula does not show the value of the individual literals, but it shows whether the associated literal exists complement or not. The indices declare only the order of literals which have to be calculated. In this case, the formula $(\bigvee_{i=1}^{n_j} \overline{x_i} = \overline{x_1} \vee x_1 \overline{x_2} \vee .. \vee x_1 x_2 \cdots \overline{x_n})$ formed from [18] is applied to describe the orthogonalizing difference-building in a mathematically easier way:

$$m_M \ominus m_S = \left(\bigwedge_{m=0}^{n-1} x_{n-m} \right) \ominus \left(\bigwedge_{s=0}^{n-1} x_{n-s} \right) := \left(\bigwedge_{m=0}^{n-1} x_{n-m} \right) \wedge \left(\bigvee_{s=0}^{n_j-1} \bar{x}_{n-s} \right) =$$

$$= (x_n \cdot x_{n-1} \cdots x_1)_M \wedge (\bar{x}_n \vee x_n \bar{x}_{n-1} \vee .. \vee x_n \cdot x_{n-1} \cdots x_1)_S \qquad (9)$$

The result may differ depending on the starting literal. There are many equivalent options. They only differ in the form of coverage. The minterms of the difference are pairwise disjoint to each other, so that the result is already availabe in an orthogonal form. The the result corresponds to the minuend if the subtrahend is already orthogonal to the minuend ($m_M \perp m_S$):

$$\textbf{if}: m_S \nsubseteq m_M \textbf{then}: m_M \ominus m_S = m_M \qquad (10)$$

The number of the minterms (blocks) in the result called n corresponds to the number of the variables presented in the subtrahend and are not presented in the minuend at the same time. The number of the possible results can be defined by $n!$ for $n > 0$. The orthogonalizing difference-building of two TVs (TV_M, TV_S), is defined by a corresponding Equation (11):

$$TV_M \ominus TV_S = [t_n, t_{n-1}, .., t_1]_M \ominus [t_n, t_{n-1}, .., t_1]_S :=$$

$$= [t_n, t_{n-1}, .., t_1]_M \wedge \begin{bmatrix} \bar{t}_n & - & \cdots & - \\ t_n & \overline{t_{n-1}} & \cdots & - \\ \vdots & \vdots & \vdots & \vdots \\ t_n & t_n & \cdots & \bar{t}_1 \end{bmatrix}_S \qquad (11)$$

By the use of this new method, two calculation procedures - building the difference and the subsequent orthogonalization - can be performed in one step. That means that the orthogonalizing difference-building $\ominus$represents the composition of the difference-building \and the subsequent orthogonalization *Orth* [6].

3.2. Analysis

3.2.1. Mathematical

The general validity is proved by mathematical induction:

<u>Basis: n = 1</u>

$$\left(\bigwedge_{m=0}^{1-1} x_{1-m} \right) \wedge \left(\bigvee_{s=0}^{1-1} \bar{x}_{1-s} \right) = (x_1)_M \wedge (\bar{x}_1)_S$$

$$(x_1)_M \wedge (\bar{x}_1)_S = (x_1)_M \wedge (\bar{x}_1)_S$$

<u>Statement is true:</u>

<u>Inductive step: n = n + 1</u>

$$\left(\bigwedge_{m=0}^{n} x_{(n+1)-m} \right) \wedge \left(\bigvee_{s=0}^{n_j} \bar{x}_{(n+1)-s} \right) = \left(x_{(n+1)} x_n x_{n-1} .. x_1 \right)_M \wedge \left(\bar{x}_{(n+1)} \vee x_{(n+1)} (\bar{x}_n \vee x_n \bar{x}_{n-1} \vee .. \vee x_n \cdots x_1) \right)_S$$

$$\left(x_{(n+1)} \bigwedge_{m=0}^{n-1} x_{n-m} \right) \wedge \left(\bar{x}_{(n+1)} \vee x_{(n+1)} \bigvee_{s=0}^{n_j-1} \bar{x}_{n-s} \right) = \left(x_{(n+1)} \bigwedge_{m=0}^{n-1} x_{n-m} \right) \wedge \left(\bar{x}_{(n+1)} \vee x_{(n+1)} \bigvee_{s=0}^{n_j-1} \bar{x}_{n-s} \right)$$

The new Equation (3) is equated with the usual Equation (9) to show the equivalence. Since the term of the minuend ($\bigwedge_{m=0}^{n-1} x_{n-m}$) is equal on both sides, it can be neglected:

$$\left(\bigvee_{s=0}^{n-1} \bar{x}_{n-s}\right) = \left(\bigvee_{s=0}^{n_j-1} \bar{x}_{n-s}\right)$$

$$\bar{x}_n \vee \bar{x}_{n-1} \vee ..\vee \bar{x}_1 = \bar{x}_n \vee x_n\bar{x}_{n-1} \vee ..\vee x_n x_{n-1} \cdots x_2\bar{x}_1$$

Due to the axiom of absorption with $\bar{x}_\iota \vee x_i\bar{x}_j = \bar{x}_\iota \vee \bar{x}_j$, the equivalence between two methods is confirmed. They only differ in their form of coverage. The right side is the orthogonal form of the left side. This proves that the orthogonalizing difference-building supplies equivalent results as the difference-building derived out of the set theory.

3.2.2. Computing Time

A study between usual and new method in computing time depending on the length (dimension $dim[i]$) of the minuend and subtrahend indicates almost identical computing times. The new method ($orth\_Diff$) has a slightly longer computing time with increasing dimension than the method of difference-building ($Diff$). The comparison of the average values of both methods shows a maximum difference of just 1.2μs, which can be explained by the consideration complexity. The supplementation instruction per loop iteration results in a multiplicative constant which explains the minimal difference. Against it, the results are orthogonal. The comparison in Figure 3 illustrates that the method $orth\_Diff$ has faster computing time with increasing dimension $dim[i]$ as the composition of method $Diff$ and the subsequent method of orthogonalizing $Orth$ [19].
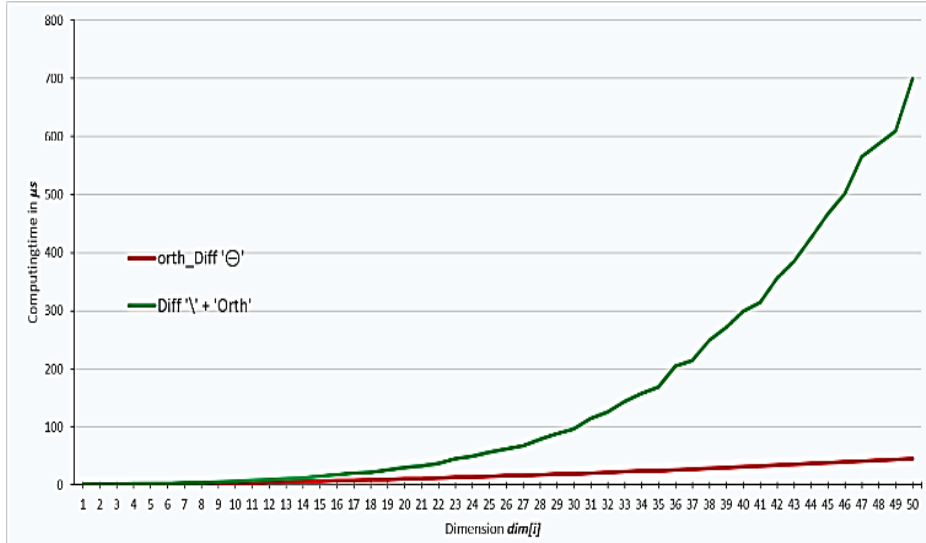
Figure 3: Comparison of computing time of $orth\_Diff\ominus$ and composition of $Diff$ and $Orth$

In Diagram (Fig. 4) the explanation for the higher computing time of the composition is provided. With increasing dimension the computing time of the method $Orth$ increases, whereas the method $Diff$ decreases with respect to the composition. In dimension $dim[14]$ both functions have the same percentage value of the whole computing time. Thereafter, the percentage value of $Diff$ runs low and at the same time the percentage value of $Orth$ increases.
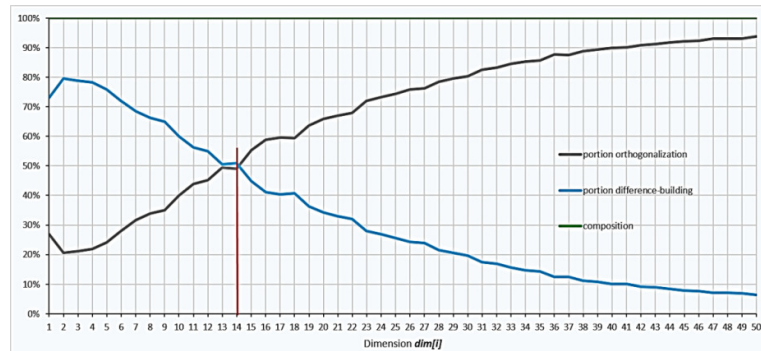
Figure 4: Percentage value of whole computing of *Diff* and *Orth*

At the dimension $dim[50]$ the share of computing time of *Orth* is already 93.7 % and the percentage value of *Diff* is 6.7 %. Finally, the measurement of computing time demonstrates that the new method *orth_Diff* brings a significant advantage due to the orthogonal results [19].

3.2.3. Complexity

The complexity of both methods is analyzed by the evaluation of their implemented functions. The dimension of the inserted minterms respectively ternary-vectors is determined as the input size j. Figure 5 shows the pseudocode of *orth_Diff*. For the function *orth_Diff* is obtained according to the rules of the O-calculus many complexities of $O(1)$ for instructions and accordingly comparisons and two interleaved nested complexities of $O(1)$. As the result, the total complexity is $O(j^2)$ [19]. The complexity of the function *Diff* is $O(j)$ because the inner loops are not required and may be replaced by three linear operations with the complexity of $O(1)$ (Fig. 6).

```
(1)     function orth_Diff(Minuend, Subtrahend) begin
(2)         dim = Subtrahend.Length
(3)         ret = TVL()
(4)         for j = 0 to dim step 1
(5)             if Subtrahend[j] == 2 then next
(6)             else if Subtrahend[j] == Minuend.Negieren(j) then
(7)                 ret.leeren()
(8)                 ret.add(Minuend)
(9)                 return ret
(10)            else if Subtrahend[j] == Minuend[j] then next
(11)            else
(12)                tmp = TV()
(13)                for z = 0 to j step 1
(14)                    tmp[z] = Verunden(Minuend[z], Subtrahend[z])
(15)                next
(16)                tmp[z] = Subtrahend.Negieren(j)
(17)                for z = j+1 to Subtrahend.Length step 1
(18)                    tmp[z] = Minuend[z]
(19)                next
(20)                ret.add(tmp)
(21)            end if
(22)        next
(23)        return ret
```

Figure 5: Pseudocode of *orth_Diff* with complexity

However, the additive and multiplicative constants have no influence on the complexity and are omitted for this reason. Although the total complexity of the usual method is minor but it increases in total by an additional code for orthogonalization.

```
(1)      function Diff(Minuend, Subtrahend) begin
(2)         dim = Subtrahend.Length
(3)         ret = TVL()
(4)         for j = 0 to dim step 1
(5)            if Subtrahend[j] == 2 then next
(6)            else if Subtrahend[j] == Minuend.Negieren(j) then
(7)               ret.leeren()
(8)               ret.add(Minuend)
(9)               exit                                            O(j)
(10)           else if Subtrahend[j] == Minuend[j] then next
(11)           else
(12)              tmp = Minuend
(13)              tmp[j] = Subtrahend[j]
(14)              ret.add(tmp)
(15)           end if
(16)        next
(17)        return ret
```

Figure 6: Pseudocode of *Diff* with complexity

3.2.4. Data Memory Request

First, the memory request of the data structures TV and TVL is determined for this comparison [19]. A TV consists of an array of $m \cdot$ byte elements. Each byte represents an element of the ternary variables. This means that the memory request of the data structures depends on the dimension of the ternary variables. In *C#* the data type *Byte* consists of 8Bits. For a 64 bit system a minimum memory request $Sb_{TV}(m)$ for the data structure of TV with an addressing pointer is calculated by:

$$Sb_{TV}(m) = \underbrace{m \cdot 8\ Bits}_{m \cdot Byte-Elements} + \underbrace{32\ Bits}_{Pointer\ at\ TV} \tag{12}$$

Since a TVL consists of a list of TVs. The memory request of a TVL depends on the number of ternary variables and also the number of the consisting TVs called k. The easiest way to implement a list is the use of a linked list. Each entry in the list has a pointer at its follower. Thus, a minimal memory request $Sb_{TVL}(k, m)$ is calculated by:

$$Sb_{TVL}(k, m) = k \cdot Sb_{TV}(m) + \underbrace{32\ Bits}_{Pointer\ at\ TVL} \tag{13}$$

For this reason, the minimal total memory request Sb_{tot} for the operation of two TVs, the difference-building and orthogonalizing difference-building in this case is:

$$Sb_{tot} = \underbrace{2 \cdot Sb_{TV}(m)}_{TV_M \& TV_S} + \underbrace{Sb_{TVL}(k, m)}_{Result-TVL} \tag{14}$$

The theoretical memory request has to be calculated equally for both function. Because of that, they have the same minimal memory request. Addional memory request for a function for orthogonalization is not needed, because an orthogonal TVL of the orthogonalizing difference is already provided. That shows that the new method, which is the composition of two functions has the advantage of reducing the memory request in addition to faster computing time. The minimal theoretical memory request depending on the dimension $dim[i]$ and the number of n which primarily affects the memory usage is illustrated in Figure 7. Thereby, $dim[i]$ is varied at a constant n and afterwards in the reverse case and analyzed on memory request. For constant $dim[i]$ and changing n the memory usage is higher than in the reverse case. It applies $\Delta Sb_{tot}(n) = \Delta Sb_{tot}(dim[i])$.
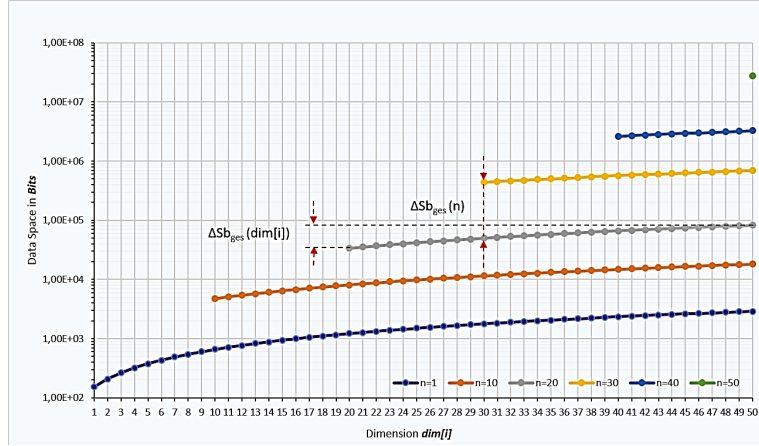
Figure 7: Minimal theoretical memory request

3.3. Applications of Orthogonalizing Difference-Building

3.3.1. Orthogonalizing Difference-Building of two Functions

By Equation (14) (presented in TVL-arithmetic) based on the orthogonalizing difference-building it is possible to calculate an orthogonal difference of two functions respectively two TVLs. One of them is the minuend function f_M^{orth} which has to be orthogonal and the other one is the subtrahend function f_S. The minuend is subducted the common set of subtrahend and minuend; the result is represented in an orthogonal form. The associated minterms (TVs) have the corresponding index-notation, which means M stands for the minuend and S stands for the subtrahend:

$$D\left(f_M^{orth}\right) \ominus D(f_S) = \begin{bmatrix} TV_{1M} \\ TV_{2M} \\ \vdots \\ TV_{mM} \end{bmatrix} \ominus \begin{bmatrix} TV_{1S} \\ TV_{2S} \\ \vdots \\ TV_{mS} \end{bmatrix} :=$$

$$= \begin{bmatrix} (TV_{1M} \ominus TV_{1S}) \wedge (TV_{1M} \ominus TV_{2S}) \wedge ..\wedge (TV_{1M} \ominus TV_{mS}) \\ (TV_{2M} \ominus TV_{1S}) \wedge (TV_{2M} \ominus TV_{2S}) \wedge ..\wedge (TV_{2M} \ominus TV_{mS}) \\ \vdots \\ (TV_{mM} \ominus TV_{1S}) \wedge (TV_{mM} \ominus TV_{2S}) \wedge ..\wedge (TV_{mM} \ominus TV_{mS}) \end{bmatrix} \quad (14)$$

In this case it is important to stress that any outcome of each individual $\ominus$-linkings has to be considered. That means, if the combination of $(TV_{1M} \ominus TV_{1S}) = 0$ arises for example, this will complete the appropriate row to 0 because of: $x_i \wedge 0 = 0$. Since the orthogonalizing difference-building has already been proved in general validity, there is no need proof for generel validity in this case, because all the single links are generally valid. Therefore, out of logical conclusion, Equation (14) is generally valid. If two functions are disjoint to each other $(f_M \perp f_S)$ then a difference cannot be formed and it follows:

$$\textbf{if}: f_S \nsubseteq f_M \textbf{then}: f_M \ominus f_S = f_M \quad (15)$$

3.3.2. Orthogonal Complement of a Function

A further application is the building of an orthogonal complement of a function of the disjunctive normal form. In the set theory the set difference of universal set G and a set A corresponds to set $\bar{A}$ which is the complement of the given set A:

$$G \setminus A = \bar{A} \quad \rightarrow \quad G \cap \bar{A} = \bar{A} \tag{16}$$

Transferred to the switching (Boolean) algebra the complement of a function $\bar{f}(\underline{x})$ can be determined from the difference of a unit function $f(\underline{x}) = 1$ and the function $f(\underline{x})$:

$$f(1) \setminus f(\underline{x}) = \bar{f}(\underline{x}) \quad \rightarrow \quad f(1) \wedge \bar{f}(\underline{x}) = \bar{f}(\underline{x}) \tag{17}$$

As both methods are equivalent, Equation (17) can be formulated with $\ominus$ to find out an orthogonal complement:

$$f(1) \ominus f(\underline{x}) = \bar{f}(\underline{x})^{orth} \text{ or } \quad D(1) \ominus D(f) = D(\bar{f}^{orth}) \tag{18}$$

3.3.3. Orthogonal EXOR of two Functions

The EXOR-operation of two minterms $m_{i,j}$ and also of two functions $f_{i,j}$ can be calculated as:

$$\begin{aligned} f_i \oplus f_j &= f_i\bar{f}_j \vee \bar{f}_if_j = \\ &= (f_i \vee f_j) \wedge (\bar{f}_i \vee \bar{f}_j) = \\ &= (f_i \vee f_j) \wedge \overline{(f_i \wedge f_j)} = \\ &= (f_i \vee f_j) \setminus (f_i \wedge f_j) = \end{aligned} \tag{19}$$

It may be formulated with $\ominus$ due to the equivalence to get orthogonal result:

$$f_i \oplus f_j = (f_i \vee f_j) \ominus (f_i \wedge f_j) \tag{20}$$

By AND-ing and OR-ing of the functions $f_{i,j}$ an inherent relation between these two functions is constructed. Thus, the minuend function does not need to be orthogonal. The advantage is the orthogonal result again. By using $\ominus$ the difficulties which arise with the building of the complement of a function are circumvented. The form of the function changes by building the complement and the transformation back to its original form requires a more sophisticated calculation.

3.3.4. Orthogonal EXNOR of two Functions

The EXNOR-operation of two minterms $m_{i,j}$ or two function $f_{i,j}$ is basically the complement of the EXOR-operation of the same minterms or functions. The EXNOR can be expressed by the complement of EXOR, which also can be formulated by using $\ominus$. Accordingly, the result is orthogonal:

$$f_i \odot f_j = \overline{f_i \oplus f_j} = f(1) \ominus (f_i \oplus f_j) = f(1) \ominus [(f_i \vee f_j) \ominus (f_i \wedge f_j)] \tag{21}$$

3.3.5. Orthogonalizing OR-ing

Another method, the orthogonalizing OR-ing, which is based on the orthogonalizing difference-building, is formed in the following. The orthogonalizing OR-ing Ⓥ is a variant of building the disjunction of two summand-minterms (S_1, S_2) whereby the result is orthogonal. Orthogonalizing OR-ing is going to be explained by an example in a K-map with 4 variables (Fig. 8). Two summand-minterms (a group of 8 and a group of 4) are orthogonalizing OR-ed and a result consisting of several blocks appears; the several blocks are pairwise orthogonal to each other.
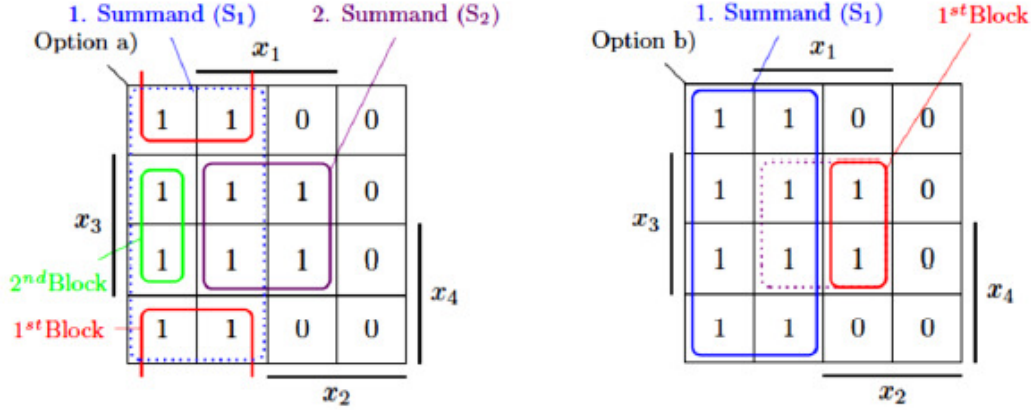
Figure 8: Orthogonalizing OR-ing in K-map

Option a):

$$\bar{x}_2 \; Ⓥ \; x_3 x_1 = \underbrace{\bar{x}_3 \bar{x}_2 \lor x_3 \bar{x}_2 \bar{x}_1 \lor x_3 x_1}_{\text{Sum of Blocks}}$$

Option b):

$$x_3 x_1 \; Ⓥ \; \bar{x}_2 = \underbrace{x_3 x_2 x_1 \lor \bar{x}_2}_{\text{Sum of Blocks}}$$

The idea out of the K-map is noted as propositional logic which a Boolean form is going to be derived. By the use of orthogonalizing OR-ing, the intersection set of first and second summand-minterm (S_1, S_2) is removed from the first summand-minterm S_1 and the second summand-minterm S_2 is linked by a disjunction to that subtraction:

$$S_1 Ⓥ S_2 = [S_1 \setminus (S_1 \land S_2)] \lor S_2 = [S_1 \land \overline{(S_1 \land S_2)}] \lor S_2 =$$
$$= [S_1 \land (\overline{S_1} \lor \overline{S_2})] \lor S_2 = S_1 \overline{S_2} \lor S_2 \qquad (22)$$

As $S_1 \setminus (S_1 \land S_2) = S_1 \setminus S_2$ applies, ⊖ is substituted. Consequently, the orthogonalizing OR-ing Ⓥ can also be expressed by the orthogonalizing difference-building ⊖:

$$S_1 Ⓥ S_2 = [S_1 ⊖ S_2] \lor S_2 \qquad (23)$$

The Boolean form of orthogonalizing OR-ing of two minterms $m_{S_1} = \bigwedge_{S_1=0}^{n-1} x_{n-S_1}$ and $m_{S_1} = \bigwedge_{S_2=0}^{n-1} x_{n-S_2}$ is defined by Equation (24), which does not give the value of separate literals, but it shows whether the value of the variable is complement or not. The indices only indicate the order of the variables to be calculated:

$$m_{S_1} \ \textcircled{V} \ m_{S_2} = \left(\bigwedge_{S_1=0}^{n-1} x_{n-S_1} \right) \ \textcircled{V} \ \left(\bigwedge_{S_2=0}^{n-1} x_{n-S_2} \right) :=$$

$$= \left[\left(\bigwedge_{S_1=0}^{n-1} x_{n-S_1} \right) \wedge \left(\bigvee_{S_2=0}^{n_j-1} \bar{x}_{n-S_2} \right) \right] \vee \left(\bigwedge_{S_2=0}^{n-1} x_{n-S_2} \right) = \tag{24}$$

By swapping the two summands in their position, the result changes. Both solutions, however, are equivalent because the same set is covered. They only differ in the form of coverage which also be seen in Figure 8, in which both possible solutions are presented. But in order to represent the orthogonal result with a minimum of minterms to work in the TVL-representation with low memory request, the summand-minterm with more literals has to be accepted as the first summand-minterm,because the commutativity applies for $\textcircled{V}$ and also for $\ominus$. The following Equation (25) is used for the orthogonalizing OR-ing of two ternary vectors $TV_{S_1} = [t_n, t_{n-1}, .., t_1]_{S_1}$ and $TV_{S_2} = [t_n, t_{n-1}, .., t_1]_{S_2}$:

$$TV_{S_1} \ \textcircled{V} \ TV_{S_2} = [t_n, t_{n-1}, .., t_1]_{S_1} \ \textcircled{V} \ [t_n, t_{n-1}, .., t_1]_{S_2} =$$

$$= \left[[t_n, t_{n-1}, .., t_1]_{S_1} \wedge \begin{bmatrix} \bar{t}_n & - & \cdots & - \\ t_n & \bar{t}_{n-1} & \cdots & - \\ \vdots & \vdots & \vdots & \vdots \\ t_n & t_n & \cdots & \bar{t}_1 \end{bmatrix}_{S_2} \right] \vee [t_n, t_{n-1}, .., t_1]_{S_2} \tag{25}$$

The number of the minterms in the result called n corresponds to the number of the variables presented in the second summand m_{S_2} and are not presented in the first summand m_{S_1} at the same time; plus*1* for the second summand m_{S_2} as the last linked minterm. The number of the possible results can be charged by $n!$ for $n > 0$. Depending on the starting literal the result may differ. There are many equivalent options which only differ only in the form of coverage. If both minterms are disjoint (orthogonal) to each other, the result corresponds to the disjunction of both minterms:

$$\textbf{if: } m_{S_1} \not\subseteq m_{S_2} \qquad \textbf{then: } m_{S_1} \textcircled{V} m_{S_2} = m_{S_1} \vee m_{S_2} \tag{26}$$

By the use of orthogonalizing OR-ing $\textcircled{V}$, two calculation procedures - OR-ing and subsequent orthogonalizing - can be performed in one step. That means that the orthogonalizing OR-ing $\textcircled{V}$ is the composition of OR-ing (V) and the subsequent orthogonalization Orth.

It is not necessary to prove this method for general validity because it includes the already general method of orthogonalization difference-building. The Equations (23) and (2) are equalized to indicate the equivalence of orthogonalization OR-ing and usual OR-ing:

$$S_1 \bar{S_2} \vee S_2 = S_1 \vee S_2 \tag{27}$$

Due to the axiom of the absorption the equivalence is verified. The right side is the orthogonal form of the left side which means they only differ in the form of coverage. So, the results of both sides are equal.

4. CONCLUSION

This work shows that the method of the orthogonalizing difference-building is generally valid and is also equivalent to the usual method of difference-building. In contrast to the composition, orthogonalizing difference-building has faster computing time with increasing dimension. In addition, the method does not require additional memory request for an additional function for orthogonalization because this method already provides orthogonal results. The orthogonalizing difference-building is used to calculate the orthogonal difference of two minterms respectively two TVs, two functions or two TVLs. It is also employed to determine the complement of a function as well as the EXOR and EXNOR of two functions to achieve an orthogonal result. Another method, the orthogonalizing OR-ing of two minterms or TVs, is developed out of the orthogonalizing difference-building. The application of ternary-vector-list is amplified by these new methods to implement simple and quick elementary functions. Due to the inner orthogonalization further processing steps in the TVL arithmetic are considerably simplified because the orthogonal form of disjunctive normal form has the advantage to consider it as antivalence normal form. Thus, additional calculation such as the differential calculus are remarkably facilitated.

REFERENCES

[1]   Zander, H. J.: Logischer Entwurf binärer Systeme. 3. bearb. Auflage. Berlin, Germany: Verl.Technik, 1989, ISBN 3-341-00526-9.

[2]   Bronstein, I.N.; Semendjajew, K.A.; Musiol, G.; Mühlig, H.:Taschenbuch der Mathematik. 7. vollständig überarbeitete und ergänzte Auflage. Frankfurt am Main, Germany: wissenschaftlicher Verl. Harri Deutsch GmbH, 2008, ISBN 978-3-8171-2007-9.

[3]   Popula, L.: Mathematik für Ingenieure und Naturwissenschaften, Band 1. 13. durchgelesene Auflage. Wiesbaden, Germany: Viewer + Teubner Verlag | Springer Fachmedien. Wiesbaden GmbH, 2011, ISBN 978-3-8348-1749-5.

[4]   Matthes, W.: Datenzugriffsprinzipien in objektorientierten Rechnerarchitekturen. Preprint. Technische Universität Karl-Marx-Stadt (Chemnitz), 1989.

[5]   Matthes, W.: Spezielle Hardware zur Verarbeitung von Ternärvektorlisten. Dissertation. Technische Universität Karl-Marx-Stadt (Chemnitz), 1987.

[6]   Posthoff, Ch.; Steinbach, B.: Binäre Gleichungen - Algorithmen und Programme.wissenschaftliche Schriftreiche. Technische Universität Karl-Marx-Stadt (Chemnitz), 1979.

[7]   Posthoff, Ch.; Steinbach, B.: Binäre dynamische Systeme - Algorithmen und Programme. wissenschaftliche Schriftenreihe. Technischen Hochschule Karl-Marx-Stadt (Chemnitz), 1979.

[8]   Posthoff, Ch.; Steinbach, B.: Binäre dynamische Systeme. Berlin: Oldenbourg R. Verlag GmbH, 1981, ISBN 348625071X.

[9]   Bochmann, D.; Zakrevskij, A.D.; Posthoff, Ch.: Boolesche Gleichungen. Theorie - Anwendungen - Algorithmen.Berlin: VEB Verlag Technik, 1984, ISBN 3211958150.

[10] Kühnrich, M.: Ternärvektorlisten und deren Anwendung auf binäre Schaltnetzwerke. Dissertation. Technische Hochschule Karl-Marx-Stadt (Chemnitz), 1979.

[11] Posthoff, C.; Bochmann, D.; Haubold, K.: Diskrete Mathematik. 1. Auflage. Leipzig, DDR: BSB Teubner, 1986, ISSN 0465-3769.

[12] Kempe, G.: Tupel von TVL als Datenstruktur für Boolesche Funktionen. Dissertation (A). Technische Universität Bergakademie Freiberg, 2003.

[13] Whitesitt, J.E.: Boolesche Algebra und Ihre Anwendungen. Band 3. Braunschweig, Germany: Friedr. Vieweg + Sohn GmbH, 1969.

[14] Pomberger, G.; Dobler, H.: Algorithmen und Datenstrukturen - eine systematische Einführung in die Programmierung. 1. Auflage. München, Germany: Addison-Wesley Verlag, 2008, ISBN-10: 3827372682.

[15] Saake, G.; Sattler, K.-U.: Algorithmen und Datenstrukturen - Eine Einführung mit Java. 3. Auflage. Heidelberg, Germany: dpunkt.verlag GmbH, 2006, ISBN-10: 3898643859.

[16] Wagenknecht, Ch.: Algorithmen und Komplexität. 1. Auflage. Leipzig, Germany: Carl Hanser Verlag GmbH \& Co. KG, 2003, ISBN-10: 3446223142.

[17] Bochmann, D.: Binäre Systeme. Ein BOOLEAN Buch. Hagen, Germany: LiLoLe-Verlag GmbH, 2006, ISBN 3-934447-10-4.

[18] Crama, Y.; Hammer, P.L.: Boolean Functions. Theory, Algorithms, and Applications.New York, USA: Cambridge University Press, 2011, ISBN 978-0-521-84751-3.

[19] Kassim,H.; Can, Y.; Sattler, M.S.:Untersuchung eines neuen Algorithmus zur Berechnung orthogonalisierter Differenz. Bachelor-Thesis, Lehrstuhl für Zuverlässige Schaltungen und Systeme, Friedrich-Alexander-Universität Erlangen Nürnberg, Germany, 2014.

[20] Can, Y.; Fischer, G.: Orthogonalizing Boolean Subtraction of Minterms or Ternary Vektors. International Conference on Computational and Experimental Science and Engineering (ICCESEN2014), 24-29 October, 2014, Antalya-Turkey.

AUTHORS

Yavuz Canwas born in Erlangen, Germany, in 1979. He received his Diploma (Dipl.-Ing.) degree in me-chatronics from Friedrich-Alexander-University Erlan-gen-Nürnberg, Germany, in 2010. He is currently a Research Assistant of Prof. Georg Fischer working toward his Ph.D. degree in the Institute for Electronics Engineering at Friedrich-Alexander University in Erlangen. His research interest include orthogonality of Boolean functions and Ternary-Vector-List.

Georg Fischer was born in Lower Rhine region, Germany, in 1965. He received the Diploma degree in electrical engineering with focus on communications, micro-wave and electro-dynamics from RWTH Aachen University, Aachen, Germany, in 1992, and the Dr.-Ing. degree in electrical engineering from the University of Paderborn, Paderborn, Germany, in 1997. From 1993 to 1996, he was a Research Assistant with the University of Paderborn, where he was involved with adaptive antenna array systems for mobile satellite communications. From 1996 to 2008, he performed research with Bell Laboratories, Lucent (later Alcatel-Lucent), where he focused on the RF and digital architecture of mobile communication basestations for global system for mobile communications (GSM), Universal Mobile Telecommunications System (UMTS), and features for network coverage and capacity enhancements. In 2000, he became a Bell Labs Distinguished Member of Technical Staff (DMTS), and in 2001, a Bell Labs Consulting Member of Technical Staff (CMTS). He was also a Chairman with the European Telecom-munications Standards Institute (ETSI) during the physical layer standardization of the GSM-EDGE system. From 2001 to 2007, he was a Part-Time Lecturer with the University of Erlangen–Nuremberg, Erlangen, Germany, during which time he lectured on basestation RF and digital technology. Since April 2008, he has been a Professor of electronics engineering with the University of Erlangen–Nuremberg. He holds over 50 patents concerning microwave and communications technology. His research interests are in transceiver design, analog/digital partitioning, digital signal processing, con-verters, enhanced amplifier architectures, duplex filters, metamaterial structures, GaN transistor technology and circuit design, and RF microelectromechanical systems (MEMS) with specific emphasis on frequency agile, tunable, and reconfigurable RF systems for software-define radio (SDR) and cognitive radio (CR) applications. His new research interests concentrate on medical electronics like using microwaves for detection of vital parameters. Georg Fischer is a Senior Member of the IEEE Microwave Theory and Techniques Society (MTT-S)/Antennas and Propagation Society (AP-S)/Communications Society (COMSOC)/ Vehicular Technology Society (VTC) and Engineering in Medicine and Biology Society (EMBS). He is a member of VDE-ITG and the European Microwave Association (EUMA). He was the co-chair of the European Conference on Wireless Technology (ECWT) of European Microwave Week Conference (EUMW 2007). For EUMW 2013, Nuremberg, Germany and GeMiC 2015, Nuremberg, Germany he has served as the General Technical Program Committee (TPC) chairman.

AN ANDROID APPLICATION FOR VOLUNTEER BLOOD DONORS

Sultan Turhan

Department of Computer Engineering,
GalatasarayUniversity, İstanbul, TURKEY
`sturhan@gsu.edu.tr`

ABSTRACT

There is an expectation that the blood will always be there when it is really needed. Blood donor volunteers constitute the main supply source in an effective blood supply chain management. They feed blood stocks through their donation. In an emergency situation, if the stocks are insufficient, the only source of blood supply will be the people who come to the health center and donate the blood on a voluntary basis. It is certain that time is a very important component in such situation. For this reason, the health care center should call the nearest available donor in order to ensure to get the service as quickly as possible. A smart phone application is developed to facilitate the identification of the nearest available blood donor volunteer and the communication with him/her in the emergency situations where the blood can't be supplied through the blood banks' stocks. In this paper this application will be presented.

KEYWORDS

Smart phone application, Blood donation information system, optimization, distance calculation, java development platforms

1. INTRODUCTION

Despite all the advances in medicine and technology, an alternative medical way to substitute blood, blood components or blood-derived products have not been found yet [2]. Blood can only be supplied by living donors. Blood transfusion has been responsible for saving millions of lives each year around the world. Yet the quantity and quality of blood pool available for transfusion is still a major concern across the globe, especially in the developing countries [3]

According to Turkish Red Crescent Society (TürkKızılayı)[13], the estimated blood requirement for Turkey is about 2.100.000 unit per year and the organization can fulfill 85% of this requirement thanks to volunteer blood donors. There are three types of blood donors in Turkey: a) voluntary donors, b) replacement donors, and c) professional donors [3]. Most donations are as a result of replacement donations provided by the relatives of patients. In case of operation or treatment, health care center employees asked to patient's relatives to donate blood even they don't have the same blood group. Professional donors are those who donate blood in exchange for money. Unfortunately these two types of blood donors don't help maintaining a reliable stock of blood. Moreover, they do not provide rare blood groups. These points indicate the need and importance of voluntary blood donations [6]. Voluntary donors are non-remunerated donors and donate blood voluntarily without any inducements such as money or any other substitute of

money [10]. Such donations are providing adequate round the clock availability during emergency situations and they are the source of safe blood also. The sustainability of a safe and adequate blood supply is still provided by volunteer blood donors [9].

Almost all over the world, blood recruitment services are conducted in a systematic way. Compared to worldwide applications, Turkey has not been reached yet to a desired level. There are 297 blood centers across Turkey [10]. They are mostly in big cities like İstanbul, Ankara and İzmir. In little cities and provinces, Turkish Red Crescent Society manages the blood recruitment. Unfortunately there isn't any coordination between all these institutions and the blood recruitment and supply management keep still their unsteadiness between shortage and wastage. On the other hand, in an emergency situation or a chronic disease case, when the blood stocks are insufficient to fulfill the requirement, the blood is supplied via public announcement through the traditional or social media, which may not always result in a successful way [12].

In order to maintain continuously blood recruitment, blood banks should implement different campaign for encouraging people to become a registered volunteer donor [5]. Also, they have to facilitate blood donation process with help of new technology in order to increase the volunteer donors' willingness and accessibility [3]. Market research experts define the different generations found in the blood donor population. The common characteristic of a large part of this population is to be digital natives whose smart phones are ubiquitous.

Rh++ [1] is an integrated information system whose aim is to manage the blood donation and blood supply chain. The proposed system is used by the patients and/or relatives of the patients to notify their blood requirements and by the living donors to be aware of these requirements. When the blood request is defined into the system, the system checks the stock of the blood banks in the neighborhood. In this way, the system can declare quickly weather the blood needed can be supplied or not. If the blood needed cannot be supplied from the neighborhood blood banks, the system send request to living donors of the same region, via the mobile application installed in donors' smart phones and interprets the response coming from the donors. If there is any affirmative answer from the living donors, the system informs the related health care provider.
Rh++ seeks to provide services to different kind of user such as healthcare institutions, blood banks, volunteer donors, patients or their relatives. Therefore, its infrastructure is designed as flexible as possible, in order to ensure data exchange easily between the organizations, each of them has different infrastructure architecture. For this reason, Service Oriented Architecture principles [13] are adopted directly in infrastructure architecture as well as process designs. The system consists of a backbone which provides data exchanges between different kind of information systems, a web portal and mobile phone's applications which ensure the user interaction.

2. SMART PHONES' APPLICATION FOR VOLUNTEER DONORS

2.1. Operating System

We chose to use the Android software stack produced by Google. Android is an open source framework designed for mobile devices that packages an operating system, middleware, and key applications [15]. The Android SDK provides libraries needed to interface with the hardware and make/deploy an Android application [16]. Applications are written in Java. Android uses a SQLite database to store persistent data. Unlike dedicated systems, our software is intended to integrate with the phone's existing applications. Our application must share resources with the other application. To make for a pleasant integration, the application runs as inconspicuously as possible while using limited resources. We launch a background service that constantly listens to the web services. Only when the algorithm described in the following section receive a request,

will the service wake up and interrupt the user. By only waking up the activity when a request is defined, we allow applications to run on top of our application while we minimize our memory consumption and user interaction.

2.2. General Description

The application is developed for the smart phones using Android operating system. The main duty of the application is to notify regularly the donor's location to Rh++. The process of being a registered donor of Rh++ consists of following tasks: Volunteer uses Rh++'s web portal interface to be registered as donor into the system. Rh++ is a smart information system, and all users' interventions are minimized. Nevertheless, following the registration, volunteer is called by blood recruitment center's staff and required to pass the routine health checks. If his/her health condition is suitable to be a donor, blood center gives the username and password to the volunteer and s/he downloads the application into his/her smart phones. The following picture illustrates the login screen of the Android application.
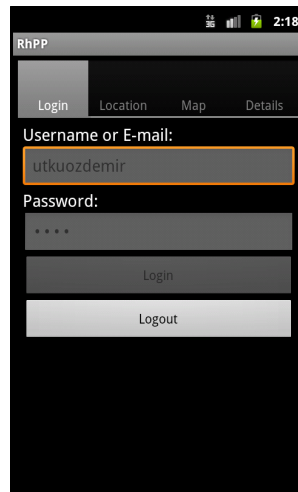
Figure 1.Login Screen of Mobile Application

Once donor connects to the system, the application downloaded into his/her smarts phone starts to send his/her location information to the main system. Two of the common problems in mobile phones' applications are the optimization of battery usage and mobile data exchange. In order to control these two usages, the update frequency of volunteer donor's location information is let under the control of the user. Donor may decrease the update frequency if s/he is idle in a specific location (work, home, theater etc.) or increase it, if s/he is on the move. Also, volunteer donor may update his/her own location information using "Update My Location Now" button on the above mentioned screen whenever s/he wants. Figure 2 illustrates the "Update Frequency" screen.
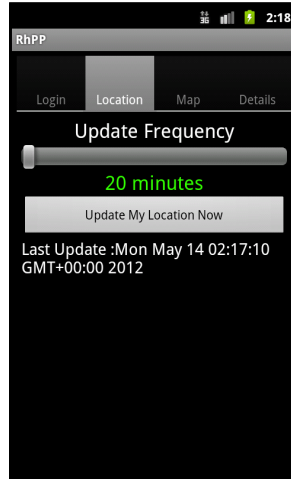
Figure 2.Update Frequency Period Screen

When a blood request comes to the system and if it cannot be supplied from the stocks, the system checks the donors' location, determines the nearest ones and sends them the alerts. Donors receive this alert and respond affirmatively or negatively via their application. Thanks to this rapid communication, the system may first dedicate if the blood can be supplied or not, and secondly calculate the acquisition time of the necessary blood. When the volunteer donor answers affirmatively, the requester health center's information is sent to donor's smart phones

2.3. Technical Details

The application has been initially developed on Android Studio using ANT. Application has two major tasks: a) sending periodically donor's location information to Rh++ and b) receiving alerts from Rh++. To determine the donor's latitude and longitude values, we included maps feature to our applications and we chose to implement Google's maps services. Figure 3 illustrates the map screen.
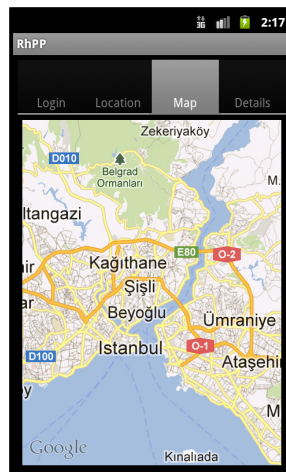
Figure 3. Map Screen

To find the nearest available donor to the requester health center, the system calculates the distance between the requester health center and the donor.   It is Euclidian distance calculation. The data exchange between the application and the main system is provided by the web services.

Two web services are generated for this operation:

- **public String getNearestDonorsToHospital (intuserID, inthospitalID, int radius, intcheckedInLastNHours, String bloodGroup, String bloodType):** This web service is responsible to determine the nearest donors to the requester healthcare center. In order to make this calculation, the web service gets the healthcare center location's information, requested blood information and the radius value determined by the user. With all this input, the web service determine the name, surname and location information of the donors who are suitable, close to the requester healthcare center andupdated his/her location at most checkedInLastNHours time period ago.

- **public StringgetNearestDonors (intuserID, double latitude, double longitude, int radius, intcheckedInLastNHours, String bloodGroup, String bloodRH):** This web service realize the same operation as the above one. The only difference between them is that in this web service the central point is not a registered healthcare center but it is a specific point of which the longitude and latitude value are defined by the users.

It is important to specify that Android cannot support natively the communication established with web services. For this reason, we use a specific open source WSDL library, ksoap2-android, to call web service s from the Android applications.

3. OPTIMIZATION

3.1. Distance Calculation

In the initial system, to find the nearest volunteer donor in order to send the call request, the system calculates the distance as crow flies using both the healthcare centers and the living donors' coordinates. To determine the location of nearest donor, calculating the distance as crow flies was effectual on the paper, but some problems -such as traffic jam, physical conditions, or road conditions- appear when it is put into real life practice. Calculating the distance as crow flies isn't sufficient for the system. So an optimization in distance calculation is decided. Previously, to calculate the distance, requester healthcare coordinates was taken as the coordinates of the circle drawn virtually on the map with the radius of which the length is determined by the user. Then, appropriate ones among the donors whose coordinates are in the area covered by this circle are selected. As it can be easily understand from Figure 4, being in the area covered by the virtual circle does not mean that the donor can reach the hospital in real life. For this reason, an adjustment is made by taking actual road conditions into account and the Google maps functions are inserted into distance calculation. This calculation helps to determine not only nearest volunteer living donor but the most realistic one. It should be noticed that one of the most important constraints in this system is the time. Any optimization or improvement on nearest donor determination will be vital in emergency cases.

3.2. Development Environment

The second improvement on the system is made for Android's users. As everyone agrees, many software and hardware products cannot meet the requirements of the new technologies because of its fast development. Even in the case they are sufficient to meet the requirements, ensuring the

integration between different kind of material and software requires really a hard work and the process takes time.

The initial system has been developed on Android Studio using ANT[18]. It is one of the oldest Java library and command-line tool whose mission is to drive processes described in build files as targets and extension points dependent upon each other.There are several building tools such as ANT, Maven or Ivy. Each of them has its own strengths and weaknesses. ANT has a very flexible structure but it is the developer's job to determine every development step. Also, it doesn't have the library management. For example, if the project needs a library with .jar extension, finding these related .jar files and copying them to relevant directories is entirely under developer's responsibilities. Maven [19] enforces a standardized project layout. This saves time getting new people working on the projects. It provides automatic resolution of dependencies which are cached so that if there are multiple projects, there is no need to keep making local copies. Maven also promotes the concept of a resource repository and can create complex products with minimal direction. Also, the build environment is pretty portable but it has a very standardized layouts. This causes a real problem if the project has an unusual type. It's very easy to make the library management and move the project on Maven in case if the project complies totally with Maven's compelling structure. Gradle [17] is a build automation framework which brings together the strengths of all above mentioned building tools. It offers both ANT's flexibility and library management without Maven's coercion. It easily allows the developer to create .jar and/or .war files using the project's source code, run the test and integrate the add-on or other programs. As it uses Maven's library management, it ensures simple organization of .jar files' part. Another advantage provided by Gradle is the ability of managing multiple projects in a simple way. It has a solid infrastructure to enable to integrate any existing project without structural changes.

Therefore, in order to guarantee a successful result of the improvements made on the project, the system has been carried on integrated development framework Gradle. Besides, as the technology evolves, many of old systems have been removed on Gradle which offers better performance.
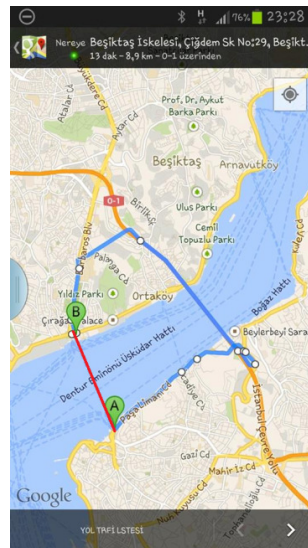

Figure 3. The difference between distance as crow files and real distance

4. CONCLUSIONS

In this study, we presented a smart phone's application for the volunteer blood donor to increase the willingness and accessibility with the purpose of providing a continuous blood supply. This application helps health care centers to provide the blood as quick as possible when their stocks are insufficient. The application sends periodically actual location information of available donors to main system and the blood requests to the donors. In this way, it provides an uninterrupted communication between the health care centers and volunteer donors. The distance of the volunteer donors to the healthcare center is an important criterion in the determination of the donors. Therefore an optimization is also realized on this process. In the initial system, the distance calculation is made by taking the distance as crow flies. In the optimized system, it is converted to the actual distance. This optimization makes the system more realistic. The second improvement is performed on the system's infrastructure. Especially, by taking into consideration the rapid development of mobile device technology which uses Android operating system, the system has been carried from the from ANT building environment onto Grade build automation platform. In further studies, we aim the add evaluation of traffic density between living donors' locations and healthcare centers to the living donor selection criteria

REFERENCES

[1] Turhan, S., Ozdemir, U, Yaşar, A., 2012, Kan Bağışı Ve Temini Bilgi Sistemi: Türkiye İçin Mobil Modelleme, Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu, Trabzon, Turkey, pp.192-198

[2] Dutta, D.A. et al., 2011,Artificial Human Blood , Inventi Impact: Pharm Tech, Vol. 2011,No.1

[3] Gillespie, T. W., & Hillyer, C. D. (2002). Blood donors and factors impacting the blood donation decision. Transfusion Medicine Reviews, 16(2), 115-130.

[4] V. Bosnes, M. Aldrin, H. E. Heier, "Predicting Blood Donor Arrival." , Transfusion, Cilt: 45, Sayı:2, 2005, s. 162-170.

[5] Prcmasudha, B.G., et al., 2010, Application of Spatial Decision Support System to Blood Bank Information Systems, International Journal of Geoinformarics, Vol.6, No. 2, pp.51 – 58.

[6] Sime, S..L. et al., 2005, Strengthening The Service Continuum Between Transfusion Providers and Suppliers: Enhancing the Blood Services Network., Transfusion, Vol.45, No.4, pp.206S-223S.

[7] France, C. R., France, J. L., Wissel, M. E., Kowalsky, J. M., Bolinger, E. M., & Huckins, J. L. (2011). Enhancing blood donation intentions using multimedia donor education materials. Transfusion, 51(8), 1796-1801.

[8] Stanger, S. H., Yates, N., Wilding, R., & Cotton, S. (2012). Blood inventory management: hospital best practice. Transfusion medicine reviews, 26(2), 153-163.

[9] Williamson, L. M., & Devine, D. V. (2013). Challenges in the management of the blood supply. The Lancet, 381(9880), 1866-1875.

[10] Türkiye Kan Merkezleri ve Transfüzyon Derneği, Ulusal Kan ve Kan Ürünleri Rehberi, Haziran 2011, www.kmtd.org.tr

[11] Ş.Hablemitoğlu, Y. Özkan, F. Yıldırım, "Bir Fedakârlık Örneği Olarak "Kan Bağışı" " , Aile ve Toplum, Ocak – Şubat – Mart 2010, s.67 – 77

[12] Nilsson Sojka, B., & Sojka, P. (2003). The blood-donation experience: perceived physical, psychological and social impact of blood donation on the donor. Vox Sanguinis, 84(2), 120-128.

[13] Erl, Thomas. Soa: principles of service design. Vol. 1. Upper Saddle River: Prentice Hall, 2008.

AUTHORS

Sultan N. Turhan received MSc degree in computational science and engineering in 2003 from Istanbul Technical University and her Ph.D degree in Engineering Management from Marmara University. Between 1992 – 1998, she worked as database administrator, IT project coordinator and IT responsible in different institutions. Between 2002 – 2006, she also worked for Intelitek– Element A.S as academic consultant for distance learning and e-learning platforms. Since 1998, she is working as senior lecturer in Computer Engineering department of Galatasaray University. Currently she is working on developing mob ile software in special education area for the children with Autism Spectrum Disorder.

DESIGN AND IMPLEMENTATION OF VARIABLE RADIUS SPHERE DECODING ALGORITHM

Wu Di, Li Dezhi and Wang Zhenyong

School of Electronics and Information Engineering,
Harbin Institute of Technology, Harbin, China
wudi45_2012@163.com, lidezhi@hit.edu.cn, ZYWang@hit.edu.cn

ABSTRACT

Sphere Decoding (SD) algorithm is an implement decoding algorithm based on Zero Forcing (ZF) algorithm in the real number field. The classical SD algorithm is famous for its outstanding Bit Error Rate (BER) performance and decoding strategy. The algorithm gets its maximum likelihood solution by recursive shrinking the searching radius gradually. However, it is too complicated to use the method of shrinking the searching radius in ground communication system. This paper proposed a Variable Radius Sphere Decoding (VR-SD) algorithm based on ZF algorithm in order to simplify the complex searching steps. We prove the advantages of VR-SD algorithm by analyzing from the derivation of mathematical formulas and the simulation of the BER performance between SD and VR-SD algorithm.

KEYWORDS

MIMO, Signal Detecting, Maximum Likelihood Solution, Sphere Decoding, Variable Radius

1. INTRODUCTION

As the communication demand growing constantly, researchers have a stricter requirement for transform rate and BER performance of signal detection. Maximum Likelihood (ML) algorithm has the best BER performance among the MIMO signal detecting algorithms. The ML algorithm searches the feasible signal location in all possible areas, so it is seldom used in practical project. SD algorithm is an optimization algorithm of ML algorithm. The main strategy of SD algorithm is searching the feasible signal location within the super-sphere volume, when the radius is *R*. After finished one circulation, change the numerical value of *R* and do the same thing with the new super-sphere, whose radius is new *R*. Finally find the signal located nearer than any other signals. The classical SD algorithm has two categories. One is called the SD algorithm based on breadth first search, the other is based on depth first search. The results of both search algorithms make no differences.

The main deficiency of classical SD algorithm is that the complexity is related to the original value of the convergence radius. If the original value of *R* is too small, there will be no signal location in the super-sphere area, and the system will have to re-choose the original value of *R* again. If we try to increase *R* in some algorithms, it is necessary to take the step size into consideration. If the original value of R is too large, the system will take all the possible signal location into consideration, which makes the algorithm complexity comparable with ML's. In

addition, if the convergence algorithm of *R* does not suit the situation, the computational complexity will make no difference with ML's. In that way, the SD algorithm may lose its significance of existence. In conclusion, whether the original value of the convergence radius is too large or too small, it will waste a lot of operands before it gets a suitable radius.

Considering all the analysis above, it demonstrates that the original value of the convergence radius has a great influence on the performance of classical SD algorithm. An excellent convergence strategy can greatly decrease the computational complexity of classical SD algorithm. An algorithm is proposed in this paper to calculate the original value of convergence radius to avoid a too large or small original value of the convergence radius. Besides, this paper proposed the VR-SD algorithm to find the target signal location after calculating the original convergence radius.

2. AN IMPROVED STRATEGY OF SD ALGORITHM

In this paper, the Multiple Input Multiple Output (MIMO) system deploys *M* antennas to transmit the signal, and *N* antennas to receive the signal (*M<N*). Firstly, preprocess the channel state information (CSI) matrix *H* to a square matrix. If *M=N*, the CSI matrix *H* will be a square matrix without pre-processing. This paper only discuss the *M=N* system.

In the assumption, the CSI matrix H can be transformed into real number by the functions below:

$$\bar{H} = \begin{bmatrix} \mathrm{Re}(H), -\mathrm{Im}(H) \\ \mathrm{Im}(H), \mathrm{Re}(H) \end{bmatrix}$$

The size of $\bar{H}$ is $2N \times 2N$.

If *X* and *Y* are respectively transmit and receive matrix in complex number field, the transmit and receive matrix in real number field can be transformed in the following functions:

$$\bar{X} = \begin{bmatrix} \mathrm{Re}(X) \\ \mathrm{Im}(X) \end{bmatrix} \quad \bar{Y} = \begin{bmatrix} \mathrm{Re}(Y) \\ \mathrm{Im}(Y) \end{bmatrix}$$

If Z is the noise matrix, and $\bar{Z}$ is the noise matrix in real number field, the equation $\bar{Y} = \bar{H}\bar{X} + \bar{Z}$ still exists. The target of SD algorithm is to find the maximum likelihood solution in the convergence sphere, which is expressed as follows.

$$\hat{x} = \arg\min_{\bar{x} \in \Lambda} \left\| \bar{y} - \bar{H}\bar{x} \right\|^2 = \left(\bar{x} - \hat{\bar{x}} \right)^T \bar{H}^T \bar{H} \left(\bar{x} - \hat{\bar{x}} \right) \leq R_{SD}^2 \tag{1}$$

R_{SD} is the convergence radius. Let the center of the super-sphere be $\hat{\bar{x}}$:

$$\hat{\bar{x}} = (\bar{H}^H \bar{H})^{-1} \bar{H}^H \bar{y} \tag{2}$$

QR decomposes the matrix $\bar{H}$, and the R matrix is an upper triangular matrix in real number field. The equation can be transformed as following:

$$\left(\bar{x}-\hat{x}\right)^T \bar{H}^T \bar{H}\left(\bar{x}-\hat{x}\right)=\left\|R\left(\bar{x}-\hat{x}\right)\right\|^2$$

$$=\left\|\begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} & L & r_{1,2N} \\ 0 & O & L & L & r_{2,2N} \\ 0 & 0 & O & L & r_{3,2N} \\ 0 & 0 & 0 & O & M \\ 0 & 0 & 0 & 0 & r_{2N,2N} \end{bmatrix}g\begin{bmatrix} \bar{x}_1-\hat{x}_1 \\ \bar{x}_2-\hat{x}_2 \\ \bar{x}_3-\hat{x}_3 \\ M \\ \bar{x}_{2N}-\hat{x}_{2N} \end{bmatrix}\right\|^2 \tag{3}$$

$$=\left|r_{2N,2N}(\bar{x}_{2N}-\hat{x}_{2N})\right|^2+\left|r_{2N-1,2N-1}(\bar{x}_{2N-1}-\hat{x}_{2N-1})+r_{2N-1,2N}(\bar{x}_{2N}-\hat{x}_{2N})\right|^2+L \leq R_{SD}^2$$

In the functions below, R_{SD0} stands for the original radius of convergence. In VR-SD algorithm, R_{SD0} is not set by empirical value but a serious of calculation. The function of $f(\bar{x}_{2N},\bar{x}_{2N-1},L,\bar{x}_1)$ is set to be:

$$f(\bar{x}_{2N},\bar{x}_{2N-1},L,\bar{x}_1)=\left|r_{2N,2N}(\bar{x}_{2N}-\hat{x}_{2N})\right|^2+\left|r_{2N-1,2N-1}(\bar{x}_{2N-1}-\hat{x}_{2N-1})+r_{2N-1,2N}(\bar{x}_{2N}-\hat{x}_{2N})\right|^2+L \tag{4}$$

If $\tilde{x}_i=\bar{x}_i-\hat{x}_i$, the function of $f(\bar{x}_{2N},\bar{x}_{2N-1},\cdots,\bar{x}_1)$ can be transformed into:

$$\tilde{f}(\tilde{x}_{2N},\tilde{x}_{2N-1},\cdots,\tilde{x}_1)=\left|r_{2N,2N}\cdot\tilde{x}_{2N}\right|^2+\left|r_{2N-1,2N-1}\cdot\tilde{x}_{2N-1}+r_{2N-1,2N}\cdot\tilde{x}_{2N}\right|^2+\cdots \tag{5}$$

The original radius of convergence is chosen as $R_{SD0}^2=\tilde{f}[\min(\tilde{x}_{2N},\tilde{x}_{2N-1},\cdots,\tilde{x}_1)]$. There are some parameters of matrix R mixed in the function f, so the value of f has little chance to be the minimum value or the maximum value on condition that each $\tilde{x}_i$ is in their own minimum value. In other words, the value of $\tilde{f}[\min(\tilde{x}_{2N},\tilde{x}_{2N-1},\cdots,\tilde{x}_1)]$ varies from the maximum value to the minimum value. In that case, the original value of the convergence radius cannot be set larger than the maximum value of $\tilde{f}[\min(\tilde{x}_{2N},\tilde{x}_{2N-1},\cdots,\tilde{x}_1)]$, or smaller than the minimum value, which can avoid many unnecessary calculations.

The variable radius algorithm will be used after R_{SD0} is calculated. The searching steps start from the bottom. After calculating $\tilde{f}(\tilde{x}_{2N},0,\cdots,0)$ in different $\tilde{x}_{2N}$, compare the results with R_{SD0}^2. Record all the possible values of $\tilde{x}_{2N}$ which satisfied the inequation $\tilde{f}(\tilde{x}_{2N},0,\cdots,0)\leq R_{SD0}^2$. After that, choose one random value of $\tilde{x}_{2N}$ from its all possible values. Then calculate the value $\tilde{f}(\tilde{x}_{2N},\tilde{x}_{2N-1},0,\cdots,0)$ in different $\tilde{x}_{2N-1}$ based on every possible $\tilde{x}_{2N}$, and then compare the values between $\tilde{f}(\tilde{x}_{2N},\tilde{x}_{2N-1},0,\cdots,0)$ and R_{SD0}^2. Record every combination of $\tilde{x}_{2N}$ and $\tilde{x}_{2N-1}$ which satisfied the inequation $\tilde{f}(\tilde{x}_{2N},\tilde{x}_{2N-1},0,\cdots,0)\leq R_{SD0}^2$. Iterate this algorithm until no $\tilde{x}_i$ can be found to satisfy the inequation $\tilde{f}(\tilde{x}_{2N},\tilde{x}_{2N-1},\cdots,\tilde{x}_i,\cdots,0)\leq R_{SD0}^2$. Then step back to the upper layer. When the inequation $\tilde{f}(\tilde{x}_{2N},\tilde{x}_{2N-1},\cdots,\tilde{x}_{i+1},\cdots,0)\leq R_{SD0}^2$ is satisfied, change the value of $\tilde{x}_{i+1}$, and restart the comparison in layer i again. If an i satisfies the inequation $\tilde{f}(\tilde{x}_{2N},\tilde{x}_{2N-1},\cdots,\tilde{x}_i,\cdots,0)\leq R_{SD0}^2$, it means the original convergence radius R_{SD0} needs to be increased so that the inequation can establish. On the contrary, if after one circulation, there are more than one group of $[\tilde{x}_{2N},\tilde{x}_{2N-1},\cdots,\tilde{x}_1]$ satisfy the inequation $\tilde{f}(\tilde{x}_{2N},\tilde{x}_{2N-1},\cdots,\tilde{x}_1)\leq R_{SD}^2$, it means

the R_{SD0} need to be decreased. After decreasing the R_{SD0} to R_{SD}, repeat the comparison circulation till there only exists one group of $[\tilde{x}_{2N}, \tilde{x}_{2N-1}, \cdots, \tilde{x}_1]$ that make the inequation $\tilde{f}(\tilde{x}_{2N}, \tilde{x}_{2N-1}, \cdots, \tilde{x}_1) \leq R_{SD}^2$ established. Then the solution, $[\tilde{x}_{2N}, \tilde{x}_{2N-1}, \cdots, \tilde{x}_1]$, can transform into the real number signal matrix $\hat{\bar{x}} = [\bar{x}_{2N}, \bar{x}_{2N-1}, \cdots, \bar{x}_1]$, and the target signal in complex number area can be expressed as:

$$\hat{x} = [\bar{x}_{2N} + j\bar{x}_N, \bar{x}_{2N-1} + j\bar{x}_{N-1}, \cdots, \bar{x}_N + j\bar{x}_1]_{1 \times N} \tag{6}$$

3. THE EXPERIMENTS OF VR-SD ALGORITHM

This paper compares the performance between classical SD algorithm and VR-SD algorithm in 2×2 MIMO system using QPSK modulation. Knowing from the aforementioned analysis, if the convergence circulation is free from loop times, the performance of classical SD algorithm will be the same as ML algorithm's. The circulation of classical SD algorithm may not keep doing until it finds the final target signal, because the CSI keeps changing during the process of implementation. If the time overhead of signal detection algorithm is too large, its CSI may change before it finishes the calculation. Furthermore, since the signal is being transferred all the time, if the system wastes too much time in signal detection, the subsequent signal will jam together. Considering all the situation above, the times of convergence circulation of classical SD algorithm will be constant during the simulation in this paper. After the times of convergence circulation reach the setting number, the system will choose a signal in the super-sphere of the last circulation as the target signal randomly, and force the algorithm to exit the circulation.

To get the convergence time of different convergence circulations of classical SD algorithm, this paper did the simulation of classical SD algorithm within the limited convergence times of 4, 5, 6 and 7. On the condition of the same hardware and the same version of MATLAB software, compare the convergence time between the limited times classical SD algorithm and the unknown convergence times VR-SD algorithm. The simulation is done in 2×2 MIMO system using QPSK modulation with the 2^{20} code length. As the variance of each kind of simulation is small, we do each kind of simulation ten times and calculate the average time. The results are shown in table 1:

Table 1. Calculation time of VR-SD and classical SD

Algorithm	Calculation time (s)	Convergence times
VR-SD	136.2	N/A
Classical SD	120.7	4
	142.1	5
	148.5	6
	150.3	7

By comparing the convergence time of classical SD algorithm, it is concluded that the calculation time rises with the rising convergence times. The calculating time difference between 4 and 5 times is nearly 22 seconds, but the time difference between 5 and 6 times sharply decreases to just over 6 seconds. The time difference between 6 and 7 times is even less. This shows that with the increasing convergence times, the percent of classical SD algorithm finishes their calculation before reaching the limit convergence times is increasing. From Table 1, the calculating time of VR-SD algorithm is close to the classical SD algorithm when the convergence times is 5. In the following simulation, we compare the BER between VR-SD and the classical SD algorithm with 5 convergence times.
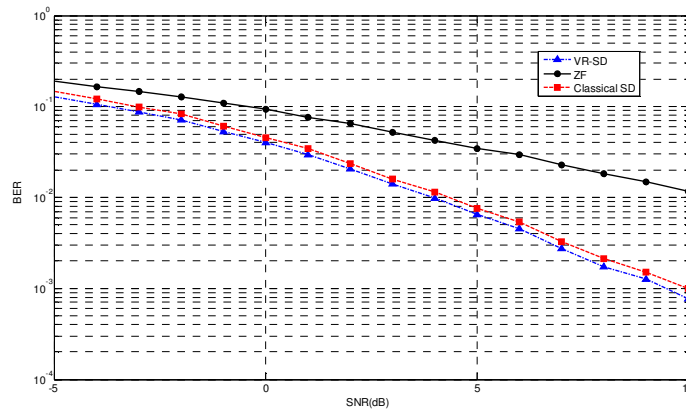
Figure 1. Comparison of BER between VR-SD, classical SD and ZF

Figure 1 shows the BER performance of VR-SD compared with classical SD algorithm and ZF algorithm. The computational complexity of ZF is much lighter than VR-SD algorithm, while the BER of VR-SD and classical SD is much better than ZF's. Because of the limitation of convergence times, the BER of classical SD algorithm doesn't have the same performance as the one which without limitation. Though the limitation of convergence times decreases the computational complexity, it still has some bad influence on its BER. Considering this, it is not strange that the BER of classical SD is not as well as the VR-SD's. We still cannot ignore the fact that if the times of convergence circulation keep rising, the BER of classical SD algorithm will be better than VR-SD.

4. CONCLUSIONS

This paper solve the problem of the algorithm complexity of convergence radius by designing a SD algorithm strategy—VR-SD algorithm. The VR-SD algorithm can avoid the unnecessary computational complexity caused by an unsuitable value of original convergence radius. Subjected to the same BER, both the classical SD and the VR-SD algorithm have a nearly 3 dB gain from the ZF algorithm. While the gain between the classical SD algorithm and the VR-SD algorithm is about 0.5dB. Considering the time overhead of classical SD algorithm, which convergence times is 5, is 4% more than the VR-SD algorithm's. The gain of 0.5dB from classical SD to VR-SD is a really great improvement. It demonstrates that the use of variable radius strategy raises the BER of SD algorithm in a considerable degree.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Vikalo H, Hassibi B & Kailath T, (2004) "Iterative decoding for MIMO channels via modified sphere decoding", IEEE Transactions on Wireless Communications, Vol. 3, No. 6, pp2299-2311.

[2]   Lampe L, Schober R & Pauli V, (2004) "Multiple-symbol differential sphere decoding", IEEE Transactions on Communications, Vol. 53, No. 12, pp1981-1985.

[3]   Ding Y, Wang Y R & Li N, (2014) "Widely linear sphere decoding by exploiting the hidden properties of PSK signals", IEEE Globecom 2014- Signal Processing for Communications Symposium (GC14 SPC).

[4]   Soltani. M.D., Alimadadi. M & Amindavar, H, (2014) "A hybrid method to reduce the complexity of
      k-best sphere decoding algorithm", 2014 22nd Iranian Conference on Electrical Engineering (ICEE),
      pp1765-1770.
[5]   El-Khamy. M, Vikalo. H., Hassibi, B & McEliece, R.J, (2009) "Performance of sphere decoding of
      block codes", IEEE Transactions on Communications, Vol. 57, No. 10, pp2940-2950.
[6]   Viet-Hoa Nguyen, Berder. O. & Scalart, P, (2013) "On the efficiency of sphere decoding for linearly
      precoded MIMO systems", Wireless Communications and Networking Conference (WCNC), 2013
      IEEE, pp4021-4025.
[7]   Li Shi-ping, Wang Long & Chen Fang-chao, (2012) "Ordered sphere decoding detection algorithm
      for MIMO systems", 2012 24th Chinese Control and Decision Conference (CCDC), pp3322-3325.
[8]   Xu Ren & Huahua Wang, (2012) "Research of Sphere Decoding Detection Algorithm in LTE-A
      System", 2012 8th International Conference on Wireless Communications, Networking and Mobile
      Computing (WiCOM), pp1-4.
[9]   Singh. A, Elia. P & Jalden. J, (2013) "Rate-reliability-complexity tradeoff for ML and lattice
      decoding of full-rate codes", 2013 IEEE International Symposium on Information Theory
      Proceedings (ISIT), pp1267-1271.

AUTHORS

Wu Di was born in 1990, and get his B.S degree in Harbin Institute of Technology in
2013.He is now getting his M.S degree in the same school. His studying field is about the
wireless communication and the signal detecting of MIMO.
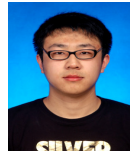
Li Dezhi was born in 1981. He has received his PhD degree in Harbin Institute of
Technology. He is the lecturer in school of electronics and information Engineering. His
main research fields are satellite communication, wireless multimedia transmission and
wireless communication.

Wang Zhenyong received his B.S. and M.S. degrees in communication and information
system from Harbin Institute of Technology in 2000 and 2002, respectively. He is the
lecturer in school of electronics and information Engineering. His main research interests
are cross-layer design, satellite communication, wireless multimedia transmission, and
wireless heterogeneous network

PERFORMANCE OF THE MAXIMUM STABLE CONNECTED DOMINATING SETS IN THE PRESENCE OF STATIC NODES IN A MOBILE ADHOC NETWORK

Natarajan Meghanathan

Jackson State University, 1400 Lynch St, Jackson, MS, USA

ABSTRACT

The topology of mobile ad hoc networks (MANETs) change dynamically with time. Connected dominating sets (CDS) are considered to be an effective topology for network-wide broadcasts in MANETs as only the nodes that are part of the CDS need to broadcast the message and the rest of the nodes merely receive the message. However, with node mobility, a CDS does not exist for the entire duration of the network session and has to be regularly refreshed (CDS transition). In an earlier work, we had proposed a benchmarking algorithm to determine a sequence of CDSs (Maximum Stable CDS) such that the number of transitions is the global minimum. In this research, we study the performance (CDS Lifetime and CDS Node Size) of the Maximum Stable CDS when a certain fraction of the nodes in the network are static and compare the performance with that of the degree-based CDSs. We observe the lifetime of the Maximum Stable CDS to only moderately increase (by a factor of 2.3) as we increase the percentage of the static nodes in the network; on the other hand, the lifetime of the degree-based CDS increases significantly (as large as 13 times) as we increase the percentage of static nodes from 0 to 80.

1. INTRODUCTION

A mobile ad hoc network (MANET) is a dynamic distributed system of wireless nodes that move arbitrarily with time [1]. The wireless nodes operate in a limited transmission range and are battery-charged. Two nodes can communicate directly only if they are within the transmission range of each other. Hence, communication between any two nodes in a MANET is typically through one or more intermediate nodes (multi-hop paths). Several communication protocols have been proposed for unicast [2-3], multicast [4-5] and broadcast [6-7] communication in MANETs. In this paper, we focus on broadcast communication in MANETs. Specifically, our focus is on connected dominating sets (CDS), typically considered the graph-theoretic equivalent for a communication topology that can facilitate network-wide broadcasts. A CDS is a subset of the nodes in the network such that every node in the network is either in the CDS or is a neighbor (i.e., has a wireless link) of a node in the CDS.

Broadcasts are considered to be resource-intensive operations in terms of both energy consumption at the nodes as well as the volume of traffic generated due to redundant retransmissions. If each node in the network broadcasts the message exactly once in its

neighborhood, then every node receives a copy of the message from each of its neighbors (referred to as flooding). Though flooding guarantees that the broadcast message reaches every node in the network, each node loses energy to receive message broadcast by each of its neighbors (in addition to the energy lost at each node to transmit/ broadcast the message once in its neighborhood). It is not required for every node in the network to broadcast the message in its neighborhood. Instead, if we use a CDS like communication topology for broadcast, then it would be sufficient if only the constituent nodes of the CDS broadcast (retransmit) and the rest of the nodes only spend energy to merely receive the message. The lower the number of nodes constituting the CDS (referred to as CDS Node Size), the lower the number of retransmissions. Unfortunately, the problem of determining a minimum node size CDS is NP-hard [8]. Several heuristics have been proposed to determine CDSs with approximation ratios for the CDS Node Size as low as possible. A common thread among all of these heuristics is to give preference to include nodes that have a larger degree as part of the CDS so that the CDS Node Size is as low as possible. Apparently, nodes with a larger degree (larger number of neighbors) were the preferred candidates for inclusion to the CDS.

In [9], we observed that the degree-based heuristics are quite unstable in the presence of node mobility in MANETs. We had then proposed a benchmarking algorithm to determine a sequence of connected dominating sets over the duration of a network session such that the number CDS transitions is the global minimum. Referred to as the Maximum Stable CDS algorithm, the benchmarking algorithm assumes the entire sequence of future topology changes is known a priori and determines a sequence of long-living stable CDSs as follows: Whenever a CDS is required at time instant *t*, we determine a connected graph of the network (called a mobile graph) whose constituent edges exist for the longest possible time starting from time instant *t*; we then simply run a CDS algorithm/heuristic on the mobile graph and use the CDS for the duration of the mobile graph and repeat the above procedure for the entire network session. The algorithm can be used to arrive at a sequence of long-living stable CDSs such that the number of transitions (from one CDS to another) during a network session is the global minimum and the average CDS lifetime is the global maximum (benchmarks).

As in many MANET simulation studies, the performance comparison of the Maximum Stable CDS proposed in [9] vis-a-vis the degree-based CDS was only conducted when all the nodes in the network are mobile and not conducted when a certain percentage of the nodes in the network are static. This formed the motivation for our research in this paper. Our hypothesis is that the stability of the degree-based CDS would significantly increase with increase in the percentage of static nodes in the network (compared to the CDS lifetime incurred when all nodes were mobile) as there are good chances that an appreciable fraction of the static nodes are also part of the degree-based CDS and contribute towards its stability and cover the non-CDS nodes that could be mobile; on the other hand, the lifetime of the Maximum Stable CDS would only marginally increase, as it would be difficult to find a connected mobile graph that exists for a longer time even in the presence of a certain fraction of static nodes.

The rest of the paper is organized as follows: Section 2 presents a heuristic to determine degree-based CDS and evaluates its run-time complexity. Section 3 presents the Maximum Stable CDS algorithm to determine a sequence of stable CDSs, explains its proof of correctness and illustrates its working with an example. Section 4 presents the simulation results evaluating the lifetime and node size of the Maximum Stable CDS vis-a-vis the degree-based CDS under an extensive set of mobility scenarios varying the maximum node velocity and the percentage of static nodes for each level of node mobility. Section 5 differentiates our work from related work. Section 6 concludes the paper. Throughout the paper, the terms 'node' and 'vertex', 'link' and 'edge' are used interchangeably. They mean the same.

2. HEURISTIC FOR DEGREE-BASED CONNECTED DOMINATING SET

In this section, we first describe a heuristic that could be used to determine the CDS for a network graph based on the degree (the number of neighbors of a vertex) in the graph. We then illustrate examples to show the execution of the CDS construction heuristic. We finally describe an algorithm that can be used to validate the existence of a CDS at any time instant.

2.1. Generic Heuristic to Determine a CDS

The overall idea is to give preference (for inclusion to the CDS) for vertices that have a larger degree. As indicated in the pseudo code of Figure 1, we maintain three lists: *CDS Node List* - vertices that are part of the CDS; *Uncovered Node List* - vertices that are yet to be covered by a node in the CDS Node List; *Candidate Node List* - vertices that are covered by a node in the CDS Node List, but not yet considered for inclusion to the CDS Node List. Initially, the CDS Node List is empty and the Uncovered Node List is the set of all the vertices in the graph; to start with, the Candidate Node List has a single entry corresponding to the vertex that has the largest degree. The Candidate Node List is implemented as a priority queue and the vertices in the Candidate Node List are stored in the decreasing order of their degree (the vertex with the largest degree is the first vertex to be removed from this list). In each iteration, we remove a vertex from the Candidate Node List and if it has one or more uncovered neighbor nodes, then the vertex is added to the CDS Node List and the newly covered neighbor nodes are removed from the Uncovered Node List and included in the Candidate Node List. If a vertex removed from the Candidate Node List has no uncovered neighbors, then the vertex is not included in the CDS Node List. We repeat the above procedure until the Uncovered Node List gets empty or there are no more vertices in the Candidate Node List. If the Candidate Node List gets empty while the Uncovered Node List is not yet empty, then it implies the underlying graph is not connected (i.e., all the vertices are not in one component). If the underlying graph is connected, the iterations stop when there are no more vertices in the Uncovered Node List (this implies, all the vertices in the graph are covered by at least one node in the CDS Node List).

Input: Graph G = (*V*, *E*); Degree Score Vector $D(V)$ for all the vertices
Output: *CDS Node List*

Initialization: *CDS Node List* = ϕ; $\forall$ $x \in V$, *Uncovered Node List* = $\{x; D(x)\}$

$\qquad$ *Candidate Node List* = $\{s; D(s)\}$ where $s \in V$ and $D(s) = \underset{i \in V}{Max}\{D(i)\}$

Begin *CDS Construction*

$\qquad$ **while** (*Candidate Node List* $\neq \phi$ AND *Uncovered Node List* $\neq \phi$) **do**
$\qquad\qquad$ Vertex *candidateNode* = **Dequeue**(*Candidate Node List*)
$\qquad\qquad$ *uncoveredNodes* = $\{x \mid x \in$ Neighbors(*candidateNode*) AND
$\qquad\qquad\qquad$ $x \in$ *Uncovered Node List*$\}$
$\qquad\qquad$ **if** (*uncoveredNodes* $\neq \phi$) **then**
$\qquad\qquad\qquad$ *CDS Node List* = *CDS Node List* $\cup$ $\{candidateNode\}$
$\qquad\qquad\qquad$ $\forall$ $x \in$ *uncoveredNodes*,
$\qquad\qquad\qquad\qquad$ *Uncovered Node List* = *Uncovered Node List* - $\{x; D(x)\}$
$\qquad\qquad\qquad\qquad$ *Candidate Node List* = *Candidate Node List* $\cup$ $\{x; D(x)\}$
$\qquad\qquad$ **end if**
$\qquad$ **end while**

$\qquad$ **if** (*Uncovered Node List* = ϕ) **then**
$\qquad\qquad$ **return** *CDS Node List*
$\qquad$ **end if**

End *CDS Construction*

Figure 1. Pseudo Code for a CDS Construction Heuristic based on Node Degree

Note that the degree score for a node is determined offline (prior to the execution of the heuristic) and is not updated during the execution of the heuristic. In other words, the ordering of the nodes in the Candidate Node List in each iteration is based on the initial (static) degree values input to the heuristic. Thus, for example, if two nodes *u* and *v* with degree scores of 4 and 6 respectively are in the Candidate Node List, but node *v* has only two uncovered neighbors whereas node *u* has three uncovered neighbors, node *v* with a relatively larger degree score would still be ahead of node *u* in the Candidate Node List.

The time complexity for the CDS construction heuristic is O(Vlog$V$ + Elog$V$), where V and E are respectively the number of vertices and edges in the network graph; O(log$V$) is the time complexity for a dequeue operation in a priority queue (Candidate Node List) maintained as a Max-Heap [8] as well as the time complexity to include a vertex in the Candidate Node List by visiting the neighbors of the vertex that is added to the CDS Node List. We could run the while loop at most V times, once for each vertex, and across all such iterations, the entire set of edges are traversed and the end vertices of the edges are considered for inclusion to the Candidate Node List. Hence, the overall time complexity of the generic heuristic to construct a degree-based CDS is O(($V+E$) log$V$).

2.2. Example to Construct Degree-based CDS

In this section, we show the execution of the degree heuristic on a sample network graph. For each vertex in Figure 2, the ID is indicated inside the circle and the degree of the vertex is indicated outside the circle. The last graph in each of these figures is a CDS graph comprising of only edges between two CDS nodes (indicated as solid lines) and edges between a non-CDS node and a CDS node (indicated as dotted lines).
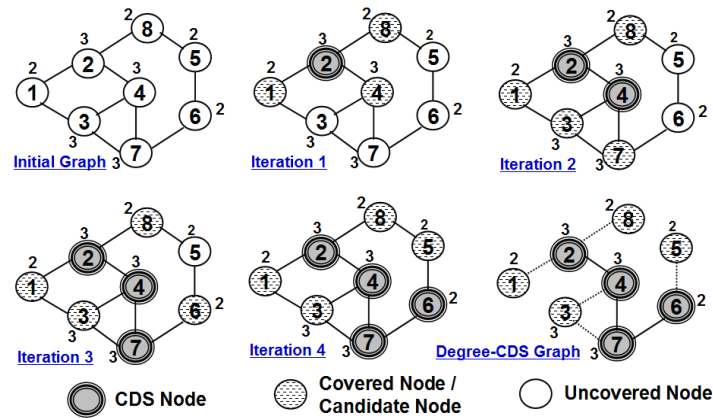
Figure 2. Example to Illustrate the Construction of a Degree-based CDS

Figure 2 illustrates the execution of the heuristic to determine degree-based CDS: in iteration 3, even though vertices 3 and 7 have the same larger degree (3), vertex 3 is not considered for inclusion to the CDS because all of its three neighbors are already covered (i.e., none of the neighbors of vertex 3 are in the Uncovered Node List); on the other hand vertex 7 has one uncovered neighbor and is hence included to the CDS Node List. The sequence of vertices included in the Degree-based CDS are 2, 4, 7 and 6. One significant observation could be made in the toy example shown in Figure 2: Vertex 3 with a relatively higher degree (3) is connected to two high-degree vertices (vertices 2 and 7), but does not contribute much in terms of covering nodes that are yet in the Uncovered Nodes List once vertices 2 and 7 get into the CDS Node List. Thus, vertex 3 is a highly enclosed vertex that is not exposed much to vertices outside its own

community. On the other hand, even if vertex 2 gets into the CDS Node List, vertex 7 could also get into the CDS Node List, because they are connected to two different vertices/communities (vertices 8 and 6) that only either of them could cover, but not both.

2.3. Algorithm to Validate the Existence of a CDS

We now present the algorithm that we use to validate the existence of a CDS at any time instant. This algorithm would be very useful in discrete-event simulations when we tend to use a CDS that existed at an earlier time instant (say, t-1) at a later time instant (t). That is, given a CDS Node List that was connected (i.e., the CDS nodes are reachable from one another directly or through one or more intermediate CDS nodes) and covered the non-CDS nodes in the network graph at time instant t-1, we want to validate whether the same holds true at time instant t. We do so by first constructing a CDS graph based on the CDS Node List at time instant t. All the vertices in the network are part of the CDS graph and the edges in the CDS graph are those that exist between any two CDS nodes as well as between a CDS node and a non-CDS node at time instant t. We check if the CDS Node List is connected at time instant t by running the Breadth First Search (BFS) [8] algorithm, starting from an arbitrarily chosen CDS node and see if every other CDS node could be reached as part of the BFS traversal only on the CDS-CDS edges. If all the CDS nodes could be visited, the CDS Node List is considered to be connected. We then check whether every non-CDS node has an edge to at least one CDS node at time t. If both the validations (connectivity of the CDS Node List and coverage of every non-CDS node by at least one CDS node) are true, then we consider the CDS to exist at time instant t.

3. BENCHMARKING ALGORITHM FOR MAXIMUM STABILITY CDS

In this section, we describe a benchmarking algorithm [9] proposed to determine a sequence of longest-living connected dominating sets for MANETs such that the number of transitions is the global minimum and the average lifetime of the CDS is the maximum. We refer to the algorithm as the Maximum Stable CDS algorithm; it works on the basis of graph intersections and assumes the availability of information about topology changes for the entire duration of the network session. The algorithm introduces the notion of a mobile graph $G(t, ..., t+k) = G(t) \cap G(t+1) \cap G(t+2) \cap ... \cap G(t+k)$, wherein $G(t)$, $G(t+1)$, $G(t+2)$, ..., $G(t+k)$ are static graphs (snapshots) of the network at time instants t, $t+1$, $t+2$, ..., $t+k$ respectively. A mobile graph $G(t, ..., t+k)$ is thus a graph that essentially captures the vertices and edges that exist in the network for all the time instants t, $t+1$, $t+2$, ..., $t+k$. The Maximum Stable CDS algorithm works as follows: Whenever a CDS is required at time instant t, we determine a mobile graph $G(t, ..., t+k)$ such that $G(t, ..., t+k)$ is connected and $G(t, ..., t+k+1)$ is not connected. We determine a CDS on such a longest-living connected mobile graph $G(t, ..., t+k)$ and repeat the above procedure for the duration of the network session to obtain a sequence of longest-living CDSs such that the number of transitions needed to change from one CDS to another is the optimum (minimum). We say the optimum number of transitions incurred is the global minimum because the algorithm assumes the knowledge of the entire topology changes and always chooses the longest-living mobile graph since the time instant starting from which a CDS is needed. Later, we also prove that the number of transitions accomplished by any other CDS construction algorithm cannot be below the value obtained for the Maximum Stable CDS algorithm. Accordingly, the average of the lifetimes of the longest-living connected dominating sets in the sequence determined by the Maximum Stable CDS algorithm for the duration of the network session is the global maximum and can serve as a benchmark to compare the average lifetime incurred with any other sequence of connected dominating sets (like the degree-based CDSs) under identical conditions.

As seen in the pseudo code for the Maximum Stable CDS algorithm (Figure 3), it does not matter what CDS construction heuristic we use to determine a CDS in each of the longest-living mobile

graphs G(i, ..., j), because the CDS would not exist in the mobile graph G(i, ..., j+1) as the latter would not be a connected graph. In this research, we determine the connected dominating sets in the longest-living mobile graph G(i, ..., j) based on the degree of the vertices in the mobile graph G(i, ..., j). The average lifetime of the Maximum Stable CDS would not be affected because of this; however, the size of the Maximum Stable CDS is likely to be affected. As degree-based CDSs are likely to incur a lower CDS Node Size, it would be only appropriate to determine the connected dominating sets in each of the longest-living mobile graphs G(i, ..., j) using the degree of the vertices in the mobile graph so that the Maximum Stable CDS would incur a lower CDS Node Size and at the same time incur the maximum lifetime. The time complexity of the Maximum Stable CDS algorithm depends on the time complexity of the underlying heuristic used to determine the CDSs in each of the mobile graphs. As one can see in the pseudo code (Figure 3), there could be at most T mobile graphs (where T is the number of static graph snapshots, number of graph samples) on which a CDS heuristic needs to be run; if we use a $\Theta(V^2)$ heuristic to determine degree-based CDSs, then the overall time complexity of the Maximum Stable CDS algorithm is $O(V^2 T)$.

Input: Static graphs G$_1$, G$_2$, ..., G$_T$, where T is the duration of the network session
Output: *Maximum Stable CDS*
Auxiliary Variables: *i, j*
Initialization: *Maximum Stable CDS* = ϕ; $i = 1$; $j = 1$

Begin *Maximum Stable CDS Construction*

      while ($i \leq T$) **do**
            while G(i, ..., j) is connected AND $j \leq T$ **do**
                  $j \leftarrow j + 1$
            end while
            if ($i < j$) **then**
                  $j \leftarrow j - 1$
                  Maximum Stable CDS = Maximum Stable CDS $\cup$ {CDS in G(i, ..., j) }
                  $i \leftarrow j + 1$
            end if
            else
                  $i \leftarrow i + 1$
            end else
            $j \leftarrow i$
      end while

    return *Maximum Stable CDS*

    End *Maximum Stable CDS Construction*

Figure 3. Pseudo Code for the Benchmarking Algorithm to Determine Maximum Stable CDS

3.1. Proof of Correctness

We now present an informal proof of correctness of the Maximum Stable CDS algorithm. For a more formal proof, the interested reader is referred to [9]. To prove that the Maximum Stable CDS algorithm finds a sequence of long-living CDSs that incur the minimum number of transitions (say m), assume the contrary - i.e., there exists a hypothetical algorithm that determines a sequence of CDSs that undergo n transitions such that $n < m$. We will now explore whether this is possible. For $n < m$ to exist, there has to be an epoch of time instants [p, ..., s] that is a superset of an epoch of time instants [q, .., r] such that $p < q < r < s$ and that the hypothetical algorithm was able to find a CDS that existed during time instants [p, ..., s], but the Maximum Stable CDS algorithm can only find a CDS that existed during time instants [q, ..., r] and had to

go a transition at time instant $r + 1$. This implies the Maximum Stable CDS algorithm could not find a connected mobile graph $G(q, ..., r+1)$ and could only find a connected mobile graph $G(q, ..., r)$. This means there existed no connected mobile graph from time instants $[q, ..., s]$ and hence there is no connected mobile graph from time instants $[p, ..., s]$. If that is the case, it would not be possible to find a CDS that exists from time instants $[p, ..., s]$; thus, the sequence of connected dominating sets determined by the hypothetical algorithm has to undergo at least as many transitions as those determined by the Maximum Stable CDS algorithm, which is a contradiction to our initial assumption that $n < m$. Hence, the number of transitions (m) incurred by the sequence of long-living CDSs determined by the Maximum Stable CDS algorithm serves as a lower bound for the number of transitions (n) incurred by any other algorithm/heuristic to determine a sequence of CDSs.
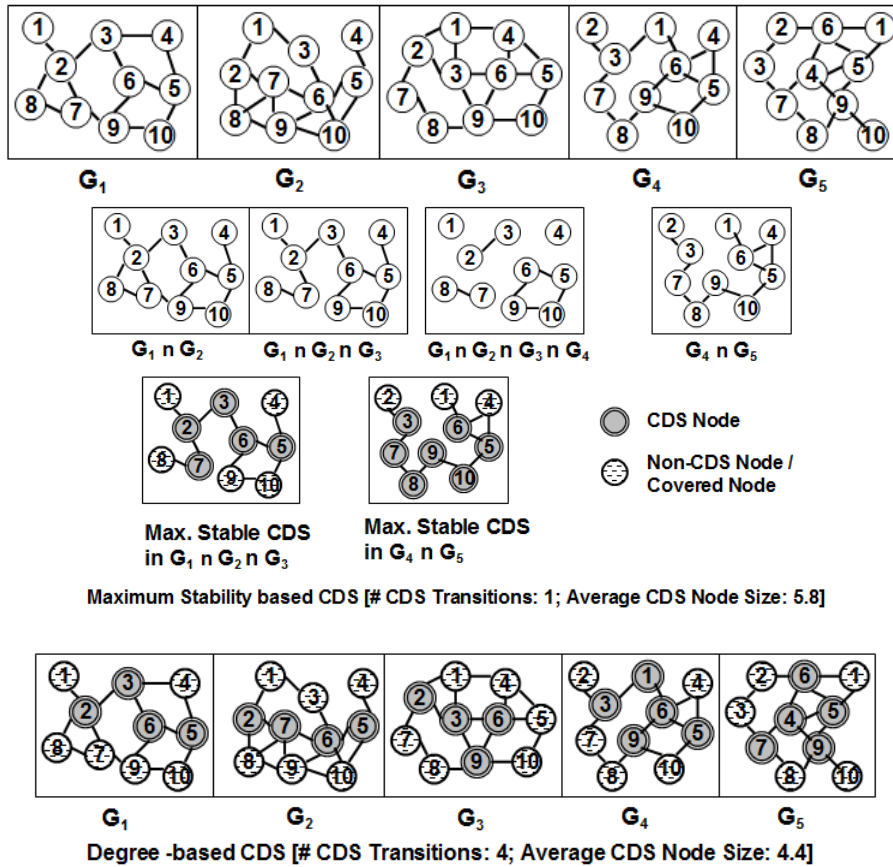
Figure 4. Example to Illustrate the Execution of the Maximum Stable CDS Benchmarking Algorithm and a Comparison of Maximum Stable CDS vs. Degree-based CDS

3.2. Example

Figure 4 presents an example to illustrate the execution of the Maximum Stable CDS algorithm on a sequence of five static graphs $G_1, ..., G_5$ and a comparison of the number of transitions (changes from one CDS to another) and CDS Node Size incurred for Maximum Stable CDS vis-a-vis a degree-based CDS on the same sequence of static graphs. We see that there exists a connected mobile graph $G(1, 2, 3)$ and a connected mobile graph $G(4, 5)$. Hence, it would suffice to use one CDS determined on the mobile graph $G(1, 2, 3)$ and another CDS determined on the mobile graph $G(4, 5)$ - resulting in only one CDS transition across the five static graphs. On the

other hand, the degree-based CDS determined on an individual static graph does not exist in the subsequent static graph. Hence, we end up determining a new degree-based CDS on each of the five static graphs - resulting in a total of four CDS transitions across the five static graphs. In Figure 4, we use a degree-based heuristic to determine a CDS on each of the two mobile graphs. We do observe a tradeoff between CDS Node Size and the number of CDS transitions. The average CDS Node Size incurred for the Maximum Stable CDS is (3*5 + 2*7) / 5 = 5.8, whereas the average CDS Node Size incurred for degree-based CDS is (3*4 + 2*5) / 5 = 4.4. The mobile graphs G(1, 2, 3) and G(4, 5) have relatively fewer links than the individual static graphs. Hence, it is not possible to cover a larger number of nodes with the inclusion of few nodes in the CDS. As a result, the Maximum Stable CDS incurs a relatively larger CDS Node Size compared to that of the degree-based CDS.

4. SIMULATIONS

We conducted the simulations in a discrete-event simulator implemented in Java. We assume a network of dimensions 1000m x 1000m; the number of nodes is fixed as 100 and the transmission range of the nodes is fixed at 250m; for these values, on average, there are $\pi*250*250*100/(1000*1000) = 20$ neighbors per node. The overall connectivity of the network for the above conditions is observed to be more than 99.5%. Initially, the nodes are uniform-randomly distributed throughout the network. The number of static nodes in the network is varied with values of 0, 20, 40, 60 and 80. Since the total number of nodes in the network is 100, the above values for the number of static nodes also represent the percentage of static nodes in the network. A node designated to be static in the beginning of the simulation does not move at all for the duration of the simulation; likewise, a node designated as mobile - keeps moving for the entire simulation.

The node mobility model used is the Random Waypoint Model [10], wherein the maximum velocity of each mobile node is v_{max} and it is varied with values of 5 m/s (low mobility), 25 m/s (moderate mobility) and 50 m/s (high mobility). According to the Random Waypoint Model, each mobile node starts from its initial location and chooses to move to an arbitrary location (located within the network boundaries) with a velocity that is uniform-randomly selected from the range $[0, ..., v_{max}]$; after moving to the chosen location, the node chooses another arbitrary location (within the network) and moves to the chosen location with a new velocity that is uniform-randomly chosen from the range $[0, ..., v_{max}]$. In other words, a node moves in a straight line from one location to another chosen location with a particular velocity and after reaching the targeted location, the node chooses another target location to move with a velocity that could be different from the one used before. A mobile node moves like this for the duration of the simulation session and generates a mobility profile for a particular combination of values for the number of static nodes and maximum node velocity. The collection of the mobility profiles of all the nodes is stored in a mobility profile file and is feed in to the heuristic/algorithm for determining a degree-based CDS and the Maximum Stable CDS. We generate 100 such mobility profile files for each of the combinations of the values for the number of static nodes (0, 20, 40, 60 and 80) and the maximum node velocity (5 m/s, 25 m/s and 50 m/s). The decision of whether a node will be static or mobile is made prior to creating each of the mobility profile files.

We run the simulations for each mobility profile file and average the results observed for the CDS Lifetime and CDS Node Size for the degree-based CDS as well as the Maximum Stable CDS. We start the simulations at time 0 sec (based on the initial distribution of the nodes in a particular mobility profile file) and run the simulations for a period of 1000 seconds, sampling the network for every 0.25 seconds of the simulation. We take a snapshot of the network at each of these sampling time instants (referred to as the static graphs). We use the following approach to run the simulations for the degree-CDS: Whenever a CDS is needed at a particular time instant t, we

determine the degree scores of the nodes based on the static graph snapshot of the network at time instant t and feed in these scores to the CDS construction heuristic (described in Section 2.1). The CDS determined at time instant t is used for the subsequent sampling time instants as long as it exists (validated using the algorithm presented in Section 2.3). When the currently known CDS is observed to no longer exist at a particular time instant, we repeat the above procedure. We continue like this for the duration of the simulation and measure the lifetime and node size for the sequence of degree-based connected dominating sets used for the particular mobility profile file. We average the results for these two metrics (Figures 5-7) observed for all the mobility profile files generated for a particular combination of values for the number of static nodes and v_{max}. In the case of the Maximum Stable CDS algorithm, for a particular simulation run under a mobility profile file, we determine a sequence of connected mobile graphs for the duration of the simulation session and run the degree-based CDS construction heuristic to determine a sequence of connected dominating sets. We repeat this procedure for all the mobility profile files generated and determine the average of the lifetime and node size for the sequence of Maximum Stable CDSs for the particular combination of values of the number of static nodes and v_{max}.

As expected, the Maximum Stable CDS incurred a larger CDS lifetime compared to that of the degree-based CDS. However, for any given level of node mobility (v_{max}), the difference in the CDS lifetimes between the Maximum Stable CDS and degree-based CDS significantly decreases with increase in the percentage of static nodes in the network. For $v_{max} = 5$ m/s, the ratio of the lifetime of the Maximum Stable CDS to that of the degree-based CDS decreases from a factor of about 10 to 2.5 as we increase the percentage of static nodes from 0 to 80. As the value of v_{max} increases, the reduction in the difference is even more prominent. For $v_{max} = 25$ m/s, the ratio of the lifetime of the Maximum Stable CDS to that of the degree-based CDS decreases from a factor of about 7.5 to 1.8; and for $v_{max} = 50$ m/s, the ratio of the lifetimes decreases from about 7 to 1.4. Thus, the lifetime of degree-based CDS becomes very comparable to that of the Maximum Stable CDS when 80% of the nodes in the network are static and only 20% are mobile. In general, when more than half of the nodes in the network are static, the lifetime of the degree-based CDS starts to swiftly approach the lifetime of the Maximum Stable CDS. Thus, as predicted in our hypothesis, the presence of static nodes in the network has a very positive effect on the stability of the degree-based CDSs, but only a marginal effect on the lifetime of the Maximum Stable CDS.
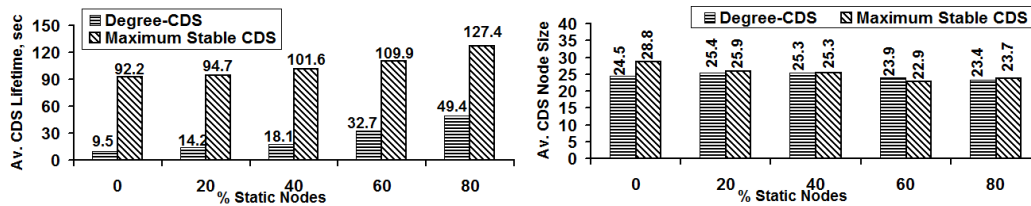

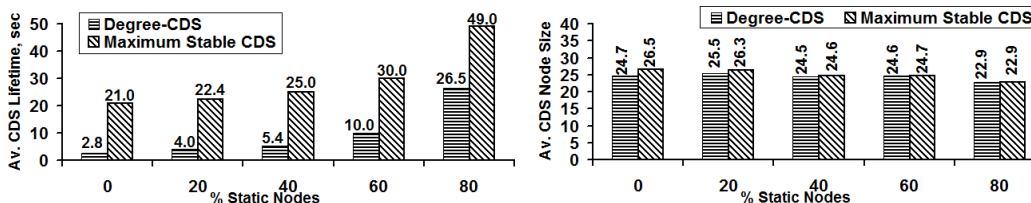Figure 5. Average CDS Lifetime and Average CDS Node Size: Maximum Node Velocity, $v_{max} = 5$ m/s

Figure 6. Average CDS Lifetime and Average CDS Node Size: Maximum Node Velocity, $v_{max} = 25$ m/s
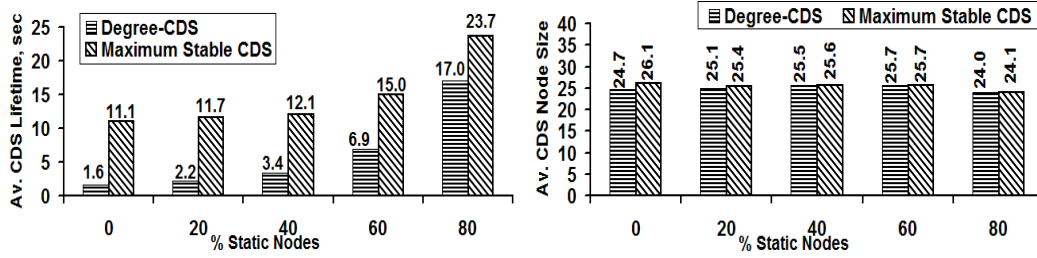
Figure 7. Average CDS Lifetime and Average CDS Node Size: Maximum Node Velocity, v_{max} = 50 m/s

The lifetime of the degree-based CDSs increases by a factor of 5 to 13 as we increase the percentage of static nodes from 0 to 80; the increase is substantial with increase in the level of node mobility. At v_{max} = 5 m/s, the increase in the lifetime of the degree-based CDSs is by a factor of 5 as we increase the percentage of static nodes from 0 to 80; however at v_{max} = 25 m/s and 50 m/s, the lifetime of the degree-based CDSs increases by a factor of 10 to 13. At the same time, the node size for the degree-based CDSs remains slightly high when the percentage of static nodes is low (0, 20 or 40) and then again decreases by a factor of at most 20% when the percentage of static nodes increases to 60 and 80. Thus, we observe the presence of static nodes in the network to significantly increase the lifetime and even marginally reduce the node size for the degree-based CDSs, thereby reducing the stability-tradeoff appreciably. We observe the lifetime of the Maximum Stable CDS to increase at most by a factor of 2.3 with increase in the percentage of static nodes. Like it was observed for the degree-based CDSs, the increase is only about 40% when v_{max} = 5 m/s and is higher for v_{max} = 25 m/s and 50 m/s; similarly, the node size for the Maximum Stable CDSs decreases by about 20% with increase in the percentage of static nodes. We observe the node size for the Maximum Stable CDS to converge towards the node size of the degree-based CDS with increase in the percentage of the static nodes.

5. RELATED WORK

MANET simulation studies have typically used the Random Waypoint model as the mobility model. As per this mobility model, a mobile node moving in a particular direction could pause for a certain time after reaching a targeted location and then continue to move. The pause time is one of the parameters of the Random Waypoint mobility model. The influence of static nodes on the performance of a communication protocol has been so far evaluated only on the basis of this pause time. Also, such studies have been typically conducted to analyze the performance of unicast routing protocols (e.g., [11]) and multicast routing protocols (e.g., [12]) for MANETs. Even in such works, the focus had been on the impact of pause time on metrics such as delay, throughput, energy consumption, etc, but not explicitly on the lifetime of the communication topologies like paths and trees. To the best of our knowledge, we have not come across any studies that consider the influence of static nodes of any form (including pause time) on the performance of network-wide communication topologies, like that of connected dominating sets for MANETs. In this paper, we chose to designate certain nodes as static throughout the simulation and let the other nodes to move (rather than letting all mobile nodes to move and pause alternately). We wanted to see if the presence of static nodes (nodes that do not move at all) could improve the overall stability of connected dominating sets (as it is seen in the simulations) so that we could henceforth come up with a network design strategy wherein even though certain nodes in a network are bound to be mobile all the time, one may deploy certain percentage of static nodes that could significantly increase the stability of network-wide communication topologies such as connected dominating sets.

6. CONCLUSIONS

The simulation results confirm our hypothesis: We observe that for a given level of node mobility, the lifetime of the Maximum Stable CDS does not increase substantially with increase in the percentage of static nodes; on the other hand, the lifetime of the degree-based CDS increases significantly (and approaches the lifetime of the Maximum Stable CDS) with increase in the percentage of static nodes. As we increase v_{max} from 5 m/s to 50 m/s, the average lifetime incurred for the Maximum Stable CDS increases by a factor of 1.4-2.3 as we increase the percentage of static nodes from 0 to 80. This could be attributed to the fact that the mobility of certain percentage of nodes is sufficient to disrupt the connectivity of the mobile graphs spanning over several time instants. On the other hand, as we increase v_{max} from 5 m/s to 50 m/s, the lifetime of the degree-based CDS increases substantially (by factors as large as 13) with increase in the percentage of the static nodes from 0 to 80. Thus, even if a certain fraction of nodes move faster, the presence of a significant fraction of static nodes helps to boost the lifetime of the degree-based CDS. In this paper, the degree-based CDS heuristic did not take into account the presence of static nodes while considering nodes for inclusion in the CDS. With the results observed in this paper, we anticipate that if the heuristic is adapted to take into account both the node velocity (or some form of node mobility) and node degree while forming a CDS, the stability of such CDSs could be significantly higher than that incurred with CDSs determined using a degree-based heuristic without any appreciable increase in the CDS node size. We thus anticipate the results of this paper to open new avenues of research for improving the stability of degree-based heuristics for MANETs.

REFERENCES

[1]   C. Siva Ram Murthy and B. S. Manoj, Ad Hoc Wireless Networks: Architectures and Protocols, Prentice Hall, 1st edition, March 2012.

[2]   D. B. Johnson and D. A. Maltz, "Dynamic Source Routing in Ad Hoc Wireless Networks," Mobile Computing - The Kluwer International Series in Engineering and Computer Science, vol. 353, pp. 153-181, 1996.

[3]   C. E. Perkins and E. M. Royer, "Ad Hoc On-demand Distance Vector Routing," Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications, pp. 90-100, New Orleans, LA, USA, Feb 25-26, 1999.

[4]   C. W. Wu and Y. C. Tay, "AMRIS: A Multicast Protocol for Ad hoc Wireless Networks," Proceedings of the IEEE Military Communications Conference, vol. 1, pp. 25-29, Oct 31-Nov 3, 1999.

[5]   A. B. Mnaouer, L. Chen, C. H. Foh and J. W. Tantra, "OPHMR: An Optimized Polymorphic Hybrid Multicast Routing Protocol for MANET," IEEE Transactions on Mobile Computing, vol. 6, no. 5, pp. 551-562, May 2007.

[6]   F. Dai and J. Wu, "Performance Analysis of Broadcast Protocols in Ad Hoc Networks Based on Self-Pruning," IEEE Transactions on Parallel and Distributed Systems, vol. 15, no. 11, pp. 1-13, November 2004.

[7]   S. Saha, S. R. Hussain and A. K. M. Ashikur Rahman, "RBP: Reliable Broadcast Protocol in Large Scale Mobile Ad Hoc Networks," Proceedings of the 24th IEEE International Conference on Advanced Information Networking and Applications, pp. 526-532, Perth, WA, USA, April 20-23, 2010.

[8]   T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, Introduction to Algorithms, 3rd ed., MIT Press, July 2009.

[9]   N. Meghanathan and A. Farago, "On the Stability of Paths, Steiner Trees and Connected Dominating Sets in Mobile Ad Hoc Networks," Ad Hoc Networks, vol. 6, no. 5, pp. 744-769, July 2008.

[10]   C. Bettstetter, H. Hartenstein and X. Perez-Costa, "Stochastic Properties of the Random-Way Point Mobility Model," Wireless Networks, vol. 10, no. 5, pp. 555-567, September 2004.

[11]   S. R. Das, C. E. Perkins, E. M. Royer, "Performance Comparison of Two On-demand Routing Protocols for Ad hoc Networks," Proceedings of the IEEE International Conference on Computers and Communications, vol. 1, pp. 3-12, Tel Aviv, Israel, March 26-30, 2000.

[12] K. Kavitha and K. Selvakumar, "Analyzing Multicast Routing Protocols with Different Mobility Models," International Journal of Emerging Technology and Advanced Engineering, vol. 3, no. 1, pp. 33-41, January 2013.

INVESTIGATION OF CHAOTIC-TYPE FEATURES IN HYPERSPECTRAL SATELLITE DATA

Osman Kocal and Mufit Cetin

Department of Computer Engineering, Yalova University, Yalova, Turkey
osman.kocal@yalova.edu.tr, mufit.cetin@yalova.edu.tr

ABSTRACT

Hyperspectral images provide detailed spectral information with more than several hundred channels. On the other hand, the high dimensionality in hyperspectral images also causes to classification problems due to the huge ratio between the number of training samples and the features. In this paper, Lyapunov Exponents (LEs) are used to determine chaotic-type structure of EO- 1 Hyperion hyperspectral image, a mixed forest site in Turkey. Experimental results demonstrate that EO-1 Hyperion image has a chaotic structure by checking distribution of Lyapunov Exponents (LEs) and they can be used as discriminative features to improve classification accuracy for hyperspectral images.

KEYWORDS

Hyperspectral Image, EO-1 Hyperion, Phase Space, Chaos, Lyapunov Exponents

1. INTRODUCTION

Hyperspectral sensors collect the electromagnetic spectrum of reflected light from objects and combine spectral information for each pixel in the image of a scene. Every pixel in the image represents to a unique 'fingerprints' of the object, also known as a spectral signature and provides more spectral information to characterize different materials [1]. Therefore, hyperspectral images are much more functional to identify different objects than multispectral images. With the recent advances in sensor technology, hyperspectral images including AVIRIS and EO-1 HYPERION have become available and used in many application areas such as forestry, mining, oil industry, agriculture, ecology, environmental monitoring and geology etc [2, 3, 4, 5].

Hyperspectral images provide detailed spectral information with more than several hundred channels. On the other hand, the high dimensionality in hyperspectral images also causes to classification problems due to the huge ratio (the Hughes phenomenon) between the number of training samples and the features [6, 7]. Besides, high dimensionality of hyperspectral images causes the failures in common image classification algorithms and high computational complexity due to hundreds of bands. Therefore, dimension reduction is an important research topic for classification of hyperspectral images on remote sensing.

The purpose of dimension reduction ideally is to preserve as much as possible the information of original hyperspectral data set. Thus, feature extraction is an important way to reduce the

dimension of the hyperspectral image without a subspace from the original data like feature selection method [8, 9]. In this context, the purpose of the study is to reduce the dimension of the hyperspectral data from hundreds of bands to only a few features using Lyapunov Exponents (LEs) of original reflectance curves.

2. CHAOTIC STRUCTURE ANALYSIS USING LYAPUNOV EXPONENTS

Phase spaces of spectral signals are crucial to identify relationship between spectral signals and chaotic structure of image. Phase spaces allow observing signal dynamics in a spectral band series and providing understandable important information related to the chaotic dynamics of the original signal. The hyperspectral reflectance curves for each class have similar patterns and they can be used as a time series depend on the number of spectral bands used in the imaging system. The constant spectral band difference between the neighboring elements of each reconstructed vector, which is the embedding dimension of the phase space, is represented by the spectral band for each pixel [10]. The phase-space vector of the spectral signal is reconstructed as follows:

$$n = p.k + b$$
$$s[n] = (\, x[n], x[n+1], \ldots \ldots x[n+p.k-1])$$

Where n denotes index of data point in 1-D spectral signal, x[n] is the nth sample of the spectral signal, k indicates the total number of spectral bands, b denotes index of the spectral band for selected pixel on hyperspectral image and p is number of pixel, as numbers between 0 and total number of samples.

Lyapunov exponents provide a rapid effective method to determine chaotic structure in dynamic systems [11, 12]. Furthermore, Lyapunov exponents are a quantitative measure of the rate at nearby orbits converge and diverge. In the system, having more than one positive and negative Lyapunov exponents indicate chaotic behaviour [13, 14, 15]. There are many algorithms proposed in the literature for estimating the Lyapunov exponents. The most popular from these algorithms is from Abarbanel et al [16].

$$\lambda = \lim_{n \to \infty} \frac{1}{N} \sum_{n=1}^{N} ln \frac{d(s(n+1), s(m+1))}{d(s(n), s(m))}$$

where, s(n) denotes the reference point, s(m) is the nearest neighbour of s(n) on a nearby trajectory, d(s(n),s(m)) is the initial distance between the nearest neighbours, d(s(n+1),s(m+1)) is the distance between s(n+1) and s(m+1) which are the next pair of neighbours on their trajectories [17, 18].

3. TEST DATASET

Earth Observing 1 (EO-1) HYPERION, launched on 21 November 2000 by NASA, is the first hyperspectral satellite. The Hyperion sensor detects a total of 242 channels from 356 to 2577 nm at approximately a 10-nm spectral resolution with a 30 m spatial resolution and scans a ground area approximately 7.7 km in the across-track direction, and 42 km in the along-track direction. Test image covers a portion of Sundiken Mountains, a mixed forest site, in Eskisehir province, Turkey. In the study, atmospheric absorption bands and low signal-to-noise ratio bands were eliminated from the dataset and remained 153 bands are used for analysis [19, 20]. In the study area, four object classes including pine and oak trees, grassland and water are selected to evaluate Lyapunov exponents as a feature tool.

4. RESULTS

Lyapunov Exponents (LEs) are used to demonstrate the chaotic-type features for EO-1 Hyperion hyperspectral data set in the study. First, 1-D hyperspectral spectral vectors are created as a chaotic series and combined from manually selected pixels for each class in the image, remaining in 153 healthy channels of Hyperion data (Figure 1).
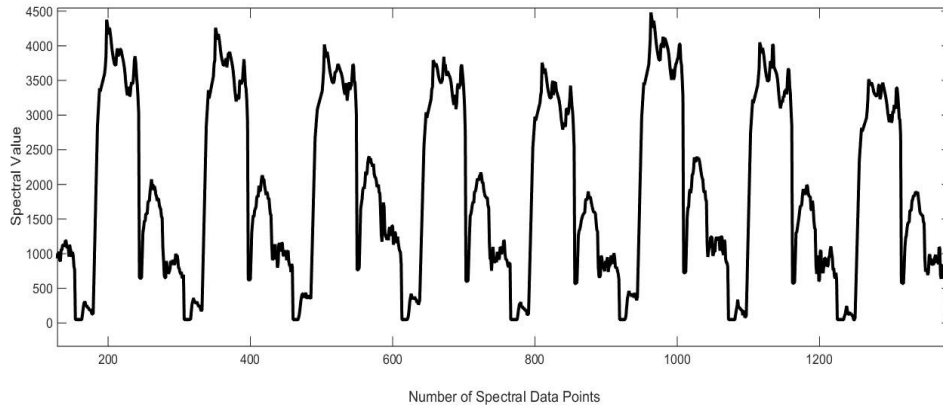
Figure 1. Spectral series of pixels for class Oak trees

Then, phase spaces of spectral signals are initially produced to explore the chaotic structure of hyperspectral image using 1-D spectral vectors (Figure 2). Next, Lyapunov Exponents (LEs) of the spectral signals are calculated for four objects including pine, oak, grassland and water and it is observed different negative or minus signs among them.
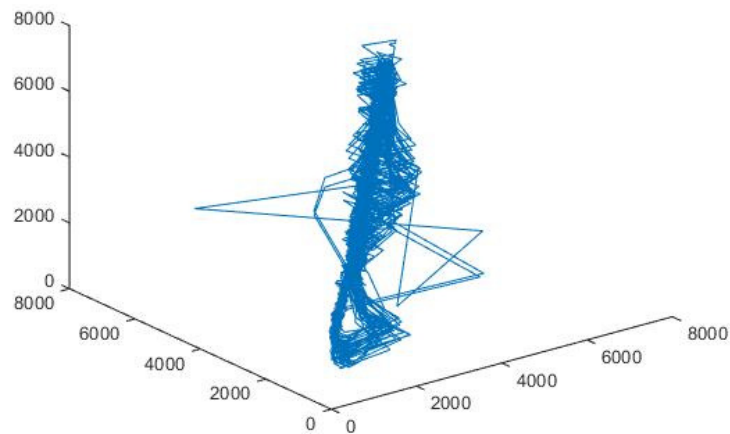
Figure 2. Reconstructed phase space of the spectral signal in 3-D

A positive and negative Lyapunov exponents are computed in all the hyperspectral signals setting total number of data points (n=3060). Positive and negative Lyapunov exponents indicate the possible existence of chaotic behaviour in hyperspectral data (Table1). Kaplan-Yorke dimension (DKY) indicates the fractal dimension of the reconstructed signal corresponding Lyapunov spectrum and values of the embedding dimension (m) [13, 11].

Table 1. The Lyapunov exponents for 4738 trials each with embedding dimension 5 calculated for each object

Type of Object	Number of Iteration	m	Lyapunov Exponents					Average Absolute Forecast Errors	Average Neighborhood Size	DKY
			Λ1	Λ2	Λ3	Λ4	Λ5			
pine	4738	5	0.1632	-0.0272	-0.1756	-0.3697	-0.8537	106	117	2.77
oak	4738	5	0.1851	-0.0155	-0.1831	-0.3611	-0.8469	177	185	2.93
grassland	4738	5	0.1408	-0.0031	-0.1791	-0.3603	-0.7431	207	187	2.62
water	4738	5	0.3241	0.0529	-0.1093	-0.3118	-0.7257	54	39	3.86

5. CONCLUSION

The experimental results demonstrate that EO-1 Hyperion image has a chaotic structure by checking distribution of Lyapunov Exponents (LEs) and they can be used as discriminative features to improve classification accuracy for hyperspectral images. Future work, therefore, includes the investigation and evaluation of the effects of chaotic features on hyperspectral image classification.

REFERENCES

[1]   Sarath, T. & Nagalakshmi, G., (2014) "A Land Cover Fuzzy Logic Classification By Maximumlikelihood", International Journal of Computer Trends and Technology, Vol. 13, No.2, pp56-60.

[2]   Carter, G.A. & Miller, R.L., (1994) "Early detection of plant stress by digital imaging within narrow stress-sensitive wavebands", Remote Sensing of Environment, Vol. 50, pp295–302.

[3]   Crosta, A.P., Sabine, C. and Taranik, J.V., (1998), "Hydrothermal Alteration Mapping at Bodie, California, Using AVIRIS Hyperspectral Data", Remote Sensing of Environment, Vol.65, No.3,pp 309-319.

[4]   Ellis, J.M., Davis, H.H., Zamudio, J.A., (2001), "Exploring for onshore oil seeps with hyperspectral imaging", Oil and Gas Journal, Vol. 99, No. 37, pp49-56.

[5]   Horig, B., Kuhn, F., Oschutz, F. and Lehmann, F., (2001), "HyMap hyperspectral remote sensing to detect hydrocarbons", Int. J. Remote Sensing, Vol. 22, No.8, pp1213-1422.

[6]   Mianji, F.A. & Zhang, Y., (2011), "Robust hyperspectral classification using relevance vector machine", IEEE Trans. Geosci. Remote Sens., Vol.49, pp2100–2112.

[7]   Zhang, X., He, Y., Zhou, N., and Zheng, Y., (2013), "Semisupervised dimensionality reduction of hyperspectral images via local scaling cut criterion", IEEE Geoscience Remote Sensing Letters, Vol. 10, No.6, pp1547–1551.

[8]   Shaw, G. & Manolakis, D., (2002), "Signal processing for hyperspectral image exploitation", IEEE Signal Process. Mag. , Vol. 19, No.1, pp12–16.

[9]   Semeniv, O. V., Shatokhina, Y. V., and Yatsenko, V. A., (2008), "Validation of hyperspectral data classification models", J. Autom. Inf. Sci. , Vol. 40, No.5, pp46–51.

[10] Jihao, Y., Chao, G. and Xiuping, J, (2012), "Using Hurst and Lyapunov Exponent For Hyperspectral Image Feature Extraction", IEEE Geoscience and Remote Sensing Letters, Vol. 9, No. 49, pp705 – 709.

[11] Banbrook, M. & McLughlin, S., (1994), "Is speech chaotic?: Invariant geometrical mesures for speech data", IEEE Colloquiunz on Exploiting Chaos in Signals Processing, Digest No 1994/193, pp8/1-8/10.

[12] Eckmann, J.P. , Kamphorst, O.S., Ruelle, D. and Ciliberto, S., (1986), "Lyapunov exponents from time series", Phys. Rev. A, Vol. 34, pp4971-4979.

[13] Lorenz, E.N., (1963), "Deterministic nonperiodic flow", J. Atmos. Sci. , Vol.20, No.2, pp130–141.

[14] Wolf, A., Swift, J.B., Swinney, H.L. and Vastano, J.A. , (1985), "Determining Lyapunov exponents from a time series", Physica D, Vol. 16, pp285-317.

[15] Kocal, O.H., Yuruklu, E., Avcibas, E., (2008), "Chaotic-type features for speech steganalysis", IEEE Transactions on Information Forensics and Security, Vol. 3,No.4, pp651–661.

[16] Abarbanel, H.D.I., (1996), Analysis of Observed Chaotic Data , 1st ed. New York: Springer-Verlag.

[17] Kokkinos, I. & Maragos, P., (2005), "Nonlinear speech analysis using models for chaotic systems", IEEE Trans. Speech Audio Process., Vol. 13, No.6, pp1098–1109.

[18] Martinez, F., Guillamon, A., Alcaraz, J.C., and Alcaraz, M.C., (2002), "Detection of chaotic behavior in speech signals using the largest Lyapunov exponent", in Proc. 14th Int Conf. Digital Signal Processing, Jul. 1–3, 2002, 1, pp317–320.

[19] Parkinson, C.L., Ward, A., King, M.D., and editors, (2006), Earth Science Reference Handbook: A Guide to NASA's Earth Science Program and Earth Observing Satellite Missions. Washington, D.C.: National Aeronautics and Space Administration.

[20] Cetin, M. & Musaoglu, N., (2009), "Merging hyperspectral and panchromatic image data: qualitative and quantitative analysis", Int. J. Remote Sens., Vol. 30, No.7, pp1779–1804.

AUTHORS

Osman Hilmi Kocal was born in Istanbul, Turkey, in 1967. He received the B.Sc., M.Sc., and Ph.D. degrees in electronics and telecommunication engineering from the Technical University of Istanbul, Istanbul, in 1989, 1992, and 1998, respectively. He was a Research Assistant with the Technical University of Istanbul and Turkish Air Force Academy, Istanbul, respectively. Currently, he is an Assistant Professor with the Department of Computer Engineering, Yalova University, Yalova, Turkey. His research interests include chaotic signals and adaptive signal processing.

Mufit Cetin is an associate professor in the department of Computer Engineering at Yalova University. He received his BS, MS and PhD degrees in Geomatics Engineering from Yildiz Technical University in 1996, Institute of Gebze Technology in 2001 and Istanbul Technical University in 2007, respectively. His current research interests include image processing and remote sensing.

INTENTIONAL BLANK

BANDWIDTH-PERFORMANCE TRADEOFFS FOR A TRANSMISSION WITH CONCURRENT SIGNALS

Aminata A. Garba

Dept. of Electrical and Computer Engineering, Carnegie Mellon University
aminata@ece.cmu.edu

ABSTRACT

We consider a bandwidth-efficient transmission scheme, where two signals are sent concurrently. The BER and the achievable minimum distances for the signals' constellations at the receivers are derived as functions of the signals' energies and their input probability distributions. Finally, trade-offs between bandwidth, signals' energies and achievable performances are discussed.

KEYWORDS

Bandwidth efficiency, Energy, Interference, Minimum Distance, BER Performance.

1. INTRODUCTION

The demand of bandwidth is largely increasing due to the explosion of smart devices and bandwidth-greedy applications. This has led to the development of several bandwidth efficient technologies such as Code Division Multiplexing Access, Multiple-Input, Multiple-Output (MIMO), ultra wideband (UWB), cognitive radio (CR) technologies. Nonetheless, a major limitation of bandwidth efficient technologies is the interference caused by simultaneous transmissions [1]-[5]. It is then important to analyze the impact of simultaneous signal transmissions on the system performance. In [6], symbol error probability is analyzed for a multi-user detector using successive cancellation, when signals differentiated by their powers are transmitted simultaneously.

In this paper, we consider simultaneous transmission of two signals and analyze the impact on the system performance. We present performance analysis and compare the resulting performance with the achievable performance if the signals were sent separately over different channels. We also discuss impact of interference and trade-offs between performance and bandwidth efficiency.

2. SYSTEM MODEL & ASSUMPTIONS

We consider a communication scheme, where two signals are transmitted concurrently and independently (i.e., without collaboration). For example, this scheme can model a cognitive transmission where the cognitive signal interferes with the primary signal, either in a controlled interference scenario or because the cognitive user sensed an idle channel while the primary user was actually transmitting, due for example to challenges related to sensing in cognitive transmissions [7]. It can also model a different transmission scheme whereby; users with no or

small cognition capabilities can transmit simultaneously without time, frequency or code division multiplexing, thus interfering with each other.

The two signals, which may first be encoded using forward error correcting code, are modulated using BPSK modulation and sent concurrently over an AWGN channel.

Under these assumptions, the received symbol can be written as follows:

$$Y = X_1 + X_2 + n, \tag{1}$$

where X_1 and X_2 are the (encoded and) modulated symbols from the two sources and n is additive white Gaussian noise (AWGN) with zero mean and variance σ^2 that corrupts the received information. Furthermore, we assume that X1 has energy E1 and takes values in the input alphabet $A = \{ -a, a \}$ with probability P and (1-P) and X2, has energy E2 and takes values in its input alphabet $B = \{ -b, b \}$ with probabilities P and (1-P). Finally, a, b are assumed non-negative valued constants known to the receivers with (for simplicity) $a \geq b$, i.e., $E_1 \geq E_2$. (The results for the case where $E_1 < E_2$ can be derived from the considered case.)

At the receiver, minimum-distance symbol-by-symbol detection is used, i.e., a received symbol is detected as the closest point in the signal constellation. Finally, the received symbol Y is used to make decisions on the sent symbols X_1 and X_2 at both receivers respectively.

We refer to the minimum distance of a signal constellation at a given receiver, as the minimum of the distances between any two points of the signal constellation.

Furthermore, throughout this paper, $\widehat{d}_{min1}$ and $\widehat{d}_{min2}$ define the minimum distances for the signal constellations at the receivers 1 and 2 respectively, if each signal was sent separately (scenario 1). On the other hand, d_{min1} and d_{min2} represent the minimum distances for received constellations at receivers 1 and 2, when both symbols are sent concurrently (scenario 2). For both scenarios, we consider the same transmission conditions (symbols' energies, channel conditions, SNR).

Finally, since in AWGN channels, the performance strongly depends on the minimum distance of the received signal constellation, we next analyze the minimum distances d_{min1} and d_{min2} for the signals' constellations at the receivers when the 2 symbols are transmitted concurrently.

3. PROPOSITIONS & LEMMAS

A. Proposition 1

The achievable minimum distance for a transmission with two signals sent concurrently and independently is at most the difference of the achievable minimum distances of the constellations if the signals were sent separately under the same conditions.

In particular, the minimum distance d_{min1} and d_{min2} of the received signals constellations, are respectively

$$d_{min1} = \widehat{d}_{min1} - \widehat{d}_{min2},$$

$$d_{min2} = \min (\widehat{d}_{min2}, \widehat{d}_{min1} - \widehat{d}_{min2}) \tag{2}$$

Proof

We assume BPSK modulated symbols at the transmitters and minimum-distance detection at the receivers. Assuming that the first signal uses input alphabet {-a, a} with probability P and (1-P) and the second signal uses {-b, b} with probability P and (1-P), the minimum distances of the received constellations are given by $\hat{d}_{min1} = 2a$, $\hat{d}_{min2} = 2b$ is each signal is transmitted separately (scenario 1).

If the two signals were sent concurrently (scenario 2) and assuming a minimum-distance detector, the decision regions are shown in Figure 1. (Note that the minimum distance detection is equivalent to the ML detection for AWGN channels.)

Using the decision regions illustrated in Figure 1, the minimum distances of the signal constellations at the receivers are given by:

$$d_{min_1} = (a-b)-(-a+b) = 2(a-b) = \hat{d}_{min1} - \hat{d}_{min2},$$
$$d_{min_2} = \min\big((a-b)-(-a+b), 2(a-b)\big) = \min(2b, d_{min1})$$
$$= \min\Big(\hat{d}_{min2}, \big(\hat{d}_{min1} - \hat{d}_{min2}\big)\Big).$$

$$(3)$$

B. Proposition 2

1. If $E_1 \neq E_2$, then the minimum distance d_{min1} and d_{min2} of the signals' constellations at receivers 1 and 2, are respectively

$$d_{min_1} = \hat{d}_{min1} - \hat{d}_{min2} = 2\left(\sqrt{E_1} - \sqrt{E_2}\right)$$
$$d_{min_2} = \min\left(\hat{d}_{min2}, \hat{d}_{min1} - \hat{d}_{min2}\right)$$

$$(4)$$

$$= \begin{cases} \hat{d}_{min2} = 2\sqrt{E_2} & \text{if } E_1 \geq 4E_2 \\ \hat{d}_{min1} - \hat{d}_{min2} = 2\left(\sqrt{E_1} - \sqrt{E_2}\right) & \text{if } E_1 < 4E_2 \end{cases}$$

2. If the signals have the same energy ($E_2 = E_1 = E$), the minimum distance is zero for both constellations): $d_{min_1} = d_{min_2} = 0.$

Proof

Part 1 of the proposition results from equation (3) using the average symbol energy given as follows:

$$E_1 = mean\left(X_1^2\right) = a^2 P + a^2(1-P) = a^2$$
$$\Rightarrow a = \sqrt{E_1}; \text{ similarly } b = \sqrt{E_2}$$

$$(5)$$

Part 2 of the proposition follows from part 1 when $E_1 = E_2$.

C. Proposition 3

When 2 signals are sent concurrently as described in section II, the BER performance for the strongest signal can be determined as follows:

$$BER_1 \leq P_e = 2P \times (1-P) \times Q\left(\sqrt{SNR_1} - \sqrt{SNR_2}\right)$$

$$+ \left(P^2 + (1-P)^2\right) \times Q\left(\sqrt{SNR_1} + \sqrt{SNR_2}\right) \tag{6a}$$

$$\xrightarrow{P=1/2} \frac{1}{2}Q\left(\sqrt{SNR_1} - \sqrt{SNR_2}\right) + \frac{1}{2}Q\left(\sqrt{SNR_1} + \sqrt{SNR_2}\right) \tag{6b}$$

Signal space and decision region at the first receiver

Signal space and decision regions for the second receiver
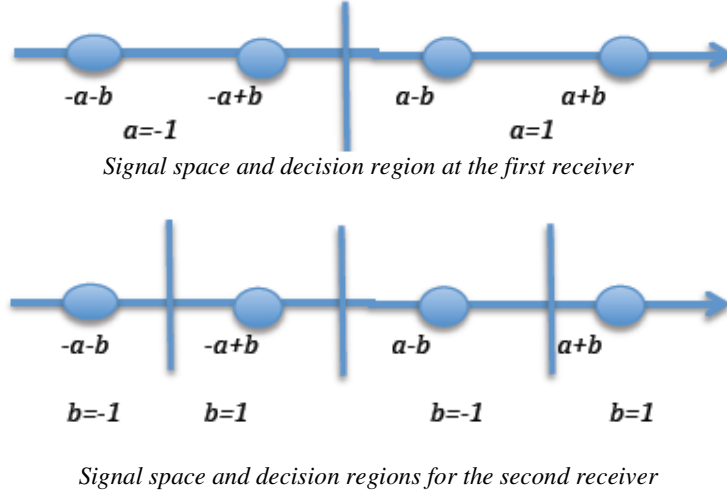
Figure 1: Signal space and decision regions

where [P, (1-P)] is the probability input distribution for the considered binary signaling, SNR_1, SNR_2 are the signal to noise ratios of the individual signals and $Q(x)$ is the standard Q function defined by $Q(x) = Pr(X>x) = Pr(X<-x)$ when X is zero mean, variance 1, Gaussian random variable. (Proof of Proposition 3 is outlined in Appendix.)

D. Lemmas

Lemma 1

If $E_2 \leq E_1 < 4E_2$,

$$d_{min_1} = d_{min_2} = 2\left(\sqrt{E_1} - \sqrt{E_2}\right) < 2\sqrt{E_2} \tag{7a}$$

$$d_{min_1} = d_{min_2} = \hat{d}_{min1} - \hat{d}_{min2} < \hat{d}_{min2} < \hat{d}_{min1} \tag{7b}$$

Lemma 2

If $E_1 > 4E_2$,

$$d_{min_1} = 2\left(\sqrt{E_1} - \sqrt{E_2}\right) > d_{min_2} = 2\sqrt{E_2} \tag{8a}$$

$$d_{\min_1} = \left(\hat{d}_{\min 1} - \hat{d}_{\min 2}\right) > d_{\min_2} = \hat{d}_{\min 2}$$

(8b)

Lemma 3

For the cross points where E1=4E2 , the minimum distances satisfy the following equalities and inequality:

$$d_{\min_1} = d_{\min_2} = 2\sqrt{E_2}$$

(9a)

$$d_{\min_1} = d_{\min_2} = \hat{d}_{\min 2} < \hat{d}_{\min 1}$$

(9b)

Lemma 4

If $E_2 = \alpha^2 E_1$, for $0 \le \alpha \le 1$, the minimum distances of the signal constellations can be written as follow:

First signal constellation:

$$d_{\min_1} = 2\left(1-\alpha\right) \times \sqrt{E_1} = \left(1-\alpha\right)\hat{d}_{\min 1}$$

(10a)

Second Signal constellation:

$$\text{If } E_1 \ge 4E_2, i.e., \left(\alpha \le \frac{1}{2}\right) \Rightarrow d_{\min_2} = \hat{d}_{\min 2} = \alpha \hat{d}_{\min 1}$$

(10b)

$$\text{If } E_1 < 4E_2, i.e., \left(\alpha > \frac{1}{2}\right) \Rightarrow d_{\min_2} = d_{\min_1} = \left(1-\alpha\right)\hat{d}_{\min 1}$$

(10c)

were $E_2/E_1 = \alpha^2$, for some $\alpha \in [0,1]$.

Lemma 5

The BER for the strongest signal is given, when $E_2 = \alpha_2 E_1$, for $0 \le \alpha \le 1$, by :

$$BER_1 = 2P\left(1-P\right)Q\left(\left(1-\alpha\right)\sqrt{SNR_1}\right) + \left(P^2 + \left(1-P\right)^2\right)Q\left(\left(1+\alpha\right)\sqrt{SNR_1}\right)$$

$$\xrightarrow{P=1/2} \frac{1}{2} \times Q\left(\left(1-\alpha\right)\sqrt{SNR_1}\right) + \frac{1}{2} \times Q\left(\left(1+\alpha\right)\sqrt{SNR_1}\right).$$

(11)

4. RESULTS AND DISCUSSION

Proposition 1 illustrates that when 2 signals are sent concurrently (scenario 2), the achievable minimum distance for any of the two signals' constellations is at most the difference of the achievable minimum distances if the 2 signals were sent separately (scenario 1). Clearly, sending the two signals simultaneously results into a lost of performance for the strongest signal, whereas the benefit is a better (double) efficiency of the bandwidth utilization.

From Proposition 2, if two signals with different energies are sent concurrently, the achievable minimum distances of the signals' constellations at the receivers are non-zero. Therefore,

transmitting concurrent signals with different energies can allow multiplexing. Whereas when two signals with the same energy are sent concurrently, the signals' constellations at the receivers will both have a zero minimum distance, thus a very poor performance. This is, for example, the case of a concurrent transmission of 2 symmetrical, *iid* (independent and identically distributed), equal-powered signals, without time, frequency or code division.

Proposition 3 expressed the BER after the strongest signal detection as a function of the SNR. Since the first term in 6b is mostly predominant, the BER depends strongly on the difference of the square root of the 2 SNRs. This is consistent with results from Propositions 1 and 2.

The result of lemma 1 shows that if the energies of two signals sent concurrently satisfies $E_2 \leq E_1 < 4E_2$, the two signals will be detected with similar performance, which is lower than the achievable performances if they were transmitted separately. It can be seen as if each signal selflessly gives up part of its energy to combat the mutual interference so that both signals are detected with similar performance. However, signal 1 with higher energy has a higher contribution to fighting the mutual interference than the contribution of the weakest signal 2.

From Lemma 2, if the energies of two signals sent concurrently satisfy $E_1 > 4E_2$, the signal with the lowest energy does not contribute to fighting the mutual interference. This is a kind of selfish scenario where only the strongest signal 1 uses part of its energy to combat the interference caused to him by the second signal. The weakest signal uses all of its energy for its actual data transmission and will be detected with a performance similar to the case where it is transmitted separately with energy E2 (scenario 2).

Moreover, when the energies of two signals sent concurrently satisfy $E_1 = 4E_2$, results (from Lemma 3) are similar to the previous (greedy) case where the strongest signal uses part of its energy to combat the interference caused by the second signal, whereas the weakest signal uses all of its energy for its data transmission. However, in this case, both signals end up having the same performance, equal to the performance of the weakest signal when sent separately (under scenario 1). In such a case, the total energy of both signals can be written as function of $d_{min} = d_{min1} = d_{min2}$, as follows:

$$E_{total} = 1.25 \times d^2_{min.}$$
(12a)

For example, to double the minimum distance d_{min}, up to 4 times the total energy would be required.

On the other hand, if the two signals were sent separately, the bandwidth efficiency would be twice smaller. However, the energy required to transmit each signal with a minimum distance d_{min} is $E_1 = E_2 = 0.25 \times d^2 min$. The total energy for a performance with $d_{min} = d_{min1} = d_{min2}$ at both receivers would then be :

$$E_1 + E_2 = 0.5 \times d^2_{min}$$
(12b)

which is 2.5 time lower than the total energy for the same minimum distance if the signals were sent concurrently.

From Lemma 4, for $E_2 \leq E_1$, the ratio E_2/E_1 can be written as $E_2/E_1 = \alpha 2$, for some positive constant α less or equal to 1. In that case, the minimum distance of the signal constellation at the receiver of the first (strongest) signal is a fraction of the minimum distance it would achieve if it was sent separately and decreases as α increases. Hence the closer the two energies are, the lower the minimum distance of the first signal constellation would be. For the second (weakest) signal, two

cases can be considered. If $\alpha \leq 1/2$, then the (weakest) signal will be detected at the receiver with similar performance as if it was sent separately with Energy E_2. However, if $\alpha > 1/2$, both signals will be detected with the same performance which is lower than both achievable individual performances if the signals were sent separately.
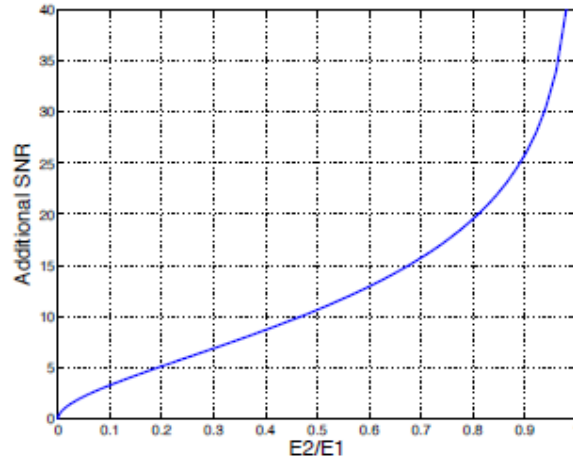
Figure 2: Additional SNR required for the first signal when sent concurrently with a second signal to achieve a performance similar to the case of separate transmission ($E_2=0$) for uncoded transmission.
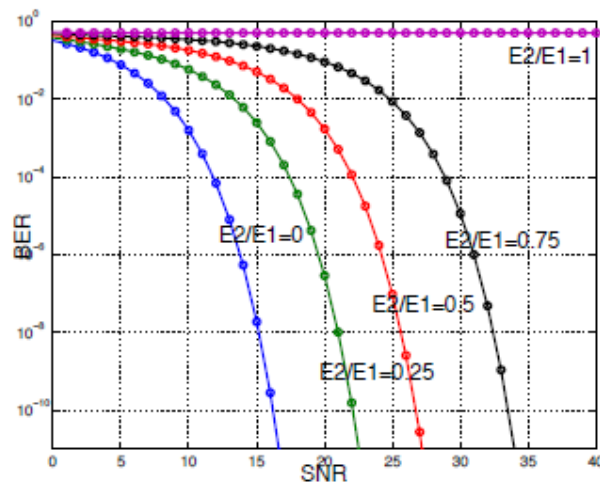
Figure 3: BER as function of SNR for the strongest signal with energy E1 when sent concurrently with another signal with energy E_2 for several values of E_2/E_1. The transmission is uncoded and uses BPSK modulation with P=1/2.

Lemma 5 results from proposition 3 when the ratio E_2/E_1 equals α^2. Figure 2 illustrates the additional SNR required for the strongest signal to achieve the same BER performance for uncoded transmission when it is sent concurrently with another signal. The additional SNR is in reference to the SNR it would require if it was sent separately in uncoded transmission with energy E_1. This additional SNR increases as E_2/E_1 increases. The additional SNR tends to infinity when $E_2/E_1=1$ and equals to 0 when $E_2/E_1 =0$ (i.e., in the case of individual transmission, $E_2=0$). Figure 3 illustrates the (uncoded) BER as a function of the SNR for the strongest (first) signal when it is sent concurrently with a second signal for several values of E_2/E_1. The performance decreases as E_2/E_1 increases, where the case $E_2/E_1 =0$ is equivalent to the case where the strongest signal is sent separately ($E_2=0$). For $E_2/E_1 =1$, it is not possible to achieve a good performance for

any value of the SNR. Note that the results in Figures 2 and 3 considered uncoded transmission and different results would be obtained if channel coding, multiplexing and/or further processing are assumed.

Finally, it is interesting to note that the system considered in this paper can also model a wireless transmission were at a given time instant, a and b are quasi-constant, known channel fading coefficients at the receivers.

5. CONCLUSION

This paper analyzes the performance of a transmission with two concurrent signals and trade-offs between the bandwidth and performance are discussed. More specifically, we present analytical expressions for the minimum distances for the two signals' constellations in various scenarios and selected BER performance results.

Our future work focuses on the design and analysis of transmision schemes for wireless systems when multiple signals are sent simultaneously and for the case of several cognitive signals transmission.

REFERENCES

[1] R. G. Gallager, "A perspective on multiaccess channels," IEEE Transactions on Information Theory, pp. 124 – 142, March 1985.

[2] T. M. Cover, J. A. Thomas, Elements of Information Theory, John Wiley & Sons, New York, 1995.

[3] R. Etkin, D. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," IEEE Trans. Inf. Theory, vol. 54, no. 12, pp. 5534–5562, Dec. 2008.

[4] H. Sato, "The capacity of the Gaussian interference channel under strong interference," IEEE Trans. Inf. Theory, vol. IT-27, no. 6, pp. 786–788, Nov. 1981.

[5] A. A. Garba, J. Bajcsy, "Capacity of OCDMA network transmission with optical hard-limiters," Proc. Canadian Workshop on Information Theory, pp. 86 – 90, Montréal, QC, June 2005.

[6] G.J.M. Janssen, S. Ben Slimane, "Symbol Error Probability Analysis of a Mutiuser Detector for MPSK Signals Based on Successive Cancellation", IEEE J. Select. Areas on Communications, pp. 330- 338, February 2002.

[7] Tevfik Yucek, Huseyin Arslan, "A Survey of Spectrum Sensing Algorithms for Cognitive Radio Applications", IEEE Communication Surveys & Tutorials, VOL. 11, NO. 1, pp. 116-130.

APPENDIX

Proof of Lemmas

Lemmas 1, 2 and 3, follow directly from Propositions 1 and 2. Lemma 4 results from proposition 2, using $E_2 = \alpha^2 E_1$ and the fact that if the signals were sent separately with energies E_1 and E_2 respectively, the minimum distances for the signal constellations at the receivers would be:

$$\hat{d}_{min1} = 2\sqrt{E_1}; \hat{d}_{min2} = 2\sqrt{E_2}.$$

Lemma 5 is obtained from Proposition 3, by replacing $SNR_2 = \alpha^2 SNR_1$ in (6a) and (6b) when $E_2 = \alpha^2 E_1$ for $0 \leq \alpha \leq 1$.

Proof of Proposition 3

For the first signal, the BER can be approximated as follow:

$$BER_1 \leq P_e = P \times \Pr\left(Y < 0 \mid X_1 = a\right)$$
$$+ \left(1 - P\right)\Pr\left(Y > 0 \mid X_1 = -a\right) \tag{13a}$$

$$\Pr\left(Y < 0 \mid X_1 = a\right) = P \times \Pr\left(n < -\left(a + b\right)\right)$$
$$+ \left(1 - P\right) \times \Pr\left(n < -\left(a - b\right)\right)$$
$$= P \times Q\left(\frac{a + b}{\sigma}\right) + \left(1 - P\right) \times Q\left(\frac{a - b}{\sigma}\right) \tag{13b}$$

The first equality (13a) is obtained by using the decision regions illustrated in Figure 1 and the law of total probability. The last equality (13b) follows by using the definition of the Q-function for a Gaussian AWGN random variable with zero mean and variance σ^2. Similarly:

$$\Pr\left(Y > 0 \mid X_1 = -a\right) = PQ\left(\frac{a - b}{\sigma}\right) + \left(1 - P\right)Q\left(\frac{a + b}{\sigma}\right). \tag{13c}$$

Proposition 3 results by replacing expressions from equations (13b) and (13c) into (13a), using Equation (5), and replacing $SNR_1 = E_1/\sigma^2$ and $SNR_2 = E_2/\sigma^2$ in the previous expression.

INTENTIONAL BLANK

THE APPROPRIATENESS OF THE FACTUAL DENSITY AS AN INFORMATIVENESS MEASURE FOR ONLINE NEWS

Filipo Sharevski

Department of Computer and Information Technology,
Purdue University, West Lafayette, USA
f.sharevski@purdue.edu

ABSTRACT

In circumstances where the receptivity of the online news is affected by the media bias in covering public attention events, the quality of the textual component is of pervasive importance for a reliable perception of their informativeness. Aware of this threat, several natural language processing techniques have been developed for the purpose of capturing the quality of the web content based on the concepts of objectivity classification and stylometric features, knowledge maturing, factual density, or simple word count. This paper explores the appropriateness of the factual density as an adequate quality measure of the information reported on the missing Malaysia Airliners Flight 370 as a public attention event. The results suggest that the factual density needs to be applied under strict conditions in terms of increased confidence level of the textual news content, if its substance is a subject of capitalization as a referent source of information.

KEYWORDS

Factual Density, Natural Language Processing, Text Informativeness

1. INTRODUCTION

Being today's most attractive content delivery platform [1], the global Internet undoubtedly is the central location where "almost all consumers search for information in order to better inform their decisions" whether for making a purchase, read the news, realize their travel plans, improve the personal health or manage their personal finances [2]. Trumping the other content and information delivery platforms or channels when it comes to influence" [2] suggest that the information retrieved from any Internet source have a significant impact on the way our world views and perceptions are shaped with an intensive consumption of the online content [3]. Thus, the information retrieval in the digital realm needs to rely not just on the plausibility of the content itself, but should also take into consideration the "credibility and reliability of the source and the quality aspects characteristic for this content" [4], [5]. Given that the online information is characterized with a high receptivity for a narrower and subjective purpose, the impact of the information produced and disseminated through the online news on the formation of the general public brings an interesting topic for discussion.

In this context, of practical interest is to assess the quality of online media coverage of a certain event of a great public attention, together with the reliability and the plausibility of the respective information for the purpose of creating a reference article (i.e. Wikipedia article). However, the quality assessment of the online media coverage in a whole is "too general and practically impossible" [5], given the various modalities included in the production of the news, such as text, video, audio, pictures or different animations. Therefore, it is useful to restrict the online news quality assessment only to the textual portion of the news' web feeds, since the text is the skeleton bearing the critical information about given event, supported by the additional set of pictorials or videos (in fact, many related news videos render the same content included in the textual part only in an audio-visual format). As such, the focus of the undertaken effort rests on the idea that "the quality of a text document can be related to its informativeness, i.e. the amount of useful information contained in a document" [5].

Unlike the other metrics designed to capture the quality of the web content based on objectivity classification and stylometric features [6], knowledge maturing [7], or simple word count [8], the concept of informativeness incorporates "complex semantic features" in order to better instrument the textual quality assessment [4]. According to [4] and [5], the "informativeness of a document can be measured through factual density of a document, i.e. the number of facts contained in a document, normalized by its length". The rationale for using the factual density comes from the fact that it assess the influential aspects of the textual news portion in more general terms compared to the objectivity classification, and that the online news reception precedes the transformation process of "contextualized information artefacts into explicitly linked and formalized learning objects", as suggested by [7]. Moreover, it has been proven that it outperforms the word count measure [4], [5] in terms of web content quality assessment.

With this in place, the online news' informativeness assessment is directed towards the estimation of the actual appropriateness of the factual density as an adequate quality measure of the information reported about an event of great public attention. Given the unpreceded public attention drawn by "the largest-ever multinational air-sea search" [9] for the missing Malaysia the information relative to this incident is set at the centre of the web news' informativeness analysis. For this purpose, the online news reports collected from two different sources – Fox News and Thompson Reuters – and compared to the factual density each of the information reporting trails disseminates in the public realm of Internet journalism. Each of the datasets includes a daily review article respective to the Malaysia Airlines Flight 370 within the period of 33 days after the aircraft disappeared from the regular route, elaborating on the progress of the multinational air-sea search by bringing additional facts and potential evidence leads. The factual density of each of the respective articles in the two data sets – automatically calculated using the ReVerb [10] – is compared relative to the overall ranking of each of the articles in two instances: one where the threshold for confidence scores of the automatic factual determination is not taken into consideration, and another where this confidence threshold is considered in context of achieving an optimal balance between the recall and the precision of the extracted factual relations. The outcome of the two comparisons is finally discussed respective to the appropriateness of the use of the factual density as a measurement instrument for assessing the quality of the online news reports.

2. FACTUAL DENSITY AS AN INFORMATIVENESS MEASURE

The factual density $f_d(t)$ of an arbitrary textual resource t is the ratio between the fact count $f_c(t)$ and the size of the text $size(t)$. As pointed in [4] and [5], the most applicative approach for operationalization of the factual density is the utilization of the "Open IE methods, justified by the fact that the extraction output brings the "relational tuples representing facts from text [subject of the extraction] without requiring a pre-specified vocabulary or manually tagged training

corpora" [5], which in turn "makes such systems scalable to a broader set of relations and to a far larger corpora as the Web" [11]. Assuming a model of textual resource "as a bag-of-relations occurring in all facts extracted from it", the relational tuple as a unit output of the Open IE systems is structured as a "triplet consisting of two arguments in the form of noun phrases and a relational phrase expressed as a verb phrase" [4], i.e. f = (*10 nations*, *failed to turn up*, *any trace of the Boeing 777 – 200R*). Having the unit dividend in the factual density formula, the actual value of the fact count can be derived by simple summation operation of all relational tuples produced as an outcome of the Open IE system.

Since the basic definition brings no explicit details on the actual granularity of the size variable, there are three possibilities for text quantization: the overall character count including the white spaces, the word count, and the number of the sentences in the text. The most appropriate text quantization considers the word count as the text size, given the fact that it mainly strikes balance between potential miscalculation of the factual density due to an excessive normalization when the character count is used as a quantifying unit, or potential factual density degradation due to a considerably coarse quantization step in case of a sentence count. In case where the factual density is calculated relative to the character count, a potential threat exists in terms of text manipulation as to decrease (or potentially) increase the number of characters, yielding to an overall text degradation both in semantic and syntactic terms. On the other side, using the sentence unit quantization may mimic an actually large factual density of texts that are semantically and syntactically rich, but contain very few sentences. Having the word count already proposed as an actual quality measurement for web contents [8], the rationale for its choice is also valid considering a potential cross-metrics comparison.

In regards to the general threat of factual density degradation, it is also useful to reconsider the Open IE systems employed for factual derivation. In fact, "the output of the Open IE systems contains a large portion of uninformative and incoherent extractions" [5], [11] that are anyway determined as a relational tuple, which are also considered as facts that add into the final fact count and lead to overestimation of the factual density in a given textual resource. Incoherent extraction refers to a relation tuple or phrase that "has no meaningful interpretation", coming as a result from the fact that "these systems often yield to an incomprehensible predictions" [10]. On the other side, the uninformative extraction refers to an extraction that omits critical information and is caused by "improper handling of relation tuples that are expressed by a combination of a verb with a noun, such as the light verb constructions" [10]. As a countermeasure to these threads, [10] introduced two syntactic constraints: (1) "every multi-word relation phrase must begin with a verb, end with a preposition, and be a contiguous sequence of words in the sentence"; and (2) "a binary relation phrase ought to appear with at least a minimal number of distinct argument pairs in a large corpus", which essentially enable factual derivation "without requiring a pre-specified vocabulary or manually tagged training corpora" from a given text resource. Incorporating these improvements intended to minimize the overestimation of the factual density, Authors in [10] developed the ReVerb extractor that takes as input "noun-phrase chunked sentences from an arbitrary text resource and produces relational tuples representing the facts included within it" [12]. Given that the ReVerb "outperforms the other Open IE systems [13], [10], by achieving considerably higher precision and recall" [5][10][12], it is chosen as tool for the automatic factual density derivation in executing the central analytical engagement.

3. RELATED WORK

The factual density as a informativeness measurement of a given textual web content was initially employed by [4] and [5] in their "performance comparisons of the factual density against the word count measurement in respect to identification of featured/good articles in Wikipedia", and the usability evaluation "of relational the features for categorizing Wikipedia articles into

featured/good versus non-featured ones". The outcome of the performance comparison provided with an evidence that the factual density "corroborates the good performance of word count in Wikipedia in case the article are with variable length", but in the case of articles of similar lengths "the word count measure fails while factual density can separate between them with an F-measure of 90.4%".

Trying to assess whether the aforementioned conclusions are applicable beyond the specific class of online textual content such as Wikipedia articles, [5] "extend the analytical effort into the sphere of the arbitrary Internet texts" retrieved from various, randomly selected online contents by comparing a human-ranked reference corpus against an analogous raking derived from an Open IE factual density estimation system, consisting of 50 arbitrary web text on the material in Spanish. Correlating the "ground truth human-annotated ranking with the ranking gathered from the automatic prediction", [5] were able to confirm the adequacy of the factual density as a measurement of the informativeness, at least in the case of Spanish web texts.

In contrast to these efforts, this work approaches the web content informativeness measurement - based on the factual density – as to assess the "actual feasibility of the automatic factual density estimation" [5] for text documents containing online news reports, instead of Wikipedia articles or arbitrary web content. By shifting the research commitment beyond the specific domains analysed by [4] and [5] towards a segment of an English-based web content of considerable interest especially in the case of public opinion formation [2], the undertaken analytical effort takes a recent event with massive public attention and news coverage - the incident involving the missing Malaysia Airlines Flight 370 and the ongoing largest-ever multinational air-sea search – as to determine whether an automatic evaluation of the informativeness of different online news feeds brings plausible and unified reporting information that can be utilized for the purpose of deriving a reliable incident briefing or creating a reference article.

4. EXPERIMENTAL SETUP AND RESULTS

Two separate datasets corresponding to the daily reporting on the investigation and search progress were compiled from the Fox News and Thompson Reuters. The Fox News feed is taken as referent since it was continuously ranked 1st in the news ranking for more than 147 months, including March, 2014 [14], which corresponds with more than 72 % of the time period of online news coverage of the incident that was initially reported on March 8, 2014. The addition of Thompson Reuters reflects the idea of balancing the potential thread of polarized online coverage [15]. Each of the datasets includes a daily review article about the Malaysia Airlines Flight 370 in a period of 33 days (between March 8 to April 9, 2014) after the aircraft disappeared, elaborating on the progress of the multinational air-sea search. Special attention was devoted to collect those daily articles that are covering the relevant topic in the overall progress in that particular point of time. The choice for this timeframe rests on the fact that the flight data recorder (able to record up to 25 hours of flight information) contains a so-called "underwater locator beacon" that can emit beacon ping signals (on 37.5 KHz) up to 30 days from the moment of the incident [16], assuming maximum operational lifetime of the flight data recorder battery (the additional reports from April 7th to April 9th as account for possible fluctuations in both the recorder's battery lifetime and the underwater locator beacon signal).

After the textual content of all the 66 web articles were manually extracted (as to ensure that the substance of the overall report is preserved), the factual density of each of these articles in the two data sets was automatically calculated. In the first instance, the threshold for confidence scores is not taken into consideration for the factual extraction in order to determine the general degree of online news informativeness. Since ReVerb also outputs a confidence score for each of the factual extractions, in the second instance a confidence threshold of 0.6 is considered in the context of

achieving an optimal balance between the recall and the precision of the extracted factual relations [17] as an improved measurement of the informativeness. Another reason for considering a considerable higher confidence threshold value is that "it reduces the need for human fact checking" [17], which corresponds with the fact that the informativeness assessment and the comparison outlined in this work does not involve human-annotated ranking. The resultant factual density rankings of the two data sets in the first and second instance are given in Table 1a and Table 1b, and Table 2a and Table 2b respectively (see Appendix). Following a similar statistical approach as taken in [5], Table 3 provides with details on Spearman's ρ rank correlation coefficient between the two data sets in both of the evaluation instances, as well as the respective values of the significance level. While [5] utilize the correlation outcome of the statistical analysis as to support its hypothesis on the adequacy of the factual density of the online web content relative to a human-annotated referent set, in context of our analysis the statistical analysis is used as to determine the feasibility of the automatic factual density calculation as informativeness metrics of the online news contents.

Table 3. Correlation tests on factual density ranking of online news articles using spearman's rank correlation coefficient

Confidence Threshold	Spearman's ρ	p-value	Significance Level
n/a	-0.013	0.471	52.9%
0.6	-0.235	0.094	90.6%

5. THE APPROPRIATENESS OF THE FACTUAL DENSITY AS AN INFORMATIVENESS MEASUREMENT FOR ONLINE NEWS

As Table 3 suggests, the factual density is far from an appropriate informativeness measurement for online reports of worldwide public interest. Given the enormous attention devoted to the investigation progress of the missing Malaysia Airlines Flight 370, an intuitive conjecture is to expect that most of the articles will yield with a factual set that can easily be utilized as a relevant and highly informative source on the overall incident. However, the general comparison suggests that this is not the case, questioning the impartibility and the reliability of the online news as an "influential source for making informed decisions" [2], [3], as well as the actual appropriateness of the ReVerb Open IE tool used in the factual count calculation.

Nevertheless, reconsidering the general factual density derivation effort in terms of increased confidence score on the factual counting brings presumably sufficient corroborating evidence on the adequacy of the factual density as an informativeness measurement of online news articles. The considerable improvement in the correlation coefficient – although still considered weak – brings the value of the significance level acceptable to a certain degree relative to the task of utilizing the online news as a referent source of information. Again, the values obtained in the analytical instance considering the confidence threshold of 0.6 by no means eliminate the general conclusion outlined above, suggesting that despite the fact that the factual density can be appropriately employed for the purpose of quality assessment under certain conditions, it still has to be taken with a degree of caution considering the inherit threat of journalistic bias.

In this context, it is also interesting to consider the rankings relative to utilization of the news articles in building a chronological overall report of the incident. Relying only on the assumed discriminative power of the factual density over the quality of a given online news recourse, the comparison of the rankings outlined in Table 2a and Table 2b suggests an unbalanced potential set of chronological references, having the Fox News articles dominating the incident briefing at the period of the very beginning and the critical end of flight data recorder battery lifetime, leaving the Thompson-Reuters articles to dominate the other portion of the chronological

reporting with a slightly better level of informativeness. Aware of the fact that behind this particular outcome various different causal relationships can be identified, it was brought to the discussion more from the information assurance and security perspective on the utilization of the informativeness metrics. As such, it is worth mentioning that the informativeness – measured using the factual density - can easily be manipulated and exploited as to produce a different, but equally plausible, set of chronological summaries of the incident by compiling daily reports to contain a high number of potentially extractable facts. Having *a priori* favourable fact count, one can easily modify the article to include facts that can lead to inferences in the final chronological report pre-programmed to correspond to its malicious intensions. Furthering the argument towards the actual appropriateness of the ReVerb tool for an informativeness assessment, the overall findings also suggest that despite the claim that ReVerb fails to utilize the semantic dimension in order to improve the robustness of the factual extraction, since no significant resilience level of the factual density can be recognized according to the overall analytical outcome.

6. CONCLUSION

The work presented in this paper reveals interesting aspects on the appropriateness of the applicability of the factual density measurement in the attempt for an automatic derivation of the online news articles quality in terms of informativeness. Considering the nature of the online journalism together with the limitations of the factual determination relative to the involvement of a strong semantic component yield to a general conclusion that the informativeness assessment is valid under strict conditions of increased confidence level of the automatic factual extraction. Since the analytical effort focused on evaluation only on the automatically-derived factual density ranking, a plausible future direction of the appropriateness assessment is to compare these results against human-annotated ranking as to determine the overall usability of the online news content, at least in case of massive event reporting. Notwithstanding the useful informativeness included within the analysed reports, a year after the incident the Malaysian Airlines Flight 370 is still missing without any optimistic lead on critical evidence – including the flight data recorder - that can possibly shed light on the overall investigation.

APPENDIX

Table 1a. Factual density ranking of the Fox News articles without the confidence score

Date	Fact Count	Word Count	Factual Density $f_d(t)$
3/8/2014	69	1345	0.051301115
3/9/2014	10	203	0.049261084
3/10/2014	70	992	0.070564516
3/11/2014	44	778	0.05655527
3/12/2014	67	1,015	0.066009852
3/13/2014	66	1,039	0.063522618
3/14/2014	44	678	0.064896755
3/15/2014	6	130	0.046153846
3/16/2014	73	1,104	0.066123188
3/17/2014	85	1,448	0.058701657
3/18/2014	15	146	0.102739726
3/19/2014	27	897	0.030100334
3/20/2014	87	1,177	0.073916737
3/21/2014	45	787	0.057179161
3/22/2014	68	1,090	0.062385321
3/23/2014	82	1,263	0.064924782

Date	Fact Count	Word Count	Factual Density
3/24/2014	72	1,150	0.062608696
3/25/2014	67	1139	0.058823529
3/26/2014	78	1315	0.059315589
3/27/2014	69	1088	0.063419118
3/28/2014	70	1170	0.05982906
3/29/2014	7	113	0.061946903
3/30/2014	67	1434	0.046722455
3/31/2014	75	989	0.075834176
4/1/2014	50	828	0.060386473
4/2/2014	73	1170	0.062393162
4/3/2014	60	1067	0.056232427
4/4/2014	20	497	0.040241449
4/5/2014	45	497	0.09054326
4/6/2014	13	283	0.045936396
4/7/2014	73	1124	0.064946619
4/8/2014	42	823	0.051032807
4/9/2014	35	619	0.056542811

Table 1b. Factual density ranking of the Thomson Reuters articles without the confidence score

Date	Fact Count	Word Count	Factual Density $f_d(t)$
3/8/2014	64	979	0.065372829
3/9/2014	30	518	0.057915058
3/10/2014	95	1,196	0.079431438
3/11/2014	70	958	0.073068894
3/12/2014	10	724	0.013812155
3/13/2014	66	941	0.070138151
3/14/2014	43	666	0.064564565
3/15/2014	85	1,253	0.067837191
3/16/2014	90	1,346	0.066864785
3/17/2014	25	517	0.048355899
3/18/2014	70	996	0.070281124
3/19/2014	72	1033	0.069699903
3/20/2014	63	1,020	0.061764706
3/21/2014	62	991	0.062563068
3/22/2014	72	1,122	0.064171123
3/23/2014	50	803	0.062266501
3/24/2014	63	903	0.069767442
3/25/2014	37	655	0.05648855
3/26/2014	71	1091	0.06507791
3/27/2014	18	340	0.052941176
3/28/2014	55	975	0.056410256
3/29/2014	36	680	0.052941176
3/30/2014	46	726	0.063360882
3/31/2014	22	418	0.052631579
4/1/2014	39	812	0.048029557
4/2/2014	45	696	0.064655172
4/3/2014	24	409	0.058679707
4/4/2014	49	712	0.068820225
4/5/2014	52	778	0.066838046
4/6/2014	50	732	0.068306011
4/7/2014	45	669	0.067264574
4/8/2014	50	935	0.053475936
4/9/2014	55	1007	0.054617676

Table 2a. Factual density ranking of the Fox News articles with a 0.6 confidence score

Date	Fact Count	Word Count	Factual Density $f_d(t)$
3/8/2014	35	816	0.042892157
3/9/2014	3	79	0.037974684
3/10/2014	25	802	0.03117207
3/11/2014	15	313	0.047923323
3/12/2014	20	578	0.034602076
3/13/2014	16	473	0.033826638
3/14/2014	13	449	0.028953229
3/15/2014	2	59	0.033898305
3/16/2014	25	750	0.033333333
3/17/2014	30	927	0.03236246
3/18/2014	0	0	0
3/19/2014	7	190	0.036842105
3/20/2014	32	799	0.040050063
3/21/2014	45	787	0.057179161
3/22/2014	21	576	0.036458333
3/23/2014	31	980	0.031632653
3/24/2014	22	671	0.032786885
3/25/2014	19	710	0.026760563
3/26/2014	31	999	0.031031031
3/27/2014	22	686	0.032069971
3/28/2014	24	563	0.042628774
3/29/2014	1	12	0.083333333
3/30/2014	24	850	0.028235294
3/31/2014	24	374	0.064171123
4/1/2014	16	487	0.032854209
4/2/2014	27	825	0.032727273
4/3/2014	25	715	0.034965035
4/4/2014	14	379	0.036939314
4/5/2014	14	391	0.035805627
4/6/2014	5	145	0.034482759
4/7/2014	15	397	0.037783375
4/8/2014	13	341	0.038123167
4/9/2014	12	207	0.057971014

Table 2b. Factual density ranking of the Thomson Reuters articles with a 0.6 confidence score

Date	Fact Count	Word Count	Factual Density $f_d(t)$
3/8/2014	33	829	0.039806996
3/9/2014	9	225	0.04
3/10/2014	68	1,238	0.054927302
3/11/2014	29	936	0.030982906
3/12/2014	4	45	0.088888889
3/13/2014	32	838	0.038186158
3/14/2014	9	291	0.030927835
3/15/2014	29	786	0.036895674
3/16/2014	32	789	0.040557668
3/17/2014	10	240	0.041666667
3/18/2014	27	640	0.0421875
3/19/2014	22	754	0.029177719
3/20/2014	24	632	0.037974684

3/21/2014	21	593	0.035413153
3/22/2014	37	776	0.047680412
3/23/2014	25	523	0.047801147
3/24/2014	24	609	0.039408867
3/25/2014	16	335	0.047761194
3/26/2014	31	713	0.043478261
3/27/2014	10	216	0.046296296
3/28/2014	19	511	0.037181996
3/29/2014	23	507	0.045364892
3/30/2014	22	492	0.044715447
3/31/2014	11	226	0.048672566
4/1/2014	19	501	0.037924152
4/2/2014	16	368	0.043478261
4/3/2014	10	228	0.043859649
4/4/2014	19	442	0.042986425
4/5/2014	18	546	0.032967033
4/6/2014	18	518	0.034749035
4/7/2014	17	491	0.034623218
4/8/2014	17	525	0.032380952
4/9/2014	18	410	0.043902439

REFERENCES

[1] Nielsen, "111 Million Americans go on the Web for Daily News," Newswire, 2013. [Online]. Available: http://www.nielsen.com/us/en/newswire/2013/todays-top-headline-111-million-americans-go-on-the-web-for.html.

[2] Fleishman Hillard, "Digital Influence Index 2012," London, UK, 2012.

[3] Reuters Institute - Oxford University, "Digital News Report 2013," Oxford, UK, Jun. 2013.

[4] E. Lex, M. Voelske, M. Errecalde, E. Ferretti, L. Cagnina, C. Horn, and M. Granitzer, "Measuring the Quality of Web Content using Factual Information," in Proceedings of the 2nd Joint WICOW/AIRWeb Workshop on Web Quality, 2012, no. iii, pp. 7–10.

[5] C. Horn, A. Zhila, A. Gelbukh, R. Kern, and E. Lex, "Using Factual Density to Measure Informativeness of Web Documents," in In Proceedings of the 19th Nordic Conference on Computational Linguistics (NODALIDA), 2013.

[6] E. Lex, A. Juffinger, and M. Granitzer, "Objectivity classification in online media," in Proceedings of the 21st ACM conference on Hypertext and hypermedia - HT '10, 2010, pp. 293–294.

[7] N. Weber, K. Schoefegger, T. Ley, S. Lindstaedt, A. Brown, and S. Barnes, "Knowledge Maturing in the Semantic MediaWiki : A Design Study in Career Guidance," Lect. Notes Comput. Sci., vol. 5794, pp. 700–705, 2009.

[8] J. E. Blumenstock, "Size Matters : Word Count as a Measure of Quality on Wikipedia," in Proceedings of the 17th international conference on World Wide Web, 2008, pp. 1095–1096.

[9] S. Neuman, "Search For Flight MH370 Reportedly Largest In History," NPR, 2014. [Online]. Available: http://www.npr.org/blogs/thetwo-way/2014/03/17/290890377/search-for-flight-mh370-reportedly-largest-in-history.

[10] A. Fader, S. Soderland, and O. Etzioni, "Identifying Relations for Open Information Extraction," in EMNLP '11 Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2011, pp. 1535–1545.

[11] D. S. Etzioni, O., Banko, M., Soderland, S., & Weld, "Open Information Extraction from the Web," Commun. ACM, vol. 51, no. 12, pp. 68–74, 2008.

[12] P. Gamallo, M. Garcia, and S. Fern, "Dependency-Based Open Information Extraction," in Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, 2012, pp. 10–18.

[13] F. Wu and D. S. Weld, "Open Information Extraction using Wikipedia," in Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 2010, pp. 118–127.

[14] D. Irvine, "Cable News Ratings: Fox Still Number One—MSNBC, CNN Shedding Viewers," Accuracy in Media, 2014. [Online]. Available: http://www.aim.org/don-irvine-blog/cable-news-ratings-fox-still-number-one-msnbc-cnn-shedding-viewers/.

[15] R. Sambrook, "Delivering trust: impartiality and objectivity in the digital age," 2012.

[16] Investigations Office of Aviation Safety, "Flight Data Recorder Handbook for Aviation Accident," 2002.

[17] F. Wu and D. S. Weld, "Autonomously semantifying wikipedia," in Proceedings of the sixteenth ACM conference on Conference on information and knowledge management - CIKM '07, 2007, pp. 41–50.

AUTHORS

Filipo Sharevski (born 1985) is a Ph.D. Candidate at the Purdue University in the domain of Cybersecurity. In 2008 and 2011 he earned the B.Sc. and M.Sc. degrees in Telecommunications from the Faculty of Electrical Engineering and Information Technologies at University "Ss. Cyril and Methodius" in Skopje, Macedonia. After graduation, he was affiliated with the Institute of Telecommunications at the Faculty of Electrical Engineering and Information Technologies, Vip Macedonia Mobile Network Operator as networking engineer, and with Purdue University as a teaching and research assistant in the domain of Cybersecurity. His research interests are primarily focused on cyber forensics, network security, advanced mobile communication technologies and cyber warfare, as well as the natural language information assurance and security.

MULISA : A New Strategy for Discovery of Protein Functional Motifs and Residues

Yi-Yuan Chiu[1], Chih-Ta Lin[1], Chien-Heng Lin[1] and Jinn-Moon Yang[1,2]

[1]Institute of Bioinformatics and Systems Biology,
National Chiao Tung University, Hsinchu, Taiwan.
`y2chiu.bi98g@g2.nctu.edu.tw`
`mcjethrolin.bi97g@g2.nctu.edu.tw`
[2]Department of Biological Science and Technology,
National Chiao Tung University, Hsinchu, Taiwan.
`moon@cc.nctu.edu.tw`

ABSTRACT

To predict and identify details regarding function from protein sequences is an emergency task since the growing number and diversity of protein sequence. Here, we develop a novel approach for identifying conservation residues and motifs of ligand-binding proteins. In this method, called MuLiSA (Multiple Ligand-bound Structure Alignment), we first superimpose the ligands of ligand-binding proteins and then the residues of ligand-binding sites are naturally aligned. We identify important residues and patterns based on the z-scores of the residue entropy and residue-segment entropy. After identifying new pattern candidates, the profiles of patterns are generated to predict the protein function from only protein sequences. We tested our approach on ATP-binding proteins and HEM-binding proteins. The experiments show that MuLiSA can identify the conservation residues and novel patterns which are really correlated with protein functions of certain ligand-binding proteins. We found that our MuLiSA can identify conservation patterns and is better than traditional alignments such as CE and CLUSTALW in some ligand-binding proteins. We believe that our MuLiSA is useful to discover ligand-binding specificity-determining residues and functional important patterns of proteins.

KEYWORDS

Multiple ligand-bound structure alignment, Functional motifs, Conserved patterns, ATP-binding proteins, HEM-binding proteins.

1. INTRODUCTION

Human genome have been sequenced and led to a flood of sequence information. On the other hand, recent developments in X-ray crystallography and NMR have made it faster in solving protein structures. These data contains a lot of information that can be extracted by techniques which were used to visualize the sequence conservation information. The residues most related to the functions of a protein are often the most conserved (1). Many studies have demonstrated that most protein domains of same protein families, such as PROSITE (2) and Pfam (3), share conserved peptide patterns, called motifs, and some critical residues. The fundamental problems in proteomics include both identifying and understanding the role of the essential sites that determine that structure and proper function of the proteins. After solving these problems,

researchers can apply this useful information as a clue to predict protein functions without protein structure information.

Many groups have used the identification of conserved patterns as a method to predict protein function. Some of these groups predict protein motifs using principle component analysis (4-7). Other groups use structure alignment (8) or sequence alignment (9) as a method to identify conservation sites. Evolutionary trace analysis was used to predict functional patterns in different phylogenic trees and look for functional important residues (8, 10-13). However, these methods always use protein structure or protein sequence information to predict protein conservation patterns and may miss these conservation patterns because of the noises from other protein structures which are far apart from ligand-binding site.
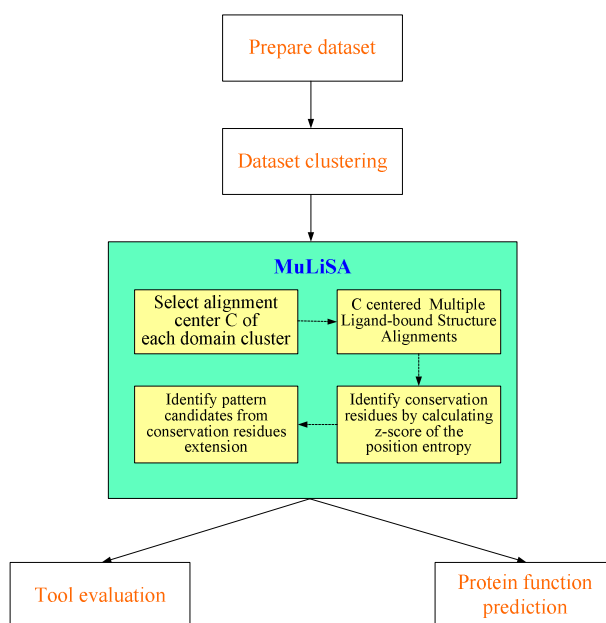

Figure 1. The workflow of analysis and identification of conservation patterns and residues in proteins by MuLiSA. This flow starts from dataset preparation and clustering, followed by multiple ligand-bound structure alignments (MuLiSA), tool evaluation and protein function prediction.

2. MATERIALS AND METHODS

Identification of conservation patterns and residues in proteins by multiple ligand-bound structure alignments encompasses a variety of sequential computational phases, including dataset preparation, dataset clustering, multiple ligand-bound structure alignments, post-alignment analysis and entropy calculation, tool verify and protein function prediction (Figure 1). In dataset preparation, we first select one kind of ligand-binding protein that we are interested and get ligand-binding protein list from PDBsum (18) database. Because we need precise protein structures to identified conservation residues and motifs, we only select protein structures resolved by X-ray diffraction. Then we select ligand-binding domains using programs from SCOP database (19). In data clustering, we generate all-against-all multiple ligand-bound structure alignments of these selected ligand-binding domains and generate one structure similarity matrix and one un-gapped sequence identity matrix for each kind of ligand-binding proteins. Once we have these two matrixes, we select non-redundant protein domains, and undergo protein domain clustering.

In the main step of MuLiSA, first we choose the alignment center domains C of each domain cluster based on structure similarity. Second, we undergo C centered multiple ligand-bound structure alignment. After we generate the alignments, z-score calculation of position entropy can help us to identify conservation residues of each domain cluster. For we believed that the functional important motifs mostly composed of functional important residues, we identified pattern candidates by conservation residues extension. Finally, we used SCOP (19) and PROSITE (2) databases to verify our results; and then we generate profiles of pattern candidates and use them to search for protein sequence with these patterns in SWISS-PROT database (20).

2.1. Preparation of ligand-binding proteins

We have applied MuLiSA to three kinds of ligand-binding proteins, which are ATP-binding proteins, ADP-binding proteins, and HEM-binding proteins. The ligand-binding protein lists were taken from PDBsum database (18). In order to get ligand-binding domains, first we need to get ligand-binding protein structures. Protein structure three-dimensional information was downloaded from Protein Data Bank (PDB) database(21) according to ligand-binding protein lists getting from PDBsum database (18). The ligand-binding domains were chosen downloaded from Structure Classification of Proteins (SCOP) database (19).

Ligand-binding domains were chosen with four criteria, they are as follows: 1) When one of distances between atoms of residues of the domain and atoms of ligands is near than 5Å, we think that this domain is a ligand-binding domain. 2) Because multiple ligand-bound structure alignment first superimposed the ligands of aligned proteins, we only choose protein domains which only bind with one ligand. 3) We only choose ligand-binding domains which the ligand they bind is only bind by one protein domain. 4) We only choose one protein domain in one protein structure. Because the SCOP domain files do not contain ligand information, after choosing these domains we must add back ligand information from Protein Data Bank (PDB) database (21) into these protein domain files. It must be mentioned that we only choose protein domains solved by x-ray crystallography because we think that these structures are more convincing.

2.2. Datasets for verification

To verify whether our alignment results is reasonable and can reflect protein function information, we use the classification of Structural Classification of Proteins (SCOP) database (19) as the benchmark of our structure similarity matrix for non-redundant domain clustering. PROSITE patterns from PROSITE database (2) were also used to quality assessment and refinement of multiple ligand-bound structure alignments. The protein sequences and annotations were downloaded from SWISS-PROT database (20) and were used for profile verification and protein function prediction.

2.3. Method

The main idea of this tool is that we try to align together conservation residues of proteins at ligand-binding sites by ligand superimposition; and then identify conservation residues and patterns by z-score of entropy calculation. Because we have to change the three-dimensional coordinates of proteins along with superimposed ligands, we developed a structure superimpose tool to deal with this problem. We developed this program MuLiSA from ICP algorithm(22), this program can make proteins and ligands rotation and displacement on three-dimensional space. After we get the superimposed protein structures, we regard two residues are aligned together based on three order rules: 1) Rule 1: Cβ or Cα (Gly) atom of amino acid residues in 1Å; 2) Rule 2: Cβ or Cα (Gly) atom of same amino acid residues in 4Å; 3) Rule 3: Cβ or Cα (Gly) atom of

same group amino acid residues in 4Å or Cβ or Cα (Gly) atom of different group amino acid residues in 2Å.

The amino acid groups are defined as follows: 1) Basic amino acids: lysine, arginine, and histidine; 2) Acidic amino acids: aspratate, glutamate, asparagine, and glutamine; 2) Aromatic amino acids: phenylalanine, and tryptophan; 3) Aliphatic amino acids: glycine, alanine, valine, leucine, isoleucine, and methionine; 4) Hydroxyl containing amino acids: serine, threonine, and tyrosine; 5) Disulfide-bond forming amino acid: cysteine; 6) Cyclic amino acid: proline.

2.4. Sequence identity matrix and structure similarity matrix

If two protein domains have the similar function and have highly similar structures in ligand-binding sites, these two protein domain structures should fit well in three-dimensional space. We introduced structure similarities in accordance with multiple ligand-bound structure alignments to present this information. S^T_{ab} is the structure similarity of protein domain a and protein domain b. L_a is the length (residue numbers) of protein domain a, L_b is the length (residue numbers) of protein domain b, and L is the aligned residue number of protein domains a and b. S^T_{ab} is given as $S^T_{ab} = \frac{L}{\min\{L_a, L_b\}}$. We also generate un-gapped sequence identity matrix between protein domains for non-redundant protein domain selection based on only aligned residues of protein domains a and b S^E_{ab} is the un-gapped sequence identity of protein domain a and protein domain b. mt is the number of identical aligned residues of protein domain a and protein domain b; mmt is the number of non-identical aligned residues of protein domain a and protein domain b and $S^E_{ab} = \frac{mt}{mt + mmt}$.

Redundant protein domains must be removed because the profiles generated from alignments may be incredible. We regarded two protein domains are redundant protein domains when their structure similarity and sequence identity are both above 0.8; therefore, we first cluster these protein domains and only choose one with no mutation residues and with the smallest X-ray diffraction resolution. In order to generate a convincing multiple alignments, we must choose an alignment center domains C before we generate this alignments. In structure similarity matrixes, the non-redundant protein domain of one cluster which has the highest structure similarity with other protein domains than others was selected as the alignment center C of this cluster. This protein domain was used to be the alignment center of multiple ligand-bound structure alignment.

2.5. Identification of conservation residues and pattern candidates

To identify these conservation residues, we used entropy (S_p), defined as $S_P = -\sum_{i=1}^{20} f_{pi} * \ln(f_{pi})$, where i and f_{pi} denote the ith amino acid type, the probability of finding the amino acid type i at position p. The entropy is 0 when this position is totally conserved. In order to estimate the statistical significance of the position entropy, z-score was applied to identify relative conservation positions: $Z_p = \frac{X_p - \mu}{\sigma}$, where Z_p is the z-score value of position p, σ is the standard deviation of all positions entropy, μ is the average value of all positions entropy and X_p is the entropy of position p. We identified a conservation position p when $Z_p > 2.5$.

We generate alignment profiles of pattern candidates (discovered by our MuLiSA) and PROSITE patterns from multiple ligand-bound structure alignments. $PF_{pi} = \{f^i_p\}$ where $1 \le i \le 20$, where PF_p is the profile of position p; f^i_p is the probability of the ith amino acid type at position p. We

then used profiles to search for matched protein segments in protein sequences. The search window size is the length of profiles and shifts one residue each time. Each protein sequence should have $N-(n-1)$ (N is the length of this sequence and n is the length of this pattern) profile search scores, and we suppose the segment with the highest profile search score of this protein sequence should be the pattern candidate that we are looking for. The scoring function is as

follows: $S = \dfrac{\sum_{p=1}^{n}\sum_{i=1}^{20} PF_{pi}}{n}$, where S is the profile search score, n is the length of a pattern, PF_{pi} is the

profile value of amino acid type i at position p. The score is 1 when a segment perfectly matches this profile.

Table 1. Statistics of proteins, domains and pattern candidates

Ligand name	No. of proteins [a]	No. of selected domains [b]	No. of non-redundant domains [c]	Clusters [d]	Selected alignment center C [e]	No. of important residues [f]	No. of pattern candidates [g]
ATP	173	60	46	Protein kinases catalytic subunit (7)	d1phk__	10	1
				Class I aminoacyl-tRNA synthetases (RS), catalytic domain (4)	d1maua_	16	3
Heme	1145	860	131	CCP-like (13)	d1llp__	11	3
				Cytochrome P450 (13)	d1eupa_	12	3
				Cytochrome b5 (5)	d1cyo__	4	1
				Monodomain cytochrome c (23)	d1i54a_	3	1
				Cytochrome c' (4)	d1i54a_ [h]	3	1

[a] Number of ligand-binding proteins in PDBsum database.

[b] Number of ligand-binding domains selected by our program.

[c] Number of selected non-redundant domains.

[d] The domain clusters that according to structure similarity and SCOP database classification; the domain names are based on SCOP database nomenclature. We only choose domain clusters with domain number > 3 because the alignments are more statistical meaningful; and we only choose domain clusters with PROSITE patterns because we need benchmarks to verify our results. The numbers in the parentheses are the non-redundant domain numbers of each cluster.

[e] The alignment center C domain of each cluster. The alignment center C chosen is important; because in MuLiSA, the alignment center C highly affects the alignment results.

[f] Number of conservation residues with z-score > 2.5.

[g] Number of identified pattern candidates with length equal or longer than 5 residues.

[h] We choose same alignment center C of domain clusters: monodomain cytochrome c and cytochrome c', because same pattern candidates were identified in these clusters.

3. RESULTS

In order to identify the wealth of information present in protein structures, we analyzed conservation residues and patterns in multiple ligand-bound structure alignments. Most sequence and structure alignment techniques are protein-based alignment; in other words, these techniques analyze residue conservation only by comparing protein structure or protein sequence similarity.
At the present, we have applied MuLiSA to ATP- and HEM-binding proteins and identified several conservation residues and pattern candidates. We have generated sequence profiles from multiple alignments and used them to discover protein sequences which may have these profiles. We also proved that MuLiSA is better than other tools in several cases and can discover functional information when comparing with SCOP (19) and PROSITE database(2). Our major intention was to extract protein structure information from ligand-binding proteins and apply this information to protein function prediction. Table 1 shows some statistics about the dataset we used in this study. Through getting ligand-binding protein lists, selecting ligand-binding domains, domain clustering, non-redundant domains and alignment center C selection, we use MuLiSA and z-score of entropy calculation to identified conservation residues and pattern candidates of each cluster. These identified conservation residues may be functional important and we survey the literature and it proves that some of these identified conservation residues are critical to ligand-binding or correlate with conformation stability. After pattern candidate identification, we generate profiles of these pattern candidates and use these profiles predict protein functions.

3.1. ATP-binding proteins

ATP, adenosine triphosphate, is the major energy currency of the cell. It transfers energy from chemical bonds to endergonic reactions of the cell. ATP powers most of the energy-consuming activities of cells, such as muscle contraction, synthesis of polysaccharides, active transport of ions and nerve impulse. Because of ATP is a so important compound and because of the large number of experimental data, like ATP-binding protein structures and literatures, we choose ATP-binding proteins as our first research target. We have generated structure similarity matrix of non-redundant ATP-binding domains for functional-based domain clustering, and we also identified conservation residues and pattern candidates. Finally, we used profiles of pattern candidates to undergo protein function prediction.

3.1.1 Structure similarity matrix and alignment center selection

Figure 2 shows the structure similarity matrixes and SCOP classifications of 25 non-redundant ATP-binding domains. When comparing with classifications of SCOP database (19), protein domains with higher structure similarities are usually clustered together and they are always belong to same SCOP families. As we all agree that SCOP database (19) is a convincing domain structural and functional classification database, it tells us that the multiple ligand-bound alignment and structure similarity calculation is reasonable and can reflect structural and functional information.

In Figure 2A, the domains belong to the same SCOP families are with same colors. The bold values means the structure similarity is larger than the average value of the row; in other words, the domain in this row is much similar with these compared domains than others. In this matrix, we find that most domains of same SCOP family usually have higher structure similarity with each other (see the regions with red frame), it tell us that the multiple ligand-bound structure alignment and structure similarity calculation is reasonable and can reflect structural and functional information. Figure 2B shows the SCOP classification of protein domains.

The alignment center C chosen is important; because in MuLiSA, the alignment center C highly affects the alignment results. The alignment center C of one cluster is chosen when one domain has higher structure similarity with other domains than others in this cluster. For example, the first cluster in Figure 2A, "Class I aminoacyl-tRNA synthetases (RS), catalytic domain family", contains four domains: d1gtra2, d1h3ea1, d1maua_and d1n77a2. The structure similarity summation of d1gtra2 with others is 0.3+0.39+0.34=1.03; d1h3ea1 is 0.3+0.44+0.36=1.1; d1maua_is 0.39+0.44+0.36=1.19; and d1n77 is 0.34+0.36+0.36=1.06. Therefore, we choose d1maua_ as the alignment center C of this cluster.

Figure 2 (A) Structure similarity matrix of 25 non-redundant ATP-binding domains; (B) SCOP classification of 25 non-redundant ATP-binding domains. The domains belong to same SCOP families are with same colors. The bold values means the structure similarity is larger than the average value of the row; in other words, the domain in this row is much similar with these compared domains than others. In this matrix, we find that most domains of same SCOP family usually have higher structure similarity with each other (see the regions with red frame), it tells us that the multiple ligand-bound structure alignment and structure similarity calculation is reasonable and can reflect structural and functional information. The protein domains were classified according to SCOP classification hierarchy: class, fold, superfamily, and family. The protein domains were named by SCOP database nomenclature.

3.1.2 Protein functional motifs and residues

Figure 3 shows the multiple ligand-bound structure alignment results and the identified conservation residues in "Protein kinases, catalytic subunit family" of ATP-binding domains. The identified conservation residues, aligned positions with z-score of entropy calculation > 2.5, are

close to ATP in three-dimensional space. It implies that these conservation residues may play important role in ATP-binding. In Figure 3B, the labeled residue numbers belong to protein domain d1phk__, which is the selected alignment center C of this cluster; and the red framed region means the PROSITE patterns. We observed that most identified conservation residues were on these PROSITE pattern region, it tell us that identifying pattern candidates from conservation residues extension may be a reasonable approach.

Figure 3. MuLiSA result and identified conservation residues in "Protein kinases, catalytic subunit family" of ATP-binding domains. (A) Three-dimensional distributions of identified conservation residues and the ligand superimposition. Yellow: d1phk__; blue: d1atpe_; green: d1qmza_; red: d1csn__; grey: d1hck__; pink: d1gol__; light blue: d1h1wa_; (B) Multiple ligand-bound structure alignment result of "Protein kinases, catalytic subunit family" domains. The identified conservation residues, aligned positions with z-score of entropy calculation > 2.5, are close to ATP in three-dimensional space. It implies that these conservation residues may play important role in ATP-binding. The labeled residue numbers belong to protein domain d1phk__, which is the selected alignment center C of this cluster; and the red framed region means the PROSITE patterns. We observed that most identified conservation residues were on these PROSITE pattern region, it tell us that identifying pattern candidates from conservation residues extension may be a reasonable approach.

3.1.3 Protein functional pattern verifications

In order to verify the effectiveness of profiles generated from our alignments in protein function prediction, we compare the performance in profile search between dataset 1, which contains protein sequences with PROSITE pattern; and dataset 2, which contains protein sequences not only with PROSITE pattern but also have "ATP-binding" annotations in SWISS-PROT database. Dataset 1 contains protein sequences contain PROSITE pattern: aminoacyl-transfer RNA synthetases class-I signature and dataset 2 contains protein sequences contain not only PROSITE pattern: aminoacyl-transfer RNA synthetases class-I signature but also have "ATP-binding" annotations in SWISS-PROT database. We observed that the area under curves of dataset 2 is larger than the area under curves of dataset 1. Because the profile of pattern candidates were generated from alignments of ATP-binding domains and the protein sequences in dataset 1 are not all have "ATP-binding" annotations in "KW" of SWISS-PROT database, we suppose that the

profile of pattern candidate is more convincing in ATP-binding proteins but not proteins only with PROSITE patterns.

Table 2. Hit rate comparison of dataset difference in profile verification of ATP-binding proteins

Family	PROSITE patterns and pattern candidates [a]	Dataset 1 [c]		Dataset 2 [d]	
		No. of sequence [e]	Hit rate [f]	No. of sequence	Hit rate
Protein kinases catalytic subunit	Protein kinases ATP-binding region signature	859	85.15%	773	89.18%
	Serine/ Threonine protein kinases active-site signature.		85.73%		86.67%
	Pattern candidate 1 [b]		84.79%		86.76%
	Pattern candidate 2		64.19%		68.35%
	Pattern candidate 3		71.37%		75.43%
Class I aminoacyl-tRNA synthetases (RS), catalytic domain	Aminoacyl-transfer RNA synthetases class-I signature	1129	26.61%	1056	50.42%
	Pattern candidate 1		20.18%		37.43%

[a] PROSITE patterns and pattern candidates that we identified.
[b] Pattern candidate 1 of "Protein kinases catalytic subunit family".
[c] Dataset 1: sequences only with PROSITE patterns
[d] Dataset 2: sequences with PROSITE patterns and SWISS-PROT annotations
[e] Number of sequences recorded which have PROSITE patterns in this cluster. For example, in "Protein kinases catalytic subunit family", number of sequences which have Protein kinases ATP-binding region signature or Serine/ Threonine protein kinases active-site signature is 859; hence the number of sequences of pattern candidates in this cluster is 859.
[f] Average hit rate when true positive rate are 50%, 60%, 70%, 80%, 90% and 100%.

Table 2 summarize the average hit rate of true positive rates 50%, 60%, 70%, 80%, 90% and 100% in dataset 1: sequences with PROSITE pattern, and database 2: sequences with PROSITE pattern and SWISS-PROT annotations for profile verification. We observed that whether in dataset 1 or dataset 2, the hit rate of PROSITE patterns are all higher than pattern candidates. Thus, the PROSITE pattern is really meaningful for protein sequences which have these PROSITE patterns. However, we also observed that the hit rates in dataset 2 are generally higher than hit rates in dataset 1. Because dataset 1 only contains sequences with PROSITE patterns but database 2 contains sequences with PROSITE pattern and SWISS-PROT annotations, it tell us that the profiles we generated from multiple alignments of ATP-binding proteins may be more meaningful for protein sequences with "ATP-binding" annotations in SWISS-PROT database.

In Table 3, we summarized the true-positive rates, profile scoring scores, and z-score of profile scoring scores of top 100, 500, 1000, 1500, 2000, 2500 and 3000 ranked sequences in profile scoring ranking list. We also compare the hit rates between pattern candidates and PROSITE patterns. We observed when protein sequences with profile scoring score 0.600, the true positive rate is 82.27% and the z-score is 2.87. Thus when protein sequences with profile scoring score higher than 0.600, we can say these protein sequence may be ATP-binding proteins with 82.27% confidence. When comparing with the hit rate of our defined pattern candidates and PROSITE patterns, we observed that almost all the top 3000 ranked protein sequences with "ATP-binding" annotations were all searched by pattern candidates. Although some of pattern candidates partially overlapped with PROSITE patterns, it tells us that the pattern candidates are useful for protein function prediction in ATP-binding proteins.

Table 3. Hit rate comparison of pattern candidates and PROSITE patterns in protein function prediction of ATP-binding proteins

No. of top ranked sequence [a]	True-positive rate [b]	Profile scoring score [c]	Z-score of profile scoring score[d]	Hit rate of all pattern candidates [e]	Hit rate of PROSITE pattern[f]
100	100.00% (100)	0.840	6.52	100.00% (100)	0.00% (0)
500	98.40% (492)	0.720	4.70	100.00% (492)	0.00% (0)
1000	95.70% (957)	0.650	3.63	99.79% (955)	0.21% (2)
1500	82.27% (1234)	0.600	2.87	97.65% (1205)	2.35% (29)
2000	76.65% (1533)	0.583	2.61	80.43% (1503)	19.57% (30)
2500	70.28% (1757)	0.567	2.37	94.25% (1656)	5.75% (101)
3000	61.53% (1846)	0.556	2.20	94.53% (1745)	5.47% (101)

[a] The top ranked sequence number. For example, 100 in this column means the 100 ranked sequences with highest profile scoring score in profile scoring ranking list of ATP-binding protein prediction.

[b] True positive rate of protein sequences with "ATP-binding" annotations of top 100, 500, 1000, 1500, 2000, 2500 and 3000 ranked sequences in profile scoring ranking list of ATP-binding protein prediction.

[c] The smallest profile scoring scores of top 100, 500, 1000, 1500, 2000, 2500 and 3000 ranked sequences.

[d] Z-score of profile scoring scores. The average of all SWISS-PROT sequence scores is 0.411515; the standard deviation of all SWISS-PROT sequence scores is 0.065701.

[e] Hit rate of true positives that predicted by profiles of our defined pattern candidates.

[f] Hit rate of true positives that predicted by profiles of PROSITE patterns.

3.2 HEM-binding proteins

Heme is a member of a family of compounds called porphyrins, which consist of four pyrrole rings. Heme metabolism is an important metabolic pathway because many important hemoproteins contain heme as a prosthetic group. For example, hemoglobin is a very important hemoprotein and it is an oxygen carrier in the blood. There are also cytochromes, which participate in important electron transfer reactions, and tryptophan oxygenase which is a hemoprotein of intermediary metabolism.

3.2.1 Structure similarity matrix and conservation residues

Our results show that structure similarity matrix and SCOP classifications of 40 non-redundant HEM-binding domains. The structure similarity matrix is still similar with SCOP classification and MuLiSA can apply to different kinds of ligand-binding proteins. We have also identified several conservation residues of protein domain clusters in HEM-binding proteins. Figure 4 shows the multiple ligand-bound structure alignment result and identified conservation residues in "Cytochrome b5 family" of HEM-binding domains. In Figure 4A, the identified conservation residues are closed to heme in three-dimensional space. It implies that these conservation residues may play important role in HEM-binding. In Figure 4B, the labeled residue numbers were belonged to protein domain d1cyo__, which is the selected alignment center C of this cluster, and the red framed region means the PROSITE patterns. We observed that most identified conservation residues were on these region, it also tell us that identifying pattern candidates from conservation residues extension may be a reasonable approach.
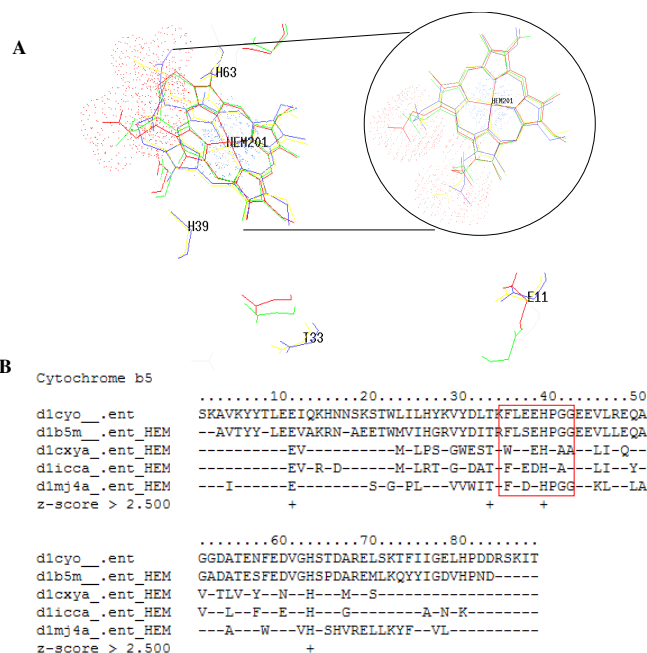
Figure 4. MuLiSA result and identified conservation residues in "Cytochrome b5 family" of HEM-binding domains. (A) Three-dimensional distributions of identified conservation residues and the ligand superimposition. Yellow: d1cyo__; blue: d1b5m__; green: d1cxya_; red: d1icca_; grey: d1mj4a_; (B) Multiple ligand-bound structure alignment result of "Cytochrome b5 family" domains. The identified conservation residues are closed to heme in three-dimensional space. It implies that these conservation residues may play important role in HEM-binding. The labeled residue numbers were belonged to protein domain d1cyo__, which is the selected alignment center C of this cluster, and the red framed region means the PROSITE patterns.

Table 4. Hit rate comparison of pattern candidates and PROSITE pattern in protein function prediction of HEM-binding proteins

Top number of sequence [a]	True-positive rate [b]	Profile scoring score [c]	Z-score of profile scoring score [d]	Hit rate of all pattern candidates [e]	Hit rate of PROSITE pattern [f]
100	92.00% (92)	0.798	4.72	100.00% (92)	0.00% (0)
200	80.50% (161)	0.744	4.00	96.27% (155)	3.73% (6)
300	69.00% (207)	0.708	3.52	97.10% (201)	2.90% (6)
400	69.75% (279)	0.692	3.30	87.81% (245)	12.19% (34)
500	70.40% (352)	0.685	3.21	90.34% (318)	9.66% (34)
600	60.33% (362)	0.685	3.21	90.61% (328)	9.39% (34)
700	57.86% (405)	0.669	2.99	91.60% (371)	8.40% (34)

[a] The top ranked sequence number.

[b] True positive rate of protein sequence with "Heme" annotations in SWISS-PROT database of top ranked sequences.

[c] The smallest profile scoring scores of top 100, 200, 300, 400, 500, 600 and 700 ranked sequences.

[d] Z-score of profile scoring scores. The average of all SWISS-PROT sequence scores is 0.436928; the standard deviation of all SWISS-PROT sequence scores is 0.071717.

[e] Hit rate of true positives that predicted by profiles of our defined pattern candidates.

[f] Hit rate of true positives that predicted by profiles of PROSITE patterns

In order to verify the effectiveness of profiles generated from our alignments in protein function prediction, we also compare the performance in profile search between datasets 1, which contains protein sequences with PROSITE pattern; and dataset 2, which contains protein sequences not only with PROSITE pattern but also have "Heme" annotations in SWISS-PROT database. In Table 4, we summarized true-positive rates, profile scoring scores, and z-score of profile scoring scores of top 100, 200, 300, 400, 500, 600 and 700 ranked sequences in profile scoring ranking list. We also compared with the hit rate of pattern candidates and PROSITE patterns. We observed that when protein sequences with profile scoring score 0.744, the true positive rate is 80.50% and the z-score is 4.00. Thus when protein sequences with profile scoring score higher than 0.744, we can say these protein sequences may be HEM-binding proteins with 80.50% confidence. When comparing the hit rate between pattern candidates and PROSITE patterns, we observed that almost all the top 700 ranked protein sequences with annotations were searched by pattern candidates. Although some of pattern candidates may partially overlap with PROSITE patterns, it tells us the pattern candidates are useful in protein function prediction of HEM-binding proteins.

3.3 Comparison with CE and CLUSTALW

Because multiple ligand-bound structure alignments only focus on ligand-binding sites, we neglect noise from protein structure apart from the ligand-binding sites and get the functional-dependent alignments of ligand-binding domains. CE and CLUSTALW are structural alignment and multiple sequence alignment tools, respectively. In Figure 5A, we find that only the alignments of MuLiSA can align together the PROSITE defined patterns together of two domains, d1maua_ and d1gtra2. In Figures 5B and 5C, we find that the shift of conservation patterns of CE alignment result. In fact, for CE uses only protein structure information to undergo structure alignment, we find that in this case the bad alignment of conservation patterns was because of a huge structure similar region apart from ATP-binding site, and it did disturb the alignment of PROSITE patterns. In other words, through ligand superimposition can only focus on ligand-binding sites and disperse noises from other region, thus the identified conservation residues and patterns will be much more related to ligand-binding.

```
A
MuLiSA:
d1maua_.ent        MKTIFSGIQ PSGVITIGNY IGALRQFVELQHEYNCYFCIVDQHAITVWQDPHELRQNIRR
d1gtra2.ent_ATP    --G---PPE PNGY-HIGH- A--I-NFFQA--G------FDD---V----------EYI--

C.E.:
d1maua_.ent        MKTIFSGIQ PSGVITIGNY IGALRQFVELQHEYNCYFCIVDQHAITVWQDPHELRQNIRR
d1gtra2.ent        --------- ---------- ----------------------------------------

CLUSTALW:
d1maua_.ent        MKTIFSGIQ PSGVITIGNY IGALRQFVELQHEYNCYFCIVDQHAITVWQDPHELRQNIRR
d1gtra2.ent        --TNFIRQI IDEDLASGKH T---TVHTRFPPEPNGYLHIGHAKSICLNFG---IAQDYKG
```
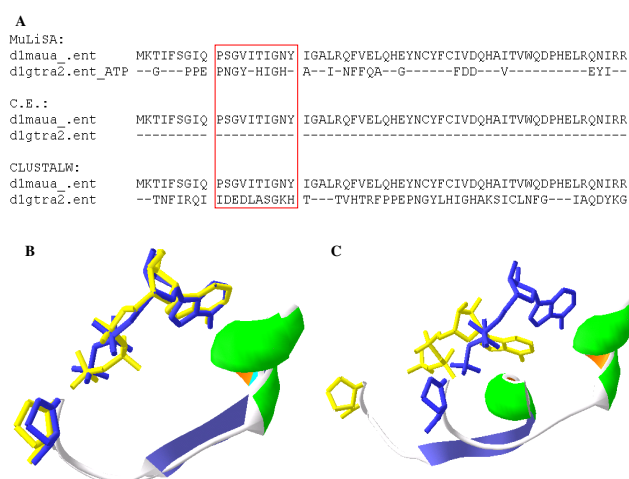

Figure 5. The comparison of MuLiSA, CE, and CLUSTALW results of two Class I aminoacyl-tRNA synthetases (RS), catalytic domains: d1maua_ and d1gtra2. (A) Alignment comparison between three methods. The shadowed region is the PROSITE defined patterns; (B) 3D structure alignment result of MuLiSA; (C) 3D structure alignment result of CE. The PROSITE defined patterns together (PROSITE pattern:

P-x(0,2)-[GSTAN]-[DENQGAPK]-x-[LIVMFP]-[HT]-[LIVMYAC]-G-[HNTG]-[LIVMFYSTAGPC]) of two domains, d1maua_ and d1gtra2. (B), two ATPs were nearly superimposed and the PROSITE patterns also aligned well. (C) The PROSITE patterns were shifted. In fact, for CE uses only protein structure information to undergo structure alignment, we find that in this case the bad alignment of conservation patterns was because of a huge structure similar region apart from ATP-binding site, and it did disturb the alignment of PROSITE patterns.

4. CONCLUSIONS

We have developed MuLiSA, a multiple ligand-bound structure alignment technique, based on functional-dependent ligand information to evaluate residue and pattern conservation. The main difference between our tool and others is that we first superimpose the ligands of proteins but not protein itself. In this way, the ligand-binding sites are superimposed naturally. Then we could identify the conservation residues and pattern candidates according to these positions and segments which were superimposed along with ligands. We have applied MuLiSA to ATP-binding proteins and HEM-binding proteins. MuLiSA can identify conservation residues and pattern candidates which play important role in ligand-binding or binding site conformation stability. For predicting protein conserved residues and patterns, our MuLiSA achieved high accuracies 80.5% and 82.3 for HEM-binding and ATP-binding proteins, respectively. We believe that our MuLiSA is very useful for prediction of conserved results, functional motifs, and protein functions.

REFERENCES

[1]   Page, R. D. M., And Holmes, E. C. (1998) Molecular Evolution: A Phylogenetic Approach, Pp 228-279, Blackwell, Oxford.

[2]   Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C. J., Hofmann, K., And Bairoch, A. (2002) The Prosite Database, Its Status In 2002, Nucleic Acids Research 30, 235-238.

[3]   Bateman, A., Coin, L., Durbin, R., Finn, R., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J., Yeats, C., And Eddy, S. R. (2004) The Pfam Protein Families Database, Nucleic Acids Research 32, D138-D141.

[4]   Casari, G., Sander, C., And Valencia, A. (1995) A Method To Predict Functional Residues In Proteins, Nature Structural Biology 2, 171-178.

[5]   Pietrokovski, S., Henikoff, J. G., And Henikoff, S. (1996) The Blocks Database--A System For Protein Classification, Nucleic Acids Research 24, 197-200.

[6]   Jones, S., And Thornton, J. M. (1997) Prediction Of Protein-Protein Interaction Sites Using Patch Analysis, Journal Of Molecular Biology 272, 133-143.

[7]   Shatsky, M., Nussinov, R., And Wolfson, H. J. (2002) Flexible Protein Alignment And Hinge Detection, Proteins: Structure Function And Genetics 48, 242-256.

[8]   hindyalov, I. N., And Bourne, P. E. (1998) Protein Structure Alignment By Incremental Combinatorial Extension (Ce) Of The Optimal Path, Protein Engineering 11, 739-747.

[9]   Higgins, D., Thompson, J., Gibson, T., Thompson, J. D., Higgins, D. G., And Gibson, T. J. (1994) Clustalw: Improving The Sensitivity Of Progressive Multiple Sequence Alignment Through Sequence Weighting,Position-Specific Gap Penalties And Weight Matrix Choice, Nucleic Acids Research 22, 4673-4680.

[10]   Sali, A., Overington, J. P., Johnson, M. S., And Blundell, T. L. (1990) From Comparisons Of Protein Sequences And Structures To Protein Modelling And Design, Trends In Biochemical Sciences 15, 235-240.

[11]   Lichtarge, O., Bourne, H. R., And Cohen, F. E. (1996) An Evolutionary Trace Method Defines Binding Surfaces Common To Protein Families, Journal Of Molecular Biology 257, 342-358.

[12]   Innis, C. A., Shi, J., And Blundell, T. L. (2000) Evolutionary Trace Analysis Of Tgf-Beta And Related Growth Factors: Implications For Site-Directed Mutagenesis, Protein Engineering 13, 839-847.

[13]   Landgraf, R., Xenarios, I., And Eisenberg, D. (2001) Three-Dimensional Cluster Analysis Identifies Interfaces And Functional Residue Clusters In Proteins, Journal Of Molecular Biology 307, 1487-1502.

[14] Notredame, C., Higgins, D. G., And Heringa, J. (2000) T-Coffee: A Novel Method For Fast And Accurate Multiple Sequence Alignment., Journal Of Molecular Biology 302, 205-217.

[15] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., And Lipman, D. J. (1990) Basic Local Alignment Search Tool, Journal Of Molecular Biology 215, 403-410.

[16] Holm, L., And Sander, C. (1993) Protein Structure Comparison By Alignment Of Distance Matrices, Journal Of Molecular Biology 233, 123-138.

[17] Gibrat, J. F., Madej, T., And Bryant, S. H. (1996) Surprising Similarities In Structure Comparison, Current Opinion In Structural Biology 6, 377-385.

[18] Laskowski, R. A., Hutchinson, E. G., Michie, A. D., Wallace, A. C., Jones, M. L., And Thornton, J. M. (1997) Pdbsum: A Web-Based Database Of Summaries And Analyses Of All Pdb Structures, Trends In Biochemical Sciences 22, 488-490.

[19] Murzin, A. G., Brenner, S. E., Hubbard, T., And Chothia, C. (1995) Scop: A Structural Classification Of Proteins Database For The Investigation Of Sequences And Structures, Journal Of Molecular Biology 247, 536-540.

[20] Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'donovan, C., Phan, I., Pilbout, S., And Schneider, M. (2003) The Swiss-Prot Protein Knowledgebase And Its Supplement Trembl In 2003, Nucleic Acids Research 31, 365-370.

[21] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., And Bourne, P. E. (2000) The Protein Data Bank, Nucleic Acids Research 28, 235-242.

[22] Besl, P. J., And Mckay, N. D. (1992) A Method For Registration Of 3-D Shapes, Ieee Transactions On Pattern Analysis And Machine Intelligence 14, 239-256.

AUTHOR INDEX