

**International Journal on  
Advances in Systems and Measurements**



The *International Journal on Advances in Systems and Measurements* is published by IARIA.

ISSN: 1942-261x

journals site: <http://www.ariajournals.org>

contact: [petre@aria.org](mailto:petre@aria.org)

Responsibility for the contents rests upon the authors and not upon IARIA, nor on IARIA volunteers, staff, or contractors.

IARIA is the owner of the publication and of editorial aspects. IARIA reserves the right to update the content for quality improvements.

Abstracting is permitted with credit to the source. Libraries are permitted to photocopy or print, providing the reference is mentioned and that the resulting material is made available at no cost.

Reference should mention:

*International Journal on Advances in Systems and Measurements, issn 1942-261x*  
vol. 15, no. 1 & 2, year 2022, [http://www.ariajournals.org/systems\\_and\\_measurements/](http://www.ariajournals.org/systems_and_measurements/)

The copyright for each included paper belongs to the authors. Republishing of same material, by authors or persons or organizations, is not allowed. Reprint rights can be granted by IARIA or by the authors, and must include proper reference.

Reference to an article in the journal is as follows:

<Author list>, "<Article title>"  
*International Journal on Advances in Systems and Measurements, issn 1942-261x*  
vol. 15, no. 1 & 2, year 2022, [http://www.ariajournals.org/systems\\_and\\_measurements/](http://www.ariajournals.org/systems_and_measurements/)

IARIA journals are made available for free, proving the appropriate references are made when their content is used.

Sponsored by IARIA

[www.aria.org](http://www.aria.org)

Copyright © 2022 IARIA

**Editors-in-Chief**

Constantin Paleologu, University "Politehnica" of Bucharest, Romania  
Sergey Y. Yurish, IFSA, Spain

**Editorial Advisory Board**

Vladimir Privman, Clarkson University - Potsdam, USA  
Winston Seah, Victoria University of Wellington, New Zealand  
Mohammed Rajabali Nejad, Universiteit Twente, the Netherlands  
Nageswara Rao, Oak Ridge National Laboratory, USA  
Roberto Sebastian Legaspi, Transdisciplinary Research Integration Center | Research Organization of Information and System, Japan  
Victor Ovchinnikov, Aalto University, Finland  
Claus-Peter Rückemann, Westfälische Wilhelms-Universität Münster / Leibniz Universität Hannover / North-German Supercomputing Alliance, Germany  
Teresa Restivo, University of Porto, Portugal  
Stefan Rass, Universität Klagenfurt, Austria  
Candid Reig, University of Valencia, Spain  
Qingsong Xu, University of Macau, Macau, China  
Paulo Esteveao Cruvinel, Embrapa Instrumentation Centre - São Carlos, Brazil  
Javad Foroughi, University of Wollongong, Australia  
Andrea Baruzzo, University of Udine / Interaction Design Solution (IDS), Italy  
Cristina Seceleanu, Mälardalen University, Sweden  
Wolfgang Leister, Norsk Regnesentral (Norwegian Computing Center), Norway

**Indexing Liaison Chair**

Teresa Restivo, University of Porto, Portugal

**Editorial Board**

Jemal Abawajy, Deakin University, Australia  
Ermeson Andrade, Universidade Federal de Pernambuco (UFPE), Brazil  
Francisco Arcega, Universidad Zaragoza, Spain  
Tulin Atmaca, Telecom SudParis, France  
Lubomír Bakule, Institute of Information Theory and Automation of the ASCR, Czech Republic  
Andrea Baruzzo, University of Udine / Interaction Design Solution (IDS), Italy  
Nicolas Belanger, Eurocopter Group, France  
Lotfi Bendaouia, ETIS-ENSEA, France  
Partha Bhattacharyya, Bengal Engineering and Science University, India  
Karabi Biswas, Indian Institute of Technology - Kharagpur, India  
Jonathan Blackledge, Dublin Institute of Technology, UK  
Dario Bottazzi, Laboratori Guglielmo Marconi, Italy  
Diletta Romana Cacciagrano, University of Camerino, Italy

Javier Calpe, Analog Devices and University of Valencia, Spain  
Jaime Calvo-Gallego, University of Salamanca, Spain  
Maria-Dolores Cano Baños, Universidad Politécnica de Cartagena, Spain  
Juan-Vicente Capella-Hernández, Universitat Politècnica de València, Spain  
Vítor Carvalho, Minho University & IPCA, Portugal  
Irinela Chilibon, National Institute of Research and Development for Optoelectronics, Romania  
Soolyeon Cho, North Carolina State University, USA  
Hugo Coll Ferri, Polytechnic University of Valencia, Spain  
Denis Collange, Orange Labs, France  
Noelia Correia, Universidade do Algarve, Portugal  
Pierre-Jean Cottinet, INSA de Lyon - LGEF, France  
Paulo Esteveao Cruvinel, Embrapa Instrumentation Centre - São Carlos, Brazil  
Marc Daumas, University of Perpignan, France  
Jianguo Ding, University of Luxembourg, Luxembourg  
António Dourado, University of Coimbra, Portugal  
Daniela Dragomirescu, LAAS-CNRS / University of Toulouse, France  
Matthew Dunlop, Virginia Tech, USA  
Mohamed Eltoweissy, Pacific Northwest National Laboratory / Virginia Tech, USA  
Paulo Felisberto, LARSyS, University of Algarve, Portugal  
Javad Foroughi, University of Wollongong, Australia  
Miguel Franklin de Castro, Federal University of Ceará, Brazil  
Mounir Gaidi, Centre de Recherches et des Technologies de l'Energie (CRTE), Tunisie  
Eva Gescheidtova, Brno University of Technology, Czech Republic  
Tejas R. Gandhi, Virtua Health-Marlton, USA  
Teodor Ghetiu, University of York, UK  
Franca Giannini, IMATI - Consiglio Nazionale delle Ricerche - Genova, Italy  
Gonçalo Gomes, Nokia Siemens Networks, Portugal  
Luis Gomes, Universidade Nova Lisboa, Portugal  
Antonio Luis Gomes Valente, University of Trás-os-Montes and Alto Douro, Portugal  
Diego Gonzalez Aguilera, University of Salamanca - Avila, Spain  
Genady Grabarnik, CUNY - New York, USA  
Stefanos Gritzalis, University of the Aegean, Greece  
Richard Gunstone, Bournemouth University, UK  
Jianlin Guo, Mitsubishi Electric Research Laboratories, USA  
Mohammad Hammoudeh, Manchester Metropolitan University, UK  
Petr Hanáček, Brno University of Technology, Czech Republic  
Go Hasegawa, Osaka University, Japan  
Henning Heuer, Fraunhofer Institut Zerstörungsfreie Prüfverfahren (FhG-IZFP-D), Germany  
Paloma R. Horche, Universidad Politécnica de Madrid, Spain  
Vincent Huang, Ericsson Research, Sweden  
Friedrich Hülsmann, Gottfried Wilhelm Leibniz Bibliothek - Hannover, Germany  
Travis Humble, Oak Ridge National Laboratory, USA  
Florentin Ipate, University of Pitesti, Romania  
Imad Jawhar, United Arab Emirates University, UAE  
Terje Jensen, Telenor Group Industrial Development, Norway  
Liudi Jiang, University of Southampton, UK



Kenneth B. Kent, University of New Brunswick, Canada  
Fotis Kerasiotis, University of Patras, Greece  
Andrei Khrennikov, Linnaeus University, Sweden  
Alexander Klaus, Fraunhofer Institute for Experimental Software Engineering (IESE), Germany  
Andrew Kusiak, The University of Iowa, USA  
Vladimir Laukhin, Institució Catalana de Recerca i Estudis Avançats (ICREA) / Institut de Ciència de Materials de Barcelona (ICMAB-CSIC), Spain  
Kevin Lee, Murdoch University, Australia  
Wolfgang Leister, Norsk Regnesentral (Norwegian Computing Center), Norway  
Andreas Löf, University of Waikato, New Zealand  
Jerzy P. Lukaszewicz, Nicholas Copernicus University - Torun, Poland  
Zoubir Mammeri, IRIT - Paul Sabatier University - Toulouse, France  
Sathiamoorthy Manoharan, University of Auckland, New Zealand  
Stefano Mariani, Politecnico di Milano, Italy  
Paulo Martins Pedro, Chaminade University, USA / Unicamp, Brazil  
Don McNickle, University of Canterbury, New Zealand  
Mahmoud Meribout, The Petroleum Institute - Abu Dhabi, UAE  
Luca Mesin, Politecnico di Torino, Italy  
Marco Mevius, HTWG Konstanz, Germany  
Marek Miskowicz, AGH University of Science and Technology, Poland  
Jean-Henry Morin, University of Geneva, Switzerland  
Fabrice Mourlin, Paris 12th University, France  
Adrian Muscat, University of Malta, Malta  
Arnaldo S. R. Oliveira, Universidade de Aveiro-DETI / Instituto de Telecomunicações, Portugal  
Aida Omerovic, SINTEF ICT, Norway  
Victor Ovchinnikov, Aalto University, Finland  
Telhat Özdoğan, Amasya University - Amasya, Turkey  
Gurkan Ozhan, Middle East Technical University, Turkey  
Constantin Paleologu, University Politehnica of Bucharest, Romania  
Matteo G A Paris, Università degli Studi di Milano, Italy  
Vittorio M.N. Passaro, Politecnico di Bari, Italy  
Giuseppe Patanè, CNR-IMATI, Italy  
Marek Penhaker, VSB- Technical University of Ostrava, Czech Republic  
Juho Perälä, Bitfactor Oy, Finland  
Florian Pinel, T.J.Watson Research Center, IBM, USA  
Ana-Catalina Plesa, German Aerospace Center, Germany  
Miodrag Potkonjak, University of California - Los Angeles, USA  
Alessandro Pozzebon, University of Siena, Italy  
Vladimir Privman, Clarkson University, USA  
Mohammed Rajabali Nejad, Universiteit Twente, the Netherlands  
Konandur Rajanna, Indian Institute of Science, India  
Nageswara Rao, Oak Ridge National Laboratory, USA  
Stefan Rass, Universität Klagenfurt, Austria  
Candid Reig, University of Valencia, Spain  
Teresa Restivo, University of Porto, Portugal  
Leon Reznik, Rochester Institute of Technology, USA

Gerasimos Rigatos, Harper-Adams University College, UK  
Luis Roa Oppliger, Universidad de Concepción, Chile  
Ivan Rodero, Rutgers University - Piscataway, USA  
Lorenzo Rubio Arjona, Universitat Politècnica de València, Spain  
Claus-Peter Rückemann, Leibniz Universität Hannover / Westfälische Wilhelms-Universität Münster / North-German Supercomputing Alliance, Germany  
Subhash Saini, NASA, USA  
Mikko Sallinen, University of Oulu, Finland  
Christian Schanes, Vienna University of Technology, Austria  
Cristina Seceleanu, Mälardalen University, Sweden  
Guodong Shao, National Institute of Standards and Technology (NIST), USA  
Dongwan Shin, New Mexico Tech, USA  
Larisa Shwartz, T.J. Watson Research Center, IBM, USA  
Simone Silvestri, University of Rome "La Sapienza", Italy  
Diglio A. Simoni, RTI International, USA  
Radosveta Sokullu, Ege University, Turkey  
Junho Song, Sunnybrook Health Science Centre - Toronto, Canada  
Leonel Sousa, INESC-ID/IST, TU-Lisbon, Portugal  
Arvind K. Srivastav, NanoSonix Inc., USA  
Grigore Stamatescu, University Politehnica of Bucharest, Romania  
Raluca-Ioana Stefan-van Staden, National Institute of Research for Electrochemistry and Condensed Matter, Romania  
Pavel Šteffan, Brno University of Technology, Czech Republic  
Chelakara S. Subramanian, Florida Institute of Technology, USA  
Sofiene Tahar, Concordia University, Canada  
Muhammad Tariq, Waseda University, Japan  
Roald Taymanov, D.I.Mendeleyev Institute for Metrology, St.Petersburg, Russia  
Francesco Tiezzi, IMT Institute for Advanced Studies Lucca, Italy  
Wilfried Uhring, University of Strasbourg // CNRS, France  
Guillaume Valadon, French Network and Information and Security Agency, France  
Eloisa Vargiu, Barcelona Digital - Barcelona, Spain  
Miroslav Velev, Aries Design Automation, USA  
Dario Vieira, EFREI, France  
Stephen White, University of Huddersfield, UK  
Shengnan Wu, American Airlines, USA  
Qingsong Xu, University of Macau, Macau, China  
Xiaodong Xu, Beijing University of Posts & Telecommunications, China  
Ravi M. Yadahalli, PES Institute of Technology and Management, India  
Yanyan (Linda) Yang, University of Portsmouth, UK  
Shigeru Yamashita, Ritsumeikan University, Japan  
Patrick Meumeu Yomsi, INRIA Nancy-Grand Est, France  
Alberto Yúfera, Centro Nacional de Microelectronica (CNM-CSIC) - Sevilla, Spain  
Sergey Y. Yurish, IFSA, Spain  
David Zammit-Mangion, University of Malta, Malta  
Guigen Zhang, Clemson University, USA  
Weiping Zhang, Shanghai Jiao Tong University, P. R. China

**CONTENTS**

*pages: 1 - 9*

**Computing Efficiency in Membrane Systems and Their Variants**

Claudio Zandron, DISCo - Universita' degli Studi di Milano-Bicocca, Italy

*pages: 10 - 19*

**Effect Analysis of Creators' Environmental Factors in Making a Tourist Map for a Familiar Place**

Yoko Nishihara, Ritsumeikan University, Japan

Ryosuke Yamanishi, Kansai University, Japan

Xinran Lin, Ritsumeikan University, Japan

Ryohei Ohyama, Ritsumeikan University, Japan

*pages: 20 - 37*

**Stock Trade Simulation Using an Average Based Trend Indicator with Heuristic Enhancements**

Yoshihisa Udagawa, Tokyo University of Information Sciences, 日本

*pages: 38 - 47*

**An FFT based Bias Voltage Control Scheme to Compensate for the Resonant Frequency Drift of CMUTs Due to Fluid Loading**

Thasnim Mohammed, University of Windsor, Canada

Sazzadur Chowdhury, University of Windsor, Canada

*pages: 48 - 59*

**The Doer Effect: Replication and Comparison of Correlational and Causal Analyses of Learning**

Rachel Van Campenhout, VitalSource Technologies, USA

Benny G. Johnson, VitalSource Technologies, USA

Jenna A. Olsen, Western Governors University, USA

# Computing Efficiency in Membrane Systems and Their Variants

Claudio Zandron

DISCo - Università degli Studi di Milano-Bicocca

Viale Sarca 336/14, 20126 Milano, Italy

Email: [claudio.zandron@unimib.it](mailto:claudio.zandron@unimib.it)

**Abstract**—Membrane systems (or P systems) are a computational model inspired by the functioning of the cell, and based upon the notion of cellular membrane. In this paper, we give a survey of some main results concerning the model, to show its potential to approach various problems in the area of the theory of computation. Some main variants are also recalled, as well as some main actual research directions.

**Keywords**—Natural Computing; Membrane systems; computational complexity.

## I. INTRODUCTION

The present paper is an extended version of [1], presented at the Thirteenth International Conference on Future Computational Technologies and Applications, April 18 - 22, 2021, in Porto, Portugal, and contains an introduction and some main results concerning membrane systems (as in the conference paper), as well as the description of some main variants of the basic model, and some main actual research directions.

Membrane systems (also known as *P systems*) have been proposed by Gh. Paun in [2] as a parallel, nondeterministic, synchronous and distributed model of computation inspired by the structure and functioning of living cells. The model consists of a hierarchical structure composed by several membranes, embedded into a main membrane called the *skin*. Membranes define *regions* that contain multisets of *objects* (represented by symbols of an alphabet) and *evolution rules*.

Using these rules, the objects evolve and are moved from a region to a neighboring one. The rules are applied using a maximally parallel and nondeterministic semantic: all objects which can evolve in a computation step must evolve; if different sets of rules can be applied in a computation step (in a maximal parallel way), then one of them is nondeterministically chosen.

A *computation* starts from an initial configuration of the system and terminates when no evolution rule can be applied. The result of a computation is the multiset of objects contained into an *output membrane*, or emitted from the skin membrane.

The model was investigated both under theoretical aspects as well as for applications to other disciplines. In particular, the formalism is suitable to model various biological systems, thanks to its features. For instance, it allows identification of separated compartmentalized spaces where different reactions can take place; it is characterized by an easy understandable writing of reactions; it can be easily simulated in a distributed and parallel computing architecture, by separating the computation carried on by each single compartment, and simply

synchronizing the exchange of information at specific time-steps. Other types of application were also investigated, such as cryptography, approximate optimization, and economy.

For a systematic introduction to P systems, we refer the reader to [3] [4]; recent information can be found in the dedicated webpage [5].

P systems with active membranes is a variant of the basic model introduced in [6]: in this variant, membranes can be multiplied by dividing existing ones, and the objects are communicated according to electrical charges associated with the membranes. Such features allow the construction of an exponential workspace in linear time, which can then be used in parallel to attack computationally hard problems.

In this paper we give a survey of some main results concerning theoretical aspects of the model, to show its potential in approaching various problems in the area of the theory of computation. In Section 2, we recall formal definitions related to P systems with active membranes. In Section 3, we give a very brief summary of results concerning their computing power, while in Section 4 we recall some results related to computing efficiency. Section 5 is devoted to recall some results concerning space complexity for membrane systems. In Section 6, we present some main variants of the basic model and, finally, Section 7 draws some conclusions and suggests a few topics for research investigation.

## II. DEFINITIONS

In this section, we recall the basic definition of P systems with active membranes.

*Definition 1:* A *P system with active membranes* of initial degree  $d \geq 1$  is a tuple  $\Pi = (\Gamma, \Lambda, \mu, w_1, \dots, w_d, R)$ , where:

- $\Gamma$  is an alphabet, a finite non-empty set of symbols, usually called *objects*;
- $\Lambda$  is a finite set of labels for the membranes;
- $\mu$  is a membrane structure (i.e., a rooted *unordered tree*) consisting of  $d$  membranes enumerated by  $1, \dots, d$ ; each membrane is labeled by an element of  $\Lambda$ , not necessarily in a one-to-one way, and possesses an *electrical charge* (or polarization), that can be neutral (0), positive (+) or negative (-).
- $w_1, \dots, w_d$  are strings over  $\Gamma$ , describing the initial multisets of objects placed in the  $d$  regions of  $\mu$ ;
- $R$  is a finite set of rules.

The rules are of the following kinds:

- *Object evolution rules*, of the form  $[a \rightarrow w]_h^\alpha$   
They can be applied if the membrane  $h$  has charge  $\alpha$  and contains an occurrence of the object  $a$ ; the object  $a$  is rewritten into the multiset  $w$ .
- *Send-in communication rules*, of the form  $a [ ]_h^\alpha \rightarrow [b]_h^\beta$   
They can be applied to a membrane labeled by  $h$ , having charge  $\alpha$  and if the external region contains an occurrence of the object  $a$ ; the object  $a$  is sent into  $h$  becoming  $b$  and, simultaneously, the charge of  $h$  is changed to  $\beta$ .
- *Send-out communication rules*, of the form  $[a]_h^\alpha \rightarrow [ ]_h^\beta b$   
They can be applied to a membrane labeled by  $h$ , having charge  $\alpha$  and containing an occurrence of  $a$ ; the object  $a$  is sent out from  $h$  to the outside region becoming  $b$ . Simultaneously, the charge of  $h$  is changed to  $\beta$ .
- *Dissolution rules*, of the form  $[a]_h^\alpha \rightarrow b$   
They can be applied to a membrane labeled by  $h$ , having charge  $\alpha$  and containing an occurrence of the object  $a$ ; the membrane  $h$  is dissolved and its contents are left in the surrounding region unaltered, except that an occurrence of  $a$  becomes  $b$ .
- *Elementary division rules*, of the form  $[a]_h^\alpha \rightarrow [b]_h^\beta [c]_h^\gamma$   
They can be applied to a membrane labeled by  $h$ , having charge  $\alpha$ , containing an occurrence of the object  $a$  but having no other membrane inside (an *elementary membrane*); the membrane is divided into two membranes having label  $h$  and charge  $\beta$  and  $\gamma$ ; the object  $a$  is replaced, respectively, by  $b$  and  $c$  while the other objects in the initial multiset are copied to both membranes.
- *Nonelementary division rules*, of the form

$$[[ ]_{h_1}^+ \cdots [ ]_{h_k}^+ [ ]_{h_{k+1}}^- \cdots [ ]_{h_n}^- ]_h^\alpha \rightarrow$$

$$[[ ]_{h_1}^\delta \cdots [ ]_{h_k}^\delta ]_h^\beta [[ ]_{h_{k+1}}^\epsilon \cdots [ ]_{h_n}^\epsilon ]_h^\gamma$$

They can be applied to a membrane labeled by  $h$ , having charge  $\alpha$ , containing the positively charged membranes  $h_1, \dots, h_k$ , the negatively charged membranes  $h_{k+1}, \dots, h_n$ , and possibly some neutral membranes. The membrane  $h$  is divided into two copies having charge  $\beta$  and  $\gamma$ , respectively; the positively charged membranes  $h_1, \dots, h_k$  are placed inside the former membrane, their charge set to  $\delta$ , while the negative ones are placed inside the latter membrane, their charges set to  $\epsilon$ . Neutral membranes inside  $h$  are duplicated and placed inside both copies.

Each instantaneous configuration of a P system with active membranes is described by the current membrane structure, including the electrical charges, together with the multisets located in the corresponding regions. A computation step changes the current configuration according to the following set of principles:

- Each object and membrane can be subject to at most one rule per step, except for object evolution rules (inside each membrane any number of evolution rules can be applied simultaneously).
- The application of rules is *maximally parallel*: each object appearing on the left-hand side of evolution, communication, dissolution or elementary division rules must be

subject to exactly one of them (unless the current charge of the membrane prohibits it). The same reasoning applies to each membrane that can be involved to communication, dissolution, elementary or nonelementary division rules. In other words, all possible rules that can be applied must be applied at each computation step; the only objects and membranes that do not evolve are those associated with no rule, or only to rules that are not applicable due to the electrical charges.

- When several conflicting rules can be applied at the same time, a nondeterministic choice is performed; this implies that, in general, multiple possible configurations can be reached after a computation step (e.g., consider two rules  $a \rightarrow b$  and  $a \rightarrow c$  in a region  $h$ ; if an object  $a$  is present in that region, then it can nondeterministically produce either  $b$  or  $c$ , by using respectively the first or the second rule).
- While all the chosen rules are considered to be applied simultaneously during each computation step, they are logically applied in a bottom-up fashion: first, all evolution rules are applied to the elementary membranes, then all communication, dissolution and division rules; then the application proceeds towards the root of the membrane structure. In other words, each membrane evolves only after its internal configuration has been updated.
- The outermost membrane cannot be divided or dissolved, and any object sent out from it cannot re-enter the system again.

A *halting computation* of the P system  $\Pi$  is a finite sequence of configurations  $\vec{C} = (C_0, \dots, C_k)$ , where  $C_0$  is the initial configuration, every  $C_{i+1}$  is reachable by  $C_i$  via a single computation step, and no rules can be applied anymore in  $C_k$ . The result of a halting computation is the multiset of objects emitted from the skin during the whole computation. A *non-halting computation*  $\vec{C} = (C_i : i \in \mathbb{N})$  consists of infinitely many configurations, again starting from the initial one and generated by successive computation steps, where the applicable rules are never exhausted. A non-halting computation produces no output.

P systems can also be used as *recognizers* (see, e.g., [7]) by employing two distinguished objects *yes* and *no*; exactly one of these must be sent out from the outermost membrane during each computation, in order to signal acceptance or rejection respectively; we also assume that all computations are halting. If all computations starting from the same initial configuration are accepting, or all are rejecting, the P system is said to be *confluent*. If this is not necessarily the case, then we have a *non-confluent* P system, and the overall result is established as for nondeterministic Turing machines: it is acceptance iff an accepting computation exists. All P systems considered in this paper are confluent.

In order to solve decision problems (i.e., decide languages), we use *families* of recognizer P systems  $\mathbf{\Pi} = \{\Pi_x : x \in \Sigma^*\}$ . Each input  $x$  is associated with a P system  $\Pi_x$  that decides the membership of  $x$  in the language  $L \subseteq \Sigma^*$  by accepting or rejecting. The mapping  $x \mapsto \Pi_x$  must be efficiently computable for each input length [8].

*Definition 2:* A family of P systems  $\Pi = \{\Pi_x : x \in \Sigma^*\}$  is said to be (*polynomial-time*) *uniform* if the mapping  $x \mapsto \Pi_x$  can be computed by two deterministic polynomial-time Turing machines  $F$  (for “family”) and  $E$  (for “encoding”) as follows:

- The machine  $F$ , taking as input *the length  $n$  of  $x$*  in unary notation, constructs a P system  $\Pi_n$ , which is common for all inputs of length  $n$ , with a distinguished input membrane.
- The machine  $E$ , on input  $x$ , outputs a multiset  $w_x$  (an encoding of  $x$ ).
- Finally,  $\Pi_x$  is simply  $\Pi_n$  with  $w_x$  added to the multiset placed inside its input membrane.

*Definition 3:* If the mapping  $x \mapsto \Pi_x$  is computed by a *single* polynomial-time Turing machine, the family  $\Pi$  is said to be *semi-uniform*. In this case, inputs of the same size may be associated with P systems having possibly different membrane structures and rules.

Any explicit encoding of  $\Pi_x$  is allowed as output of the construction, as long as the number of membranes and objects represented by it does not exceed the length of the whole description, and the rules are listed one by one. This restriction is enforced to mimic a (hypothetical) realistic process of construction of the P system, where membranes and objects are placed in a constant amount during each construction step, and require actual physical space proportional to their number. Moreover, notice that uniformity condition can also be restricted to be computed in classes below **P**, such as log-space Turing machines. We refer the reader to [8] for further details on the encoding of P systems.

Finally, we describe how time and space complexity for families of recognizer P systems are measured.

*Definition 4:* A uniform or semi-uniform family of P systems  $\Pi = \{\Pi_x : x \in \Sigma^*\}$  is said to decide the language  $L \subseteq \Sigma^*$  (in symbols  $L(\Pi) = L$ ) in time  $f: \mathbb{N} \rightarrow \mathbb{N}$  iff, for each  $x \in \Sigma^*$ ,

- the system  $\Pi_x$  accepts if  $x \in L$ , and rejects if  $x \notin L$ ;
- each computation of  $\Pi_x$  halts within  $f(|x|)$  computation steps.

*Definition 5:* Let  $\mathcal{C}$  be a configuration of a P system  $\Pi$ . The *size*  $|\mathcal{C}|$  of  $\mathcal{C}$  is defined as the sum of the number of membranes in the current membrane structure and the total number of objects they contain. If  $\vec{\mathcal{C}} = (\mathcal{C}_0, \dots, \mathcal{C}_k)$  is a halting computation of  $\Pi$ , then the *space required by  $\vec{\mathcal{C}}$*  is defined as

$$|\vec{\mathcal{C}}| = \max\{|\mathcal{C}_0|, \dots, |\mathcal{C}_k|\}$$

or, in the case of a non-halting computation  $\vec{\mathcal{C}} = (\mathcal{C}_i : i \in \mathbb{N})$ ,

$$|\vec{\mathcal{C}}| = \sup\{|\mathcal{C}_i| : i \in \mathbb{N}\}.$$

Non-halting computations might require an infinite amount of space (in symbols  $|\vec{\mathcal{C}}| = \infty$ ). The *space required by  $\Pi$*  itself is then

$$|\Pi| = \sup\{|\vec{\mathcal{C}}| : \vec{\mathcal{C}} \text{ is a computation of } \Pi\}.$$

Notice that  $|\Pi| = \infty$  occurs if either  $\Pi$  has a non-halting computation requiring infinite space, or  $\Pi$  has an infinite set of halting computations, such that for each bound  $b \in \mathbb{N}$  there exists a computation requiring space larger than  $b$ .

### III. COMPUTING POWER OF MEMBRANE SYSTEMS

The first studies of the model concerned its computing power: various types of Membrane systems have been compared with computing models like automata, Turing machines and register machines, or among themselves, also considering normal forms [9] for systems.

It is known that using a single membrane we can only generate the length sets of context-free languages, and the power cannot be extended by using an unlimited number of membranes. However, if we allow to dissolve membranes after the application of rewriting rules then the computing power is increased, when at least two membranes are used.

More formally, let us denote by  $NOP_k(\delta)$  (resp.  $NOP_k(n\delta)$ ) the family of natural numbers generated by P systems having  $k$  membranes and using (resp. not using) the dissolving membrane action. The following results can be stated [3]:

*Theorem 1:*  $NOP_1(n\delta) = NOP_*(n\delta) = NCF$   
 $NCF = NOP_*(n\delta) \subset (NE0L \subseteq) NOP_2(\delta)$   
 $NOP_*(\delta) (\subseteq ET0L) \subset NCS$

Even when dissolving membrane action is allowed, universality cannot be reached. Some more ingredients can be considered to obtain such a result like, e.g., cooperative rules, catalysts, or priorities defining the order of rules application.

A rewriting rule  $a \rightarrow w$  is said to be *cooperative* if  $a$  contains more than one symbol. This turned out to be a feature very powerful in the framework of membrane systems. In fact, when using such rules one membrane turns out to be sufficient to obtain the same power as Turing machines:

*Theorem 2:*  $NOP_1(coop) = NOP_*(coop) = NRE$

where *coop* indicates the possibility to use cooperative rules.

In order to show the functioning of the model, we summarize in the following the main ideas behind this result. Let us consider a register machine, a well-known computationally complete device [10], consisting of  $n$  registers  $r_1, \dots, r_n$ , each one containing a non-negative integer. The input is placed in register  $r_1$ , while all the other registers are initially set to zero. The machine has a set of instructions (labelled by natural numbers) of two kinds: increment instruction  $i : inc(r), j$  and conditional decrement instruction  $i : dec(r), j, k$ . A program counter (initially set to 1) define the instruction to execute.

An increment instruction increase by one the contents of register  $r$ , and set the program counter to  $j$ , while the result of a decrement instruction depends on the value of register  $r$ . If  $r$  is greater than zero, then it is decremented by one and the program counter is set to  $j$ ; if  $r$  contains zero, then no decrement is applied and the program counter is set to  $k$ . The computation halts when the program counter reaches a value associated with a specific halting instruction (setting the program counter to zero), and the contents of  $r_1$  is the output of the computation.

To simulate a register machine, we can use some symbols  $p_i$ , to store the actual value of the program counter (only one symbol of this type can be present at each time step, and a symbol  $r$  for each register  $r$ , used to store the value of the corresponding register (by considering the multiplicity of the symbol).



Using rewriting rules, an increment instruction of the register machine can be simulated by a membrane system by means of a single rewriting rule of the form  $p_i \rightarrow p_j r$ . In by applying such a rule, the actual program counter  $p_i$  is replaced by  $p_j$ , and the contents of register  $r$  is incremented by one by adding a symbol  $r$ .

A conditional decrement instruction can be simulated by the following rules. First, we apply  $p_i \rightarrow p'_i d_r$ ; the symbol  $d_r$  will be used to check whether it is possible to delete an occurrence of  $r$  (if  $r$  is greater than zero) by means of the rule  $d_r r \rightarrow d'_r$ , while  $p'_i$  is replaced by  $p''_i$ . If the decrement is possible (i.e., there is at least one symbol  $r$ ), then we have now the symbol  $d'_r$ , otherwise the symbol  $d_r$  is still present. The correct program counter can thus be set, by considering the only applicable rule between  $p''_i d'_r \rightarrow p_j$  and  $p''_i d_r \rightarrow p_k$ .

A simpler form of cooperative rules can be defined by means of catalysts. A rewriting rule with catalyst is a rule of the form  $CX \rightarrow Cw$ , where  $C$  and  $X$  are symbols and  $w$  is a string.  $C$  is said to be a catalyst: it is needed to activate the rule, but it is not changed by it. Such a feature also allows to obtain universal systems, but only when priorities defining a partial order concerning the application of the rules is also used:

*Theorem 3:*  $NOP_2(cat, pri) = NOP_*(cat, pri) = NRE$

One can also consider structured objects instead of atomic ones, by consider strings of symbols: in this case, the systems are called Rewriting P systems. Let us denote by  $RP_k$  the family of languages generated by Rewriting P systems using  $k$  membranes and context-free rewriting rules. The following theorem from [3] shows that using a single membrane only context-free languages can be obtained, but a structure with four membranes allow to obtain a strictly more powerful class.

*Theorem 4:*  $RP_1(CF) = CF \subset RP_4(CF)$

Thus, it is evident from these results that the power of such systems can be improved (as expected) by exploiting membranes to define regions to keep separated specific subsets of rules and objects. Once again, universality cannot be obtained using only this basic set of ingredients, and more features must be considered.

Further details can be found in [3] and [4].

#### IV. COMPUTING EFFICIENCY OF MEMBRANE SYSTEMS

Another interesting feature that can be considered, and already described in Section 2, is the possibility to give an active role to membranes. P systems with active membranes allow to create new membranes during the computation by division of existing membranes. In this way, we can obtain a trade off between time and space resources that allows to solve NP-complete (or even harder) problems in polynomial time and exponential space (see, e.g., [6] [11] [12] [13] [14] [15] [16]).

*Theorem 5:* The SAT problem can be solved in linear time (with respect to the number of variables and the number of clauses) by a confluent P-system with active membranes using elementary membrane division only.

In fact, consider a boolean expression  $\Phi$  in conjunctive normal form, with  $m$  clauses and  $n$  variables. We can build a P-system  $\Pi = (\Gamma, \Lambda, \mu, w_1, w_2, R)$  having initial objects

$a_1, a_2, \dots, a_n$  in region 2 and such that  $R$  is defined to contain a polynomial number of rules (with respect to the size of the input formula) that operate as it follow.

By using the variables  $a_i$  and elementary membrane division rules, in  $O(n)$  steps we generate  $2^n$  copies of membrane 2, containing all possible truth assignments of the  $n$  variables of  $\Phi$ .

Then, in  $O(m)$  steps we verify if there is at least one membrane containing a truth assignment that satisfies all the  $m$  clauses of  $\Phi$ . In this case, an object *yes* is sent out from the skin membrane; otherwise, an object *no* is sent out.

Let us denote by  $PMC_{NAM}$ ,  $PMC_{\epsilon AM}$ , and  $PMC_{AM}$  the class of problems solved in a polynomial number of steps (with respect to the input length) by P systems with active membranes without membrane division, with division for elementary membranes only, and for both elementary and non-elementary membranes, respectively.

The following results can be obtained directly from definitions:

*Theorem 6:*  $PMC_{NAM} \subseteq PMC_{\epsilon AM} \subseteq PMC_{AM}$

Moreover, it is easy to show that

*Theorem 7:*  $P \subseteq PMC_{NAM}$

In fact, the "trick" is that the deterministic Turing machine deciding  $L \in P$  is used to solve directly the problem in polynomial time. Then, we build a P system with a single membrane containing either an object *YES*, whenever an input  $x \in L$  is given, or *NO*, otherwise. This requires polynomial time, and then the P system simply send out the object in a single computation step.

The opposite inclusion is also true:

*Theorem 8:*  $PMC_{NAM} \subseteq P$

In fact, a generic P system  $\Pi$  without membrane division can be simulated by a deterministic Turing machine  $M$ , with a polynomial slowdown, as proved in [14].

Since we have shown that the NP-complete problem SAT can be solved when elementary membrane division is allowed, then we can also state the following:

*Theorem 9:*  $NP \subseteq PMC_{\epsilon AM}$

From this result and from the closure properties for  $PMC_{\epsilon AM}$  it also follows:

*Theorem 10:*  $coNP \subseteq PMC_{\epsilon AM}$

P systems with elementary membrane division can be simulated by Deterministic Turing machines using polynomial space:  $PMC_{\epsilon AM} \subseteq PSPACE$ . Hence:

*Theorem 11:*  $NP \cup coNP \subseteq PMC_{\epsilon AM} \subseteq PSPACE$

A stronger result was later proved [17]: a solution to the PP-complete problem SQRT-3SAT was obtained using P systems with active membranes and elementary membrane division. This proved that the class PP (Probabilistic Polynomial time: the class of decision problems solvable by a probabilistic Turing machine in polynomial time, with an error probability of less than 1/2 for all instances) is also included in  $PMC_{\epsilon AM}$ . A characterization of the class  $P^{\#P}$  was obtained in [18].

When division for non-elementary membranes is allowed, even harder problems can be solved, as expected. In a series of papers [19] [20] and [21] the following results were proved:

*Theorem 12:*  $PSPACE \subseteq PMC_{AM} \subseteq EXPTIME$

By limiting the nesting levels of membranes (and, as a consequence, the membrane division) to a constant depth, the problems in the class  $CH$  (Counting Hierarchy) can be solved, as proved in [22].

## V. SPACE COMPLEXITY OF P-SYSTEMS AND POLARIZATION OF MEMBRANES

In order to clarify relations between the amount of time and space needed to solve various classes of problems, in [23] a definition of space complexity for P systems has been introduced.

On the same line of what has been done for time complexity, we can define space complexity classes for Membrane systems. By  $MCSPACE_T(f)$  we denote the class of languages decided by confluent recognizer P systems (of type T) within space  $f(n)$ . In particular, by  $PMCSpace_T(=MCSPACE_T^{[*]}(p(n)))$  we denote the class of languages decided by confluent recogniser P systems using at most a polynomial number of elements.

From the definitions and from the results we recalled in the previous section, it is easy to see that:

- $P \subseteq MCSPACE_{NAM}(O(1))$
- $NP \cup co-NP \subseteq EXPMCSpace_{\varepsilon AM}$
- $PSPACE \subseteq EXPMCSpace_{AM}$

In [24] and [25] it has been shown, respectively, that the  $PSPACE$ -complete problem Quantified-3SAT can be solved by P-systems with active membranes using a polynomial amount of space, and that such P systems can be simulated by Turing machines with only a polynomial increase in space requirements, thus giving a precise characterization of the class  $PSPACE$  in terms of space complexity classes for membrane systems. A similar result to characterize the complexity class  $EXSPACE$  can be obtained by considering exponential space P systems, as showed in [26]. Thus, all types of Membrane systems with active membranes and both divisions for elementary and non-elementary membranes, and working in a polynomial space, exactly characterize  $PSPACE$ .

Investigation of classes of problems solved by P systems using sublinear space was also considered. In order to consider sublinear space, two distinct alphabets were considered in the definition of P systems: an  $INPUT$  alphabet and a  $WORK$  alphabet. Objects from the  $INPUT$  alphabet cannot be rewritten and do not contribute to the size of the configuration of a P system. The size of a configuration was defined as the sum of the number of membranes in the current membrane structure and the total number of working objects they contain; the space required by a computation is the maximum size among all configurations. Moreover, we need to define a uniformity condition for the families of P systems that is weaker than the usual P uniformity, to avoid the possibility to solve a problem directly by using the Turing machine that build the P systems we use to compute. We consider  $DLOGTIME$ -uniformity, defined on the basis of  $DLOGTIME$  Turing machines [27]. We refer the reader to [28] for formal definitions.

The efficient simulation of logarithmic space Turing machines (or other equivalent models) by employing standard techniques used in the papers previously cited seems not to

work because of two main problems: we either need to use a polynomial number of working objects (thus violating the logarithmic space condition) or to use a polynomial number of rewriting rules (thus violating the uniformity condition). Nonetheless, it has been showed in [28] that such a simulation can be efficiently done by using membrane polarization both to communicate objects through membranes as well as to store some information:

*Theorem 13:* Each log-space deterministic Turing machine  $M$  can be simulated by a  $DLOGTIME$ -uniform family  $\Pi$  of confluent recognizer P systems with active membranes in logarithmic space.

An immediate corollary of Theorem 13 is that the class  $L$  (the class of problems solved by log-space Turing machines) is contained in the class of problems solved by  $DLOGTIME$ -uniform, log-space P systems with active membranes.

The first definition of space, described above, was based on a hypothetical implementation of P systems by means of chemicals and molecules. Nonetheless, a different approach can also be considered to define space complexity for P systems, by focusing the definition of space on the simulative point of view. In fact, various implementation of P systems in silico have been proposed (see, e.g., [29], [30]) where the multiplicity of each object in each membrane are stored by means of binary numbers. This, of course, reduces the amount of requested space, thus possibly leading to different results related to complexity classes defined on the basis of this definition of space.

Such problems have been recently investigated in [31]. The first results show that, for different considered systems, the computational classes defined on the basis of this new definition of space do not differ from the corresponding classes defined on the basis of the original space definition. In particular, classes for systems using at least a linear amount of space have been considered so far. It is conjectured that differences between classes defined considering the two definitions of space can be highlighted when sublinear space is considered, like in the case of logarithmic space: an important open problem is to find specific classes where this difference exists, thus proving that efficient storing of information concerning objects can be exploited in some cases. It would also be interesting to understand which features can be used/are necessary to obtain the same result.

## VI. VARIANTS OF MEMBRANE SYSTEMS

Apart from the basic model, strictly based on the internal organization of the cell, many variants have been proposed in the literature, by considering different features of the cell, or different organizations and connection of the whole system. We recall in the following some of the main variants and their features; in particular, we will recall Tissue-like P systems and Spiking Neural P systems.

Other variants have been proposed and investigated, like *population P systems*, *P automata*, *P colonies*, *metabolic P systems*, *numerical P systems* or P systems incorporating *probabilistic* or *quantum* notions, or their application in various fields has been discussed. As an example, in [32] an overview

of research related to the design of robot controllers by means of some variants of membrane systems is presented - namely, numerical P systems, enzymatic numerical P systems, P colonies, and XP colonies - and their performances compared to traditional controllers for robot systems. In [33] the notion of energy associated to membranes has been considered, to define quantum-like systems.

We refer the interested reader to the webpage [5] and to the Handbook [4] as a starting point for the various variants.

### A. Tissue P Systems

A variant appeared quite early was that of *Tissue P systems*, proposed in [34] to highlight role of inter-cellular communication, exchanging substances between adjacent cells. A specific introduction to this variant can be found in chapter 9 of [4].

*Definition 6:* A *tissue P system* is a structure  $\Pi = (\Gamma, E, w_1, \dots, w_d, R)$ , where:

- $\Gamma$  is an alphabet, i.e., a finite non-empty set of symbols, usually called *objects*;
- $E \subseteq \Gamma$  is the alphabet of objects initially located in the external environment, in *arbitrarily many* copies (i.e., they are never exhausted during the computation);
- $d \geq 1$  is the *degree* of the system, i.e., the initial number of cells;
- $w_1, \dots, w_d$  are *finite* multisets over  $\Gamma$ , describing the initial contents of the  $d$  cells; here  $1, \dots, d$  are *labels* identifying the cells of the P systems, and 0 is the label of the external environment;
- $R$  is a finite set of rules.

The rules of  $R$  are of the following types:

- (a) *Communication rules*, denoted by  $(h, u/v, k)$ , where  $h$  and  $k$  are distinct labels (including the environment), and  $u$  and  $v$  are multisets over  $\Gamma$  (at least one of them nonempty): these rules are applicable if there exists a region with label  $h$  containing  $u$  as a submultiset and a region  $k$  containing  $v$  as a submultiset; the effect of the application of the rule is to exchange  $u$  and  $v$  between the two regions. If  $h = 0$  (resp.,  $k = 0$ ) then the rule is denoted by  $(\lambda, u/v, k)$  (resp.,  $(h, u/v, \lambda)$ ), and in that case we require multiset  $u$  (resp.,  $v$ ) to contain at least an object from  $\Gamma - E$ , i.e., an object with finite multiplicity, if  $v$  (resp.,  $u$ ) is empty
- (b) *Division rules*, of the form  $[a]_h \rightarrow [b]_h[c]_h$ , where  $h \neq 0$  is a cell label and  $a, b, c \in \Gamma$ : these rules can be applied to a cell with label  $h$  containing at least one copy of  $a$ ; the effect of the application of the rule is to divide the cell into two cells, both with label  $h$ ; the object  $a$  is replaced in the two cells by  $b$  and  $c$ , respectively, while the rest of the original multiset contained in  $h$  is replicated in both cells.
- (c) *Separation rules*, of the form  $[a]_h \rightarrow [\Gamma_1]_h[\Gamma_2]_h$ , where  $h \neq 0$  is a cell label,  $a \in \Gamma$ , and  $\{\Gamma_1, \Gamma_2\}$  is a partition of  $\Gamma$ : these rules can be applied to a cell with label  $h$  containing at least one copy of  $a$ ; the effect of the application of the rule is to separate the cell into two cells, both with label  $h$ ; the object  $a$  is consumed, while the objects from  $\Gamma_1$  in the original multiset contained

in  $h$  are placed inside one of the cells, and those from  $\Gamma_2$  in the other. All separation rules in  $R$  must share the same partition  $\{\Gamma_1, \Gamma_2\}$  of  $\Gamma$ .

A *tissue P system with cell division* only uses communication and division rules, while a *tissue P system with cell separation* only uses communication and separation rules.

A *configuration*  $\mathcal{C}$  of a tissue P system consists of a multiset over  $\Gamma - E$  describing the objects appearing with finite multiplicity in the environment, and a multiset of pairs  $(h, w)$ , where  $h$  is a cell label and  $w$  a finite multiset over  $\Gamma$ , describing the cells. A *computation step* changes the current configuration according to the following set of principles:

- Each object can be subject to at most one rule, and each cell can be subject to *any number* of communication rules or, *alternatively*, a *single* division or separation rule.
- The application of rules is *maximally parallel*: each region is subject to a maximal multiset of rules (i.e., no further rule can be applied).
- When several conflicting rules can be applied at the same time, a nondeterministic choice is performed; this implies that, in general, multiple possible configurations can be reached after a computation step.

A *halting computation*  $\vec{\mathcal{C}} = (\mathcal{C}_0, \dots, \mathcal{C}_k)$  of the tissue P system  $\Pi$  is a finite sequence of configurations, where  $\mathcal{C}_0$  is the initial configuration, every  $\mathcal{C}_{i+1}$  is reachable from  $\mathcal{C}_i$  via a single computation step, and no rules are applicable in  $\mathcal{C}_k$ .

Tissue P systems can be used as language *recognisers* by employing two distinguished objects *yes* and *no*: we assume that all computations are halting, and that one of the objects *yes* or *no* (but not both) is released into the environment, and only in the last computation step, in order to signal acceptance or rejection, respectively. If all computations starting from the same initial configuration are accepting, or all are rejecting, the tissue P system is said to be *confluent*. A confluent P system is said to *accept* if its computations are accepting, and to *reject* otherwise. Confluent tissue P systems may be locally nondeterministic, as long as all computations agree on the final result. As a special case, tissue P systems whose initial configuration generates a single computation are called *deterministic*.

Different results concerning computational aspects related to this variant have been presented in the literature (e.g. [35], [36], [37], [38] [39]). Cell division rules and cell separation rules have been considered in order to attack computationally hard problems (e.g. [40],[41], [42], [43]), exploiting the same ideas presented above for P systems with active membranes. It was shown that NP-complete problems can be solved in polynomial time, also in this case. Nonetheless, for this kind of systems it is not possible to reach enough power to solve PSPACE-complete problems, due to the missing structured organization of membranes. In fact, it turned out that such a model solves exactly all problems in the complexity class  $\mathbf{P}^{\#\mathbf{P}}$ , that is the class of problems that are solved by Turing machines using an oracle for counting problems [44].

### B. Spiking Neural P Systems

Another variant that has been widely investigated are *spiking neural* (SN) P systems [45], inspired by the structure and

functioning of spiking neural systems. In this kind of a variant, there is a single objects, called the *spike*, communicated through connections called *synapses* among *neurons*.

Formally, a Spiking Neural P system of degree  $m \geq 1$  is a construct

$$\Pi = (O, \sigma_1, \sigma_2, \dots, \sigma_m, \text{syn}, \text{in}, \text{out})$$

where:

- 1)  $O = \{a\}$  is the singleton alphabet ( $a$  is called *spike*);
- 2)  $\sigma_1, \sigma_2, \dots, \sigma_m$  are *neurons*, of the form  $\sigma_i = (n_i, R_i)$ , with  $1 \leq i \leq m$ , where:
  - a)  $n_i \geq 0$  is the *initial number of spikes* contained in  $\sigma_i$ ;
  - b)  $R_i$  is a finite set of *rules* of the following two forms:
    - (1)  $E/a^c \rightarrow a; d$ , where  $E$  is a regular expression over  $a$ , and  $c \geq 1, d \geq 0$  are integer numbers; if  $E = a^c$ , then it is usually written in the following simplified form:  $a^c \rightarrow a; d$ ;
    - (2)  $a^s \rightarrow \lambda$ , for  $s \geq 1$ , with the restriction that for each rule  $E/a^c \rightarrow a; d$  of type (1) from  $R_i$ , we have  $a^s \notin L(E)$  (where  $L(E)$  denotes the regular language defined by  $E$ );
- 3)  $\text{syn} \subseteq \{1, 2, \dots, m\} \times \{1, 2, \dots, m\}$ , with  $(i, i) \notin \text{syn}$  for  $1 \leq i \leq m$ , is the directed graph of *synapses* between neurons;
- 4)  $\text{in}, \text{out} \in \{1, 2, \dots, m\}$  indicate the *input* and the *output* neurons of  $\Pi$ .

The rules of type (1) are called *firing rules*, while rules of type (2) are called *forgetting rules*.

If a neuron  $\sigma_i$  contains  $k \geq c$  spikes, such that  $a^k \in L(E)$ , then the application of a spiking rule  $E/a^c \rightarrow a; d \in R_i$  removes  $c$  spikes from  $\sigma_i$  (leaving  $k - c$  spikes in the neuron), and prepares one spike to be sent to all neurons connected to  $\sigma_i$  by means of a synapse defined in *syn*. If the delay  $d$  is set to zero, then the spike is immediately emitted. On the contrary, the neuron waits  $d$  computation steps before sending out the spike. During the waiting time, the neuron is *closed*: it cannot receive new spikes (spikes sent to  $\sigma_i$  by other neurons are lost), and it cannot fire new rules. After  $d$  computation steps, the neuron sends the spike out and becomes open again, ready to receive spikes and to apply new rules.

To apply a *forgetting* rule  $a^s \rightarrow \lambda$  from  $R_i$ , the neuron  $\sigma_i$  must contain *exactly*  $s$  spikes. In that case, all  $s$  spikes are removed from  $\sigma_i$ .

In each time unit, if a neuron  $\sigma_i$  can use one of its rules, then a rule from  $R_i$  must be used. If two or more rules can be applied in a neuron at the same time step, then only one of them is nondeterministically chosen to be applied. This means that in each neuron the rules are applied following a sequential semantics, while all neurons work in parallel.

A configuration of a Spiking Neural system is described by both the number of spikes present in each neuron and by the number of steps to wait until it becomes open (initially all neurons are open).

A *computation* starts in the initial configuration, with a positive integer number given in input to a specific *input*

*neuron* as the number of time steps between the insertion of two spikes into that neuron, and proceeds by applying the rules in the neurons as described above. The output of the system is the number considered to be the number of time steps between the arrival of two spikes in a specific *output neuron*.

Spiking neural P systems have been widely investigated from a computational point of view ([46], [47], [48], [49], [50]), and many efforts are now concentrated on their application implementing artificial intelligence and deep learning strategies.

For instance, in [51] a Spiking Neural Network learning model is built, based on an evolutionary membrane algorithm, to solve the classical problem of supervised classification. The designed algorithm is able to automatically adjust the various learning parameters, by operating on the synaptic weight during the learning stage of the spiking neural model. The authors show that the proposed algorithm has competitive advantages, with respect to other experimental algorithms presented in the literature, in solving twelve supervised classification benchmark problems.

In [52] spiking neural P systems were used to solve the skeletonization problem: skeletonization consists in an image transformation, that simplifies the original image but preserving its topological properties. A parallel software simulating SN P systems was implemented on a Graphics Processors Units (GPU) architecture.

Fault diagnosis of power systems has been the subject of [53], where the problem is approached by means of an interval-valued fuzzy spiking neural P system: interval-valued fuzzy logic is used within a spiking neural P system, in order to deal with uncertainty. The paper shows that such a system can be used to accurately diagnose faulty sections in a power transmission networks.

## VII. CONCLUSIONS

We survey some results concerning P systems with active membranes, concerning both the computational aspects, as well as computational efficiency. Some results concerning space complexity of the model have also been recalled, as well as some references to the main variants of the basic model. For further reading on the subject, we refer the reader to the main volumes on the subject [3] [4].

A recent survey on different strategies to approach computationally hard problems (containing links to various research paper on the subject) by P systems with active membranes can be found in [54], also containing some open problems worthing investigations.

Links to various paper concerning space complexity for membrane systems are available in [24] [25]; readers interested in results concerning sublinear space or even constant amount of space can refer to [55] and [56], respectively. A recent survey concerning results obtained by considering different bounds on space can be found in [57].

There are various research topics which deserve to be studied with respect to computing power and efficiency. In particular, precise characterizations of complexity classes obtained by considering systems using specific or minimal subset

of features, restriction on their functioning (like, e.g., communication in only one direction), and relations of such classes with standard complexity classes are currently active research topics [58] [59] [60] [61] are currently under investigations.

Other research directions concern the application of variants of P systems to real-life problems, as described in the previous section. Some actual research topics concern neovascularization lesion [62], Covid-19 pandemic [63], and the development of software tools to efficiently simulate P systems by exploiting the high level of parallelism offered by the use of modern GPUs [64].

Another interesting topic can be found in [65], where self-organization and self-repairing processes for membrane systems is presented and discussed (together with other types of bio-inspired systems), including self-organizing method based on optimization of cells placement and population P systems.

### ACKNOWLEDGEMENTS

This work was partially supported by Università degli Studi di Milano-Bicocca, Fondo di Ateneo Quota Dipartimentale (FAQD-2019).

### REFERENCES

- [1] C. Zandron, Computing Efficiency in Membrane Systems, Proc. of Future Computing Conference 2021 (H. Sato, The Univ. of Tokyo, ed.), The Thirteenth International Conference on Future Computational Technologies and Applications, April 18 - 22, 2021, Porto, Portugal, 8–13.
- [2] Gh. Păun, Computing with membranes, J. of Computer and System Sciences, 61(1), 2000, 108–143
- [3] Gh. Păun, Membrane Computing. An Introduction, Springer, Berlin, 2002
- [4] Gh. Păun, G. Rozenberg, A. Salomaa (Eds.), Handbook of Membrane Computing, Oxford University Press, 2010
- [5] The P systems Web page: <http://ppage.psystems.eu/> (Retrieved 2022-05-30)
- [6] Gh. Păun, P systems with active membranes: Attacking NP-complete problems, J. of Automata, Languages and Combinatorics 6(1), 2001, 75–90
- [7] E. Csuhaj-Varju, M. Oswald, Gy. Vaszil, P automata, Handbook of Membrane Computing, Gh. Paun et al. (Eds.), Oxford University Press, 2010, 144–167
- [8] N. Murphy, D. Woods, The computational power of membrane systems under tight uniformity conditions, Natural Computing 10(1), 2011, 613–632
- [9] C. Zandron, C. Ferretti, G. Mauri, Two Normal Forms for Rewriting P systems, in Machines, Computations and Universality, Proc. of 3rd Int. Conf. MCU 2001, LNCS 2055, Springer-Verlag, 2001, 153–164
- [10] M. Minsky, Computation: Finite and Infinite Machines, Prentice-Hall, 1967.
- [11] S. N. Krishna, R. Rama, A variant of P-systems with active membranes: Solving NP-complete problems, Rom. J. of Inf. Sci. and Tech., 2, 4 (1999)
- [12] Z. Gazdag, G. Kolonits, A new method to simulate restricted variants of polarizationless P systems with active membranes, Journal of Membrane Computing, 1, 4, 2019, 251–261
- [13] A. Leporati, C. Zandron, M. A. Gutierrez-Naranjo, P systems with input in binary form, Int. J. of Found. of Comp. Sci., 17(1), 2006, 127–146
- [14] C. Zandron, C. Ferretti, G. Mauri, Solving NP-complete problems using P systems with active membranes, In I. Antoniou, C.S. Calude, M.J. Dinneen, eds., Unconventional Models of Computation, Springer-Verlag, London, 2000, 289–301
- [15] C. Zandron, A. Leporati, C. Ferretti, G. Mauri, M. J. Pérez-Jiménez, On the Computational Efficiency of Polarizationless Recognizer P Systems with Strong Division and Dissolution, Fundamenta Informaticae, 87(1), 2008, 79–91
- [16] A.E. Porreca, G. Mauri, C. Zandron, Non-confluence in divisionless P systems with active membranes, Theoretical Computer Science, 411, 6, 2010, 878–887
- [17] A. E. Porreca, A. Leporati, G. Mauri, C. Zandron, P systems with Elementary Active Membranes: Beyond NP and coNP, in Gheorghie M. et al. (eds.), CMC 2010, Jena, Germany, August 2010, LNCS 6501, Springer, 2011, 338–347
- [18] A. Leporati, L. Manzoni, G. Mauri, A.E. Porreca, C. Zandron, Simulating elementary active membranes with an application to the P conjecture, LNCS 8961, Springer, 2014, 284–299
- [19] A. Alhazov, C. Martin-Vide, L. Pan, Solving a PSPACE-complete problem by P-systems with restricted active membranes, Fund. Inf. 58, 2, 2003, 67–77
- [20] P. Sosik, The computational power of cell division in P systems: Beating down parallel computers?, Natural Computing, 2(3), 2003, 287–298
- [21] A.E. Porreca, G. Mauri, C. Zandron, Complexity classes for membrane systems, RAIRO-Theor. Inform. and Applic. 40(2), 2006, 141–162
- [22] A.E. Porreca, L. Manzoni, A. Leporati, G. Mauri, C. Zandron, Membrane division, oracles, and the counting hierarchy, Fundamenta Informaticae, 138, 1–2, 2015, 97–111
- [23] A. E. Porreca, A. Leporati, G. Mauri, C. Zandron, Introducing a space complexity measure for P systems, Int. J. of Comp. Comm. and Control, 4(3), 2009, 301–310
- [24] A. E. Porreca, A. Leporati, G. Mauri, C. Zandron, P Systems with Active Membranes: Trading Time for Space, Natural Computing 10(1), 2011, 167–182
- [25] A. E. Porreca, A. Leporati, G. Mauri, C. Zandron, P systems with active membranes working in polynomial space, Int. J. Found. Comp. Sc., 22(1), 2011, 65–73
- [26] A. Alhazov, A. Leporati, G. Mauri, A. E. Porreca, C. Zandron, Space complexity equivalence of P systems with active membranes and Turing machines, Theoretical Computer Science 529, 2014, 69–81
- [27] D.A. Mix Barrington, N. Immerman, H. Straubing, On uniformity within NC<sup>1</sup>. Journal of Computer and System Sciences 41(3), 1990, 274–306
- [28] A. E. Porreca, C. Zandron, A. Leporati, G. Mauri, Sublinear Space P systems with Active Membranes, Membrane Computing: 13th International Conference, LNCS, CMC 2012, Springer, Berlin, 2013, 342–357
- [29] J. Cecilia, J. Garcia, G. Guerrero, M. Martinez-del Amor, I. Perez-Hurtado, M.J. Perez-Jimenez, Simulating a P system based efficient solution to SAT by using GPUs, Journal of Logic and Algebraic Programming, 79, 6, 2010, 317–325
- [30] M. Garcia-Quismondo, R. Gutierrez-Escudero, M. Martinez-del Amor, E. Orejuela-Pinedo, I. Perez-Hurtado, P-Lingua 2.0: A software framework for cell-like P systems, International Journal of Computers, Communications and Control, 4, 3, 2009, 234–243
- [31] A. Alhazov, A. Leporati, L. Manzoni, G. Mauri, C. Zandron, Alternative space definitions for P systems with active membranes. Journal of Membrane Computing, 3, 2021, 87–96
- [32] C. Buiu, A.G. Florea, Membrane computing models and robot controller design, current results and challenges, Journal of Membrane Computing, 1, 4, 2019, 262–269
- [33] A. Leporati, G. Mauri, C. Zandron, Quantum Sequential P Systems with Unit Rules and Energy Assigned to Membranes, in R. Freund et al. (eds.), Membrane Computing, 6th Int. Work., WMC 2005, Vienna, Austria, LNCS 3850, Springer, 2006, 310–325
- [34] C. Martín-Vide, Gh. Păun, J. Pazos, A. Rodríguez-Patón, Tissue P systems, Theoretical Computer Science, 296, 2, 2003, 295–326
- [35] R. Freund, A. Leporati, G. Mauri, A.E. Porreca, S. Verlan, C. Zandron, Flattening in (tissue) P systems, LNCS 8340, 2014, 173–188
- [36] B. Song, X. Zeng, A. Rodriguez-Paton, Monodirectional tissue P systems with channel states, Information Sciences, 546, 2021, 206–219
- [37] Y. Luo, Y. Zhao, C. Chen, Homeostasis tissue-like P systems, IEEE Transactions on NanoBioscience, 20(1), 2020, 126–136
- [38] M. Gutierrez-Naranjo, M. J. Perez-Jimenez, A. Riscos-Nunez, F. J. Romero-Campero, Characterizing tractability by cell-like membrane systems, in K.G. Subramanian et al. (Eds.), Formal Models, Languages and Applications, Ser. Mach. Percept. Artif. Intell., vol. 66, World Scientific, 2006, 137–154
- [39] L. Valencia Cabrera, B. Song, Simulating Tissue P systems with promoters through MeCoSim and P-Lingua, Journal of membrane computing, 2 (2), 2019, 95–107
- [40] R. Ceterchi, D. Orellana-Martin, G. Zhang, Division rules for tissue P systems inspired by space filling curves, Journal of Membrane Computing, 3(2), 2021, 105–115
- [41] D. Diaz-Pernil, H.A. Christinal, M.A. Gutierrez-Naranjo, Solving the 3-COL problem by using tissue P systems without environment and proteins on cells, Information Sciences, 430, 2018, 240–246
- [42] A. Leporati, C. Zandron, C. Ferretti, G. Mauri, Tissue P systems with small cell volume, Fundamenta Informaticae, 154, 1–4, 2017, 261–275

- [43] Gh. Păun, M.J. Pérez-Jiménez, A. Riscos-Núñez, Tissue P systems with cell division, *International Journal of Computers, Communications & Control*, 3, 3, 2008, 295–303
- [44] A. Leporati, L. Manzoni, G. Mauri, A.E. Porreca, C. Zandron, Characterising the complexity of tissue P systems with fission rules, *Journal of Computer and System Sciences*, 90, 2017, 115–128
- [45] M. Ionescu, Gh. Paun, T. Yokomori, Spiking neural P systems, *Fundamenta Informaticae*, 71, 2-3, 2006, 279–308
- [46] Y. Jiang, Y. Su, F. Luo, An improved universal spiking neural P system with generalized use of rules, *Journal of Membrane Computing*, 1, 2019, 270–278. <https://doi.org/10.1007>
- [47] L. Pan, Gh. Păun, B. Song, Flat maximal parallelism in P systems with promoters, *Theoretical Computer Science*, 623, 2016, 83–91
- [48] T. Wu, S. Jiang, Spiking neural P systems with a flat maximally parallel use of rules, *Journal of Membrane Computing*, 3, 2021, 221-231
- [49] X. Zhang, B. Wang, L. Pan, Spiking Neural P Systems with a Generalized Use of Rules, *Neural Computation*, 26, 2014, 2925-2943
- [50] N. Zhou, H. Peng, J. Wang, Q. Yang, X. Luo Computational completeness of spiking neural P systems with inhibitory rules for generating string languages, *Theoretical Computer Science*, to appear
- [51] C. Liu, W. Shen, L. Zhang, Y. Du, Z. Yuan, Spike Neural Network Learning Algorithm Based on an Evolutionary Membrane Algorithm, *IEEE Access*, 9, 2021, 17071–17082
- [52] D. Diaz-Pernil, H.A. Peña-Cantillana, M.A. Gutierrez-Naranjo, A parallel algorithm for skeletonizing images by using spiking neural P systems, *Neurocomputing*, 115, 4, 2013, 81–91
- [53] J. Wang, H. Peng, W. Yu, J. Ming, M.J., Perez-Jimenez, C. Tao, X. Huang, Interval-valued fuzzy spiking neural P systems for fault diagnosis of power transmission networks, *Engineering Applications of Artificial Intelligence*, 82, 2019, 102–109
- [54] P. Sosik, P systems attacking hard problems beyond NP: a survey, *Journal of Membrane Computing* 1(3), 2019, 198–208
- [55] A.E. Porreca, A. Leporati, G. Mauri, C. Zandron, Sublinear-space P systems with active membranes. In E. Csuhaj-Varjú, M. Gheorghe, G. Rozenberg, A. Salomaa, G. Vaszil, G. (eds.) *Membrane Computing, 13th International Conference, CMC 2012, LNCS 7762*, 2013, 342–357
- [56] A. Leporati, L. Manzoni, G. Mauri, A. E. Porreca, C. Zandron, Constant-space P systems with active membranes, *Fundamenta Informaticae* 134(1–2), 2014, 111–128
- [57] C. Zandron, Bounding the space in P systems with active membranes, *Journal of Membrane Computing* 2(2), 2020, 137–145
- [58] K.C. Buño, F.G. Cabarle, M.D. Calabia, H.N. Adorna, Solving the N-Queens problem using dP systems with active membranes, *Theoretical Computer Science*, 736, 2018, 1–14
- [59] A. Leporati, L. Manzoni, G. Mauri, A.E. Porreca, C. Zandron, Monodirectional P systems, *Natural Computing*, 15, 4, 2016, 551–564
- [60] A. Leporati, L. Manzoni, G. Mauri, A.E. Porreca, C. Zandron, Characterizing PSPACE with shallow non-confluent P systems, *Journal of Membrane Computing*, 1, 2, 2019, 75–84
- [61] Y. Luo, H. Tan, Y. Zhang, Y. Jiang, The computational power of timed P systems with active membranes using promoters, *Mathematical Structures in Computer Science*, 29, 5, 2019, 663-680
- [62] J. Xue, A. Camino, S.T. Bailey, X. Liu, D. Li, Y. Jia, Automatic quantification of choroidal neovascularization lesion area on OCT angiography based on density cell-like P systems with active membranes, *Biomedical optics express*, 9,7, 2018, 3208-3219
- [63] F. Baquero, M. Campos, C. Llorens, J.M. Sempere, P Systems in the Time of COVID-19, *Journal of Membrane Computing*, 3, 4, 2021, 246–257
- [64] I. Perez-Hurtado, L. Valencia-Cabrera, A. Riscos-Nunez, M.J. Perez-Jimenez, Design of Specific P Systems Simulators on GPUs, In *Membrane Computing: 19th International Conference, CMC 2018, Dresden, Germany, September 4–7, 2018, Revised Selected Papers*, 11399, 2019, 202-207
- [65] L. Duo, L. Xiubin, Z. Qingqi, Q. Yanling, L. Yue, Achievements, challenges, and developing directions of bio-inspired self-repairing technology, *Microelectronics Journal*, 111, 2021, 105044



## Effect Analysis of Creators' Environmental Factors in Making a Tourist Map for a Familiar Place

Yoko Nishihara

*College of Information Science and Engineering  
Ritsumeikan University*

Shiga, Japan  
nisihara@fc.ritsumei.ac.jp

Ryosuke Yamanishi

*Faculty of Informatics  
Kansai University*

Osaka, Japan  
ryama@kansai-u.ac.jp

Xinran Lin

*College of Information Science and Engineering  
Ritsumeikan University*

Shiga, Japan  
is0440re@ed.ritsumei.ac.jp

Ryohei Ohyama

*College of Information Science and Engineering  
Ritsumeikan University*

Shiga, Japan  
is0400hs@ed.ritsumei.ac.jp

**Abstract**—People use maps for directions and sightseeing, but they rarely consider the process of map-making. Since maps are made by several creators cooperatively, map-making can be considered a collaborative decision-making process. If this process is analyzed, we may discover that environmental factors influence map-making. This study discovers the effects of environmental factors for map-making through experiments and analysis. We invited participants to join in the experiments, in which they made a tourist map for a familiar place, and we controlled the environmental factors in the map-making process. The environmental factors include the following: (1) the number of creators, (2) the presence of creators' conversations, and (3) the creators' space. We analyzed the created maps, tourist attractions, and participants' conversations to determine the effects of the environmental factors and found the following three effects: (a) If the creators are many, then more tourist attractions will be mapped, which corresponds to factor (1). (b) If the creators have conversations, the rate of minor tourist attractions increase, which corresponds to factor (2). (c) If the map-making environment is virtual, then the map has more tourist attractions than a physical one, which corresponds to factor (3). We discovered the above effects of the environmental factors in tourist map-making. According to the results of this study, it is expected that people can obtain their desired maps using these factors and effects.

**Keywords** - collaborative decision-making; tourist map making; number of creators; conversation; physical and virtual spaces.

### I. INTRODUCTION

People look at the map for sightseeing that introduces tourist attractions. Tourist maps can be found in guidebooks or the Internet. These maps are available at the tourist information center of the tourist location. Through this map, they can understand the outline of the tourist destination and locate exciting places. Further, they can plan a route to see and enjoy the fascinating tourist attractions.

Although the tourist maps are for the same location, the creators are different, thus the maps are different. Each map has original photos and descriptions of the tourist attractions. People rarely consider the decision-making conducted in making maps and choosing tourist attractions. Map-making is

considered a collaborative decision-making process since it is conducted by several creators cooperatively. If the process of map-making is analyzed, we may discover environmental factors influencing it, which is not for automatic generation of map but for interaction design for map-making.

This study discovers the effects of environmental factors for map-making through experiments and analysis. We invited participants to partake in the experiments of making a tourist map. If the participants are unfamiliar with an area, they may fail in an attempt to create a tourist map. Additionally, if tourist maps already existed, the participants may be confused when creating maps and choosing tourist attractions. Thus, we asked the participants to create a map of a familiar place in the experiment. They re-evaluated the familiar place and chose spots that were preferable for tourist attractions. We controlled factors in the experiments to obtain several types of tourist maps and analyzed the maps to discover how the factors affect tourist map-making.

The task in the experiment is not just find the spot we have already known but more creative where the participants have to re-evaluate and create the meaning toward the spot, thus, we consider that it should be “creator(s)” rather than just “contributor(s).” We also consider that it should be “environmental factor(s)” rather than just “context.” Because the creator's space of map-making is not context.

#### A. Contribution

The world is full of contents. The research in computer science has tried to analyze and utilize such big data of contents, e.g., text classification and music recommendation. So many papers have proposed brand new techniques for processing the data and claimed the effectiveness for some intended tasks. However, we believe that the point of the research focusing on contents should be not only the effectiveness in the designed tasks but also how the data is coming. Especially for some experience-based contents, how the data is created should be more notable.

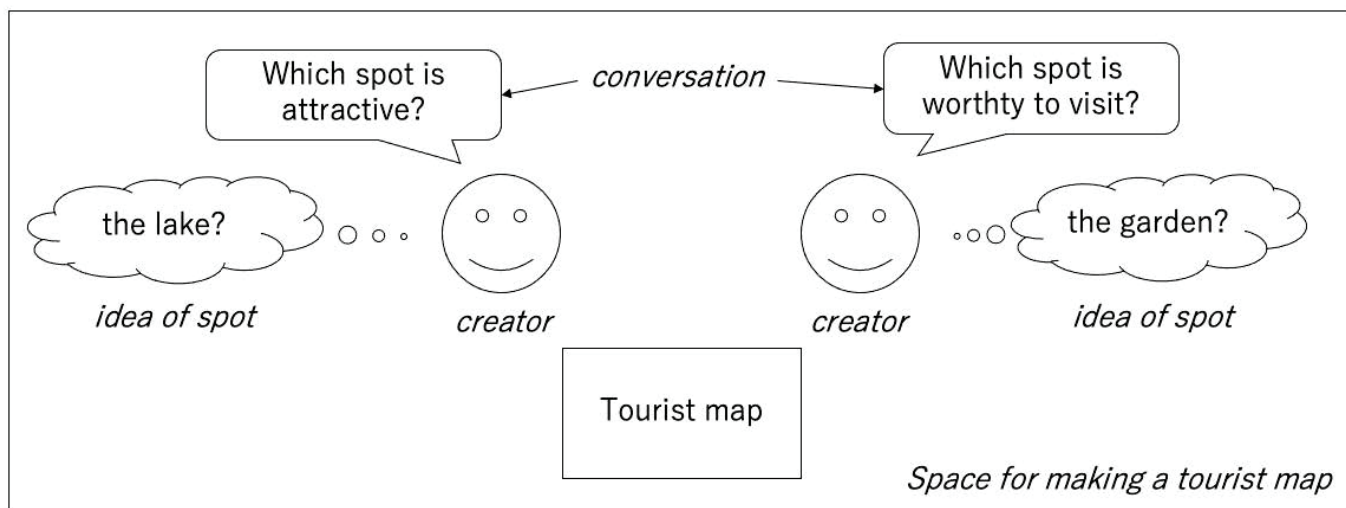


Figure 1. Interaction of tourist map-making. Creators have ideas of spots. They choose spots that would be tourist attractions and map them. They have conversations among them to find the other spots. They can work either in a physical and virtual space.

In our previous study presented at ACHI2021 [1], we conducted a preliminary analysis of two factors and their effects on tourist map-making. This study presents detailed analysis results for the previous experiment and clarifies the effects of the number of creators and the presence of creators' conversations. Further, we include another factor about the creators' space and clarify the effect on tourist map-making.

The contribution of this paper can be summarized as follows;

- 1) It has been found that the number of creators and the usage of virtual space increase the number of tourist attractions on a tourist map.
- 2) It has also been found that conversations between creators increase the rate of minor tourist attractions that may be exciting for tourists. The minor tourist attractions included personal memories and impressions in their descriptions. They re-evaluated their familiar place and found spots preferable for tourist attractions. The re-evaluation can generate a new value for a place even if it is familiar.

The rest of this paper is organized as follows. Section II describes related works. Section III demonstrates a map-making interaction and discusses the expected factors. Section IV describes the first experiment to verify the effect of the number of creators and their conversations in map-making and the results. Section V describes the second experiment to verify the effect of the creators' space in map-making and the results. Section VI concludes this paper.

## II. RELATED WORK

We introduced previous studies on map- and decision-making supports.

### A. Map-making with sensors and social media

People use map services on the Internet to search for buildings and facilities. If the name of a building in the United Kingdom and France are Big Ben and The Arch of Triumph, respectively, the service will show the location of the building on map<sup>1</sup>. The outline of a city changes continuously as new roads and buildings are built. Sensors such as GPS (Global Positioning System) [2], [3] are used to update the existing map to a new one. GPS data is used not only for updating maps but also for making new ones; for example, a cycling map [4], an evacuation map in a disaster-hit area [5], and a floor map inside [6]. The traffic volume of a city changes daily. There may be a traffic jam in the morning and evening, but not during the day. GPS in automobiles is used for making a traffic map of a city [7]–[9]. Sensors are installed across a city. Acoustic sensors are used for making a city noise map [10]. Geographic Information System (GIS) data are also sensor data that is used for making a hazard map of a landslide [11].

Recently, short messages on social media such as tweets are used for locating tourist attractions in a city [12], [13]. People can locate streets suitable for jogging by collecting tweets that include the word “jogging.” Further, by collecting short messages such as sunset, people can discover viewpoints for a beautiful sunset [14]. Tweet data can be used for detecting a neighbor's boundary [15]. The activities of pets can be collected and mapped from social media [16]. Because search results on a Web's search engine include future event information, the event information is extracted and shown on a map [17].

Our study does not create a map using sensors or social media. The creators remember suitable spots for tourist at-

<sup>1</sup>For example, Google Map is one of the services (<https://www.google.com/maps>).

tractions on the map. We analyze the characteristics of the maps to find how the desired map can be obtained.

### B. Decision-making support

Decision-making has been supported in various ways, and multi-agent simulation systems have been studied for complex tasks with numerous constraints and alternatives [18]. Further, optimization methods have been studied for interaction among agents, robots, and people [19]. Virtual reality environments are designed for specific decision-making processes, such as policymaking [20]. Health care support is provided through online social networking services [21]. Each study discovered points that should be considered in decision-making. These points differ in each decision-making task. The decision-making process would be analyzed to determine these points [22].

The volume of decision-making increases as the number of workers increases; this is called collaborative decision-making. Many studies have supported collaborative decision-making. Each worker in a collaborative decision-making process has the best solution for themselves. However, when a worker provides the best solution for a small problem, the set of solutions fails to solve the entire problem [23], [24]. Consensus-building models and methods have been studied [25], [26]. Workers in decision-making have a conflict of opinions. Thus, the degree of conflict is evaluated to eliminate consensus-building [27]. An enormous decision-making task is broken down into small decision-making tasks for consensus-building [28]. Our study focuses on map-making as a collaborative decision-making process and analyzes the process of locating the support point.

### III. EXPECTED ENVIRONMENTAL FACTORS IN TOURIST MAP-MAKING

We consider tourist map-making a collaborative decision-making process. We demonstrated a tourist map-making interaction and discussed the expected factors.

In map-making, people congregate in a space to discuss and give their opinions [29]. Previously, the creators' space was physical, but people can now use virtual space and Web services for map-making. Figure 1 illustrates the interaction of a tourist map-making process. Each creator gives an idea of a spot to map them. If there are several creators, they will share their ideas of spots that others have not provided. If the creators can have conversations, they would get others' opinions to locate the other ideas of spots. If they are in the same physical space, they can walk the area on the tourist map to choose the spots for tourist attractions. Map-making is conducted collaboratively in a physical space. Even if the creators are in a virtual space, they can walk the area and map the tourist attractions using Web services.

Figure 2 shows the relations between expected environmental factors and map-making if tourist map-making is performed. There are three expected environmental factors:

- (1) The number of creators
- (2) The presence of creators' conversations

- (3) The creators' space

Although each creator's choice of tourist attraction is different and uncontrollable, the three factors are controllable through experiments. Thus, we chose three environmental factors that would affect map-making. It is unclear whether these factors affect map-making simultaneously or individually. Further, we analyzed how the factors affect map-making in this study.

We conducted two types of experiments to determine the effects of the three factors. The first experiment was conducted to find the effects of factors (1) and (2) (in Section IV). The second experiment was conducted to find the effect of factor (3) (in Section V).

### IV. EXPERIMENT 1: DO THE NUMBER OF CREATORS AND THEIR CONVERSATIONS AFFECT TOURIST MAP-MAKING?

In this section, we set the main hypothesis as follows: **if two creators make a tourist map while walking the area and conversing, the tourist map will be different from a map created by a single person.** The hypothesis has two factors: factor (1) and (2), which are the number of creators and the presence of conversations, respectively. We divide the main hypothesis into four smaller hypotheses: H1(a), H1(b), H2(a), and H2(b).

- H1(a): **The number of tourist attractions is higher** if a tourist map is created **by two people** instead of a single person.
- H1(b): **The number of tourist attractions is higher** if a tourist map is created **by two people with conversations** instead of without conversations.
- H2(a): **The rate of minor tourist attractions is higher** if a tourist map is created **by two people** instead of a single person.
- H2(b): **The rate of minor tourist attractions is higher** if a tourist map is created **by two people with conversations** instead of without conversations.

#### A. Experimental settings

The experimental procedures are as follows:

- 1) An experimenter directs participants to a place for tourist map-making.
- 2) The participants navigate the place and choose spots for tourist attractions by taking photos. They have a limited time.
- 3) After procedure 2), the participants upload the photos to Google map<sup>2</sup> with a title and a description to map the tourist attractions.

Each individual/pair of the participants created a tourist map of the Biwako-Kusatsu Campus of the Ritsumeikan University. The campus was not a tourist destination, but it is familiar to the participants. The experimenter was the third author. The participants were 35 students who had been on the campus for over a year.

<sup>2</sup><https://www.google.com/maps>

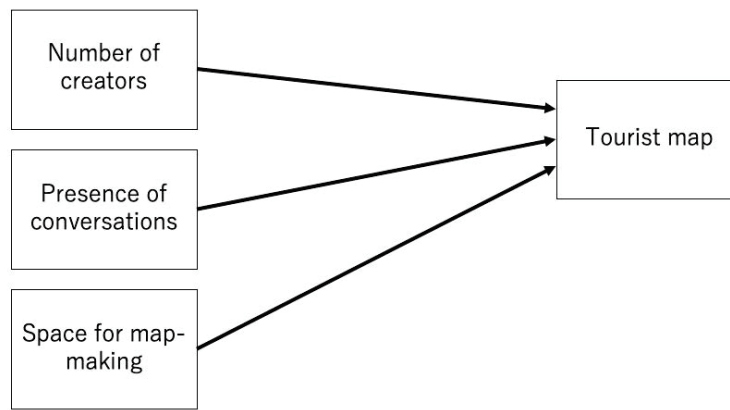


Figure 2. Relations between expected environmental factors and map-making. We assume that the environmental factors are the number of creators, the presence of conversations among creators, and the space of map-making.

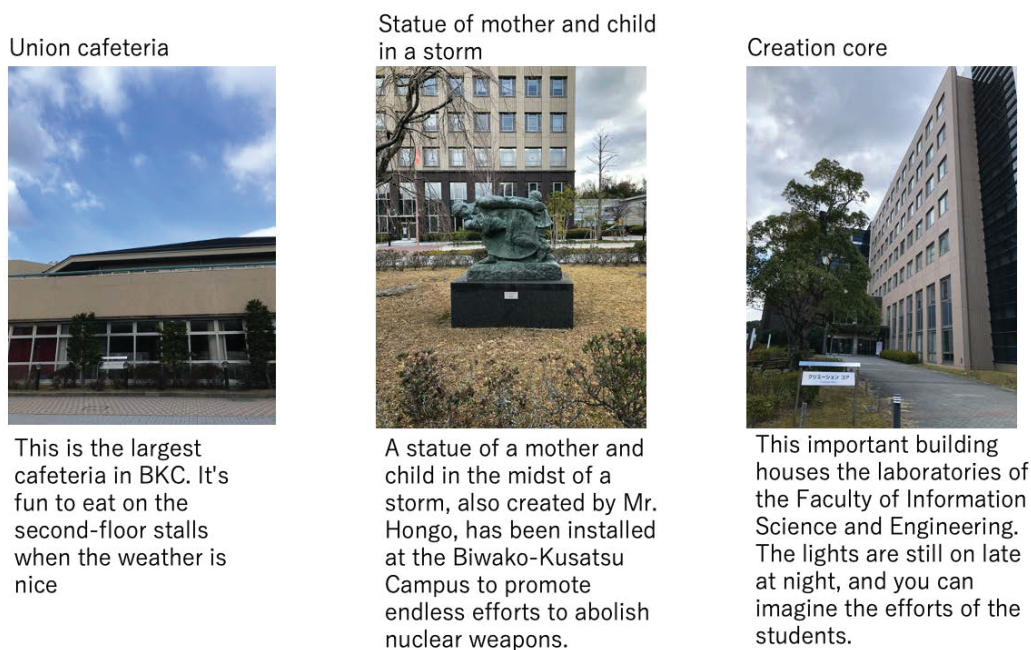


Figure 3. Example of tourist attractions by Group A in Experiment 1. Group A made a map by a single person.

We divided the participants into Groups A, B, and C. The information in each group is given as follows:

- Group A: Creating a map by a **single person**: seven participants
- Group B: Creating a map by **two people without conversations**: seven pairs and 14 participants
- Group C: Creating a map by **two people with conversations**: seven pairs and 14 participants

Group C recorded their conversations using a voice recorder. The time to navigate the campus in procedure 2) was 45 minutes. The experiments were conducted between April and September 2020.

#### B. Obtained data

We obtained the following data.

- Data 1: The number of tourist attractions of each map
- Data 2: The duration for map-making in procedure 3)
- Data 3: Transcripts of creators' conversations (if any)

#### C. Method of hypothesis verification

We used Data 1 and Data 3 to verify hypotheses H1(a) and (b) and hypotheses H2(a) and (b), respectively. The campus has buildings for classes and research. The participants might choose the buildings as tourist attractions. Further, we would like to evaluate the quality of each map based on the types of tourist attractions. We regard a tourist attraction as a major spot if it is a building/facility that is mapped on a campus map, which is published by the university<sup>3</sup>. However, if the description of a tourist spot includes personal memories and

<sup>3</sup><http://en.ritsumei.ac.jp/campusmap/>



Link meal shop



It's very convenient to shop between classes. I often go there because I can buy a drink right after I finish my lunch.

Co-learning house 1



This is where most of the classes are held. There are classrooms with a capacity of over 500 students. The toilets are very luxurious.

Co-learning house 2



There is a laboratory in the School of Food Management. This is Co-Learning House 2. The School of Information Science and Technology doesn't have many classes here.

Figure 4. Example of tourist attractions by Group B in Experiment 1. Group B made a map by two people without conversations.

View from 7<sup>th</sup> floor of Creation Core

A panoramic view of Lake Biwa

Quinss stadium



Sports watching!

Tricia



Stylish building at Ritsumeikan University

Figure 5. Example of tourist attractions by Group C in Experiment 1. Group C made a map by two people with conversations.

impressions, the above mentioned building/facility is a minor spot. If a tourist attraction does not exist on the campus map, it is a minor spot.

#### D. Experimental results

Figure 3, Figure 4, and Figure 5 show examples of the tourist attractions chosen by the participants. The figures show photos, titles, and descriptions. Buildings (Co-learning House 1 and 2), cafeterias (Union Cafeteria and Link Meal Shop), and facilities (Creation Core and Tricia) are shown.

The second column of Table I shows the number of tourist attractions corresponding to Data 1. The numbers are averages

of seven pairs in each group. The averages of Groups A, B, and C are 17.6, 18.1, and 10.3, respectively.

The third column of Table I shows the duration of map-making corresponding to Data 2. The durations are averages of the seven pairs in each group. The averages for Groups A, B, and C are 32.1, 28.6, and 22.1 minutes, respectively.

The fourth column of Table I shows the rate of minor spots. The averaged rates for Groups A, B, and C are 68.3%, 73.7%, and 86.1%, respectively.

TABLE I

EXPERIMENTAL RESULTS IN EXPERIMENT I. AVERAGED NUMBERS OF TOURIST ATTRACTIONS ON CREATED MAPS, AVERAGED DURATIONS USED FOR MAP-MAKING, AND AVERAGED RATES OF MINOR SPOTS ARE SHOWN.

Group	Averaged number of tourist attractions	Averaged duration (minutes)	Rate of minor spots (%)
A	17.6	32.1	68.3
B	18.1	28.6	73.7
C	10.3	22.1	86.1

TABLE II  
CONVERSATION EXAMPLES BY PARTICIPANTS OF GROUP C IN  
EXPERIMENT I.

Tourist attraction: A view from 7th floor of Creation Core
Participant B: You can see a nice view from the top floor of Creation Core.
Participant A: It's so nice.
Participant B: The windows are messy but you can see Biwa-lake here. How do you feel it?
Participant A: I love this. This is the highest building in the campus?
Participant B: Probably yes. This angle is also nice. You can see the whole view of Biwa-lake.
Tourist attraction: Tricia
Participant A: The building is for Department of Architecture. I hear that is a cool building. Shall we go there?
Participant B: I have never been there.
Participant A: Students' works and architecture models are displayed in the building.
Participant B: Their works are so nice. I love them.
Participant A: Me too. They are so beautiful.

### E. Results of hypothesis verification

We examined H1(a) and there was no difference between the number of tourist attractions of Groups A and B (17.6 and 18.1). We discovered that hypothesis H1(a) was not held.

We examined H1(b) and there was a difference between the numbers of tourist attractions of Groups B and C (18.1 and 10.3). The number of Group B was higher and hypothesis H1(b) was not held.

We examined H2(a) and there was a difference between the rates of minor spots of Groups A and B (68.3% and 73.7%). The rate of Group B was higher and hypothesis H2(a) was held.

We examined H2(b) and there was a difference between the rates of minor spots of Groups B and C (73.7% and 86.1%). The rate of Group C was higher and hypothesis H2(b) was held.

### F. Discussion

1) *H1(a) was not held:* The small hypothesis H1(a) occurred because the number of tourist attractions is bigger if a tourist map is created by two people instead of a single person. However, there was no difference between the two groups. Each pair of the participants in Group B consisted of two people. We assumed that Group B mapped more tourist attractions because there were more participants in a pair. The participants walked on the campus together and with a time limit. Thus, the participants of Group B did not discover more tourist attractions. Although there was no difference between the number of tourist attractions, the time for map-making of

Group B was approximately three minutes more than Group A (= 32.1– 28.6 minutes). This is because the Group B participants can split between uploading photos and writing descriptions about tourist attractions.

2) *H1(b) was not held:* The small hypothesis H1(b) occurred because the number of tourist attractions is higher if a tourist map is created by two people with conversations instead of without conversations. However, the number of Group B (two people without conversations) was greater than that of Group C (two people with conversations). We assumed that the difference occurred because of the participants' conversations. Group C participants were permitted to have conversations while walking and map-making. They suggested spots that the others might not be aware of. Table II shows examples of their conversations. Both conversations are for two different tourist attractions. The conversations show that a participant introduces a spot the other is unfamiliar with. Group C participants assumed that sharing common knowledge about the campus was meaningless. They re-evaluated the campus and discovered tourist attractions that the others are not aware of and mapped the areas on their map. Although the condition to converse positively affected finding minor tourist attractions, Group C participants required more time to locate the spots. Thus, the number of tourist attractions for Group C was smaller.

3) *H2(a) was held:* The small hypothesis H2(a) occurred because the rate of minor tourist attractions is higher if a tourist map is created by two people instead of a single person. Though the participants of Group B walked together, they chose spots individually. There are the limited number of major spots so that the two participants can cover most of them. They tried to find another spots rather than the major spots. Thus, the rate of Group B was greater than that of Group A.

4) *H2(b) was held:* The small hypothesis H2(b) occurred because the rate of minor tourist attractions is higher if a tourist map is created by two people with conversations instead of without conversations. The reason is the same as the reason of the small hypothesis H1(b). The participants of Group C were permitted to have conversations. The conversation affect positively to introduce minor spots to the others. Thus, the rate of Group C was greater than that of Group B.



## V. EXPERIMENT 2: DO THE DIFFERENCE BETWEEN PHYSICAL AND VIRTUAL SPACES AFFECT TOURIST MAP-MAKING?

In this section, we set the main hypothesis as follows: **the tourist map is different from each other if the spaces where the creators occupy for map-making are different.** The hypothesis is based on one factor (3) the space where the creators occupy (physical versus virtual). We divide the main hypothesis into two small hypotheses H3 and H4.

- H3: **The number of tourist attractions is higher** if a tourist map is created **in the virtual space** rather than the physical space.
- H4: **The rate of minor tourist attractions is higher** if a tourist map is created **in the virtual space** rather than the physical space.

### A. Experimental settings

The experimental procedures are almost the same as those in Experiment 1. The differences are explained below. The experimenter was the fourth author. The participants were also different from those in Experiment 1; they consisted of 28 students who had been on campus for over a year.

We divided the participants into Groups D and E. Groups D and E walked into the physical and virtual spaces of the location, respectively. The information in each group is given as follows:

- Group D: Walking and map-making **in a physical space**: seven pairs and 14 participants
- Group E: Walking and map-making **in a virtual space**: seven pairs and 14 participants

The physical space was the university campus, whereas the virtual space was Google's Street View<sup>4</sup>. The experiments were conducted between June and December 2021, which was the year after Experiment 1.

### B. Obtained data

We obtained the following data: Data 1, Data 2, and Data 3, which were the same as those in Experiment 1. Data 4 was acquired newly.

- Data 4: Walking routes of the participants.

### C. Method of hypothesis verification

We used Data 1 and Data 3 to verify hypotheses H3 and H4, respectively. We employed the same criteria to determine if a tourist attraction is a major or minor spot in Section IV.

### D. Experimental results

Figure 6 and Figure 7 show examples of the tourist attractions chosen by the participants. The figures show the buildings (Central Arc), facilities (Across Wing), and track field (Quinns Stadium).

The second column of Table III shows the number of tourist attractions that correspond to Data 1. The numbers are

averages of seven pairs in each group, and the averages for Groups D and E were 9.9 and 15.3, respectively. Groups D and C had the same conditions in Experiment 1. The numbers of the two groups were almost the same (10.3 and 9.9 for Groups C and D, respectively).

The third column of Table III shows the duration of map-making that corresponds to Data 2. The durations are averages of seven pairs in each group. The averages were 26 and 25 minutes for Groups D and E, respectively. The duration of Groups C and D was almost the same (22.1 and 26 minutes).

The fourth column of Table III shows the rate of minor spots. The averaged rates were 49% and 52% for Groups D and E, respectively. Although Groups D and C had the same conditions in Experiment 1, the rate of Group D was smaller than that of Group C (86.1% and 49%). In the next section, we will discuss why the rate for Group D was smaller.

### E. Results of hypothesis verification

The number of tourist attractions for Group D was lower than Group E (9.9 and 15.3, respectively). We discovered that hypothesis H3 was held.

There was no difference between the rates of minor spots for Groups D and E (49% and 52%). We discovered that hypothesis H4 was not held.

### F. Discussion

1) *H3 was held*: The small hypothesis H3 occurred because the number of tourist attractions is higher if a tourist map is created in a virtual space instead of a physical space. The participants of Group E used Google's Street View to walk and choose tourist attractions. They could walk faster than those in Group D who were walking on the actual campus. We examined the routes of Group E and discovered that they jumped from one spot to another. The routes of Group E were sets of dots, whereas those of Group D were linear lines. Because of the jump actions, the participants of Group E could check a wider portion of the campus, thus increasing the number of tourist attractions. Therefore, the small hypothesis H3 was held.

2) *H4 was not held*: It occurred because the rate of minor tourist attractions is higher if a tourist map is created in the virtual space rather than the physical space. However, there was no difference between the two rates. We assumed that the participants of both groups were students who knew the campus and would have similar experiences in either the physical or virtual spaces. The participants of both groups chose tourist attractions without effects from their spaces.

3) *Difference between Group C in Experiment 1 and Group D in Experiment 2*: The rate of the minor spots of Group D was smaller than that of Group C in Experiment 1. This might be because of the amount of personal memory of each participant. Experiment 1 in Section IV was conducted between April and September 2020. Experiment 2 in Section V was conducted between June and December 2021 after Experiment

<sup>4</sup><https://www.google.com/streetview/>

Creation core



The building where the Faculty of Information Science and Engineering is located. It is also the place where we spend most of our time.

Central arc



For international students who need to visit the International Education Center on a regular basis, there is good coffee available on the first floor.

Arcross wing



Probably the most prominent and tallest building in BKC. It is a symbolic presence in BKC that can be seen when passing through Meishin Expressway. The view from the rooftop is fantastic, but it is only open during events such as the Open Campus.

Figure 6. Example of tourist attractions by Group D in Experiment 2. Group D made a map in a physical space.

Co-learning house 1



A place where various undergraduates can study. Some departments don't use some of the facilities, but there are no departments that don't use Co-Learning House 1.

Quinns stadium



An athletic club is active in the school. Because of its large ground, it has been used for TV filming.

Union square cafeteria



A place with a cafeteria or convenience store. On days when there are classes, lunch time can get crowded, so eat early or take a seat and eat quickly to give it to someone else.

Figure 7. Example of tourist attractions by Group E in Experiment 2. Group E made a map in a virtual space.

1. The campus was occasionally closed because of COVID-19 from the beginning of 2020. The participants of Experiment 2 might have less personal memories of the campus because of the long shutdown. Thus, the rate of the minor spots of Group D is lower.

4) *Selection order of tourist attractions:* We examined the selection order of tourist attractions for both groups and assigned a turn number for each utterance. Then, we marked the utterances where the tourist attractions appeared first. The

turn number of the utterance was the number where the tourist attraction was chosen. We calculated the averaged numbers of the major and minor spots' first appearance. Each transcript had different lengths; thus, the numbers of utterances differed. We normalized the turn-numbers so that the minimum and maximum were one and 10, respectively. Table IV shows the averaged turn-numbers of the major and minor spots' initial appearance. There was no difference between the averages of major spots of Groups D and E (4.5 and 4.3). There was also no difference in the minor spots (6.6 and 7.0).

TABLE III

EXPERIMENTAL RESULTS IN EXPERIMENT 2. AVERAGED NUMBERS OF TOURIST ATTRACTIONS ON CREATED MAPS, AVERAGED DURATION USED FOR MAP-MAKING, AND AVERAGED RATES OF MINOR SPOTS ARE SHOWN.

Group	Averaged number of tourist attractions	Averaged duration (minutes)	Rate of minor spots (%)
D	9.9	26	49
E	15.3	25	52

TABLE IV

AVERAGED TURN NUMBERS INCLUDING A TOURIST ATTRACTION NAME.

Group	Major spots	Minor spots
D	4.5	6.6
E	4.3	7.0

The common point was that major spots were mentioned first, and minor spots were mentioned later. When choosing tourist attractions, the participants first remembered familiar spots that became major spots. Once they had exhausted the idea, they started considering other preferable spots for tourist attractions. During that time, they would look for spots that others might not know, which they could recall; those spots would become minor spots. The results indicate that the selection order of tourist attractions was not affected by the creators' spaces.

## VI. CONCLUSIONS

This study focused on map-making as a collaborative decision-making process. We analyzed the map-making process and conducted experiments to determine the effects of environmental factors. The expected environmental factors were (1) the number of creators, (2) the presence of creators' conversations, and (3) the space occupied by the creators. We asked participants to make a tourist map for a familiar place under controlled factors. We analyzed the created maps and other experimental results. The following three effects were discovered:

- (a) The more the creators of a tourist map, the more tourist attractions obtained.
- (b) Since conversations are allowed, the creators can change their minds while choosing tourist attractions unknown to others. Thus, the created map has more exciting tourist attractions.
- (c) The number of tourist attractions becomes higher if the space in choosing them is a virtual space.

The participants made a tourist map for a familiar place in this study. As the future work, we would like to conduct another experiment to create a tourist map for an unfamiliar place. The three factors will be examined in the next experiments to discover how they affect map-making.

## ACKNOWLEDGEMENT

This paper was partly supported by Ritsumeikan Global Innovation Research Organization and "International Collaborative Research for the Digital Archive of Japanese Cultural Resources," Art Research Center, Ritsumeikan University. We show our great appreciation.

## REFERENCES

- [1] Y. Nishihara, X. Lin, and R. Yamanishi, "Do the number of creators and their conversations affect re-evaluation of a familiar place in making a tourist map?" in The Fourteenth International Conference on Advances in Computer-Human Interactions (ACHI2021) in Nice, France, July 2021, pp. 55–56.
- [2] C. Chen, C. Lu, Q. Huang, Q. Yang, D. Gunopulos, and L. Guibas, "City-scale map creation and updating using gps collections," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: Association for Computing Machinery, 2016, pp. 1465–1474.
- [3] J. Davies, A. Beresford, and A. Hopper, "Scalable, distributed, real-time map generation," IEEE Pervasive Computing, vol. 5, no. 4, 2006, pp. 47–54.
- [4] U. Blanke, R. Guldener, S. Feese, and G. Tröster, "Crowdsourced pedestrian map construction for short-term city-scale events," in Proceedings of the First International Conference on IoT in Urban Space. Brussels, BEL: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2014, pp. 25–31.
- [5] K. Asakura, M. Takeuchi, and T. Watanabe, "A pedestrian-oriented map matching algorithm for map information sharing systems in disaster areas," International Journal of Knowledge and Web Intelligence, vol. 3, no. 4, Jan. 2012, pp. 328–342.
- [6] M. Alzantot and M. Youssef, "Crowdinside: Automatic construction of indoor floorplans," in Proceedings of the 20th International Conference on Advances in Geographic Information Systems. New York, NY, USA: Association for Computing Machinery, 2012, pp. 99–108.
- [7] R. Verma, S. Ghosh, A. Shrivastava, N. Ganguly, B. Mitra, and S. Chakraborty, "Unsupervised annotated city traffic map generation," in Proceeding of the 24th ACM SIGSPATIAL International Conference, 10 2016, pp. 1–4.
- [8] J. Biagioni, T. Gerlich, T. Merrifield, and J. Eriksson, "Easytracker: Automatic transit tracking, mapping, and arrival time prediction using smartphones," in Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems, 11 2011, pp. 68–81.
- [9] S. Nawaz and C. Mascolo, "Mining users' significant driving routes with low-power sensors," in Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems, ser. SenSys '14. New York, NY, USA: Association for Computing Machinery, 2014, pp. 236–250.
- [10] Y. Wang, Y. Zheng, and T. Liu, "A noise map of new york city," in Proceedings of the 16th ACM International Conference on Ubiquitous Computing, August 2014.
- [11] S. Lee and C. P. Poudyal, "Landslide hazard mapping using geospatial models," in Proceedings of the 2nd International Conference on Computing for Geospatial Research and Applications, ser. COM.Geo '11. New York, NY, USA: Association for Computing Machinery, 2011, pp. 1–5.
- [12] L. Ferrari, A. Rosi, M. Mamei, and F. Zambonelli, "Extracting urban patterns from location-based social networks," in Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks, ser. LBSN '11. New York, NY, USA: Association for Computing Machinery, 2011, pp. 9–16.
- [13] C. Xia, R. Schwartz, K. Xie, A. Krebs, A. Langdon, J. Ting, and M. Naaman, "Citybeat: Real-time social media visualization of hyper-local city data," in Proceedings of the 23rd International Conference on World Wide Web, ser. WWW '14 Companion. New York, NY, USA: Association for Computing Machinery, 2014, pp. 167–170.
- [14] B. Pat, Y. Kanza, and M. Naaman, "Geosocial search: Finding places based on geotagged social-media posts," in Proceedings of the 24th International Conference on World Wide Web, 05 2015, pp. 231–234.
- [15] J. Cranshaw, R. Schwartz, J. I. Hong, and N. Sadeh, "The livelihoods project: Utilizing social media to understand the dynamics of a city,"

- in Proceedings of the International AAAI Conference on Weblogs and Social Media, vol. 6, no. 1, 2012, pp. 58–65.
- [16] A. S. Jose and E. Hernandez, "City-scale mapping of pets using georeferenced images," *SIGSPATIAL Special*, vol. 8, no. 3, May 2017, pp. 5–6.
- [17] W. Li, P. Wang, and K. Yang, "Visualizing city events on search engine: Tward the search frustration for smart city," in 2015 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, 2015, pp. 1019–1026.
- [18] J. Himoff, "Decision-making support for maritime business," in Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems, ser. AAMAS '06. New York, NY, USA: Association for Computing Machinery, 2006, pp. 1455–1456.
- [19] Z. Peng, Y. Kwon, J. Lu, Z. Wu, and X. Ma, "Design and evaluation of service robot's proactivity in decision-making support process," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–13.
- [20] D. Filonik, A. Buchan, L. Ogden-Doyle, and T. Bednarz, "Interactive scenario visualisation in immersive virtual environments for decision making support," in Proceedings of the 16th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry, ser. VRCAI '18. New York, NY, USA: Association for Computing Machinery, 2018, pp. 1–4.
- [21] V. Sadovykh and D. Sundaram, "Decision making and support in healthcare online social networks," in Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, ser. ASONAM '15. New York, NY, USA: Association for Computing Machinery, 2015, pp. 1048–1052.
- [22] J. L. Mishra, D. K. Allen, and A. D. Pearman, "Information use, support and decision making in complex, uncertain environments," in Proceedings of the 76th ASIST Annual Meeting: Beyond the Cloud: Rethinking Information Boundaries, ser. ASIST '13. USA: American Society for Information Science, 2013, pp. 1–10.
- [23] G. L. Klein, J. L. Drury, and M. S. Pfaff, "Coaction: Collaborative option awareness impact on collaborative decision making," in 2011 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2011, pp. 171–174.
- [24] M. Al-Shawa, "Modeling and analyzing agents' collective options in collaborative decision making situations using the constrained rationality framework," in 2011 IEEE International Conference on Systems, Man, and Cybernetics, 2011, pp. 1626–1631.
- [25] X. Li, H. Zhang, R. Mao, and X. Wang, "A consensus reaching model for collaborative decision making in web 2.0 communities," in 2013 Sixth International Conference on Business Intelligence and Financial Engineering, 2013, pp. 53–56.
- [26] J. M. E. M. van der Werf, R. de Feijter, F. Bex, and S. Brinkkemper, "Facilitating collaborative decision making with the software architecture video wall," in 2017 IEEE International Conference on Software Architecture Workshops (ICSAW), 2017, pp. 137–140.
- [27] P. Pengfei, L. Yating, and Z. Linru, "Research on interactive collaborative decision-making method of equipment support task planning," in 2020 5th International Conference on Computer and Communication Systems (ICCCS), 2020, pp. 533–537.
- [28] P. Peng and Q. Yu, "Study on framework of interactive and collaborative decision making and design of support system based on network center," in 2015 8th International Symposium on Computational Intelligence and Design (ISCID), vol. 2, 2015, pp. 128–131.
- [29] N. Komiya, "Community safety maps for children in Japan: An analysis from a situational crime prevention perspective," *Asian Criminology*, vol. 6, no. 131, 2011, pp. 131–140.

# Stock Trade Simulation Using an Average Based Trend Indicator with Heuristic Enhancements

Yoshihisa Udagawa

Faculty of Informatics, Tokyo University of Information Sciences  
Chiba-city, Chiba, Japan  
e-mail: yu207233@rsch.tuis.ac.jp

**Abstract**— The spread of Covid-19 is making a serious impact on the world economy. As policies to maintain economic activities have been implemented in a timely manner, many stock markets have regained their stock prices to pre-Covid-19 pandemic levels. This paper describes results of statistical analyses of stock price fluctuations in the vicinity of the declaration of a state of emergency following the Covid-19 pandemic. Daily historical data concerning Dow Jones Industrial Average (DJIA), Nasdaq Composite Index (NASDAQ), France Stock Index (CAC), German Stock Index (DAX), Shanghai Composite Index (SSEC) and Nikkei Stock Average (Nikkei) are used for stock trade simulations. The results show that these stock prices plunged for approximately 20 days before the trading day that marked the lowest price, and increased for the following 25 days. Through the analyses of the results, a technical indicator defined by the difference between a stock price and the moving average is devised to predict stock price trend reversals. A stock trade simulator is developed using the devised indicator to examine the degree of stock price prediction. The simulator is enhanced by heuristics using candlestick patterns to avoid large losses. Experimental results using ten years of stock price data from nine world major stock markets show that the simulator achieved the success rate of up to 74% with trade fees for purchasing a stock. Meanwhile, the success rate remains within 25% for selling a stock, because a margin loan for borrowing a stock squeeze profits. Although the developed simulator has some limitations, it provides a tool for measuring profitability among stock markets and analyzing a stock trade opportunity that leads to a significant profit and/or loss.

**Keywords**— Covid-19 pandemic; Technical analysis; Global market comparison; Candlestick chart pattern, APAD.

## I. INTRODUCTION

This research paper is based on the previously reported contribution on analyses of the impact of Covid-19 pandemic on global stock prices [1]. More than two years have passed since the first Covid-19 pandemic was confirmed. But the spread of coronavirus is still continuing and the number of daily corona cases remains high. The spread of Covid-19 is forecasted to have a significant adverse impact on the global economy. While the global real Gross Domestic Product (GDP) grew by 2.9 percent in 2019, it was projected at -4.9 percent in 2020 [2]. The Covid-19 pandemic had a more negative impact on activity in the first half of 2020 than anticipated. Global growth was projected at 6% in 2021,

moderating to 4.4% in 2022. While the vaccine is expected to be effective for economic recovery, there are concerns about the impact of the zero-tolerance Covid-19 policy in China [3].

World stock markets experienced a large crash in the first quarter of 2020. Concerns about a further plunge of stock prices prevailed over global markets. However, due to the economic policies of each country, stock prices of world markets have turned from falling to rising in Mar. 2020. Because it is rare for the major world stock price indexes to fall all at once and then recover, we expect to deepen our understanding of stock price movements and to find an indicator for signaling stock price trend reversals by analyzing stock price fluctuations regarding the declaration of a state of emergency on Covid-19 pandemic.

It can be easily inferred that there is a difference in the degree of collapse and recovery reflecting the situation in each country. What is missing is a comparison of the stock price fluctuations of world markets from a statistical point of view. This study discusses a comparison of representative stock indexes of the U.S., European, and Asian markets [4]. Specifically, we focus on Dow Jones Industrial Average (DJIA), Nasdaq Composite Index (NASDAQ), France Stock Index (CAC), German Stock Index (DAX), Shanghai Composite Index (SSEC), and Nikkei Stock Average (Nikkei). Daily historical data are used for the research.

Dimson et al. [5] recommend investing in the U.S. markets rather than emerging ones because of the growth rate of stock prices and the stability of the investment environment. Reference [6] compares profitability of the U.S. and Asian markets by simulations that find buy-timing using a candlestick pattern model consisting of six parameters. The results of the simulation shows that the profitability of the U.S. markets outperforms other markets. However, these studies were carried out before the spread of the Covid-19 coronavirus.

Ngwakwe [7] estimates how Covid-19 pandemics affected world stock indexes, i.e., Euronext 100, SSEC, DJIA, and S&P500. The results of analyses show that SSEC has resilience to Covid-19 pandemic with profit in stock values during the first fifty days into the pandemic, while the other indexes experience adverse impact from the Covid-19 pandemic with a significant loss at that time period. All stock indexes experience a higher variability or fluctuation of stock market prices.



Verma et al. [8] statistically analyze the impact of the Covid-19 outbreak on global economic development. The indicators used in the analyses include S&P500 stock index, crude oil, gold, and 20-year treasury bond. They find that S&P500 stock index experiences high uncertainty from Feb. 2019 to Apr. 2020, i.e., the latest month of their research.

Buszko et al. [9] deal with the problem of stability of stock prices during the Covid-19 pandemic. The K-means and the Ward clustering methods are applied to the Warsaw Stock Exchange (WSE), which is one of the most important markets in the Central and Eastern European countries (CEE). The paper proposes several indicators describing the stability of stock price fluctuations, and concludes that the indicators defined in terms of profitability, volume, and volatility leads to much better results than other indicators.

This paper has two purposes. One is to analyze stock price fluctuations of approximately 245 trading days, or approximately one year, before and after the declaration of Covid-19 pandemic, and statistically compare degrees of impact on the world markets [1]. Through the analysis, we have devised an indicator for signaling a change of the direction of a price trend. The other purpose of this paper is to give definition of the devised indicator, and to discuss its prediction accuracy in terms of profits with experimental results using major global stock market's daily price data.

Obthong et al. [10] review studies on machine learning techniques and algorithms employed to improve the accuracy of stock price prediction. Of the ten studies compared, four show that prediction accuracies range from 54.41% to 90.19%. Islam et al. [11] show a comparative study for stock price prediction using three different methods, namely autoregressive integrated moving average (ARIMA), artificial neural network (ANN), and geometric Brownian motion. Empirical results using S&P500 index show that the conventional statistical model ARIMA and the Brownian motion model provide better approximation than the ANN model for next-day stock price prediction.

Stock trading consists of selling and buying pairs, and profits are fixed. What is lacking in these studies is experimental evaluations in terms of profits. Since an important objective of stock investment is to make a profit, a profit-focused evaluation is inevitable for the sake of traders. We come up with the idea of developing a simulator that implements both sell and buy timing, and performing experiments using global stock market data to examine the effectiveness of the simulator.

The findings of this research are as follows:

- I. The extent of the stock market plunge and recovery at the time of the pandemic declaration is examined using daily stock data of the six major markets. All markets show uptrend reversals in stock prices within four business days, i.e., Mar. 18 to 23, 2020.
- II. The stock price best recovered in NASDAQ, followed by Nikkei and DJIA. Stock price recoveries in European markets lag behind that of U.S. and Asian markets.
- III. It is found that the trajectory of 25-day average of difference between a stock price and the 5-day average reverses on the lowest price day. We propose an indicator to predict trend reversal named APAD, an acronym for

“Average of Price and 5-day Average Difference.” We developed stock trading simulators that use the APAD indicator to find trade opportunities and to measure the degree of profit.

- IV. The simulator is enhanced by heuristics using candlestick chart patterns. A series of enhancements improves the success rate of simulated stock trades by approximately 8.5% in a long position. The results of experiments on ten years of stock price data from nine major markets in the world show that the simulators achieve the success rate of up to 74% with trade fees in a long position, while the success rate stays within 25% in a short position because a margin for borrowing a stock squeeze profits.

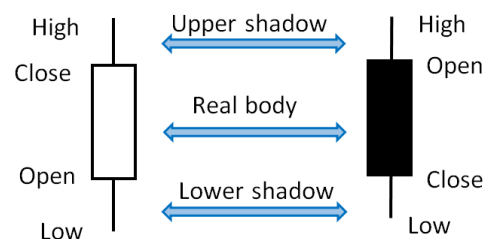
The remainder of the paper is organized as follows. Section II gives the background of the candlestick chart and introduces indicators that characterize stock price fluctuations. Section III shows the extent of stock price plunges and recoveries regarding the Covid-19 pandemic declaration. Section IV statistically examines the indicators for characterizing stock price fluctuations. Section V deals with the development of the stock trade simulator using a devised indicator named APAD, and the enhancement of the simulator using candlestick pattern heuristics. Section VI concludes the paper with our plans for future work.

## II. CANDLESTICK CHART AND PRAMETERS

This section introduces formations of a candlestick chart. The candlestick attributes to be analyzed are identified.

### A. Formation of Candlestick

As depicted in Figure 1, a daily candlestick is formed with the market's opening, high, low, and closing prices of a specific trading day [12][13]. The candlestick has a wide part, which is called *real body* representing the range between the opening and closing prices of the day of trading, as shown in Figure 1. The color of the *real body* represents whether the opening price or the closing price is higher. If the price rises, a hollow body is drawn suggesting *bullish* or buying pressure. Otherwise, a filled body is drawn suggesting *bearish* or selling pressure.



(A) Bullish candlestick (B) Bearish candlestick

Figure 1. Candlestick formation.

The thin lines above and below the body, which are named *shadows*, represent the range of prices traded in a



day. The high is marked by the top of the upper shadow and the low by the bottom of the lower shadow.

### B. Candlestick Chart and Parameters

A candlestick chart is a graph in which candlesticks are arranged in order of market dates. It is used as a tool to get information on whether the current price is higher or lower than the historical stock price movements, and what kind of price movements have been made in a certain period of time. Moving averages form a line graph by connecting the average of closing prices over a certain period of time. The line graph is useful to decide whether stock prices are in a rising or falling trend by considering the relative position between the current stock price and the moving average. As for periods of time to compute averages, each country uses its own periods. For example, the short-term average is often calculated for 5 days, the medium-term average is for 25 days, and the long-term average is for 75 days in Japan.

Figure 2 illustrates indicators including the averages that formalize a candlestick chart. In accordance with the candlestick chart notation, the following six attributes or indicators are used for analysis.

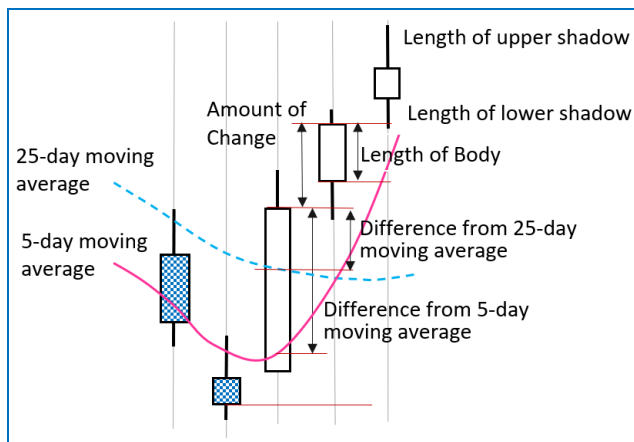


Figure 2. Candlestick chart and its parameters.

- (1) Amount of stock price change (the difference between a stock's closing price on a trading day and its closing price on the previous trading day)
- (2) Length of candlestick body
- (3) Length of upper shadow
- (4) Length of lower shadow
- (5) Difference between the stock price of a trading day and the 5-day moving average
- (6) Difference between the stock price of a trading day and the 25-day moving average

### III. STOCK PRICE PLUNGE AND RECOVERY REGARDING COVID-19 PANDEMIC

This section describes the process of data analysis to follow how this research is performed. Stock price fluctuations of the six markets are compared with the ratio of

the stock price to the lowest stock price to measure impacts of Covid-19 on each market.

### A. Data Analysis Process

Figure 3 overviews the data analysis process in this research that consists of the following operations.

- (1) Downloading daily historical data from a Web site,
- (2) Calculating the six attributes mentioned in Section II,
- (3) Extracting price data around the lowest price during Covid-19 pandemic,
- (4) Calculating the average and standard deviation of the six attributes.

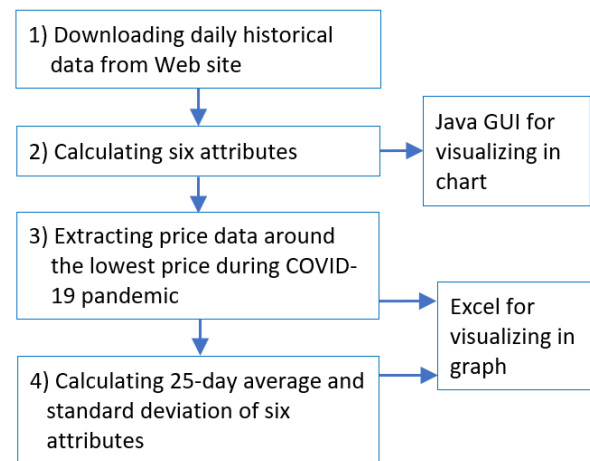


Figure 3. Overview of data analysis process.

Among the sites that provide global stock data, the Web site [4] provides daily stock data for more than 10 years in more than 40 markets. All data used in the research are downloaded manually from the site.

Because the daily stock data only consist of close, open, high, low prices, and volume of stock trading, Java programs are developed for performing operations 2) to 4), and visualizing the candlestick chart. Excel is used for visualizing data of 3) and 4) in graphs.

### B. Comparison of Stock Price Fluctuations

In order to understand overall structures of stock price fluctuations, we compare the stock price ratios to the lowest price that was recorded in Mar. 2020. Let  $CPr[n]$  be the closing price of a trading day  $n$ , and  $CPrMin$  be the lowest closing price recorded in Mar. 2020. The price ratio  $PrRatio[n]$  is calculated by the following formula;

$$PrRatio[n] = (CPr[n] - CPrMin) * 100 / CPrMin \quad (1)$$

Figure 4 is a graph of the price ratio of the six markets' stock prices to the lowest ones for 490 days, or approximately two years before and after the lowest. The lowest prices were recorded on one of the four business days from Mar. 18 to 23, 2020. Although the date of the lowest price slightly varies from market to market, stock prices continued to rise after hitting the lowest price in all markets.

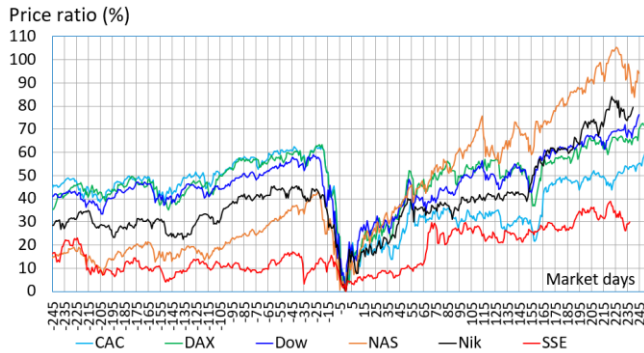


Figure 4. Ratio of stock index compared to the lowest price.

Table I summarizes the stock price profiles compared to the lowest price. The lowest prices have been recorded from Mar. 18 to Mar. 23 in the six markets under comparison. Since Mar. 22, 2020 is Sunday, they show that the stock trends reversed from downtrend to uptrend in just 4 trading days in the six markets. The degree of plunge is 63.34% in DAX (Germany), followed by 62.76% in CAC (France), and 58.95% in DJIA (US), as shown in the column “Highest price before Covid-19 pandemic.” The lowest decline was of 22.95% in SSEC (China).

TABLE I. SUMMARY OF STOCK PRICE INDICATORS COMPARED TO THE LOWEST PRICE

	Day of lowest price	Highest price before Covid-19 pandemic (%)	Highest price after Covid-19 pandemic (%)	Recovery (%)
CAC40(France)	Mar. 18	62.76	61.03	-1.73
DAX(Germany)	Mar. 18	63.34	72.59	9.25
DowJones(U.S.)	Mar. 23	58.95	76.3	17.35
NASDAQ(U.S.)	Mar. 23	43.09	105.45	62.36
Nikkei(Japan)	Mar. 19	45.49	84.06	38.57
SSEC(China)	Mar. 23	22.95	38.94	15.99

A matter worthy of note is the degree of recovery from the lowest price. NASDAQ achieves the finest recovery of 62.36% rise as the highest price ratio after the Covid-19 pandemic is 105.45%, and the highest one before the pandemic is 43.09%. Following NASDAQ, Nikkei (Japan) comes back by 38.57% rise. The slowest recovery is recorded -1.73% in CAC, followed by 9.25% in DAX. From the stock price movements of the two markets, it can be inferred that the impact of the Covid-19 pandemic in Europe is larger than the other regions.

#### IV. COMPARISON OF INDICATORS REGARDING STOCK PRICE FLUCTUATIONS

This section presents the results of comparisons of the six indicators introduced in Section II with respect to the average and standard deviation in statistics. In order to compute the meaningful standard deviation, we calculate the average and standard deviation of the past 25 days [14] including the reference trading day. This period is

commonly used for the calculation of the six attributes of a candlestick.

##### A. Amount of Stock Price Change

Figure 5 shows a graph of the 25-day moving averages and standard deviations of price changes of each stock market. As seen at the center part of Figure 5, the averages (in solid lines) of stock price changes plummet during approximately 20 days before the lowest trading day, i.e., around Feb. 20, 2020 in the five markets excluding SSEC. Stock price averages rally for approximately 25 days after the bottom, i.e., around Apr. 24, 2020.

The standard deviations (in dashed lines) over this period, also increase in the range of 3.5 to 6 times of those of the other period reflecting the high volatility of price movements.

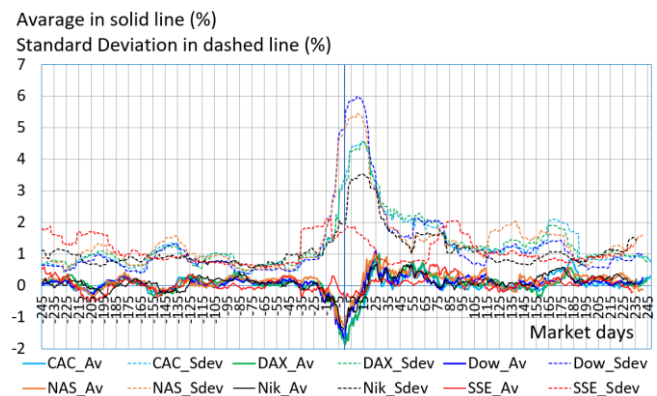


Figure 5. Average and standard deviation of price change.

Taking a closer look at the standard deviations, DJIA reaches the maximum of 6.00%, followed by NASDAQ of 5.47%, CAC of 4.56%, DAX of 4.54%, Nikkei of 3.53%, and SSEC of 2.14%, which is considered to reflect the strength of the impact on each market. The standard deviations of each market have roughly doubled after the lowest price, i.e., the right part, compared to those before the lowest price, i.e., the left part, which suggests that unstable trading has continued for roughly 180 days after the lowest price day.

##### B. Length of Candlestick Body

Figure 6 shows the 25-day averages and standard deviations of the candlestick body lengths of the six stock markets. Figure 6 shows that the averages of SSEC are positive (plus) during the plunge period (-20 to 0), which means stock prices increase on average within a market day. The 25-day average of candlestick body lengths in NASDAQ is almost zero level during the plunge period. NASDAQ has consistently positive average during the period of stock price recovery (0 to 25), which means that the stock price continued to rise during trading hours. The four other markets experience remarkable negative averages, e.g., -1.39% in CAC, -1.15% in DAX, -0.967% in Nikkei, and -0.83% in DJIA at the lowest, suggesting that colored candlesticks are prominent.

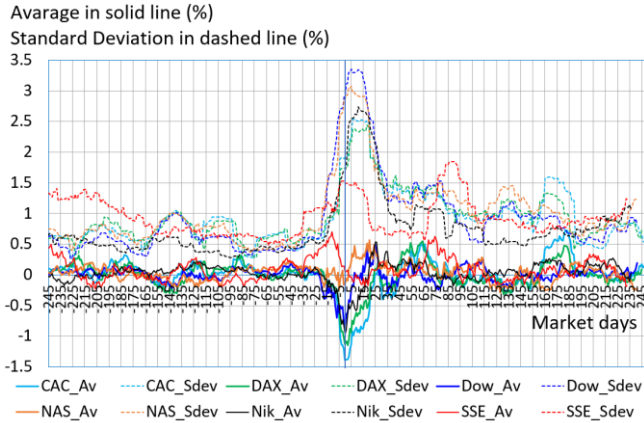


Figure 6. Average and standard deviation of candlestick body length.

The largest standard deviation of 3.36% is marked in DJIA, followed by 3.07% in NASDAQ, 2.73% in Nikkei, 2.54% in CAC, 2.44% in DAX, and 1.85% in SSEC. Trends of the average and standard deviation of SSEC are notably different from the other markets.

C. Length of Upper and Lower Shadows

Figures 7 and 8 illustrate the 25-day averages and standard deviations of the lengths of the upper and lower shadows. The two figures look similar, revealing that the lengths of upper shadows are apparently correlated with those of the lower shadows.

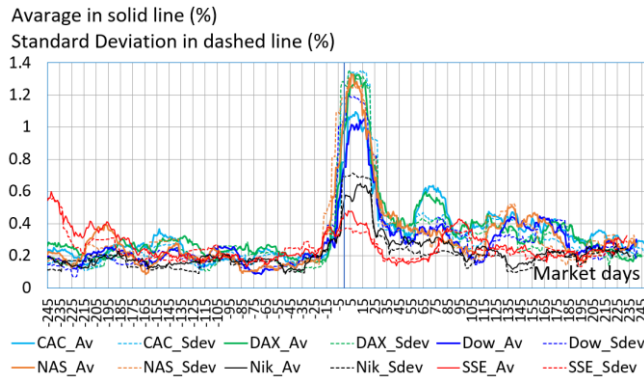


Figure 7. Average and standard deviation of upper shadow length.

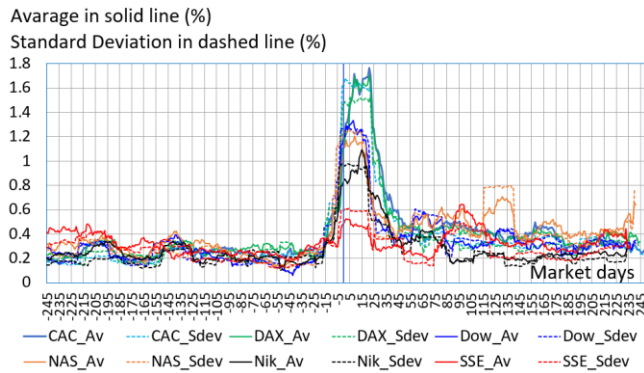


Figure 8. Average and standard deviation of lower shadow length.

In the five markets excluding SSEC, the averages and standard deviations of the upper and lower shadows increase sharply during the period from approximately 10 days before the lowest price to approximately next 25 days.

Notably in the European markets of DAX and CAC, both the averages and standard deviations have surged about six times after the lowest price compared to these before the lowest price for the upper shadows, and about eight times for the lower ones. Those surges mean that rough price movements occur during the period.

On the other hand, the Asian market is relatively stable. In SSEC, the averages and standard deviations are doubled after the lowest price than before for the upper shadows, and are tripled for the lower ones. Nikkei marks about three times higher for the upper shadows, and about five times higher for the lower ones.

Figure 9 is a scatter plot of the lengths of upper and lower shadows in the Nikkei market. It illustrates that there is a strong correlation between the lengths of the upper and lower shadows.

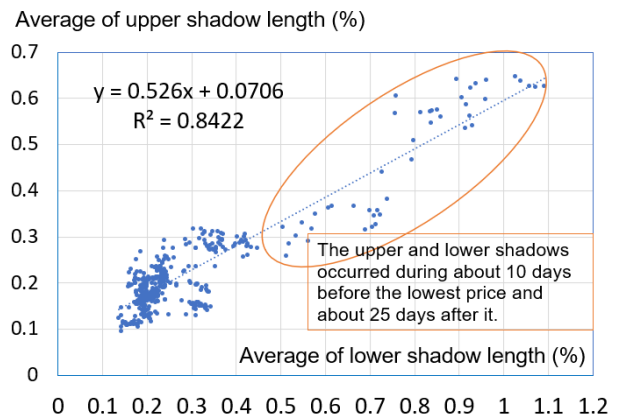


Figure 9. Scatter plot of upper and lower shadow length of Nikkei.

The points surrounded by an ellipse correspond to the shadows that occur in the period between approximately 10 days before and approximately 25 days after the lowest price. These points occupy a different portion of Figure 9 from the rest of the points. R-squared ( $R^2$ ) in statistics is 0.8422, which indicates that 84.22% of the upper shadow length can be explained by the lower shadow length, and vice versa.

D. Difference Between Stock Price and 5-day Average

Figure 10 shows the 25-day averages and standard deviations of the “difference between a stock price and 5-day moving average” for the six markets. The 25-day averages of the difference sharply decrease about 20 days before the lowest price, and sharply increase until 25 days after the bottom.

The standard deviation increased more than fourfold in the five markets except SSEC, during the period from about 20 days before the lowest price to 25 days after the lowest price. High standard deviation around 2% continues in all



markets nearly 80 days after the day recorded the lowest price.

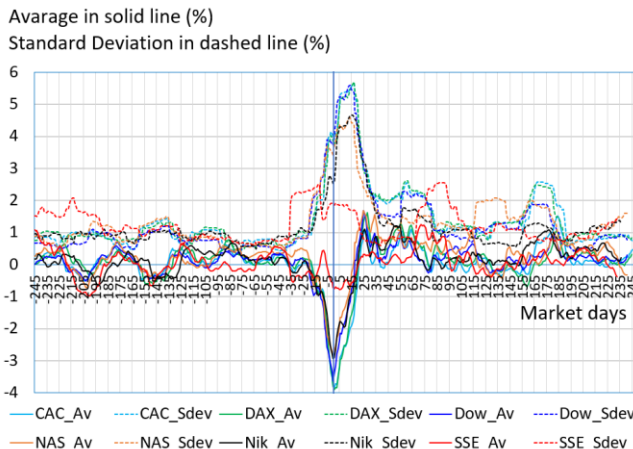


Figure 10. 25-day average and standard deviation of “difference between stock price and 5-day average.”

In the five markets excluding SSEC, the day when the 25-day average of the “difference between a stock price and 5-day moving average” reverses the trend from downward to upward roughly coincides with a day when the stock price bottoms out. As for the standard deviations, they remain high in the five markets except SSEC for the 20 days following the lowest day.

We investigate how the period for calculating the average of the “difference between a stock price and 5-day moving average” relates to the timing of the stock price reversal. Figure 11 shows the  $P$ -day average of the “difference between the 5-day average and the stock price” in NASDAQ, where  $P$  is 8, 12, 18, 25, 30, and 35. When  $P$  is 12, 25, 30, and 35, the date when this average is at its minimum coincides with the date when the stock price is at its lowest. When  $P$  is 8 and 18, the  $P$ -day average begins to rise two days before the lowest price.

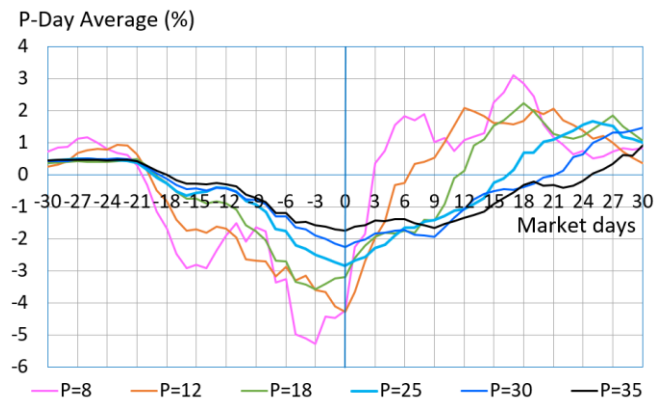


Figure 11. P-day average of “difference between stock price and 5-day average” in NASDAQ.

Figure 12 shows the results using Nikkei 225 stock data. When  $P$  is 12 and 18, the minimum value of the  $P$ -day average coincide with the day when the stock price reached

its lowest price. On the other hand, when  $P$  is 25, 30, or 35, this average reached its minimum value on the next day when the stock price reached its lowest value.

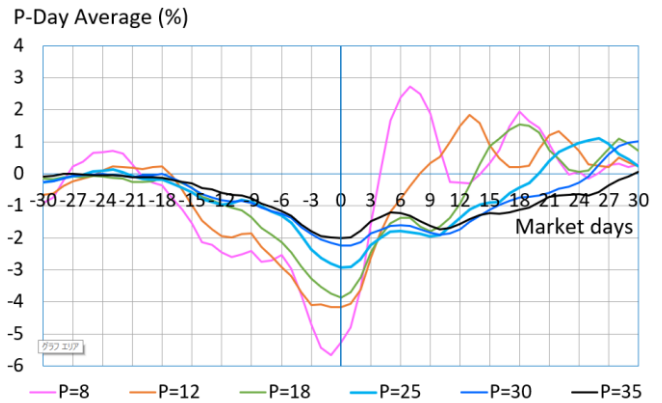


Figure 12. P-day average of “difference between stock price and 5-day average” in Nikkei.

Although there are slight differences from stock market to market, this average can be used to predict a stock price reversal with an error of less than one day if  $P$  is greater than 25. This is an important finding and detailed analyses are discussed in Section V.

#### E. Difference Between Stock Price and 25-day Average

Figure 13 shows the 25-day averages and standard deviations of the “difference between a stock price and 25-day moving average” for the six markets. This indicator shows a minimum value 4 to 14 days after the lowest stock price in the six markets, which indicates that this indicator is too slow to predict a stock trend reversal.

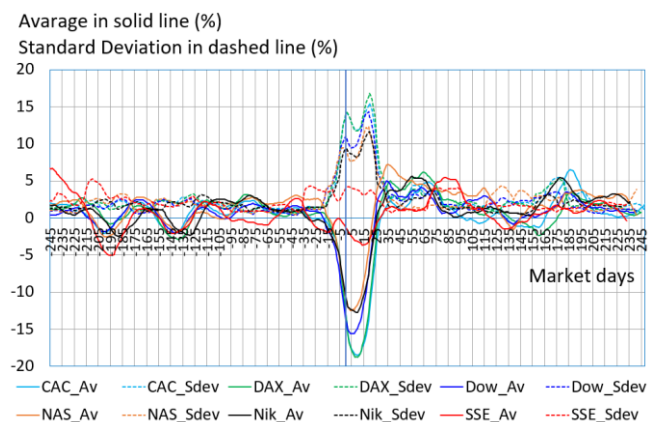


Figure 13. 25-day average and standard deviation of “difference between stock price and 25-day average.”

Again, in the five markets excluding SSEC, the averages and standard deviations fluctuate largely during a period between approximately 20 days before the day of the lowest price and the following 25 days.

The standard deviation increased more than fivefold in the five markets except SSEC. Standard deviation close to

4% continues in all markets nearly 80 days after the lowest price day.

## V. DEVELOPING STOCK TRADING SIMULATOR

In this section, we propose a new trend reversal indicator based on the analysis described in Section IV. The definition of the indicator, and the configuration of stock trading simulators using the indicator are described. Stock trading simulation are performed using nine markets around the world for ten years. The results of the simulation are analyzed in terms of profits.

### A. Trend Reversal Indicator APAD

As mentioned in Section IV-D, the 25-day average of “difference between a stock price and 5-day moving average” starts to rise just after the day when the lowest price is recorded. Based on the results of the analysis, we propose a trend reversal indicator named APAD, i.e., an acronym for “*P*-day Average of stock Price and *N*-day Average Difference.” Generally, APAD is defined using two parameters, *P* and *N*. *P* and *N* are arbitrary periods of market day.

Let  $CPr[D]$  be the closing price of a trading day *D*, and  $CPr[1]$  represent the most recent trading day as defined in (1). Let  $PAD(D, N)$  represent a ratio of the difference between the closing price and the *N*-day average of a market day *D* to the closing price as defined by the following formula:

$$PAD(D, N) = \{CPr[D] - \frac{1}{N} \sum_{k=D}^{k=D+(N-1)} CPr[k]\} * 100 / CPr[D] \quad (2)$$

$APAD(D, N, P)$  is defined as the *P*-day average of  $PAD(D, N)$  as follows:

$$APAD(D, N, P) = \frac{1}{P} \sum_{j=D}^{j=D+(P-1)} PAD(j, N) \quad (3)$$

Based on examination of Figure 10 through 12, *N* and *P* is set to five for calculating the fast APAD. *P* is set to twenty-five for the slow APAD as defined by the following formulas:

$$\text{Fast APAD: } APAD(D, 5, 5) \quad (4)$$

$$\text{Slow APAD: } APAD(D, 5, 25) \quad (5)$$

Since the APAD is a percentage price oscillator (PPO) [12] that calculates a difference relative to a closing price, the APAD can be applied to compare the world’s major stock markets.

Figure 14 illustrates the candlestick chart for 30 days before and after Mar. 18, 2020 in the DAX market. The top three-quarters of this figure are used to draw candlesticks and the rest is used to draw the APDAs.

The vertical lines in magenta color, usually centered, indicating the day of interest. Candlestick charts are displayed with 5-day and 25-day averages, in accordance with Japanese stock market conventions. APADs are shown

in the lower one-quarter of Figure 14. The magenta line shows the fast APAD ( $D, 5, 5$ ), and the blue line shows the slow APAD ( $D, 5, 25$ ).

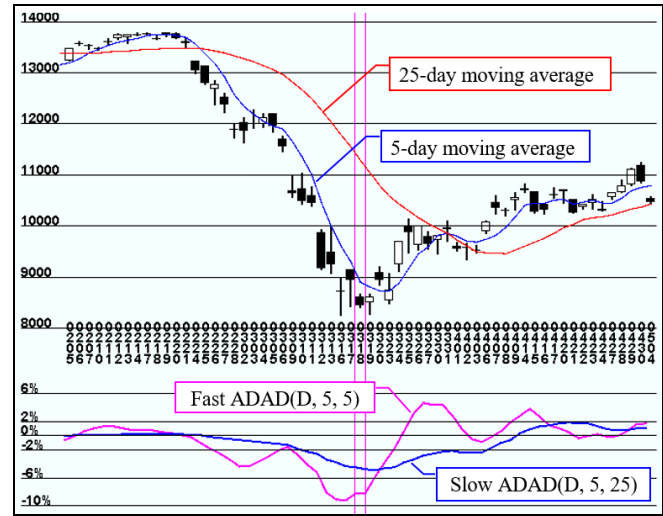


Figure 14. Candlestick chart for 30 days before and after Mar. 18, 2020 in DAX market.

The proposed two APAD lines go across up on Mar. 20, i.e., the two days after the lowest price is recorded. The two lines do not crossover on Mar. 5, which means to predict that stock prices will continue to fall. In other words, a short-term recovery from Feb. 28 to Mar. 5 is a “*dead cat bounce*,” i.e., a temporary recovery of stock prices during a prolonged decline period.

The two lines also do not crossover during the decline from Mar. 27 to April 3, which suggests that Apr. 3 is a “*buying on the dips*” type of opportunity since the slopes of the fast and slow APAD lines increase on Apr. 3. The magenta line crosses down through the blue line on Apr. 17. However, the slope of the slow APAD is positive on Apr. 17, which indicates that the uptrend of the stock price is continuing in the medium term of 25 days. During periods when the slope of the slow APAD is positive, no action should be taken for a down trend.

Similar results have been obtained in the other markets. Regarding recovery from the plunge caused by Covid-19 pandemic, it is confirmed that the fast and slow APAD lines forecast short-term trends properly.

### B. System Configuration for Stock Trading Simulator

Figure 15 shows the configuration of stock trading simulators developed in this study. The CSV files storing the stock prices are manually downloaded from the global Web site [4]. *Read\_CSV\_Data* is a Java program that converts the CSV file into an *in-memory data store*. The data store provides all the data needed for analyses, including the price change, the body length of a candlestick, the moving average of stock prices, and the APADs, etc.

A Java program named *T2L* (Trade Tracker for a Long position) is a simulator that tracks buying opportunities using APADs, and virtually performs stock trade in a long

position. Meanwhile, a program named *T2S* (Trade Tracker for a Short position) is a simulator for a trade in a short position. *T2L* and *T2S* are independent programs, each generating a log file for further analysis using a spread sheet tool including Excel. A program named *GUICC* (GUI for Candlestick Chart) is a tool for drawing candlesticks, the moving averages and the APADs, etc. and is used to draw candlestick charts, e.g., Figure 14.

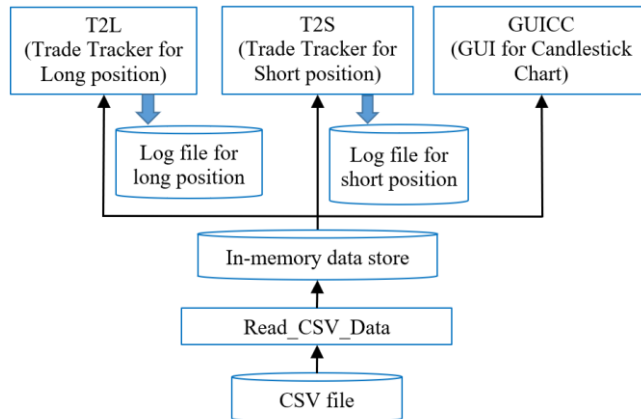


Figure 15. Configuration of developed stock trading simulator.

In the downloaded CSV file, the most recent stock prices are stored at the beginning of the file. *Read\_CSV\_Data* reads data from the beginning of the CSV file and stores the data in an internal array while maintaining the order. For example, the opening price stored in the CSV file is stored in an array named *Open[]* with the *double* data type. *Open[1]* memorizes the opening price of the most recent trading day. Since the NASDAQ market trades approximately 254 days per year, the opening price of the NASDAQ market ten years ago is stored in the vicinity of *Open[2540]*. An array named *Mday[]* with the *int* data type is used for storing market dates in integer, e.g., 20220430 to indicate Apr. 30, 2022.

Figure 16 shows the basic structure of the developed simulator. The stock market to be analyzed corresponds to the name of the CSV file, e.g., *NASDAQ.csv* for NASDAQ composite index. *Read\_CSV\_Data* method reads the data from the specified CSV file. A method named *FindIndex* is a method that takes a market date as an argument and converts it to an index of an internal array.

```

01 public class Stock_Trading_Simulator {
02     public static void main(String[] args) {
03         String filename= "c:/temp/NASDAQ.csv";
04         Read_CSV_Data( filename );
05         int IndexFrom= FindIndex( 20120501 );
06         int IndexTo= FindIndex( 20220430 );
07         T2L( IndexFrom, IndexTo);
08         T2S( IndexFrom, IndexTo);
09     }
10 }
  
```

Figure 16. Basic structure of developed simulator.

In this study, the start and end dates of the simulation are May 1, 2012 and Apr. 30, 2022, respectively. The variables *IndexFrom* and *IndexTo* store the indices of the internal array corresponding to the dates. These two variables are used as arguments of the *T2L* and *T2S* methods that perform simulations.

### C. Tracking Buy Opportunities using APAD

#### 1) Overall Algorithm for Buying Simulation

Figure 17 shows the overall algorithm for the buying simulation. The *T2L* method has two parameters, i.e., *IndexFrom* and *IndexTo*. The parameter *IndexFrom* specifies the start date of a simulation. Strictly, the value of *IndexFrom* is the index value of arrays corresponding to the start date. Similarly, the parameter *IndexTo* specifies the end date of a simulation.

```

01 T2L ( int IndexFrom, int IndexTo ) {
02     int flg= -1; // to indicate state of finding buy-opportunity
03     for ( j= IndexFrom; j<=IndexTo; j-- ) {
04         /* Buy signal is detected */
05         if ( flg < 0 & Buy_Signal_B ) {
06             traded_p = ((Open[j] + Close[j]) / 2);
07             holding_day_cnt= 0;
08             flg= 1; // to indicate buy-stock state
09         /* Sell signal is detected in long position */
10         } else if ( flg > 0 & Sale_Signal_B ) {
11             diff = ((Open[j] + Close[j]) / 2) - traded_p - buying_cost;
12             profit= profit + diff; // Total profit
13             flg= -1; // to indicate state of finding buy-opportunity
14         }
15         holding_day_cnt++;
16     }
17 }
  
```

Figure 17. Overall simulation algorithm in long position.

When  $flg < 0$  & *Buy\_Signal\_B* is *True*, the process of trading in a long position is started:

- Set the variable *traded\_p*, which indicates the trading price of the stock, to the average of the opening and closing prices.
- Set the variable *holding\_day\_cnt*, which presents the number of days to hold a stock, to 0.
- Set the variable *flg* to 1 for indicating a trade in a long position.

*Buy\_Signal\_B* is a conditional expression that signals the start of a buy, and is defined using the APADs and candlestick patterns. The details are described in the following subsections.

When  $flg > 0$  & *Sale\_Signal\_B* is *True*, the trade is closed.

- Set the profit of the trade to the variable *diff* using the expression:  $diff = ((Open[j] + Close[j]) / 2) - traded_p - buying\_cost$ .
- Set the cumulative profit to the variable *profit* using the substitution expression  $profit = profit + diff$ .
- Set the variable *flg* to -1 to indicate the state for tracking buy opportunities.

*Sale\_Signal\_B* are conditional expressions that suggests the end of a buying stock. The details are discussed in the following subsection. *Buying\_cost* is a fee of cash trade, and is defined in the following Subsection 4).

### 2) Criteria for buying stock in long position

As observed in Section V-A, the good timing to buy stock is occurred when the slope of the fast APAD is positive or when the fast APAD crosses the slow APAD above it.

Let  $APAD\_F[j]$  be the fast APAD of a trading day  $j$ , and  $APAD\_S[j]$  be the slow APAD. The larger the value of variable  $j$ , the more it represents the past, so  $j+1$  indicates stock data of one market day earlier.

The *Buy\_Signal\_B* is defined as follows:

$$Buy\_Signal\_B = (b1 \mid (b2 \ \& \ b3)) \ \& \ b4; \quad (6)$$

$$b1 = APAD\_F[j] - APAD\_F[j+1] > 0; \quad (7)$$

$$b2 = APAD\_F[j+1] - APAD\_S[j+1] < 0; \quad (8)$$

$$b3 = APAD\_F[j] - APAD\_S[j] > 0; \quad (9)$$

The Boolean variable  $b1$  shows that the slope of the fast APAD is positive.  $b2$  and  $b3$  mean that the fast APAD goes across up the slow APAD.

To improve the success rate in stock price simulation, it is necessary to reflect the characteristics of stock price fluctuations on the day of stock trading. Let  $Pchg[j]$  be the ratio of price increase, and  $Pbody[j]$  be the ratio of candlestick length to the closing price in percentages, respectively. The formula  $b4$  including parameter values of 0.15 and 0.40 is determined by a parameter tuning through experiments using the NASDAQ stock data.

$$b4 = Pchg[j] > 0.15 \ \& \ Pbody[j] > 0.40; \quad (10)$$

The Boolean variable  $b4$  means that the stock price rises on the day. The Boolean expression (6) must be *True* to begin the simulation in a long position.

### 3) Criteria for selling stock in long position

The accuracy of predicting a stock decline is generally less accurate than the accuracy of predicting its price increase [15]. Because of the difficulty of the prediction, we must rely on heuristics. The Boolean condition *Sale\_Signal\_B* in Figure 17 consists of several Boolean expressions that are derived by experiments using APAD and candlestick attributes. Specifically, *Sale\_Signal\_B* is defined as follows:

$$Sale\_Signal\_B = b11 \ \& \ ((b12 \ \& \ b13 \ \& \ b14 \ \& \ b15) \ \mid \ b16); \quad (11)$$

Let  $Avg5[j]$  be the 5-day average of stock price of a trading day  $j$ . The Boolean variable  $b11$  is defined by the following expression, which indicates that the difference between the stock price and the 5-day moving average is less than 2%, which means that the stock price is close to the 5-day average.

$$b11 = (Close[j] - Avg5[j]) * 100 / Close[j] < 2.0; \quad (12)$$

The variables  $b12$ ,  $b13$ , and  $b14$  are conditions regarding APAD. The variable  $b12$  is defined by the following expression, which indicates that the slope of the fast APAD for the previous day of the stock trade is positive.

$$b12 = (APAD\_F[j+1] - APAD\_F[j+2]) > 0.0; \quad (13)$$

The variable  $b13$  is defined by the following expression, which indicates that the slope of the fast APAD on the trading day is negative.

$$b13 = (APAD\_F[j] - APAD\_F[j+1]) < 0.0; \quad (14)$$

The variable  $b14$  specifies the slope of the slow APAD is negative, which indicates that stock prices are falling in the medium term of 25 days.

$$b14 = (APAD\_S[j] - APAD\_S[j+1]) < 0.0; \quad (15)$$

The variables  $b15$  and  $b16$  are conditions related to stock price fluctuations on the day of stock trading. The variable  $b15$  is defined by the following formula, which specifies that the ratio of candlestick length to the closing price on the day is less than 0.3%, suggesting a bearish candlestick. The value of 0.3% is determined through experiments.

$$b15 = Pbody[j] < 0.3; \quad (16)$$

The variable  $b16$  is calculated by the following formula, which indicates that if the stock price on the day is 0.2% lower than the traded price, the stock should be sell back.

$$b16 = ((Open[j] + Close[j]) / 2 - traded\_p) * 100 / traded\_p < -0.2; \quad (17)$$

$B16$  represents the degree of tolerance for loss. The tolerance values closer to zero gives better results in terms of profit, which suggests that traders in practice should be rigorous in losses.

### 4) Calculating fee of cash trade

Purchasing shares comes with trading fees. The stock trading with the lowest fee is offered by an online trading in a long position. A trading in a long position is initiated with the purchase of a stock in anticipation that its value will rise over time. The trading fees vary across brokers and an amount of a stock trade. Table II summarizes trading fees of some popular Japanese online brokers. This study assumes the trading fee is 0.1% of the traded stock price due to the fact that the fee is applicable to the trade with an amount between 1M JPY ( $\cong$  US\$7,500) and 3M JPY ( $\cong$  US\$22,500).

TABLE II. TYPICAL TRADING FEES OF JAPANESE ONLINE BROKERS

	~ ¥1M	¥1M ~ ¥3M	¥3M ~ ¥5M
SBI	0	0	¥2,576
Rakuten	0	¥3,300	¥5,500
Manex	¥550	¥2,750	¥5,500



The following formula is used to calculate the fee in this study.

$$\text{buying\_cost} = \text{traded\_p} * 0.001; \tag{18}$$

The T2L method performs simulations with and without fees in a long position.

5) Experimental results using NASDAQ historical data

Experiments using NASDAQ daily data are performed to confirm the profit characteristics of trading with the APADs. The period of the experiment is ten years, strictly between May. 1, 2012 and Apr. 30, 2022. Table III shows a sample of logging data in a long position. The log data include pairs of stock buy and sale dates, profits generated on each stock trade. Cumulative profit can be calculated by summing profits and can be used for drawing the cumulative profit graph shown later. Pairs of holding days of a stock and profits are used to draw scatter plots.

TABLE III. SAMPLE OF STOCK TRADE LOG DATA.

No.	Buy Date	Sale Date	Buy Price	Holding days	Profit	Net profit	Fee
...	...	...	...	...	...	...	...
225	20200302	20200305	8809.7	3	-45.3	-54.1	8.8
226	20200313	20200316	7742.6	1	-594	-601.7	7.7
227	20200317	20200318	7119.7	1	-173.7	-180.8	7.1
228	20200319	20200320	7073.6	1	-9.8	-16.8	7.1
229	20200324	20200501	7307	27	1336.1	1328.8	7.3
230	20200508	20200512	9089.1	2	24.8	15.7	9.1
231	20200518	20200526	9206	5	214.7	205.5	9.2
232	20200529	20200611	9436.1	9	205.9	196.4	9.4
233	20200622	20200623	10001	1	130.1	120.1	10
234	20200630	20200709	9967	6	588.7	578.7	10
...	...	...	...	...	...	...	...

Table IV shows the total profits for the combination of parameters P and N of APADs using NSADAQ daily stock price data. The largest profit of US\$20,940.8 is recorded when N is 3 and P is greater than or equal to 12. The second largest profit of US\$20,641.4 is recorded when N is 5 and P is greater than or equal to 15. It is worth noting that profits are greater when N is an odd number than when it is an even number.

TABLE IV. SIMULATED PROFITS REGARDING APAD PARAMETERS P AND N

P \ N	2	3	4	5	6
3	18,704.2	N/A	N/A	N/A	N/A
4	19,373.9	19,367.9	N/A	N/A	N/A
5	19,435.9	21,105.9	19,268.4	N/A	N/A
6	19,533.5	20,799.2	19,440.3	19,834.5	N/A
7	19,704.4	20,663.6	19,945.3	20,347.9	14,676.8
8	19,588.9	20,924.4	19,891.5	20,599.3	15,687.1
9	19,586.5	20,924.4	19,810.2	20,616.5	15,468.8
10	19,541.6	20,916.8	19,779.0	20,611.4	15,340.1
12	19,588.9	20,940.8	19,925.7	20,604.2	15,497.1
15	19,588.9	20,940.8	19,925.7	20,641.4	15,518.6
20	19,661.3	20,940.8	19,925.7	20,641.4	15,518.6
25	19,588.9	20,940.8	19,925.7	20,641.4	15,518.6
30	19,588.9	20,940.8	19,925.7	20,641.4	15,518.6

Table V shows the number of trades for the combination of parameters P and N. In general, the number of trades decreases as the parameter N increases, which means that values of APAD (D, N, P) are less sensitive to individual stock price fluctuations. For the parameter P being greater than or equal to 15, the number of trades is 355 when parameter N is 3, and 276 when parameter N is 5. The number of trades is 22.2% lower for parameter N of 5 than for parameter N of 3. This indicates that a value of 5 for the parameter N is more suitable than a value of 3 in terms of the number of trades, although slightly less profitable. In the following experiments, the value of the parameter N is set to 5 and the value of the parameter P is set to 25, which is consistent with the usage of moving average practice in Japan.

TABLE V. TRADE COUNTS REGARDING APAD PARAMETERS P AND N

P \ N	2	3	4	5	6
3	433	N/A	N/A	N/A	N/A
4	416	377	N/A	N/A	N/A
5	415	358	307	N/A	N/A
6	414	358	297	290	N/A
7	414	358	294	282	260
8	413	356	293	279	250
9	414	356	294	278	247
10	414	356	293	277	246
12	413	355	292	277	244
15	413	355	292	276	243
20	413	355	292	276	243
25	413	355	292	276	243
30	413	355	292	276	243

Figure 18 depicts the cumulative profits and the cumulative fees. The simulator generates profits in all periods since cumulated profits exceed cumulated fees. As the graph shows, the simulated profit can be divided into three periods according to the degree of profit. The first period is from the beginning of the simulation to the day of the lowest price, indicated by the arrow  $\uparrow$ , i.e., approximately Mar. 23, 2020. The simulator makes an average profit of about \$43.87 per trade including fees.

The second period is from the date of the lowest price to the date about a year later, indicated by the arrow  $\downarrow$ . The simulator earns an average profit of \$257 per trade including fees. The last period is from the last day of the second period to April 30, 2022, i.e., the last day of the simulation. The simulator generates an average profit of about \$157 per trade including fees.

The average profit in the second period is noticeably larger than those in the other periods, which suggests an impact of monetary policy. Trades in the vicinity marked by the arrow  $\diamond$  result in a loss of approximately US\$900, which motivates us to improve the simulator.

Linear regression formulas with the R-squared are added near the line graphs. R<sup>2</sup> is a statistical measure of how closely the data are fitted to the regression line. All R<sup>2</sup>s are greater than 0.9351, which indicates that the developed simulator makes profits in proportion to the number of trades throughout the simulation period.



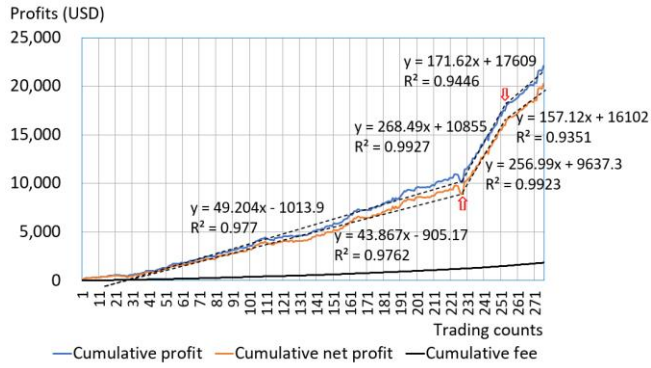


Figure 18. Cumulative profit and cumulative fee in long position.

Figure 19 is a scatter plot of all trades in a long position with the fee using the NASDAQ historical stock data. The lower left point in Figure 19 corresponds to the trade with the largest loss, meanwhile the top right dot shows the most profitable trade.

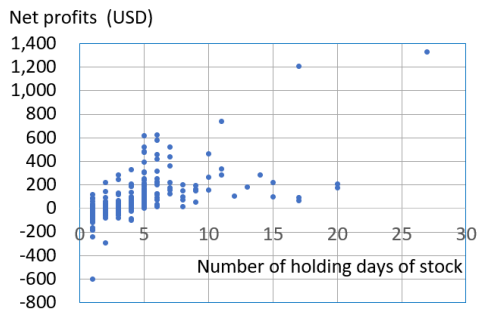


Figure 19. Scatter plot of trades in long position with fees.

Figure 20 is a candlestick chart that coincidentally contains the most profitable trade and the most lossy one. The most lossy trade begins on Mar. 13, 2020 indicated by vertical lines in orange color. The most profitable one starts on Mar. 24, 2020 indicated by vertical lines in magenta color.



Figure 20. Candlestick chart including profitable and lossy trades.

The trade on Mar. 13, 2020 is put into action because (7) and (10) are *True*. Since the stock price drops significantly the next day, (17) is set to *True*, and the simulator swiftly buys back the stock, however, resulting in the worst loss of US\$601.60.

The trade on Mar. 24, 2020 is performed because conditions (7), (8), and (9) are *True*. After that, the simulator continues to hold on the stock as the price keep exceeding the buying price. Since the fast APAD falls below the slow APAD on May 1, 2020, the Boolean conditions (12), (13), (14), (15), and (16) are set to *True*. The simulator sells back the stock, and bringing the best profit of US\$1,328.80. Meanwhile, the simulator refrains from trading on Apr. 22, 2020 because (15) is *False*, i.e., the slow APAD is positive.

#### D. Enhancing decision of buying opportunities

##### 1) Enhancement to avoid lossy trade

Stock prices sometimes move in anomalous ways for a day or so, often causing a loss-making trade. Candlestick-based trade decision processes are augmented to help the simulator avoid trades with large losses in a practical way.

Figure 21 shows an enhanced algorithm for simulating in a long position. The condition *CandlePattern\_Avoid\_Buy* is added after the condition of *Buy\_Signal\_B*. *CandlePattern\_Avoid\_Buy* consists of several candlestick patterns that signal for avoiding a buy of a stock. All patterns must be *False* for a trade in a long position to be started. Typical patterns of *CandlePattern\_Avoid\_Buy* are described below.

```

01 T2L_CP ( int IndexFrom, int IndexTo ) {
02     int flag= -1;    // to indicate state of finding buy-opportunity
03     for ( j= IndexFrom; j >= IndexTo; j-- ) {
04         /* Buy signal is detected */
05         if ( flag < 0 & Buy_Signal_B ) {
06             if ( ! CandlePattern_Avoid_Buy ) {
07                 traded_p = ((Open[j] + Close[j]) / 2);
08                 holding_day_cnt= 0;
09                 flag= 1;    // to indicate buy-stock state
10             }
11             /* Sell signal is detected in long position */
12         } else if ( flag > 0 & Sale_Signal_B ) {
13             diff = ((Open[j] + Close[j]) / 2) - traded_p - buying_cost;
14             profit= profit + diff;    // Total profit
15             flag= -1;    // to indicate state of finding buy-opportunity
16         }
17         holding_day_cnt++;
18     }
19 }
    
```

Figure 21. Overall simulation algorithm enhanced by candlestick pattern in long position.

As observed in Figure 20, the slope of the fast APAD becomes positive on Mar. 13, 2020. However, the stock purchase on this day results in unsuccessful. According to the candlestick charting, the price fluctuation on this day corresponds to the *falling window candlestick pattern*, which predicts stock prices continue to fall [13]. Let  $Pgap[j]$  be the ratio of the gap to the closing price of the trading day in percentage. This pattern is implemented by the disjunction of the two Boolean conditions below:

$$cb1 = Pbody[j+1] < 0.3 \ \& \ Pgap[j] > 0.58; \quad (19)$$

$$cb2 = Pgap[j] > 2.0; \quad (20)$$

The Boolean variable *cb1* is calculated by the previous day's candle being less than 0.3% suggesting downtrend, and the gap on the stock trading days being more than 0.58%. This is a weak upward stock price movement accompanied by a gap in a downtrend, meaning that the stock price movements are difficult to predict. The variable *cb2* is set to *True* when a gap is greater than 2.0%, which is a rare large gap, and shows signs of avoiding a stock trade. The values of 0.58, 0.3 and 2.0 are experimentally determined through trial and error using the NASDAQ stock data.

Other pattern that should be avoided for buying a stock occurred on Mar. 17, 2020 as shown in Figure 20, indicated by vertical lines in cyan color. The opening and closing prices of a stock are nearly within the range of the prior day's opening and closing prices. This pattern is named the *bullish piercing pattern*, and is considered to be a weak reversal signal [13]. Since the slope of the fast APAD turns slightly positive on Mar. 17, 2020, it is necessary to implement a process that takes into account the *bullish piercing pattern* in order to avoid the trade on the day. The variable *cb3* checks the pattern. Parameter values are heuristically determined thorough experiments using NASDAQ daily data. *Abs* means a function that calculates the absolute value of an argument.

$$cb3 = Abs((Open[j+1] - Close[j]) * 100 / Close[j]) < 0.2 \ \& \ Pbody[j+1] < -3.0 \ \& \ Pbody[j] > 3.0 \quad (21)$$

Figure 22 shows a gap-related pattern that is scarcely mentioned in the literature [12][13]. The pattern observed on May 3, 2019 in NASDAQ.



Figure 22. Sample of gap-related pattern.

The pattern consists of a three-days' bearish candlesticks followed by a bullish candlestick with a significant gap. This pattern is spotted by the following formula:

$$cb4 = Pbody[j+3] < 0 \ \& \ Pbody[j+2] < 0 \ \& \ Pbody[j+1] < 0 \ \& \ Pbody[j] > 0.75 \ \& \ Pgap[j] > 0.5 \quad (22)$$

The values of 0.75 and 0.5 are determined through experiments. The slope of the fast APAD turns positive on the day. So, without a decision using the candlestick pattern, the simulator buys stocks and incurs a loss of US\$84.00 including trading fees. Several other candlestick patterns to avoid buying stocks are incorporated into the developed simulator.

2) *Experimental results using NASDAQ historical data*

Table VI compares the main characteristics of the simulators before and after the improvement. First of all, the number of trades decreases from 276 to 223.

TABLE VI. SUMMARY OF EXPERIMENTAL RESULT IN LONG POSITION

	APAD only		APAD with enhancement	
	Without fee	With fee	Without fee	With fee
Total no. of trades	276	276	223	223
No. of profitable trades	188	182	171	166
No. of lost trades	88	94	52	57
Success rate	68.12%	65.94%	76.68%	74.44%
Amount of profit (USD)	27,783.88	26,475.82	25,432.36	24,225.54
Amount of loss (USD)	-5,180.87	-5,834.40	-1,344.64	-1,701.45
Net profit (USD)	22,603.02	20,641.42	24,087.72	22,524.10
Most profitable trade (USD)	1,336.10	1,328.80	1,515.70	1,502.90
Most lossy trade (USD)	-594.00	-601.70	-206.50	-221.30

The success rate of trading in the study is defined by the following formula:

$$Success\ rate = \frac{\text{the number of profitable trades}}{\text{the total number of trades}} \quad (23)$$

The number of profitable trades indicates the number of trades with positive profit. The total number of trades is 276. Success rate of trades is 68.12% (= 188 / 276) without the trading fee, and is 65.94% (= 182 / 276) with the fee.

After the enhancement using candlesticks, the success rate increases from 65.94% to 74.44% (= 166 / 223) with the fee. Net profit increases from US\$20,641.42 to US\$22,524.10. It is noteworthy that the maximum loss is significantly reduced from US\$-601.70 to US\$-221.30 with the trade fee after the enhancement.

Figure 23 shows the cumulative profit and the cumulative trade fee. Comparison with Figure 18 shows that there is no noticeable drop in the graphs. All R<sup>2</sup>s are greater than 0.9523, maintaining significant correlation between the cumulative profit and the number of trades.

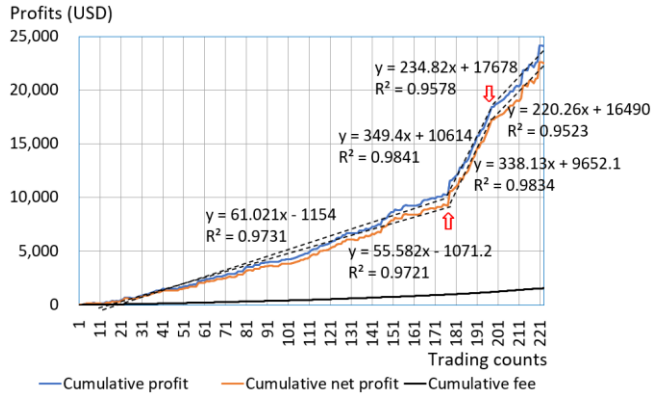


Figure 23. Cumulative profit and cumulative fee after enhancement.

Figure 24 shows a scatter plot of trades with the fee after enhancement. In comparison with Figure 19, it can be seen that the number of trades that resulted in a loss are decreased significantly, while the number of the profitable trades is roughly the same before and after the enhancement.

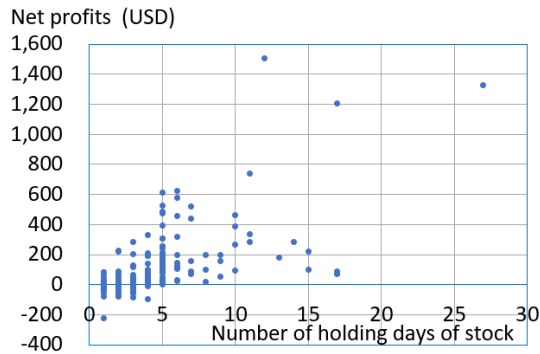


Figure 24. Scatter plot of trades with fees after enhancement using candlestick pattern.

Figure 25 shows a histogram of the number of holding days of a stock and the number of occurrences.

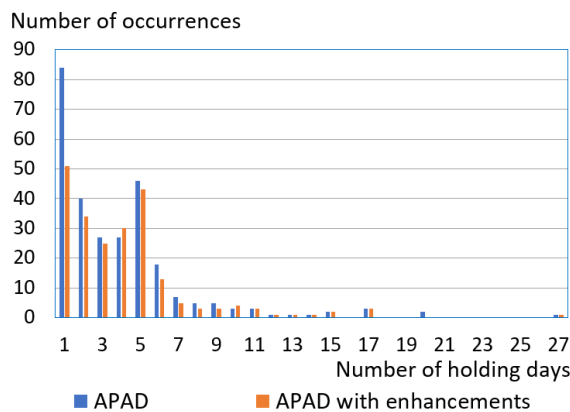


Figure 25. Histogram of the number of holding days in long position.

A trade within two days' stock holding period often results in a loss, because the stock price moves contrary to

trader's expectations. Figure 25 indicates that the proposed enhancement using candlestick patterns has the effect of noticeably reducing the number of stock trades with one-holding day, i.e., from 84 to 51.

### E. Tracking Sell Opportunities using APAD

#### 1) Overall Algorithm for Selling Simulation

Figure 26. shows the overall algorithm for the selling simulation, or simulation in a short position. The T2S method has two parameters, i.e., *IndexFrom* and *IndexTo*. Strictly, the value of *IndexFrom* is the index value of arrays corresponding to the simulation start date. Similarly, the value of *IndexTo* specifies the end date of the simulation.

```

01 T2S ( int IndexFrom, int IndexTo ) {
02   int flg= 1; // to indicate state of finding short-sell-opportunity
03   for ( j= IndexFrom; j>=IndexTo; j-- ) {
04     /* Short-sell signal is detected */
05     if ( flg > 0 & Sale_Signal_S ) {
06       traded_p = ((Open[j] + Close[j]) / 2);
07       holding_day_cnt= 0;
08       flg= -1; // to indicate short-sell-stock state
09       /* Buy signal is detected in short position */
10     } else if ( flg < 0 & Buy_Signal_S ) {
11       diff = traded_p - ((Open[j] + Close[j]) / 2) - selling_cost;
12       profit= profit + diff; // Total profit
13       flg= 1; // to indicate state of finding short-sell-opportunity
14     }
15     holding_day_cnt++;
16   }
17 }

```

Figure 26. Overall simulation algorithm in short position.

When the Boolean condition  $flg > 0 \ \& \ Sale\_Signal\_S$  is *True*, the process of trading in a short position is started:

- Set the variable *traded\_p*, which indicates the trading price of the stock, to the average of the opening and closing prices.
- Set the variable *holding\_day\_cnt*, which represents the number of days to hold the traded stock, to 0.
- Set the variable *flg* to -1 to indicate the trade in a short position.

When  $flg < 0 \ \& \ Buy\_Signal\_S$  is *True*, the trade is closed.

- Set the profit of the trade to the variable *diff* using the expression:  $diff = traded\_p - ((Open[j] + Close[j]) / 2) - selling\_cost$ .
- Set the cumulative profit to the variable *profit* using the substitution expression  $profit = profit + diff$ .
- Set the variable *flg* to 1 to indicate the state for tracking sell opportunities.

*Sale\_Signal\_S* and *Buy\_Signal\_S* are Boolean expressions that signals the start, and the end of the trade in a short position, respectively. The details are discussed in the following sections.

#### 2) Criteria for selling stock in short position

In the following formulas, parameter values are experimentally determined using the NASDAQ stock data.

Let  $CP\_5avg[j]$  be the difference between the closing price and the 5-day average of a trading day  $j$  in percentage.  $CP\_25avg[j]$  is for the 25-day average that is defined in analogy with  $CP\_5avg[j]$ . The  $Sale\_Signal\_S$  is defined as follows:

$$Sale\_Signal\_S = s1 \& s2 \& s3 \& s4 \& s5 \& (s6 / s7) \quad (24)$$

$$s1 = (Pbody[j] + Pchg[j] + CP\_5avg[j]) < -1.2; \quad (25)$$

$$s2 = (CP\_5avg[j] + CP\_25avg[j]) > -9.0; \quad (26)$$

$$s3 = CP\_5avg[j] < 2.0; \quad (27)$$

$$s4 = Pbody[j] < 0.6; \quad (28)$$

$$s5 = APAD\_F[j] > 0.2 \& APAD\_S[j] < 1.0; \quad (29)$$

$$s6 = (APAD\_F[j] - APAD\_F[j+1]) < 0; \quad (30)$$

$$s7 = (APAD\_S[j] - APAD\_S[j+1]) < 0; \quad (31)$$

The Boolean variable  $s1$  indicates the price notably falls on the trading day. The variable  $s2$  means that the stock price is within a certain range from the moving averages, i.e., the stock price is not in the bottom range.  $s3$  is *True* if the stock price is less than 2% of the 5-day average.  $s4$  represents that the ratio of the length of the body of a candlestick to the stock price is less than 0.6%, i.e., a slight price increase at most in a trading day.

The Boolean variables  $s5$ ,  $s6$ , and  $s7$  are those concerning the APADs.  $s5$  is set to *True* if the fast APAD is greater than 0.2 and if the slow APAD is less than 1.0, which suggests a weak trend of price increase. The variables  $s6$  and  $s7$  are *True* if the slopes of the fast APAD and the slow APAD are negative, which indicates a downtrend.

### 3) Criteria for buying stock in short position

The timing to buy the traded stock back in a short position, i.e.,  $Buy\_Signal\_S$  in Figure 26, is defined as follows:

$$Buy\_Signal\_S = (s11 \mid s12 \mid s13);$$

$$s11 = \frac{(trade\_p - (Open[j] + Close[j]) / 2) * 100}{trade\_p} < -0.5; \quad (32)$$

$$s12 = Pchg[j+1] > 0.0 \& Pchg[j] > 0.0 \& Pbody[j] > 0.4; \quad (33)$$

$$s13 = Pchg[j] > 1.2 \& Pbody[j] > 0.8; \quad (34)$$

The Boolean variable  $s11$  is set to *True* if the current stock price is 0.5% higher than the traded value. The condition is essential to keep losses small.  $s12$  is set to *True* if the stock price rises for two consecutive days and the length of the candlestick on the trading day is greater than 0.4%.  $s13$  is set to *True* if the stock price is greater than 1.2% and the candlestick length is greater than 0.8%. The

variable  $s12$  and  $s13$  are used to check whether the stock price for the day increase notably.

### 4) Calculating margin interest in short position

In a short selling trade, a trader must borrow a stock from an investment firm with a payment of margin interest. The leverage of trades is set to one in order to maintain the same conditions regarding the amount of the trades in long and short positions. Margin interest slightly vary across brokers. Typically, it consists of an annual interest rate and a fixed one. In this study, the margin interest is calculated using the following formula:

$$(Stock\ price) * (2.80\% / 365 * M + 1.15\%) \quad (35)$$

$M$  is the total number of days for borrowing stocks. 1.15% indicates the fixed rate.

### 5) Experimental results using NASDAQ historical data

The line graphs in Figure 27 present the cumulative profit, net profit, and the margin interest calculated by (35) in a short position. The graph of profit in this figure can be approximated by three segmented lines. The first segment is the period from the beginning of the simulation to the 78th trade marked by X, which corresponds to Sep. 17, 2018. The second segment ranges from the 79th trade to the 114th trade marked by the arrow  $\Phi$ , which corresponds to Mar. 18, 2021, approximately one year after the date of the lowest price. The day of the lowest price marked by the arrow  $\Phi$  is located roughly in the middle of the second segment. The third segment is the period from the 115th trade to the last trade, namely Apr. 25, 2022.

The minimum value of R-squared is 0.5534 that shows the degree of approximation is worse than a long position. This suggests that trades in a short position is less stable than those in a long position. The cumulated net profit is always negative because the margin calculated by (35) squeezes the profit generated by the simulator.

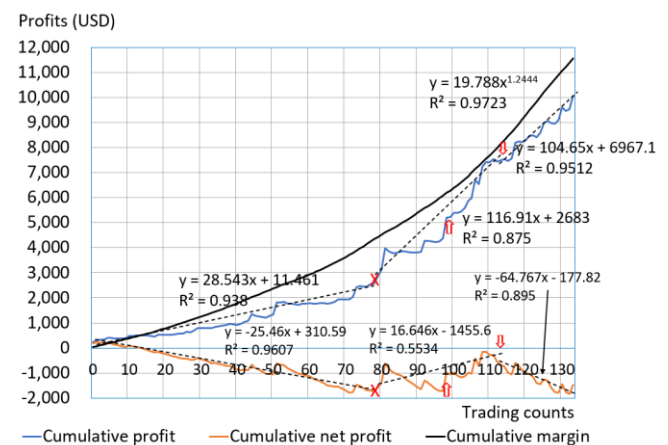


Figure 27. Cumulative profit, net profit, and margin interest in short position.

Figure 28 shows a scatter plot of simulated trades in a short position with the margin interest using the NASDAQ



daily stock data. Significant losses are incurred in the trades in which the number of days to hold stock is one.

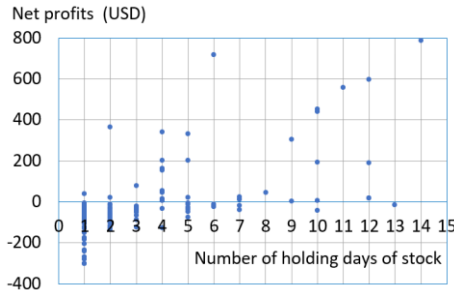


Figure 28. Scatter plot of trades in short position with margin interest.

The trade with the largest amounts of profit, which starts on Dec. 4, 2018 and ends on Dec. 26, 2018, is represented by the dot in the upper right corner of Figure 28. Figure 29 shows the candlestick chart of this lucrative trade.



Figure 29. Candlestick chart including successful trades.

The decision to buy the stock back based on the slope of the fast APAD and candlestick patterns is working profitably in this trade. It is noteworthy that Dec. 12, 2018, marked by the two dashed vertical lines, is not a day to buy the stock back although the slope of the fast APAD turns to positive. The variable *s13* defined by (34) is *False*, because the length of the candlestick body on Dec. 12, 2018 is -0.4%.

F. Improving decision of selling opportunities

1) Improvement to avoid lossy trade

Similar to the simulation in a long position, the simulation of short selling can be improved using candlestick patterns. Figure 30 shows an overall algorithm enhanced by candlestick pattern. The condition *CandlePattern\_Avoid\_Sale* is added after checking the *Sale\_Signal\_S*. *CandlePattern\_Avoid\_Sale* consists of several candlestick patterns that signal for avoiding a short sale of a stock. All patterns must be *False* for a trade in a short position to be performed.

```

01 T2S_CD ( int IndexFrom, int IndexTo ) {
02   int flg= 1;    // to indicate state of finding short-sell-opportunity
03   for ( j= IndexFrom; j>=IndexTo; j-- ) {
04     /* Short-sell signal is detected */
05     if ( flg > 0 & Sale_Signal_S ) {
06       if ( ! CandlePattern_Avoid_Sale ) {
07         traded_p = ((Open[j] + Close[j]) / 2);
08         holding_day_cnt= 0;
09         flg= -1;    // to indicate short-sell-stock state
10       }
11     /* Buy signal is detected in short position */
12   } else if ( flg < 0 & Buy_Signal_S ) {
13     diff = traded_p - ((Open[j] + Close[j]) / 2) - selling_cost;
14     profit= profit + diff;    // Total profit
15     flg= 1;    // to indicate state of finding short-sell-opportunity
16   }
17   holding_day_cnt++;
18 }
19 }
    
```

Figure 30. Overall simulation algorithm improved by candlestick pattern in short position.

Figure 31 shows a temporary-drop pattern that occurs after about a week of stock price rally, which is scarcely mentioned in the literature [12][13]. The pattern observed on Nov. 10, 2020 when the slope of the fast APAD changes the direction from up to down. This pattern is also observed on May 21, 2020, Jan. 3, 2019, Feb. 23, 2016, etc. Since this is a temporary decline, short selling on this day lead to losses.



Figure 31. Temporary-drop pattern after strong rally.

The following Boolean condition *cs1* is implemented in our simulator to avoid short selling on a temporary-drop after noticeable stock price rally. Unfortunately, because the condition involves more than six candlesticks, it is difficult to be defined by a Boolean expression, and it needs to be processed by a program.

$$\begin{aligned}
 cs1 = & (Prices \text{ rise more than four days in the past six days}) \\
 & \& (The \text{ amount of drop on the day is less than half of the} \\
 & \text{cumulative increase in the past six days}) \quad (36)
 \end{aligned}$$

The *Doji* candlestick [12][13], which has the opening price and the closing prices are equal or almost the same, is a sign of indecision. Our simulator is also implemented to refrain from a selling stock on the *Doji* candlestick. In addition, several other candlestick patterns to avoid a lossy buying stock trade are enhanced to the simulator.

2) Experimental results using NASDAQ historical data

Table VII summarizes the main characteristics of the simulators before and after the improvements in a short position. The number of trades decreases from 134 to 107, while the success rate increases from 67.91% to 71.03% without the margin interest, and from 24.63% to 25.23% with the margin interest. While profits decline slightly from US\$3,503.68 to US\$3,340.08, losses reduce notably from US\$-4,984.35 to US\$-3,270.74 with the margin interest.

TABLE VII. Summary of experimental results in short position

	APAD only		APAD with enhancements	
	Without margin	With margin	Without margin	With margin
Total no. of trades	134	134	107	107
No. of profitable trades	91	33	76	27
No. of lost trades	43	101	31	80
Success rate	67.91%	24.63%	71.03%	25.23%
Amount of profit (USD)	11,580.85	3,503.68	9,893.46	3,340.08
Amount of loss (USD)	-1,508.59	-4,984.35	-868.65	-3,270.74
Net profit (USD)	10,072.26	-1,480.68	9,024.81	69.34
Most profitable trade (USD)	877.10	785.50	877.10	785.50
Most lossy trade (USD)	-172.00	-300.30	-98.70	-280.30

Figure 32 shows line graphs of the cumulative profit, net profit, and margin interest in a short position after the improvement using candlesticks. The graph of cumulated profit in Figure 32 can be approximated by three segmented lines. The trade dates separating the three segments are the same as in Figure 27. The improvement boosts the net profit. However, the cumulative profits remain negative for most of the period.

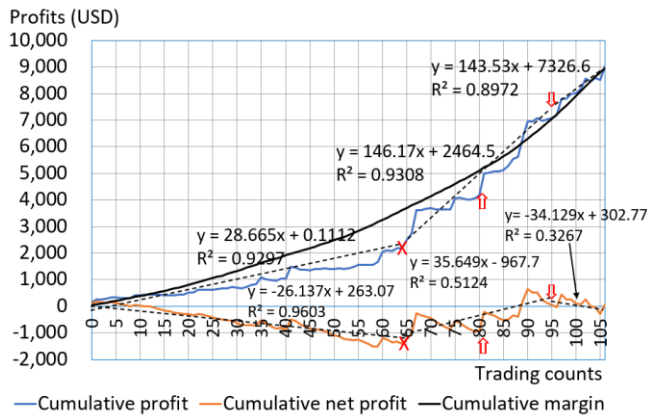


Figure 32. Cumulative profits and margin interest after improvement in short position.

Figure 33 shows a scatter plot of simulated trades with the margin interest after the improvement. Improvements have reduced the number of non-profitable trades, but not sufficient. There are a lot of loss-making trades of which stock holdings are one to seven days.

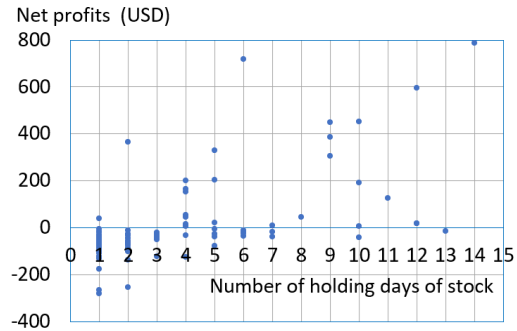


Figure 33. Scatter plot of trades with margin interest after improvement in short position.

Figure 34 shows a histogram of the number of holding days of a stock and the number of occurrences.

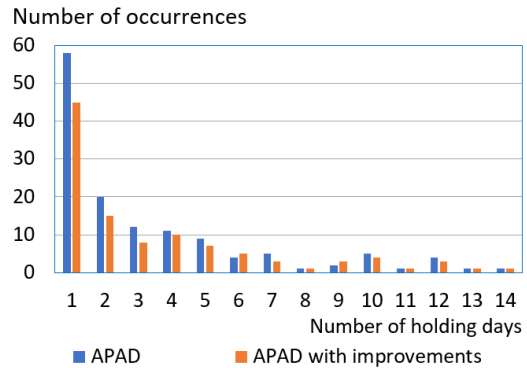


Figure 34. Histogram of the number of holding days in short position.

Similar to the histogram of a long position shown in Figure 25, the proposed improvement using candlestick patterns leads to a reduction in the number of lossy-making trades with a one-day holding period, i.e., from 58 to 45.

G. Experiments using stock data of major stock markets

Stock trade simulations are performed over the ten years' stock data using the nine markets, i.e., NASDAQ and Dow Jones Industrial Average (U.S.), Bovespa (Brazil), CAC (France), DAX (Germany), Nikkei 225 (Japan), S&P ASX (Australia), Sensex (India), and SSEC (China).

Table VIII shows the success rates for each market in both long and short positions, and with and without the fee or margin interest. Because this simulator is customized for the NASDAQ market, NASDAQ achieves the highest success rate of 74.44% with the fee in a long position. As for a short position, the Bovespa market has the highest success rate of 34.44% with the margin interest, followed by NASDAQ's success rate of 25.23%.

Table VIII. Success rates for each market in long and short positions

	Long Position		Short Position	
	Without fee	With fee	Without margin	With margin
NASDAQ	76.68%	74.44%	71.03%	25.23%
DJIA	68.80%	65.25%	66.67%	22.55%
Bovespa	68.75%	63.60%	79.47%	34.44%
CAC	61.57%	55.69%	57.94%	21.50%
DAX	61.99%	60.18%	62.50%	20.83%
Nikkei_225	61.19%	59.70%	42.73%	21.82%
S&P ASX	69.09%	66.18%	80.99%	20.66%
Sensex	77.51%	74.16%	67.91%	22.39%
SSEC	65.74%	63.75%	75.00%	22.12%

Figure 35 shows the line graphs of the cumulative profits with the fee for each market in a long position. Since stock prices in each market have risen significantly over the ten-year period of the simulation, the ratio of the profits, i.e., the Y-axis in Figure 35, is calculated by dividing the amount of profit by the average stock price of the market.

In NASDAQ, for example, the 25-day average price at the beginning of the simulation, i.e., May 1, 2012, is US\$3,031.64. The 25-day average price at the end of the simulation, i.e., Apr. 30, 2022, is US\$13,394.16. The average stock price of NASDAQ is US\$8,212.90 by calculating the average of these two prices. The maximum ratio of the amount of profit to the average stock price is 275.5% as shown in Figure 35.

All graphs in Figure 35 show an upward trend. Cumulative profit in NASDAQ become larger with increasing slope after Covid-19 pandemic. Cumulative profit in SSEC rapidly increases in the period from the 63rd to the 95th trades, which correspond to Oct. 28, 2014 and Mar. 30, 2016, respectively. The other graphs are roughly linear. This indicates that the proposed simulator constantly generates profits in the nine markets over the ten-year simulation period.

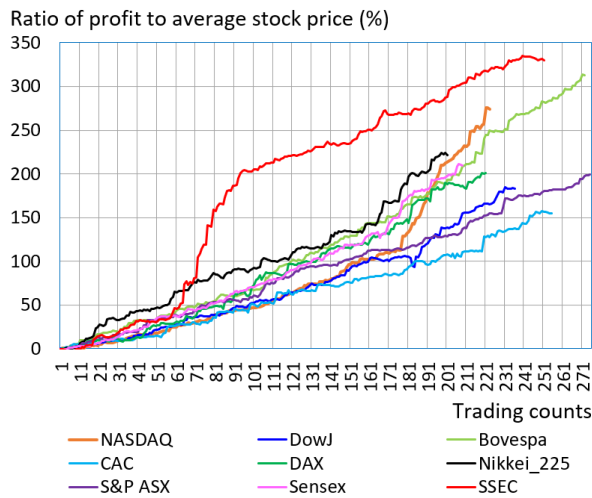


Figure 35. Line graph of profit for each market in long position.

Figure 36 shows the line graphs of the profits with the margin interest for each market in a short position. As the Bovespa’s success rate of 34.44% suggests, Bovespa apparently achieves higher profits than NASDAQ. Other markets take losses.

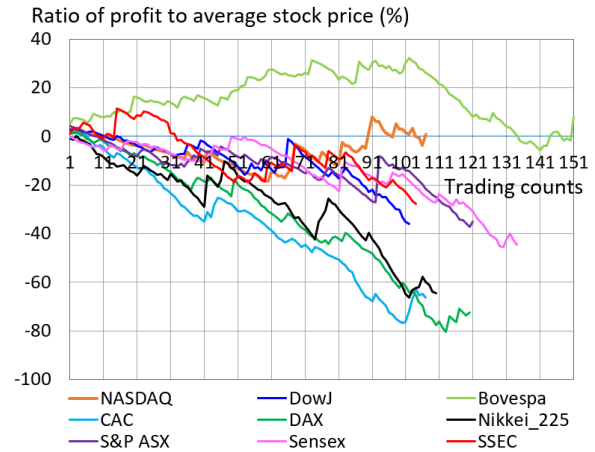


Figure 36. Line graph of profit for each market in short position.

The prediction accuracy of the developed simulator is still insufficient for a short position, as large portions of the graphs in Figure 36 are in the negative region.

## VI. CONCLUSION AND FUTURE WORK

This paper describes the results of analyses of stock price fluctuations in European, U.S. and Asian markets with special focus on the effect of Covid-19 pandemic. In general, thanks to the timely implementation of monetary measures of each country, all stock prices under study have kept rising after the lowest price recorded in Mar. 2020.

Observed in the 490 business days, i.e., approximately two years, NASDAQ see a 62.36% increase in its stock price compared to its highest price before the declaration of Covid-19 pandemic. CAC (France) decreased by -1.73%, followed by 9.25% increase in DAX (Germany). Analyses of the average and standard deviation of the six attributes of a candlestick chart reveal that CAC and DAX are deemed to experience a larger impact on stock prices than the other markets.

We propose a trend reversal indicator named APAD, an acronym for “Average of Price and 5-day Average Difference.” The indicator is devised in the process of investigating how the difference between the stock price and 5-day average is related to the reversal of stock trends affected by Covid-19 pandemic.

Since the APAD is an indicator based on a 5-day average, the APAD can be difficult to deal with short-term fluctuations over one or two days. Candlestick chart patterns are used to compensate for this shortcoming. We developed a stock trading simulator that implements the APAD enhanced by heuristic candlestick chart patterns. Experiments using ten years of stock price data from nine major stock markets in the world were conducted, and analyzed the profit characteristics.



It is confirmed that the simulator achieved the success rate of up to 74.44% in NASDAQ including fees in a long position. However, as for in a short position, the success rate remains within 25.23% percent, because the approximately 1.15% margin interest to borrow stocks squeezes profits.

We are planning researches to improve the success rate and profitability of the developed simulator through the addition of candlestick patterns and the sophisticated use of the APADs. Further experiments using stock price data from various global markets and individual companies are planned to verify the functionality of the proposed simulators.

#### REFERENCES

- [1] Y. Udagawa, "Analyzing Impact of COVID-19 Pandemic on Global Stock Prices," The Seventh International Conference on Big Data, Small Data, Linked Data and Open Data (ALLDATA 2021), IARIA, pp. 23-29, ISBN: 978-1-61208-842-6, Apr. 2021.
- [2] A. Chudik, K. Mohaddes, M. H. Pesaran, M. Raissi, and A. Rebucci, "Economic consequences of Covid-19: a counterfactual multi-country analysis," Available from: <https://voxeu.org/article/economic-consequences-Covid-19-multi-country-analysis/> Oct. 2020.
- [3] "World Economic Outlook," International Monetary Fund, <https://www.imf.org/en/Publications/WEO/> Apr. 2022.
- [4] "Major World Market Indices," Available from: <https://www.investing.com/indices/major-indices/> Apr. 2022.
- [5] E. Dimson, P. Marsh, and M. Staunton, "Should you invest in emerging markets?" London Business School, Available from: <https://www.london.edu/think/emerging-markets/> Apr. 2019.
- [6] Y. Udagawa, "Statistical Analysis of Stock Profits to Evaluate Performance of Markets," The Sixth International Conference on Big Data, Small Data, Linked Data and Open Data (ALLDATA 2020), IARIA, pp. 14-21, ISBN: 978-1-61208-775-7, Feb. 2020.
- [7] C. C. Ngwakwe, "Effect of COVID-19 Pandemic on Global Stock Market Values: A Differential Analysis," Acta Universitatis Danubius. (Economica, No 2, Vol 16, pp. 255-269, 2020.
- [8] P. Verma, A. Dumka, A. Bhardwaj, A. Ashok, M. C. Kestwal, and P. Kumar, "A Statistical Analysis of Impact of COVID-19 on the Global Economy and Stock Index Returns," Springer Nature journal Computer Science, 2:27, pp. 13, Jan. 2021.
- [9] M. Buszko, W. Orzeszko, M. Stawarz, "COVID-19 pandemic and stability of stock market - A sectoral approach," PLoS ONE 16(5): e0250938, pp. 1-26, doi:10.1371/journal.pone.0250938, May, 2021.
- [10] M. Obthong, N. Tantisantiwong, W. Jeamwathanachai, and G. Wills, "A Survey on Machine Learning for Stock Price Prediction: Algorithms and Techniques," The second International Conference on Finance, Economics, Management and IT Business, pp. 63-71, doi:10.5220/0009340700630071, Feb. 2020.
- [11] M. R. Islam and N. Nguyen, "Comparison of Financial Models for Stock Price Prediction," Journal of Risk and Financial Management, 13(8), pp. 1-19, doi:10.3390/jrfm13080181, Aug. 2020.
- [12] "Technical Analysis," Cambridge Univ., Available from: [http://www.mrao.cam.ac.uk/~mph/Technical\\_Analysis.pdf](http://www.mrao.cam.ac.uk/~mph/Technical_Analysis.pdf), pp. 1-179, Feb. 2011.
- [13] S. Nison, "Japanese Candlestick Charting Techniques," Second edition, New York Institute of Finance, 2001.
- [14] S. Glen, "Statistics How To, T-Distribution/Student's T: Definition, Step by Step Articles," <https://www.statisticshowto.com/probability-and-statistics/t-distribution/> 2020.
- [15] Y. Udagawa, "Developing a Stock Trading Simulator Using Candlestick Chart Patterns for Estimating Profitability of Global Markets," The Eighth International Conference on Big Data, Small Data, Linked Data and Open Data (ALLDATA 2022), IARIA, pp. 1-7, ISBN: 978-1-61208-945-4, Apr. 2022.

## An FFT based Bias Voltage Control Scheme to Compensate for the Resonant Frequency Drift of CMUTs Due to Fluid Loading

Thasnim Mohammed, Sazzadur Chowdhury  
 Department of Electrical and Computer Engineering  
 University of Windsor  
 Windsor, Ontario, Canada  
 mohamm43@uwindsor.ca, sazzadur@uwindsor.ca

**Abstract**—This paper presents a Fast Fourier transform (FFT) based bias voltage control scheme to compensate for the resonant frequency drift of Capacitive Micromachined Ultrasonic Transducers (CMUTs) due to fluid loading. A unified mathematical model for the resonant frequency of a CMUT that includes the electrostatic spring softening effect and the fluid loading effect due to the coupled fluidic layer has been developed that provides the basis of the proposed approach. The actual resonant frequency drift has been obtained by comparing the center frequencies of transmitted and received signals extracted through analog-to-digital conversion (ADC) and subsequent FFT of both signals. The frequency drift is then compensated by dynamically adjusting the DC bias voltage that modifies the electrostatic spring softening parameter. Analytical and COMSOL based 3D Finite Element Analysis (FEA) results show that the drift in the resonant frequency of a 6 MHz CMUT operated in water can be compensated by adjusting the bias voltage by 2% from its 75% pull-in voltage value to render an improvement of 4% in lateral and axial resolutions in imaging applications. A bias voltage adjustment of 9% of the 75% pull-in voltage value is necessary to achieve an improvement of 20.74% when the CMUT is operated in glycerol. The scheme can be realized using standard commercially available microelectronic components to improve the accuracy and reliability of diagnostic imaging and non-destructive evaluation (NDE).

**Keywords**—CMUT; electrostatic spring softening; fluid loading effect; resonant frequency drift; FFT-based microelectronic compensation.

### I. INTRODUCTION

Fluid loading induced resonant frequency drift of capacitive micromachined ultrasonic transducers (CMUTs) challenge their applications in biomedical imaging, non-destructive evaluation (NDE), and high intensity focused ultrasound (HIFU) applications [1]–[4]. Specifically, the phenomenon contributes to lower imaging resolution and beamforming functionality compared to the piezoelectric ultrasonic transducers. A suitable solution to the problem is necessary to exploit the significant advantages of CMUTs over the piezoelectric ones, such as low cost fabrication, easier CMOS integration, improved signal-to-noise ratio, higher fractional bandwidth, and thermal stability.

The CMUT is basically a reciprocal electrostatic transducer that relies on capacitance change between a

movable and a fixed electrode separated by a small gap to generate or receive ultrasound [5][6]. The space between the electrodes is filled with either vacuum or a thin film of air. The movable electrode (typically a thin diaphragm) is supported at the edges by dielectric support posts.

When a CMUT array operates in a fluidic medium, the coupled fluidic layer manifests itself as an inertial mass onto the diaphragm and as a viscous damper depending on the viscosity of the fluidic medium. As these fluid loading effects alter the effective mass of the CMUT diaphragm, a drift of the resonant frequency of the CMUT occurs. Since the elements in a CMUT array are designed to have a fixed pitch to satisfy the Nyquist criteria of spatial sampling in beamforming operations, a change in the CMUT resonant frequency compromises the functional characteristics of the array.

Attempts to mitigate the fluid loading effects and thereby minimize the center frequency drift of CMUT arrays are diverse in nature but not too many [7]–[17]. Notably, the effects of fluid loading and waveguiding on the performance of a CMUT array were addressed in an analytical model presented in [7], where frequency dependent equivalent density was used as the effective CMUT diaphragm material density to account for the distributed fluid loading. The authors in [7] also presented the dimensional dependence of the equivalent density using a finite element method (FEM) model based data fitting technique. In [8], the acoustic parameters for an array consisting of parallel driven CMUT cells were derived as a complex mechanical fluid impedance that depends on the diaphragm shape function. As the CMUTs in [8] were modeled as piston radiators, the model lacks sufficient accuracy. The authors in [9] presented a generalized model for surface acoustic wave devices considering the mass and viscous fluid loading. The authors presented the frequency and phase changes occurring when the device is operated in immersion and the dependence of the frequency drift on the fluid properties was included to analyze the fluid loading effect. Numerical modeling of fluid interaction with oscillating surfaces is available in [10] from which reduced order models depicting the fluid interaction were developed based on the effective fluid volume, effective area of shear wave interaction, and effective length of viscous interaction. The sensitivity of CMUT equivalent circuit parameters to the fluid properties was investigated in [11], where a surface acoustic wave (SAW) device was

designed for sensing the fluid properties like mass density and viscosity. The phenomenon of drift in frequency of CMUTs operated in immersion or in contact with a fluid medium has been exploited to develop various biological and chemical sensing devices [12]–[17].

In this perspective, this paper presents a method to compensate the resonant frequency drift of a CMUT due to fluid loading by dynamically adjusting the DC bias voltage. The scientific basis of the proposed method lies in the spring softening effect associated with the DC bias voltage applied to a CMUT. Due to an applied DC bias, necessary for CMUT operation in both transmit and receive modes, the effective stiffness of the diaphragm reduces to lower the resonant frequency. As the amount of spring softening (lowering of the stiffness) depends on the DC bias, the DC bias voltage can be dynamically controlled using a suitable microelectronic circuit to adjust the spring softening amount to offset the effect of fluid loading on the resonant frequency. Excellent agreements between lumped element model analysis and 3D electromechanical finite element analysis conducted by the authors show that the method can effectively offset the resonant frequency drift due to fluid loading to validate the hypothesis. A microelectronics based compensation scheme to implement the proposed hypothesis has been designed and simulated using MATLAB and Simulink. The simulation results validate the efficacy of the proposed method, thereby proving the hypothesis.

The rest of the paper has been organized in the following manner: Section II provides the problem statement and theoretical background of the proposed approach, Section III investigates the effect of fluid loading on CMUT resonant frequency in different fluids, Section IV presents the proposed frequency drift compensation method with simulation results of analytical and FEA analysis using COMSOL, Section V presents the design of microelectronics based compensation scheme along with simulation results and finally, Section VI provides the concluding remarks.

## II. THEORETICAL BACKGROUND

A CMUT array is comprised of several transmit and receive elements as shown in Figure 1(a). Each element in turn is designed to have several CMUT cells connected in parallel [10][18]–[21]. A conceptual cross-section of a typical CMUT cell is shown in Figure 1(b).

Assuming the CMUT as a lumped element mass-spring-dashpot system with a linear spring constant, the resonant frequency of the CMUT in air can be calculated following

$$f_{\text{res}} = \frac{1}{2\pi} \sqrt{\frac{k}{m}} \quad (1)$$

where,  $k$  and  $m$  are the spring constant (stiffness) and mass of the CMUT diaphragm as shown in Figure 1(b), respectively.

The array pitch, defined as the distance between corresponding points in the neighboring elements (Figure 1(a)), is optimized to satisfy Nyquist criteria of spatial sampling as expressed in (2) to minimize grating lobes.

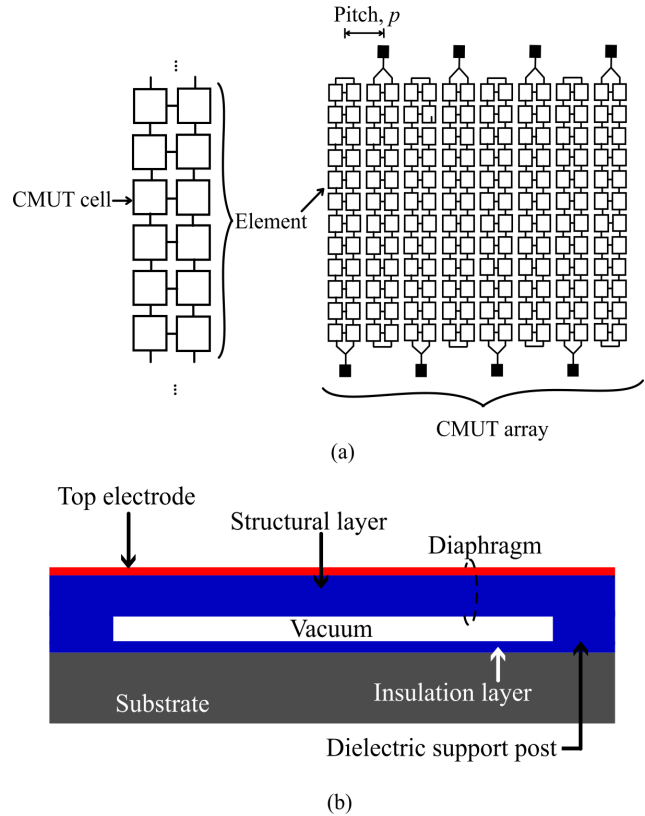


Figure 1. (a) Layout of a CMUT array comprising of array elements. (b) Typical cross-section of a single CMUT cell.

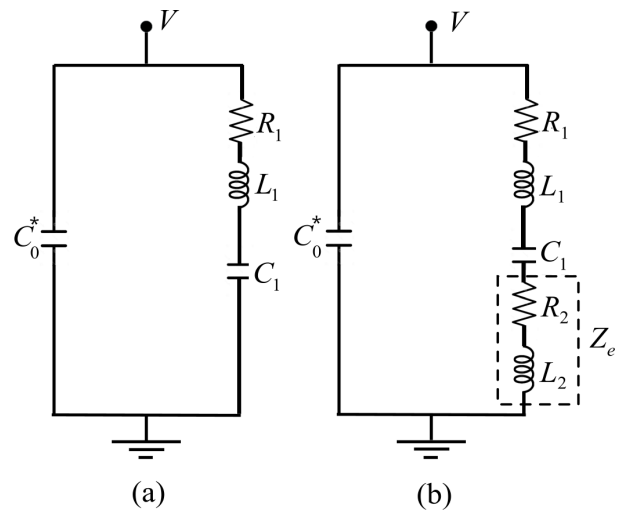


Figure 2. (a) The BVD model of a CMUT without fluid loading. (b) The BVD model of a CMUT with fluid loading. The fluid loading effect is represented by inductance  $L_2$  and resistance  $R_2$ .

$$p \leq \frac{\lambda}{2} \quad (2)$$

In (2),  $\lambda$  is the wavelength corresponding to the center frequency of operation.

The frequency response of a CMUT cell with and without the fluid loading effect can be obtained using the Butterworth-Van Dyke (BVD) model as shown in Figure 2. [22]. The BVD model depicts an analog electrical equivalent circuit of a CMUT that includes both electrical and equivalent mechanical lumped element circuit parameters.

In the BVD model of a CMUT, the electrical circuit element is represented by a lumped capacitor  $C_0^*$ , and the electrical equivalent of mass  $m$ , stiffness  $k$ , and damping element  $b$  is represented by  $L_1$ ,  $C_1$ , and  $R_1$  respectively.

Typically, the CMUT diaphragms are constructed to have a circular, hexagonal, or square shape geometry to satisfy the design requirements. For a CMUT with a square diaphragm in the small deflection regime, the stiffness  $k$  of the diaphragm can be calculated following

$$k = C_r \frac{t_d \sigma_0}{(L/2)^2} + C_b \frac{12D_{\text{eff}}}{(L/2)^4} \quad (3)$$

where  $C_r = 3.45$  and  $C_b = 4.06$  [23]. The parameters  $D_{\text{eff}}$ ,  $L$ ,  $\sigma_0$ , and  $t_d$  are the effective flexural rigidity, sidelength, residual stress, and thickness of the diaphragm, respectively. The mass of the diaphragm can be calculated from the volume and density of the CMUT diaphragm materials. The parameters  $R_1$ ,  $L_1$ , and  $C_1$  (zero bias equivalent stiffness) can be calculated from [22],

$$\begin{cases} C_1 = \frac{n^2}{k} \\ L_1 = \frac{m}{n^2} \\ R_1 = \frac{b}{n^2} \end{cases} \quad (4)$$

where,  $n$  is the electromechanical transformation ratio given by:

$$n = \frac{\epsilon_0 A V_{\text{DC}}}{(d_{\text{eff}} - x)^2} \quad (5)$$

In (5),  $\epsilon_0$ ,  $A$ ,  $V_{\text{DC}}$ ,  $x$  and  $d_{\text{eff}}$  denote the free-space permittivity, coupling area of the electrodes, DC bias voltage applied to the CMUT, average diaphragm deflection, and the effective electrode gap, respectively.

During immersion operation in a liquid medium, the coupled liquid layer contributes an additional equivalent damping  $R_2$  due to the fluid viscosity and an additional equivalent fluid mass  $L_2$ . This phenomenon alters the

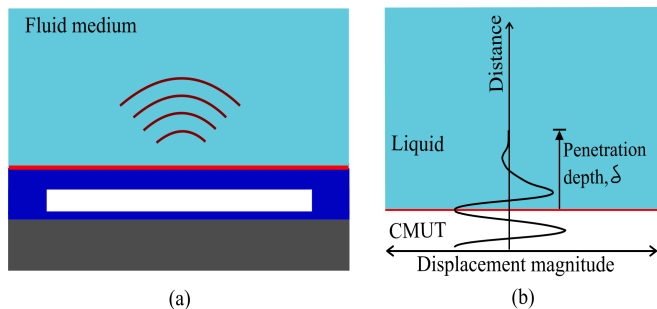


Figure 3. (a) Cross sectional view of CMUT contacted by liquid on one side (b) The penetration depth of shear wave propagated into the liquid medium.

CMUT motional impedance contributed originally by  $L_1$ ,  $C_1$ , and  $R_1$ .

The inclusion of these additional mass and damping parameters in the BVD model as shown in Figure 2(b) alters the frequency response of the CMUT. The amount of alteration (drift) depends on the density and viscosity of the fluidic medium.

The fluid loading quantities  $L_2$  and  $R_2$  can be calculated by accounting the penetration depth  $\delta$  of the damped shear wave propagating into the liquid due to the coupled oscillating surface (CMUT diaphragm) as shown in Figure 3. Following [22][24], the penetration depth  $\delta$  can be calculated from the relation  $\delta = \sqrt{2\eta / \omega_s \rho}$  where,  $\rho$ ,  $\eta$ , and  $\omega_s$  represents the liquid density, shear viscosity of the liquid, and resonant frequency of the CMUT in rad/sec, respectively. The effective height  $\delta/2$  of the entrained liquid layer can thus be used to calculate the fluid loading quantities  $L_2$  and  $R_2$  following [22] as

$$L_2 = \frac{\pi}{4k_s^2 \omega_s C_0} \left( \frac{\rho \eta}{2\omega_s \mu_q \rho_q} \right)^{\frac{1}{2}} \quad (6)$$

$$R_2 = \frac{\pi}{4k_s^2 C_0} \left( \frac{\rho \eta}{2\omega_s \mu_q \rho_q} \right)^{\frac{1}{2}} \quad (7)$$

where  $\mu_q$ ,  $\rho_q$ , and  $k_s^2$  represents the diaphragm shear stiffness, mass density of diaphragm, and coupling coefficient, respectively.

Thus, the motional impedance  $Z_m$  in the BVD model of a CMUT including the fluid loading effect (Figure 2(b)) can be calculated from:

$$Z_m = R_1 + R_2 + j\omega L_1 + j\omega L_2 + \frac{1}{j\omega C_1} \quad (8)$$

On the other hand, assuming the small-signal model of a CMUT where the AC actuation signal is much smaller compared to the DC bias voltage ( $V_{DC} \gg V_{AC}$ ), the spring softening effect due to an applied DC bias voltage reduces the zero bias CMUT diaphragm stiffness  $k$  by a factor  $k_{soft}$  as given by

$$k_{soft} = \frac{\epsilon_0 A V_{DC}^2}{d_{eff}^3} \quad (9)$$

Consequently, the spring softening effect causes a downshift of the resonant frequency of the CMUT diaphragm following

$$f_{res-ss} = \frac{1}{2\pi} \sqrt{\frac{k - k_{soft}}{m}} \quad (10)$$

During an immersion operation, the resonant frequency  $f_{res-ss}$  as expressed in (10) is further affected by the inertial mass  $m_f$  contributed by the coupled fluid layer adjacent to the CMUT diaphragm.

Thus, including both the spring softening effect and the fluid loading effect, the resonant frequency  $f_{res-ss-fl}$  of the diaphragm can be expressed as

$$f_{res-ss-fl} = \frac{1}{2\pi} \sqrt{\frac{k - k_{soft}}{m + m_f}} \quad (11)$$

A careful examination of (9) and (11) reveals that the drift in the resonant frequency due to the fluid loading effect  $m_f$  alone can be minimized by adjusting the spring softening factor  $k_{soft}$ , which is a function of the bias voltage  $V_{DC}$  in small-signal approximation.

A microelectronic circuit can be designed to sense the resonant frequency offset and adjust the bias voltage dynamically to control  $k_{soft}$  to bring the resonant frequency back to its design value while generating the desired acoustic outputs.

### III. INVESTIGATION OF FLUID LOADING EFFECTS

To investigate the proposed approach, a linear CMUT array with a center frequency of 6 MHz has been considered. This center frequency is suitable for cardiac diagnostic imaging as reported in [25].

#### A. Analytical CMUT Resonant Frequency Calculation with Fluid Loading Effects

Individual CMUT cells in the array have been designed to have the same resonant frequency as the array [26]. The element pitch  $p$  has been selected following (2) to satisfy the Nyquist criteria of spatial sampling to obtain better

directivity with minimum side lobes. The lateral and axial resolutions of the CMUT array are given by [2] as:

$$\Delta y = \left( \frac{1.22\lambda}{A_a} \right) S \quad (12)$$

$$\Delta x = \frac{n\lambda}{2} \quad (13),$$

respectively, where  $A_a$ ,  $S$ , and  $n$  are the array aperture, the distance between the focus and the array surface at the center of the array, and the number of scanning pulses, respectively. Following (12) and (13), it is necessary for the CMUT and the array physical dimensions to correspond to the same wavelength or center frequency to achieve setpoint axial and lateral resolutions and avoid any deviation from Nyquist criteria of spatial sampling [27][28].

To investigate the effects of viscosity and density of the fluidic medium on the resonant frequency of the CMUT diaphragms, a square shaped CMUT diaphragm with the same 6 MHz center frequency in air as for the array has been investigated.

Typical CMUT diaphragms are fabricated as a multilayer laminate where a thin insulating material is used as the structural layer to provide the necessary mechanical strength and a thin conducting layer is used on the top of the structural layer to act as the top electrode.

Following [27], bisbenzocyclobutene (BCB) has been used as the structural layer for this analysis. The BCB layer is coated with a thin layer of gold to form the top electrode. Low resistivity silicon has been used to constitute the bottom electrode (ground). BCB has also been used to fabricate the dielectric support posts. The major physical properties of the selected CMUT materials are listed in Table I.

TABLE I. CMUT MATERIAL PROPERTIES

Material	Young's modulus (GPa)	Poisson's ratio	Density (kg.m <sup>-3</sup> )	Residual stress (MPa)
BCB	2.9	0.34	1050	28
Gold	70	0.44	19300	106
Silicon	165.9	0.26	2329	55

TABLE II. GEOMETRICAL SPECIFICATION OF THE CMUT CELL.

Parameter	Value	Unit
Sidelength ( $L$ )	28	$\mu\text{m}$
Diaphragm thickness ( $d_m$ )	1.5	$\mu\text{m}$
Top electrode thickness ( $t_m$ )	0.1	$\mu\text{m}$
Cavity height ( $d_0$ )	0.75	$\mu\text{m}$
Insulation layer thickness ( $d_i$ )	0.1	$\mu\text{m}$

A CMUT cell has been designed using the materials listed in Table I to obtain the target 6 MHz resonant frequency in air. The design incorporates the electrostatic spring softening effect due to the bias voltage. Analytical and 3D electromechanical FEA methods were used to optimize the geometric dimensions as listed in Table II.

The drift in resonant frequency due to the spring softening effect as given by (9) and (10) is shown in Figure 4. The pull-in voltage of the CMUT has been determined from 3D electromechanical FEA using COMSOL as 395 volts. The frequency responses of the CMUT biased at 294.75 volts (74.68% of the pull-in voltage) in air, water, engine oil, and glycerol are shown in Figure 5. The physical properties of the fluids are listed in Table III.

The frequency response in air without any DC bias voltage has also been shown in Figure 5. As it can be seen, there is a significant drift in the resonant frequency due to

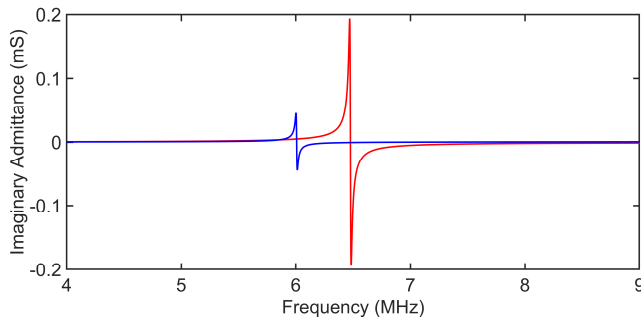


Figure 4. Drift in resonant frequency due to the spring softening effect.

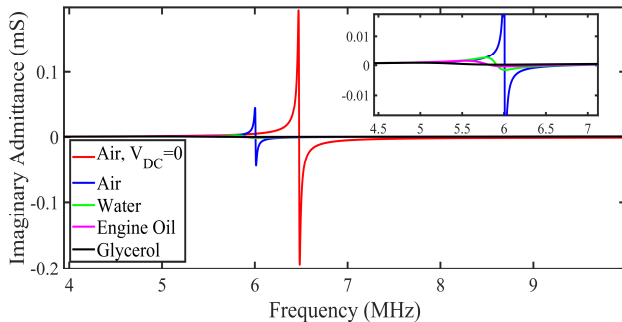


Figure 5. Frequency response of the CMUT cell with bias voltage  $V_{DC} = 0.75V_{PI}$  for unloaded (air) and loaded (water, engine oil, glycerol) conditions. (Inset y-axis unit is in  $\mu S$ .)

TABLE III. RESONANT FREQUENCY OF THE SIMULATED CMUT IN DIFFERENT FLUIDS

Medium	Resonant frequency (Analytical) (MHz)	Resonant frequency (FEA) (MHz)	Percentage deviation between analytical and FEA (%)	Drift from the resonant frequency in air (FEA)(%)
Air	6.002	6.19	3.1	-
Water	5.783	5.945	2.8	3.9
Engine oil	5.422	5.545	2.3	10.42
Glycerol	4.975	4.8788	1.9	21.18

the spring softening effect caused by the bias voltage, which is further deteriorated by the fluid loading effects contributed by the fluid mass and viscosity as shown in the zoomed-up inset.

**B. 3D FEA Based CMUT Resonant Frequency Simulation with Fluid Loading Effects**

A 3D FEA of a single CMUT cell was conducted in COMSOL Multiphysics environment to investigate the fluid loading effects on the frequency response of the CMUT cell. The cell was modeled using the electro-mechanical module as shown in Figure 6. The air/fluid column was designed using the pressure acoustics frequency domain physics module.

The CMUT diaphragm was set as a linear elastic material. The boundary between the electro-mechanics and pressure acoustics domain was defined as a fluid-solid coupled boundary that acts as a pressure load on the electro-mechanics domain. The maximum element size was set to one-sixth of the smallest wavelength in a frequency sweep while meshing. An AC perturbation signal superimposed with DC bias voltage was applied to the CMUT cell using the default frequency domain modal analysis method in COMSOL.

The COMSOL based 3D FEA generated modal frequency responses of the CMUT model for the fluids listed in Table III are shown in Figure 7. Table IV compares the analytical and 3D FEA results for the resonant frequencies of the CMUT in the investigated fluids. As Table IV reveals, the analytical and 3D FEA results are in excellent agreement with a maximum deviation of 2.8% for the liquids that validate the efficacy of the analytical and simulation methods.

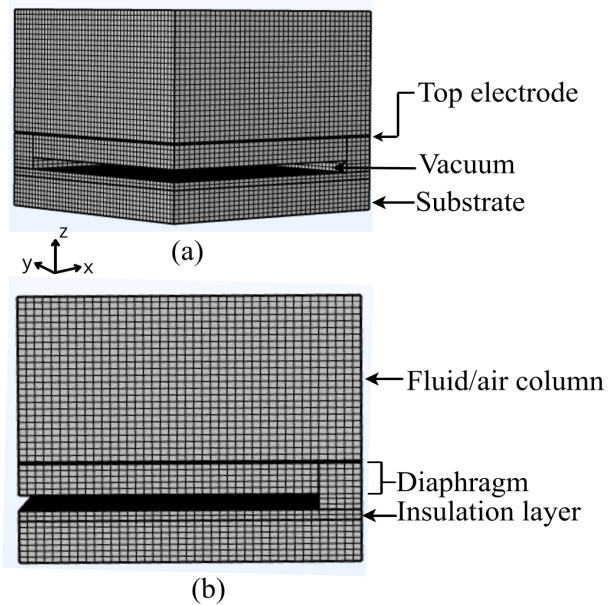


Figure 6. The CMUT model in COMSOL Multiphysics environment (As the structure is symmetric, only a quarter model of the CMUT was simulated).



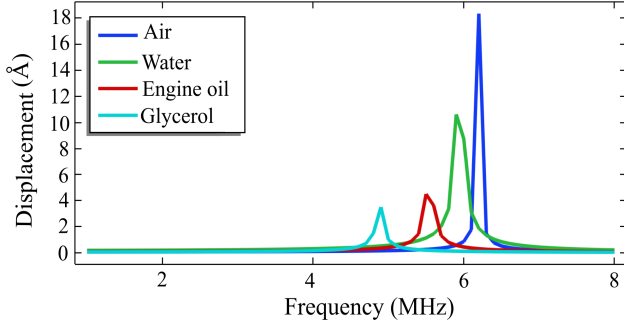


Figure 7. FEA simulation based frequency response of the CMUT cell in different fluids.

TABLE IV. CMUT RESONANT FREQUENCIES IN INVESTIGATED FLUIDS

Fluid type	Fluid Physical Properties		Resonant frequency (MHz)
	Viscosity (mPa/s)	Density (g/cm <sup>3</sup> )	
Air	0.01825	0.001204	6.002
Water	1.0016	0.998	5.783
Engine oil	287.23	0.8787	5.422
Glycerol	1412.8	1.2608	4.975

#### IV. COMPENSATING THE FLUID LOADING EFFECTS

Investigation by the authors reveals that the resonant frequency drift due to the fluid loading effect can be compensated effectively if it is possible to offset the effects of fluid loading by dynamically adjusting the DC bias voltage following (9) and (11). Precisely, the spring softening parameter  $k_{\text{soft}}$  as expressed in (9) needs to be dynamically adjusted by adjusting the bias voltage  $V_{\text{DC}}$ , to offset the fluid loading induced resonant frequency drift  $\Delta f$  expressed in (14).

$$\Delta f = f_{\text{res-ss}} - f_{\text{res-ss-fl}} \quad (14)$$

Figures 8-10 compare uncompensated (red) and DC bias compensated (black) frequency responses of the CMUT when operated in different fluidic media. From the figures, it is clear that a slight adjustment of the DC bias, (e.g.,  $0.75 V_{\text{PI}}$  to  $0.68 V_{\text{PI}}$  for glycerol) is sufficient to alter the electrostatic spring softening parameter to offset the resonant frequency drift due to the mass loading.

COMSOL based 3D FEA simulation results after adjusting the bias voltage to offset the resonant frequency drift  $\Delta f$  for water, engine oil, and glycerol are shown in Figures 11-13. The reference bias voltage was kept at 75%, of the pull-in voltage  $V_{\text{PI}}$  of the CMUT structure as in the Section III. When the CMUT was operated in water, engine

oil, and glycerol, the change in bias voltage applied was 2%, 4.2%, and 9%, respectively. The simulation results confirm that the drift  $\Delta f$  in the resonant frequency can effectively be compensated by a small change in the DC bias  $V_{\text{DC}}$  to validate the hypothesis.

The improvement in the lateral and axial resolution that could be achieved with the proposed method is 4%, 10.78%, 20.74% when operated in water, engine oil, or glycerol, respectively. Thus, the proposed method helps to achieve the expected resolution even when the array is operated in different fluidic mediums, thereby improving the imaging accuracy of a CMUT array.

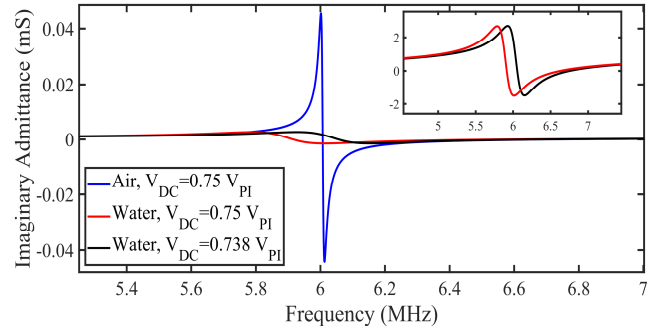


Figure 8. Simulation result of CMUT small-signal analysis for frequency compensation in water. (Inset y-axis unit is in  $\mu\text{S}$ .)

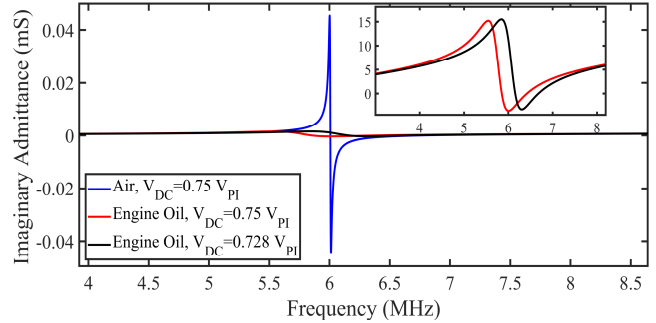


Figure 9. Simulation result of CMUT small-signal analysis for frequency compensation in engine oil. (Inset y-axis unit is in  $\mu\text{S}$ .)

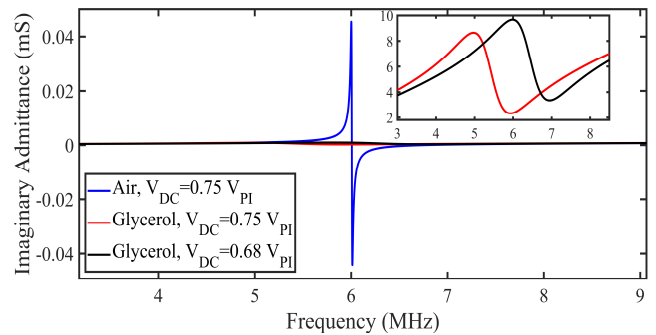


Figure 10. Simulation result of CMUT small-signal analysis for frequency compensation in glycerol. (Inset y-axis unit is in  $\mu\text{S}$ .)



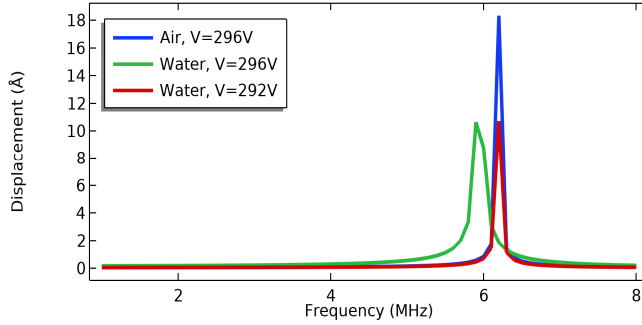


Figure 11. FEA simulation result for frequency compensation in water

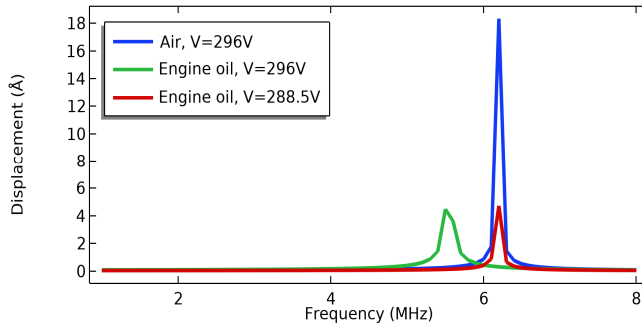


Figure 12. FEA simulation result for frequency compensation in engine oil.

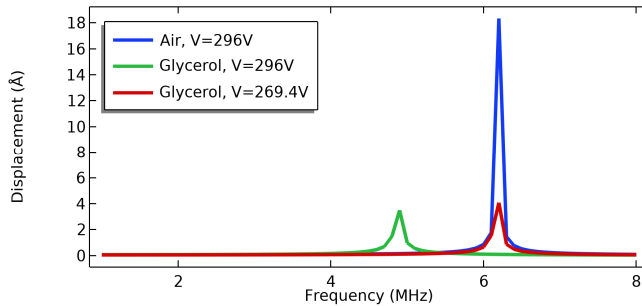


Figure 13. FEA simulation result for frequency compensation in glycerol.

## V. COMPENSATION SCHEME IMPLEMENTATION

An implementation method as shown in Figure 14 to realize the proposed compensation scheme has been developed and simulated using a Simulink based model. The model incorporates commercially available analog-to-digital converters (ADC), fast Fourier transform (FFT) modules, and bandpass filters among others, to dynamically determine the drift amount and apply necessary compensation to stabilize the center frequency using a digital signal processing (DSP) based algorithm.

### A. Implementation Strategy

The actual drift amount in the resonant frequency of a CMUT during an array operation can be determined by comparing the center frequencies of the transmitted and the

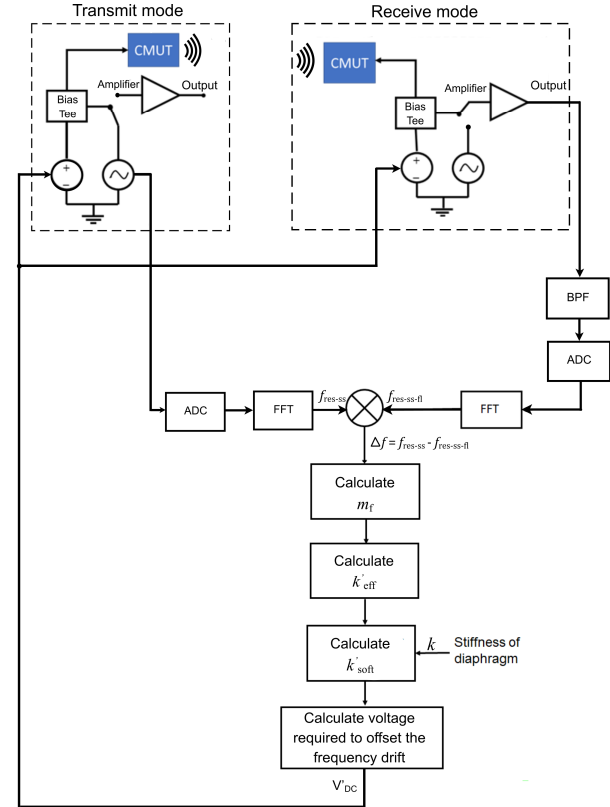


Figure 14. Block diagram of the compensation scheme.

received signals. While the center frequency of the transmitted signal can easily be extracted from the excitation AC signal applied to the CMUT, appropriate conditioning of the received signal is necessary to extract the center frequency component of the received signal. Accordingly, a bandpass filter is necessary to signal-condition the received signal before converting to frequency domain. To extract the center frequencies, the AC excitation signal and the filtered received signal are digitized using high-speed analog-to-digital converters (ADC) with a suitable sampling rate.

The time-domain digitized signals are then transformed into frequency domain using fast Fourier transforms (FFT) to obtain center frequencies in both signals ( $f_{res-ss}$  and  $f_{res-ss-fl}$  in Figure 14). A comparator then compares  $f_{res-ss}$  and  $f_{res-ss-fl}$  to determine the frequency drift  $\Delta f$  in real time.

The mass  $m_f$  of the loaded fluidic layer is then extracted from  $\Delta f$  using (10) and (11) as the other parameters in (10) and (11) are already known. The compensation scheme then calculates a modified effective stiffness  $k'_{eff}$  to offset the effects of  $m_f$  in (11). A new value of the spring softening parameter  $k'_{soft}$  is then extracted from  $k'_{eff}$  following

$$k'_{eff} = k - k'_{soft}. \quad (15)$$

Finally, the modified value of DC bias voltage  $V'_{DC}$  is extracted from  $k'_{soft}$  following (9) as

$$V'_{DC} = \sqrt{\frac{k'_{soft} d_{eff}^3}{\epsilon_0 A}} \quad (16)$$

The compensation module applies the modified bias voltage  $V'_{DC}$  to the CMUT through the bias tee as shown in Figure 14.

### B. Scheme Validation

The proposed scheme as shown in Figure 14 was implemented and simulated in MATLAB and Simulink. First, a bandpass filter was used to remove the noise from the received signal. Next, both the AC excitation signal and the filtered received signal are sampled at a frequency,  $f_{sampling} \geq 2f_{res-ss}$  (15 MHz for the designed 6 MHz array) to digitize them. The digitized signals are then fast Fourier transformed to extract the frequency components. The spectra of the digitized signals after performing FFT for operation in water are compared in Figure 15. Similarly, the spectra of the digitized signals after performing FFT for operation in engine oil and glycerol are compared in Figures 16-17, respectively. The drift in center frequency due to fluid loading is clearly visible in the figures with higher density and viscosity medium (glycerol) showing higher drift compared to lower density and viscosity medium (water).

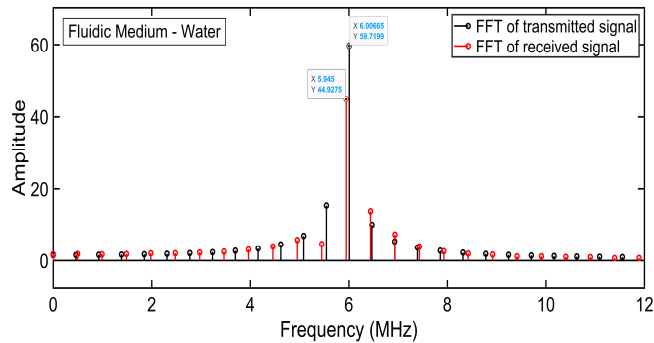


Figure 15. FFT of the filtered received signal when operated in water.

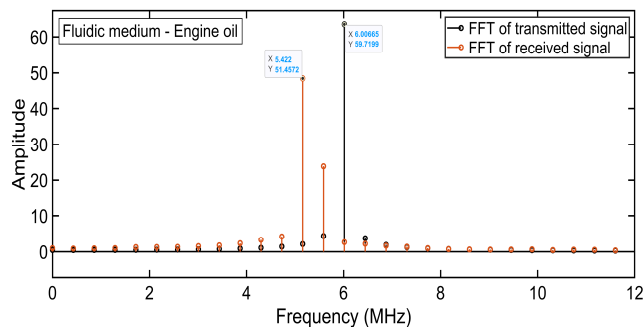


Figure 16. FFT of the filtered received signal when operated in engine oil.

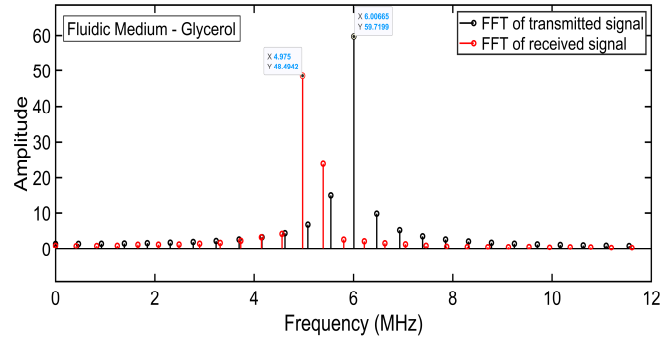


Figure 17. FFT of the filtered received signal when operated in glycerol.

Based on the magnitude of drift in center frequency, a modified DC bias voltage  $V'_{DC}$  is then calculated for each case following (14) – (16) to adjust the spring softening parameter to offset the fluid loading effect. The spectra of the filtered and digitized received signals with  $V'_{DC}$  after FFT are shown in Figure 18. for all three cases along with the spectra of the transmit AC signal. As it can be seen, the adjustment of the DC bias compensates the frequency drift as expected.

The amount of  $V'_{DC}$  in different fluidic media is different corresponding to the different amount of drift in different fluidic media. TABLE V. compares MATLAB and Simulink predicted  $V'_{DC}$  following the proposed scheme with the uncompensated DC bias  $V_{DC}$  for different fluidic media. Similarly,  $V'_{DC}$  obtained from 3D FEA simulation in COMSOL are compared in Table VI with uncompensated DC bias for different fluidic media.

The tables show that the variations in DC bias are very small to minimally affect the target acoustic pressure. Furthermore, Tables V and VI show that the analytical results obtained from the simulated scheme in Matlab and Simulink are in excellent agreement with 3D FEA results to validate the implementation scheme.

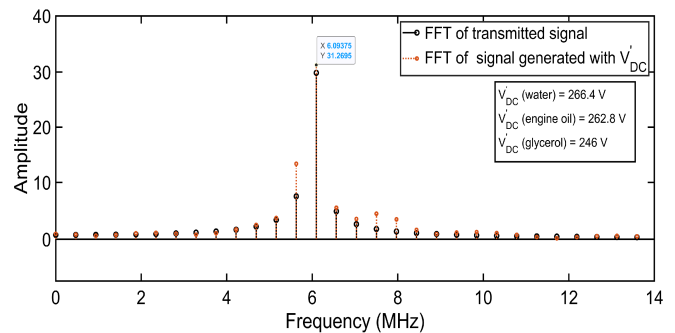


Figure 18. Comparison of FFT of the transmit and received signals after DC bias adjustment from  $V_{DC}$  to  $V'_{DC}$ .

TABLE V. COMPARISON OF DC BIAS ADJUSTMENT REQUIRED TO OFFSET THE FREQUENCY DRIFT IN MATLAB/SIMULINK

Medium	DC bias voltage ( $V_{DC}$ ) (V)	Modified DC voltage ( $V'_{DC}$ ) (V)	Percentage deviation (%) $\frac{V_{DC} - V'_{DC}}{V_{DC}}$
Water	270.675	266.43	1.56
Engine oil	270.675	262.8125	2.9
Glycerol	270.675	246	9.11

TABLE VI. COMPARISON OF DC BIAS ADJUSTMENT REQUIRED TO OFFSET THE FREQUENCY DRIFT IN COMSOL

Medium	DC bias voltage ( $V_{DC}$ ) (V)	Modified DC voltage ( $V'_{DC}$ ) (V)	Percentage deviation (%) $\frac{V_{DC} - V'_{DC}}{V_{DC}}$
Water	294.75	292	0.93
Engine oil	294.75	288.5	2.12
Glycerol	294.75	269.4	8.6

## VI. CONCLUSION

Microelectronic and digital signal processing techniques have been used to compensate fluid loading induced resonant frequency drift of a CMUT by dynamically adjusting the DC bias voltage. Analytical and 3D FEA based simulation studies revealed that the magnitude of DC bias voltage adjustment is small and is not expected to affect the acoustic power/pressure output of a CMUT array while improving lateral and axial resolution. The investigation shows that the method can improve the lateral and axial resolutions in water by approximately 4% to improve the accuracy of CMUT based biomedical diagnostic imaging. For NDE applications, the axial and lateral resolution improvements can be as high as 20.74% when glycerol is used as the coupling agent.

MATLAB and Simulink based simulation of the proposed scheme shows that commercially available high-speed data converters and digital signal processing (DSP) integrated circuits can be used to determine the frequency drift using Fast Fourier Transform (FFT) of the digitized transmit and the processed received signal. A DSP based controller can then dynamically compute and apply the modified DC bias to offset the frequency drift. High-speed signal processing components are necessary to minimize the latency time. The compensation scheme will be realized and tested once necessary funding is available.

## ACKNOWLEDGMENT

This research work was supported by the Natural Science and Engineering Research Council of Canada (NSERC)'s discovery grant number RGPIN 293218. The authors also

acknowledge the collaborative research support provided by the IntelliSense Corporation, Lynnfield, MA, Angstrom Engineering, ON, and the CMC Microsystems, Canada.

## REFERENCES

- [1] T. Mohammed and S. Chowdhury, "A Method to Minimize Resonant Frequency Drift of CMUTs Due to Fluid Loading", *SENSORDEVICES 2021, The Twelfth International Conference on Sensor Device Technologies and Applications*, ISSN: 2308-3514, pp. 88-91, November 2021. Available online: [https://www.thinkmind.org/index.php?view=article&articleid=sensor\\_devices\\_2021\\_2\\_120\\_20053](https://www.thinkmind.org/index.php?view=article&articleid=sensor_devices_2021_2_120_20053)
- [2] J. Song et al. "Capacitive Micromachined Ultrasonic Transducers (CMUTs) for Underwater Imaging Applications," *Sensors*, vol. 15, no. 9, pp. 23205–23217, Sep. 2015.
- [3] D. Zhao, S. Zhuang, and R. Daigle, "A commercialized high frequency CMUT probe for medical ultrasound imaging," in *Proc. of 2015 IEEE International Ultrasonics Symposium (IUS)*, Oct. 21-24, Taipei, Taiwan, 2015, pp. 1-4.
- [4] A. S. Ergun, G. G. Yarlioglu, O. Oralkan, and B. T. Khuri-Yakub, "Chapter 7: Techniques & Applications of Capacitive Micromachined Ultrasonic Transducer," in *MEMS/ NEMS Handbook[ Techniques and Applications]*; Leondes, C.T., Ed.; Springer Science + Business Media Inc.: Los Angeles, CA, USA, 2006; Volume 2, pp. 222–332.
- [5] A. S. Ergun, G. G. Yarlioglu, and B. T. Khuri-Yakub, "Capacitive micromachined ultrasonic transducers: Theory and technology." *Journal of aerospace engineering*, vol 16, no. 2, pp. 76-84, 2003.
- [6] Y. Huang, A. S. Ergun, E. Haeggstrom, M. H. Badi, and B. T. Khuri-Yakub, "Fabricating capacitive micromachined ultrasonic transducers with wafer-bonding technology," *Journal of Microelectromechanical Systems*, vol. 12, no. 2, pp. 128-137, April 2003.
- [7] B. Azmy, M. El-Gamal, A. El-Henawy, and H. Ragai, "An accurate model for fluid loading on circular CMUTs," in *Proc. of 2011 IEEE International Ultrasonics Symposium*, Orlando, Florida, Oct. 18-24, 2011, pp. 584-587.
- [8] A. Lohfink, P. -. Eccardt, W. Benecke, and H. Meixner, "Derivation of a 1D CMUT model from FEM results for linear and nonlinear equivalent circuit simulation," in *Proc. of 2003 IEEE Ultrasonics Symposium*, Honolulu, Hawaii, Oct. 5-8, 2003 pp. 465-468.
- [9] M. Tsai and J. Jeng, "Development of a generalized model for analyzing phase characteristics of SAW devices under mass and fluid loading," in *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 57, no. 11, pp. 2550-2563, November 2010.
- [10] E. K. Reichel, M. Heinisch, B. Jakoby, and T. Voglhuber-Brunnmaier, "Efficient numerical modeling of oscillatory fluid-structure interaction," in *Proc. of IEEE SENSORS 2014*, Valencia, Spain, Nov. 2-5, 2014, pp. 958-961,
- [11] M. Thränhardt, P. Eccardt, H. Mooshofer, P. Hauptmann, and L. Degertekin, "A resonant CMUT sensor for fluid applications," in *Proc. of IEEE SENSORS 2009, Christchurch, New Zealand, 25-28 October, 2009*, pp. 878-883.
- [12] S. Eaimkhong, M. Steiert, T. Harper, M. Cable, and J. Gimzewski, "Label-Free Biodetection Using Capacitive Micromachined Ultrasonic Transducers (CMUTs) and Its Application for Cardiovascular Disease Diagnostics". *J Nanomed Nanotechol.* 3. No. 144, 2012.
- [13] B. T. Khuri-Yakub et al., "6D-1 The Capacitive Micromachined Ultrasonic Transducer (CMUT) as a Chem/Bio Sensor," 2007 IEEE Ultrasonics Symposium Proceedings, New York, NY, 2007, pp. 472-475, doi: 10.1109/ULTSYM.2007.127.
- [14] Z. Li, L. Zhao, Z. Jiang, Z. Ye and Y. Zhao, "Capacitive micromachined ultrasonic transducer for ultra-low pressure detection," The 9th IEEE International Conference on Nano/Micro

- Engineered and Molecular Systems (NEMS), Waikiki Beach, HI, 2014, pp. 600-603, doi: 10.1109/NEMS.2014.6908883.
- [15] K. K. Park, H. J. Lee, G. G. Yaralioglu, A. S. Ergun, Ö. Oralkan, M. Kupnik, and J. K. Gimzewski, "Capacitive micromachined ultrasonic transducers for chemical detection in nitrogen". *Applied Physics Letters*, 2007, 91(9), 094102.
- [16] L. Zhao et al., "A Novel CMUT-Based Resonant Biochemical Sensor Using Electrospinning Technology," in *IEEE Transactions on Industrial Electronics*, vol. 66, no. 9, pp. 7356-7365, Sept. 2019, doi: 10.1109/TIE.2018.2878123.
- [17] H. J. Lee, K. K. Park, O. Oralkan, M. Kupnik and B. T. Khuri-Yakub, "CMUT as a chemical sensor for DMMP detection," 2008 IEEE International Frequency Control Symposium, Honolulu, HI, 2008, pp. 434-439, doi: 10.1109/FREQ.2008.4623034.
- [18] H. Wang, X. Wang, C. He, and C. Xue, Design and performance analysis of capacitive micromachined ultrasonic transducer linear array. *Micromachines*, vol. 5, no. 3, pp.420-431, 2014.
- [19] T. Zure, "Characterization of a CMUT Array," Master's Thesis, Dept. of Electrical and Computer Engineering, Univ. of Windsor, Windsor, ON, Canada, 2012.
- [20] V. Yashvanth, "CMUT Crosstalk Reduction Using Crosslinked Silica Aerogel," Master's Thesis, Dept. of Electrical and Computer Engineering, Univ. of Windsor, Windsor, ON Canada, 2018.
- [21] R. Zhang et al., "Design and performance analysis of capacitive micromachined ultrasonic transducer (CMUT) array for underwater imaging." *Microsystem Technologies*, vol. 22, no. 12, pp. 2939-2947, 2016.
- [22] D. S. Ballantine Jr et al., *Acoustic wave sensors: theory, design and physico-chemical applications*. San Diego, CA, USA: Elsevier; 1996.
- [23] M. Rahman and S. Chowdhury, "A new deflection shape function for square membrane CMUT design," in *Proc. of 2010 IEEE International Symposium on Circuits and Systems*, Paris, France, 2010, pp. 2019-2022.
- [24] S. J. Martin, Gregory C. Frye, Antonio J. Ricco, and Stephen D. Senturia, "Effect of surface roughness on the response of thickness-shear mode resonators in liquids," *Analytical Chemistry*, vol. 65, pp. 2910-2922, 1993.
- [25] E. Ashley and J. Niebauer, *Cardiology Explained*. London: Remedica; 2004. Chapter 4, Understanding the echocardiogram. [Online]. Available: <https://www.ncbi.nlm.nih.gov/books/NBK2215/> [retrieved: August, 2021].
- [26] M. Rahman, J. Hernandez, and S. Chowdhury, "An improved analytical method to design CMUTs with square diaphragms," in *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 60, no. 4, pp. 834-845, April 2013.
- [27] R. Manwar, "A BCB Diaphragm Based Adhesive Wafer Bonded CMUT Probe for Biomedical Application," Ph.D. dissertation, Dept. of Electrical and Computer Engineering, Univ. of Windsor, Windsor, ON, Canada, 2017.
- [28] F. Pavlo, "Capacitive micromachined ultrasonic transducer (cMUT) for biometric applications," Master's Thesis, Dept. of Nanoscience and Technology, Chalmers University of Technology, Göteborg, Sweden, 2012.
- [29] R. Manwar, L. Arjunan, M. Ahmadi, and S. Chowdhury, "Resonant frequency calculation of square diaphragms: A comparison," in *Proc of IEEE 6th Latin American Symposium on Circuits & Systems (LASCAS)*, Montevideo, Uruguay, 2015, pp. 1-4.

# The Doer Effect: Replication and Comparison of Correlational and Causal Analyses of Learning

Rachel Van Campenhout & Benny G. Johnson  
 Research and Development  
 VitalSource Technologies  
 Pittsburgh, USA  
 Email: rachel.vancampenhout@vitalsource.com

Jenna A. Olsen  
 Learning Analytics  
 Western Governors University  
 Salt Lake City, USA  
 Email: jennaanneolsen@gmail.com

**Abstract** - The doer effect is a learning science principle that proves students who engage with formative practice at the point of learning have higher learning gains than those who only read expository text or watch video. This principle has been demonstrated through both correlational and causal analysis. It is imperative that learning science approaches capable of increasing student learning gains be rigorously tested and replicated to confirm their validity before wide-scale use. Previously we replicated causal doer effect results using student data from courseware used at a major online university. In this paper, we will replicate both the correlational doer effect analysis as well as the causal analysis using both unit tests from the courseware and the course final exam. These multiple analyses of the doer effect on the same course data provide a unique comparison of this method and the impact of the doer effect on near and intermediate learning assessments. Findings of the correlational doer effect analyses confirmed doing was more significant to outcomes than reading, and further analysis determined these results could not be attributed to student characteristics. Results of the causal analysis verified doing was causal to learning on both the unit tests and final exam. The implications of these doer effect replication results and future research will be discussed.

**Keywords** - doer effect; learn by doing; causal discovery; replication; external validity; learning outcomes; course effectiveness; courseware.

## I. INTRODUCTION

Students deserve digital learning resources that actually help them learn. And yet, verifying which methods are effective for learning is a challenging task. A benefit of courseware as a comprehensive learning environment is the wealth of data available for analysis. As students move through the courseware, their page visits, engagement and accuracy on formative practice, summative assessment scores and more can be collected to paint a picture of what students are doing both in real time and for post hoc analysis. The large-scale data from courseware run in natural settings can be used as a basis for investigating the effectiveness of learning methods. The courseware data can provide many insights, if the right questions are asked. One such question is: Are we able to identify if courseware's formative practice questions cause increased learning?

The doer effect is the learning science principle that the amount of interactive practice a student does (such as answering practice questions) is much more predictive of learning than the amount of reading or video watching the student does [11]. Studies have shown correlational support for this principle [10]. The total amount of reading and total amount of doing are used in a linear regression to identify the doer effect coefficient as a means of quantifying the doer effect. Koedinger et al. [11] found that doing had a median of six times the relationship to learning than reading.

However, in order to recommend this approach with high confidence in its effectiveness, it is necessary to know that there is a causal relationship between doing practice and better learning. This requires ruling out the possibility of a third variable being a common cause of both, since in that case the relationship between doing and learning would merely be correlational. For example, a frequently cited external variable that could account for the doer effect is student motivation. A highly motivated "go-getter" student may do more practice and also obtain better learning outcomes, but this would not necessarily mean better outcomes were *caused* by doing the practice.

Koedinger et al. [10] used data collected from students engaged with a MOOC course paired with courseware developed by Carnegie Mellon's Open Learning Initiative (OLI) to investigate the doer effect. In their initial research, they found the learning effect of doing the formative practice was about six times larger than that of reading. Follow-up analysis [11] [12] sought to determine whether this effect was causal. A statistical design involving within- and outside-unit reading, watching and doing (described in more detail below), was able to demonstrate causal impact of doing on learning and rule out the possibility that this effect was entirely the result of a factor such as individual student motivation. There is no better explanation of the importance of causal relationships than was stated in [11]: "It should be clear that determining causal relationships is important for scientific and practical reasons because causal relationships provide a path toward explanatory theory and a path toward reliable and replicable practical application."

Replication research is critical in the learning sciences to provide additional evidence to support—or refute—claims made about effective learning practices. A large fraction of published research in the social sciences has not been replicated, and studies that cannot be reproduced are cited

more frequently than those that can [16]. Methods for increasing learning should be broadly shared to benefit as many students as possible, and those methods should be grounded in substantial evidence of their validity.

Reproducible research is not only necessary for the research community, but for practical application in educational technology. The courseware analyzed in this study was developed using the methods and approach of learning engineering—a practice that supports learners and their development through the application of the learning sciences to human-centered engineering design methods and data-driven decision making [7]. Proposed by Herbert Simon [17] and fostered at Carnegie Mellon University [6], learning engineering developed as a role to further the application of learning science for students and instructors. Learning engineering was applied at Acrobatiq after its emergence from OLI to apply learning science and a student-centered approach to developing courseware [19]. The Learning Engineering Process (LEP) outlines an iterative cycle that includes the identification of the context and problem, design and instrumentation, implementation, and data analysis and results [8]—a development process appropriate for many contexts. While the application of learning science research was a critical component of the LEP for the development of the courseware, equally vital is the analysis of data and sharing results. To fully engage the LEP is to iteratively improve through the insights data can reveal, and to share these findings with the broader research community. A goal of this paper is to further the LEP by collaborating with an institutional partner to replicate learning science research foundational to the courseware through the analysis of data gathered from students in a natural learning context. By replicating and sharing the data analysis and findings as part of the LEP, the researchers and developers maintain transparency and accountability to the learner [19].

Furthermore, replicating findings that are based on large-scale data mining provides valuable verification of the results, as the volume and type of data analyzed can be difficult to obtain. Through the courseware described in this paper and institutional collaboration, we have the data required to evaluate the relationship between doing practice and learning outcomes. Replicating this causal doer effect study adds to the body of evidence that this learning by doing methodology—and the doer effect it produces—are effective in a variety of learning situations, and supports a practical recommendation that students can increase their learning outcomes by increasing the amount of formative practice they do.

For this study, the data set came from students enrolled in a Macroeconomics course, C719, at Western Governors University. There are many benefits of analyzing student data from courseware used in a real university setting. Students engaged with the course without any external influences that might alter their natural behavior. This allows us to study their engagement and learning outcomes in as authentic a way as possible; students worked through this course as they would any other in their program, which contributes to the generalizable nature of the study. Benefits of utilizing real course data include lower costs and fewer ethical concerns as compared to controlled experiments. A controlled experiment

in a laboratory setting would allow researchers to, for example, deliver the treatment (doing practice interleaved with content) to one randomly selected set of students while delivering static content to a control group. Performance on a standard assessment would provide a measure of the effect of the treatment. This controlled experimental method would have a high internal validity, but would also have a high cost, ethical concerns, and low external validity. Instead, due to the availability of detailed data generated by courseware as students progress through their course, post hoc studies of natural learning contexts can be done with minimal cost and without ethical concerns that can come with randomized experiments, such as withholding potentially beneficial treatment from some learners.

The value of this replication study is that it extends the external validity of the doer effect findings. The Macroeconomics courseware used was designed on the Acrobatiq platform based on the principles established at OLI. This courseware utilizes the same key features of interleaved practice, immediate targeted feedback, etc., as the OLI courses previously analyzed (Introduction to Psychology, Introduction to Biology, Concepts in Computing, Statistical Reasoning) [11]. These similarities are important for confirmatory results, as it is important to have as many common variables as possible for the replication of the statistical model [12]. Investigating courseware in an entirely different subject domain built independently—yet using the same learning science principles—strengthens the external validity of a causal relationship.

This study extends our previous doer effect replication research [1] by replicating Koedinger et al.'s [11] correlational and causal doer effect analysis, using both the unit test summative assessments from the courseware as well as the WGU final exam. This analysis provides a direct comparison of both the correlational and causal analysis on the same course, providing insight into the comparison of outcomes between assessment types. Additionally, demographic information collected by WGU will be used to extend the correlational model—as done in Koedinger et al. [10]—to investigate how additional variables impact the doer effect coefficient.

	Correlational	Causal
Unit Tests	Courseware data	Courseware data
Final Exam	Courseware + WGU data	Courseware + WGU data

Figure 1. The doer effect analyses in this paper.

Given the intention of this study to replicate doer effect findings—both correlational and causal—our research questions are:

1. Can the correlational doer effect be replicated using both courseware unit tests and final exam scores?
2. Can the doer effect be accounted for by student characteristics?
3. Can the causal doer effect be replicated using both courseware unit tests and final exam scores?

To answer these questions, we will outline the required parallel features for this replication study in Section II—from the learning by doing courseware environment, to the description of the regression model and its inputs. Section III will provide the methods, results, and discussion on the correlational doer effect analyses. Section IV will outline the methods, results, and discussion for the causal doer effect analyses. Section V concludes the paper with remarks on the importance of these replication findings for the learning science methods used herein, and the implications of these findings for future research.

## II. STUDY 1: CORRELATIONAL DOER EFFECT REPLICATION

This section will provide the methods, results, and discussion for the correlational doer effect models, using both the unit tests and final exam as the outcome. This section also includes additional analysis of the correlational doer effect ratio when controlling for student characteristics.

### A. Methods

In order for this replication research to be parallel with the original study [11], the learning resource needed to be similar in the learning by doing approach. The term “learning by doing” has been broadly used to describe various kinds of learning engagement (and not all use or encourage the use of scaffolding or feedback [9]), so it is important to clarify how learning by doing is applied in this courseware. Learning by doing is a method of actively engaging the learner in the learning process by providing formative practice at frequent intervals. It has been shown that formative practice increases learning gains for students of all ages and in diverse subjects, and while this method benefits all students, it can benefit low-performing students most of all [4]. The formative practice questions integrated with the content essentially act as no-stakes practice testing, which increases learning gains and retention [5]. In Acrobatiq courseware, students can answer practice questions as many times as they like, and typically students continue to answer until they get the correct answer [20]. Feedback that explains why that choice is correct or incorrect is provided for each answer option to give additional guidance and another opportunity for learning (Figure 2). Immediate, targeted feedback was shown to reduce the time it took students to reach a desired outcome [2] [13], and feedback in practice testing outperforms no-feedback testing [5] [15]. Formative practice with targeted feedback provides scaffolding and examples that support cognitive structures for effective learning [9] [15] [18].

The courseware contains many features similar to those used in the courses for the original study [11]. Modules are made up of lesson pages, and each lesson contains readings, images, and formative practice questions all tied to a central learning objective. Learning objectives are student-centered and measurable, and the practice questions are tagged with the learning objective to feed data to the platform’s learning analytics engine [21], as well as to inform post hoc analysis. The formative practice questions are interleaved with small chunks of content to provide practice to students at the point of learning that content. Question types vary but entail both recognition and recall and most frequently include multiple choice, pull-down, text or numeric input, drag and drop, and true/false. Questions were created to target the foundational Bloom’s Taxonomy category, *remembering*, of which recognition and recall are both cognitive processes [3] [20].

In addition to formative practice questions integrated within the content, there are adaptive activities and summative assessments in the courseware [21]. The adaptive activities are placed at the end of the module and cover all learning objectives included in that module. The questions in the adaptive activities are personalized to the needs of each student at the time they enter the activity. Student performance on the formative practice generates a learning estimate for each learning objective, which is used to determine the scaffolded questions students receive in the activity. The adaptive activities are also formative in nature, so students received immediate, targeted feedback and could make multiple attempts. The questions on these activities are not scored. At the end of the module, students have a quiz—comprehensive to all learning objectives included in that module—that produces a grade. At the end of the unit is a unit test that includes questions on all learning objectives from all modules and also produces a grade. While these unit test grades were not included as part of the course grade (see below), students see their scores in the courseware. These unit tests are used as the summative assessments for the courseware-based doer effect analyses.


By partnering with Western Governors University, the doer effect analysis can also be done using data from the final exam. Students enrolled in the course were able to review the course content (the courseware) and work with faculty at their own pace in preparation for a final exam that comprised 100% of the course grade. Students had a six-month window to complete the course by passing the final exam, which they could retake as needed during that time frame. This learning science-based courseware was developed to fit WGU’s curriculum needs.

Passing the WGU course depended solely on passing the final exam. The courseware content and final exam content were written by independent development teams; however, the course learning objectives were provided to the WGU final exam development team for alignment purposes. For the final exam-based doer effect analyses, the student’s score on the first attempt at the final exam was used as the learning outcome.



## Fiscal Policy and Government

### Learning Objectives

 Explain how government uses the tools of fiscal policy to stabilize the economy.

There are two types of fiscal policy. One is put into place and left to respond automatically to changes in the level of economic activity. These policies are called *automatic stabilizers*. The second is deliberate action to change tax laws or enact new spending programs, so as to influence the level of output, employment, and prices. Even if governments change their levels of spending or taxes for other reasons, policymakers are very conscious of the effects these actions will have on output, employment, and the price level. As discussed in module 7, most economists in the Classical tradition consider fiscal policy to be of limited benefit, sometimes even harmful. Keynesians, however, regard active fiscal policy as a valuable tool for stabilizing economic activity.

Congressional legislation over the years, much of it enacted during the Great Depression, has created a system of tax collections and transfer payments that change automatically in response to changes in national income. These automatic stabilizers partially offset changes in private spending and tend to reduce fluctuations in output and employment. They primarily include changes in income tax collections, Social Security and welfare benefits, and unemployment compensation claims. Because these automatic stabilizers are triggered by changes in the economy, they do not require further action by Congress.

**Did I Get This**

Automatic stabilizers are

- destabilizing, because they tend to make recessions more severe and inflation harder to control.
- put into place and left to respond automatically to changes in the level of economic activity
- deliberate actions to enact new spending programs.
- deliberate actions to change tax laws.

✓ Correct. Automatic stabilizers respond to changes in economic activity without any need for direct action by policymakers.

Discretionary fiscal policies  require further action by Congress.

✓ Correct. Automatic stabilizers happen automatically in response to changing economic conditions.

✗ Incorrect. Discretionary fiscal policies can be large enough to completely offset changes in spending that cause equilibrium income to rise or fall.

Figure 2. A lesson page with formative questions from Macroeconomics.

The model used by Koedinger et al. [11] to determine the doer effect coefficient is an ordinary linear regression that expresses the assessment outcome as a function of total reading and total doing. The ratio of doing and reading coefficients from the model determines the overall doer effect coefficient. Following Koedinger et al. [11], the reading variables were defined as all visits to lesson pages where the student did not engage in any practice available on that page. The doing variables were defined as the number of formative practice opportunities a student attempted, including in the adaptively generated practice activities described earlier. The courseware's module quizzes and unit tests were not included as practice because of their presentation as scored summative assessments, even though in this case they made no contribution to the student's grade in the course; inclusion of these as practice did not materially affect the results of the

analysis. Unlike in some of the previous studies in which video lectures were used [11], video watching was not investigated here, as video was not a critical component of the courseware.

The following analyses use different assessments as the outcome (unit tests or final exam) and therefore the selection criteria for each analysis will have different numbers of students, as depicted in Figure 3. The initial data set included historical data from 3,513 students who enrolled in the Macroeconomics course from March 2017 to April 2019 (WGU courses have rolling enrollments). There are selection criteria relevant to both unit test and final exam analyses. Only students who completed the course (defined here as taking the final exam) were included. As the study we intend to replicate included only students who made some use of the course materials, we likewise excluded students who did not use the

courseware at all. WGU allowed students to take the course’s final exam more than once (if necessary) to pass. Only the first attempt at the final exam was included in the analysis, and student engagement with the courseware was filtered to include only that which occurred before the first attempt at the final exam. This resulted in 3,120 students in the final exam data set. For this group of students, there were 224,072 total reading page visits and 1,143,601 total first attempts on 1,162 available formative practice questions. However, for the courseware unit test analyses, to be consistent with the selection criteria of the original study, only students who completed nearly all of the assessments (in this case at least 5 of 6 unit tests) were included. This resulted in a smaller subset of data that included 493 students.

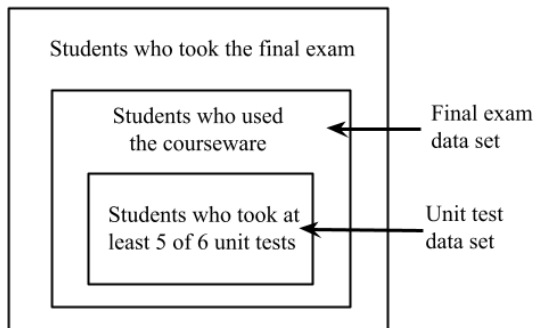


Figure 3. The subsets of student data used for the analysis.

**B. Results**

The first analysis uses data solely from the courseware platform. Using the reading and doing data from the courseware, we can use a linear regression to calculate the doer effect coefficient on the courseware unit tests. The model to compute the doer effect coefficient is replicated from Koedinger et al. [11], so we follow the same procedure for this analysis. The unit test data set of 493 students described above was used. Following [11], since some students did not take all unit tests, the total of the unit test scores was used as the outcome. Only reading and doing relevant to the assessments taken was included. Reading, doing, and score values were converted to z-scores before regression to better enable comparison of the reading and doing effects, since reading and doing are measured in different units (pages visited vs. questions answered). The R formula for the regression model is:

$$\text{lm}(z\_total\_unit\_test\_score \sim z\_total\_reading + z\_total\_doing, \text{data}=df)$$

The results in Table I show that the standardized reading coefficient is not significant ( $p = 0.838$ ) while the standardized doing coefficient is highly significant ( $p < 0.001$ ). The doer effect coefficient is the ratio of doing to reading. Previous work by Koedinger et al. [10] [11] found the effect of doing on outcomes was about six times greater than reading. In cases

TABLE I. UNIT TEST CORRELATIONAL DOER EFFECT REGRESSION ANALYSIS.

	Estimate	Standard Error	t-Value	Pr(> t )
(intercept)	0.0000	0.0403	0.000	1.000
Total Reading	0.0088	0.0429	0.205	0.838
Total Doing	0.4472	0.0429	10.420	<2e-16 ***

where Koedinger et al. [11] could not compute a size for the doer effect because reading was not significant or negatively significant, they reported such cases as an effect ratio of  $\infty$ . In this case, because reading is not significant, the confidence interval of the reading coefficient includes zero. Therefore, the doing coefficient is effectively divided by zero, giving a ratio of  $\infty$ .

As was done with the unit tests, we can use a linear regression to determine the doer effect coefficient with the final exam score as the outcome. There are 3,120 students in this data set, which is considerably more than in the previous analysis of the courseware unit tests due to the change in selection criteria. Because the unit tests in the courseware were not required to pass the course, fewer students completed them, whereas all students completing the course had to take the final exam.

$$\text{lm}(z\_final\_exam\_score \sim z\_total\_reading + z\_total\_doing, \text{data}=df)$$

TABLE II. FINAL EXAM CORRELATIONAL DOER EFFECT REGRESSION ANALYSIS.

	Estimate	Standard Error	t-Value	Pr(> t )
(intercept)	0.0000	0.0170	0.000	1.000
Total Reading	-0.1069	0.0209	-5.105	3.51e-07 ***
Total Doing	0.3655	0.0209	17.450	< 2e-16 ***

The results of this model using the final exam in Table II show that total reading is negative and significant, while total doing is positive and significant. Because total reading was negative, following Koedinger et al. [11] the doer effect in this case is also reported as  $\infty$ .

In their initial research on the doer effect, Koedinger et al. [10] wanted to verify that certain student characteristics weren’t accounting for the doer effect results. They created a linear regression model that accounted for: pretest score, Quiz 1 score, occupation, age, education and gender. The only significant coefficients in the model were the Quiz 1 score and education, with the OLI courseware usage still significant to final exam scores. No other student characteristics were significant.

Through the student data WGU collects, we were able to do a similar analysis to determine if any of the available student characteristics were significant to student outcomes, especially to the extent that they accounted for the doer effect findings. The student characteristics recorded by WGU were gender, underrepresented status, first-generation status, Pell eligible status, and age. These covariates were added to the linear regression model along with total reading and total doing. The R formula used for the unit test as the assessment was:

```
lm(z_total_unit_test_score ~ z_total_reading
+ z_total_doing
+ male + underrep
+ first_gen
+ pell_eligible
+ c_age,
data=df)
```

The results of this linear regression are in Table III. The same model was also fit using the WGU final exam data set. The results of that model are in Table IV.

TABLE III. UNIT TEST CORRELATIONAL DOER EFFECT REGRESSION ANALYSIS INCLUDING STUDENT CHARACTERISTICS.

	Estimate	Standard Error	t-Value	Pr(> t )
(intercept)	0.1017	0.0673	1.513	0.131
Total Reading	0.0511	0.0413	1.237	0.217
Total Doing	0.4355	0.0407	10.710	< 2e-16 ***
Male	0.4525	0.0842	5.373	1.2e-07 ***
Under-represented	-0.4007	0.1030	-3.888	0.000115 ***
First Generation	-0.2246	0.0781	-2.875	0.00422 **
Pell Eligible	-0.1927	0.0816	-2.360	0.0187 *
Age	-0.0029	0.0042	-0.683	0.495

TABLE IV. FINAL EXAM CORRELATIONAL DOER EFFECT REGRESSION ANALYSIS INCLUDING STUDENT CHARACTERISTICS.

	Estimate	Standard Error	t-Value	Pr(> t )
(intercept)	0.0831	0.0296	2.810	0.00498 **
Total Reading	-0.0789	0.0205	-3.843	0.000124 ***
Total Doing	0.3445	0.0203	16.969	< 2e-16 ***
Male	0.3037	0.0349	8.697	< 2e-16 ***
Underrepresented	-0.4214	0.0406	-10.377	< 2e-16 ***
First Generation	-0.1569	0.0338	-4.642	3.59e-06 ***
Pell Eligible	-0.0986	0.0351	-2.857	0.00430 **
Age	0.0051	0.0018	2.764	0.00575 **

### C. Discussion

The doer effect coefficient was  $\infty$  for both the unit test and final exam regression model. The total doing estimate was positive and highly significant for both models. The total

doing estimate was slightly smaller for the final exam compared to the unit test. This may be related to the proximity of the assessment being used as the outcome. Doing practice may have a slightly stronger effect on the outcomes of the unit test immediately following each unit as opposed to the final exam at the end of the coursework.

For the unit tests model, the reading coefficient was not significantly different from zero, and therefore the ratio was  $\infty$ . For the final exam model, the reading coefficient was negative and significant, producing the same result. Whereas the total doing estimate became slightly smaller but still significant from the unit test to the final exam model, the total reading coefficient went from not significant to negatively significant from unit test to the final exam model. This trend is interesting when comparing the results from a near proximal assessment to a distant one, and investigating this trend on additional courses would be a valuable future study.

The correlational doer effect analyses show that doing practice has a much larger effect size on learning outcomes than reading. However, as these are the only two variables evaluated as of yet, there may be a question of whether these results could be due to another variable, such as prior knowledge or demographics. The student characteristics regression models for both unit tests and the final exam show that total doing was still significant even when controlling for student characteristics. Total reading was not significant for the unit tests, and was negative and significant for the final exam score. The total doing and total reading results for the student characteristics models mirror those from the unit test and final exam correlational models. The doer effect coefficient of doing over reading can also be calculated using these student characteristics linear regressions. Just as in the previous unit test correlational model, the reading coefficient was not significant and therefore the coefficient was  $\infty$ . The results of the final exam score model that controls for student characteristics also had a negatively significant reading estimate, so the doer effect ratio is  $\infty$  in this case as well.

The unit test and final exam student characteristics models shared similarities on which covariates were significant. In both models, being male was positively significant. Underrepresented status, Pell eligible status, and first-generation status were all negatively significant in both models. Age was not significant in the unit test model but was on the final exam model. While interpreting the trends in demographic characteristics is outside the scope of this paper, what is key from this analysis is that the doer effect is still present even when controlling for student characteristics. A recent cluster analysis study on data from Western Governors University sought to understand which student characteristics provided the most value for predicting student success [14]. Results found that student activity attributes were more valuable as a stand-alone category, exceeding the value of both student readiness and demographics. Combined with the verification that the doer effect ratio remains unchanged when controlling for student characteristics, this suggests that future research should focus on how to increase engagement with formative practice for all students to maximize the benefits of the doer effect for all.

### III. CAUSAL DOER EFFECT REPLICATION

This section will provide details on the methods, results, and discussion of causal doer effect analyses using both unit tests and the final exam.

#### A. Methods

The correlational doer effect results obtained through the traditionally used ordinary linear regression approach are useful and informative; however, the old maxim "correlation is not causation" still applies. We earlier discussed the importance of whether the relationship between doing and outcomes is in fact causal, not merely correlational. Is it possible to go beyond a correlational model and answer this key question? For this purpose, Koedinger et al. [11] developed a regression model that analyzed the relationship of student doing, reading, and video watching in each unit of course content to scores on that unit's summative assessment. The key innovation in their model was to control for the total amounts of doing, reading, and watching in *other* units of the course. Student doing outside the unit can act as a proxy for a third variable like motivation that can lead to correlation between level of effort and outcomes. In this way, if the doer effect is causal, then the amount of doing within a unit should be predictive of the student's score on that unit's assessment, even when accounting for doing outside that unit. If there is not a causal relationship between doing and outcomes, we would not expect to see a statistically significant within-unit effect beyond the outside-unit effect.

Replicating this causal model using the unit tests from the courseware is a fairly straightforward task. As was done for the correlational models, the reading variable was identified as visits to lesson pages where students did not do practice (when available) and the doing variable was defined as attempts on formative practice. The courseware unit becomes the container for determining within-reading and within-doing, while all pages not in the unit become outside-reading and outside-doing. The unit test is the summative assessment for the unit and serves as the outcome variable for the model. Just like for the correlational model, using the unit tests as the outcome had the additional selection criteria of completing 5 of 6 unit tests, creating a subset of 493 students.

Unlike in the original study, where a summative assessment immediately followed each unit of course content, the final exam was obviously taken after all relevant student usage of the courseware. Furthermore, the units in the *Acrobatiq* courseware did not have a direct correspondence with the categorization of questions on the final exam. As previously discussed, all courseware resources, e.g., lesson readings and formative practice questions, were mapped to the course learning objectives. These learning objectives in turn mapped to six course competencies developed by WGU, to which final exam questions were also coded. The six competencies for the course are:

1. 3003.2.1 – The Economic Way of Thinking - The graduate analyzes economic behavior by applying

fundamental economic principles, including scarcity, opportunity cost, and supply and demand analysis.

2. 3003.2.2 – Macroeconomic Measurements and Theories - The graduate analyzes unemployment, inflation, economic growth, business cycles, and related economic theories.
3. 3003.2.3 – Federal Budget and Fiscal Policy - The graduate explains fiscal policy and its effects on the federal budget, national debt, and economy.
4. 3003.2.4 – Money, Financial Markets, and Monetary Policy - The graduate analyzes the monetary system, including the influence of monetary policy on the economy.
5. 3003.2.5 – Economic Growth and Development - The graduate explains how macroeconomic policies affect economic growth and development.
6. 3003.2.6 – International Trade - The graduate explains how trade policies influence international markets.

In order to apply the Koedinger et al. regression model [11] using the WGU final exam, these course competencies were used as the analysis units, as this provided a way to group both the courseware content and the final exam questions into a common set of logical units. Thus, when referring to a *unit* of course content for the final exam model, we specifically mean all content corresponding to one of these six competencies, with the unit summative assessment consisting of all corresponding final exam questions that assess that competency.

The competencies were used to compile the unit-based reading and doing data required for the model from the clickstream usage events logged by the courseware. Within-unit resource use (reading or doing) was defined as all use associated with a unit's content, and outside-unit resource use was defined as all resource use not designated as within-unit. In total, 47 finer-grained courseware learning objectives were mapped to the six course competencies. The learning objectives were not uniformly distributed across competencies, as the number varied according to the amount of content coverage. The mapping of the courseware's formative practice to the learning objectives was used to aggregate practice by competency.

#### B. Results

Using the unit test data set, we can replicate the causal doer effect model from [11] using the courseware summative assessments as the outcome. The same procedures in preparation for applying that regression model were followed here as well, such as confirming that there is sufficient variation in individual student reading and doing across course units for the analysis [11]. The score distribution for each of the unit tests in the courseware is shown in Figure 4. The boxplots were generated using R's boxplot function, which uses the interquartile range rule to determine outliers, shown as circles.

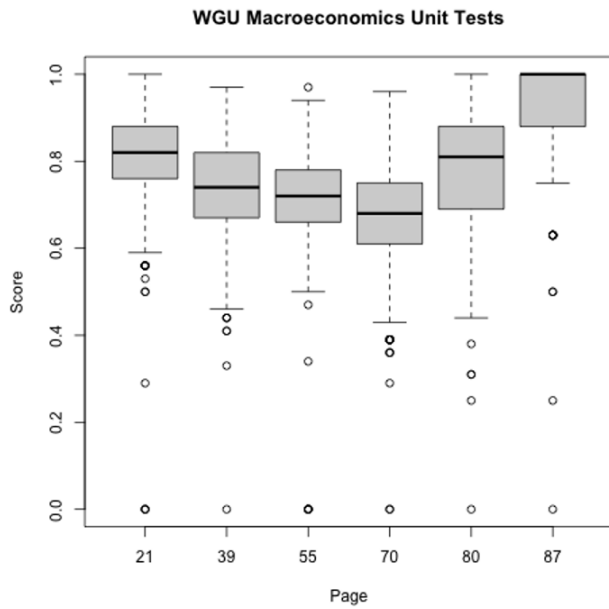


Figure 4. Score distribution for the six Macroeconomics unit tests.

Unlike in the correlational studies, here we have an observation (row) for each individual unit test each student took, giving 2,807 total observations of the 493 students in this data set. The multiple observations per student are not independent and therefore an ordinary linear regression model—which assumes independence—cannot be used. The

lack of independence can be handled by using a mixed effects linear regression model. Following Koedinger et al. [11], we use a mixed effects model to investigate the within-unit and outside-unit reading and doing relationships with learning outcomes. The R formula used to fit the model is below. This shows that a linear mixed effects regression model was fit using the lmer function. The regression formula shows unit test score modeled as a function of within- and outside-unit reading and doing, with a random intercept per student and unit test to address the lack of independence of the observations noted above.

```
lmer(z_unit_test_score ~ z_within_reading
+ z_outside_reading
+ z_within_doing
+ z_outside_doing
+ (1|student)
+ (1|unit_test),
data=df)
```

The reading and doing coefficients were tested for statistical significance using a likelihood ratio test, in which the likelihood of the full model is compared to a model with one of the variables of interest omitted. The R code below illustrates this test for the within-reading coefficient.

The results of the regression model in Table V show that both within- and outside-unit doing are positive and significant at  $p < 0.001$ . Within-unit reading is positive and significant at  $p < 0.05$ ; however, outside-unit reading is negative at  $p < 0.05$ .

```
lme.model <- lmer(z_unit_test_score ~ z_within_reading + z_outside_reading + z_within_doing
+ z_outside_doing + (1|student) + (1|unit_test),
data=df, REML=FALSE)
lme.null <- lmer(z_unit_test_score ~ z_outside_reading + z_within_doing + z_outside_doing
+ (1|student) + (1|unit_test),
data=df, REML=FALSE)
anova(lme.null, lme.model)
```

TABLE V. UNIT TEST CAUSAL DOER EFFECT REGRESSION ANALYSIS.

	Location	Estimate	Std. Error	t-Value	Pr(> t )
	(intercept)	0.0106	0.2778	0.038	0.967
<b>Reading</b>	within-unit	0.0483	0.0198	2.441	0.0146 *
	outside-unit	-0.0559	0.0268	-2.085	0.0370 *
<b>Doing</b>	within-unit	0.1629	0.0276	5.902	5.17e-09 ***
	outside-unit	0.1272	0.0271	4.697	2.70e-06 ***

Next, we can repeat this same causal doer effect analysis using the final exam as the outcome. Prior to creating the linear regression model, we examined the score distribution for each of the six competencies on the final exam, shown in Figure 5. It is seen that the competencies have differing student score distributions.

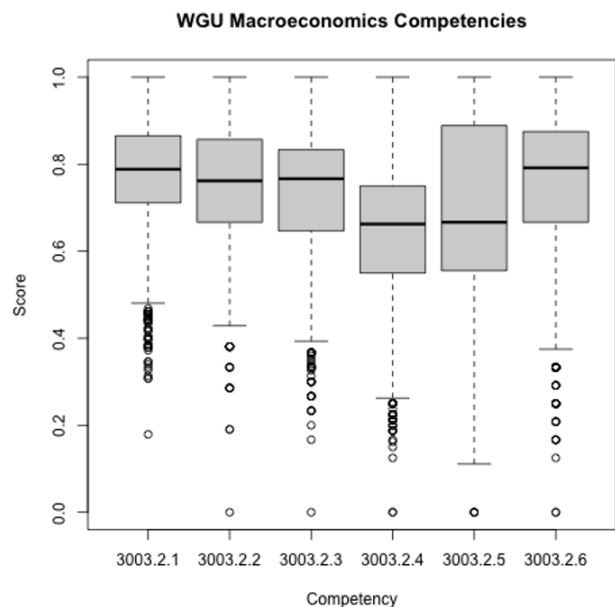


Figure 5. Score distributions for the six competencies of the Macroeconomics final exam.

For each of the 3,120 students in the data set, there is an observation for each of the six competencies, bringing the total number of observations to 18,720. As with the unit test mixed effects regression model, the model was fit with the lmer function, and shows competency score modeled as a function of within- and outside-unit reading and doing.

The R code for this model is:

```
lmer(z_WGU_COMPETENCY_SCORE ~ z_within_reading
+ z_outside_reading
+ z_within_doing
+ z_outside_doing
+ (1|student)
+ (1|competency),
data=df)
```

The reading and doing coefficients were tested for statistical significance using a likelihood ratio test as done for the unit test model.

The results of the regression analysis are presented in Table VI. There are significant effects for within-unit doing, outside-unit doing, and outside-unit reading, while within-unit reading is not significant. The within-unit and outside-unit doing coefficients are larger in magnitude than both the reading coefficients, and doing also had much larger *t*-values than reading. The reading coefficients are also negative, which we will discuss further below.

C. Discussion

The course analyzed in Koedinger et al. [11] had eleven total content unit/assessment pairs. Within-unit doing and watching were significant, as well as outside-unit doing. Reading and outside-unit watching were not significant. Outside-unit doing significance indicates that there is a variable that influences how students who generally do a lot of practice also score higher on assessments. However, the larger and more significant predictor was within-unit doing, meaning that even when controlling for outside-unit doing, within-unit doing had a statistically significant relationship with learning outcomes, indicating a causal doer effect.

For both the unit test and final exam model, both within-unit doing and outside-unit doing were strongly, positively significant. We initially discussed how significant within-unit doing would be indicative of a causal relationship between doing practice and better learning outcomes. But since outside-unit doing is also significant, does that mean that a causal doer effect is *not* supported? No. We would likely

TABLE VI. FINAL EXAM CAUSAL DOER EFFECT REGRESSION ANALYSIS.

	Location	Estimate	Std. Error	t-Value	Pr(> t )
	(intercept)	0.0000	0.1256	0.000	1.000
Reading	within-unit	-0.0125	0.0091	-1.367	0.173
	outside-unit	-0.0604	0.0130	-4.645	3.43e-06 ***
Doing	within-unit	0.1146	0.0099	11.613	< 2.2e-16 ***
	outside-unit	0.1556	0.0132	11.773	< 2.2e-16 ***



expect outside-unit doing to almost always be significant (regardless of whether the doer effect is causal), as it is well known that students who do more practice tend to get better outcomes. Significance of outside-unit doing simply reflects that, for example, students who are go-getters typically do well. What matters is that within-unit doing is *additionally* significant, which means the relationship of within-unit doing to its own unit's assessment score cannot be accounted for by the amount of outside-unit doing, indicating that relationship is causal in nature. Otherwise, we would expect outside-unit doing to be significant but not within-unit doing. But this is not the case: within-unit doing matters to learning outcomes in a way that cannot entirely be explained by a third variable—such as motivation—that leads to both greater doing and better learning. The most important finding is therefore that within-unit doing is a highly significant predictor of learning even after controlling for outside-unit doing, and this is consistent with a causal doer effect. That this finding is consistent between the unit test model and final exam model provides additional confirmation that the doer effect is present on both near transfer assessments (the unit test) as well as medium to far transfer assessments (the final exam).

Within-unit reading in the unit test model had a smaller estimate than doing but was still positively significant, indicating that reading the unit content was beneficial for the unit test. Because within-unit reading was positive and significant, we can use within-unit doing to calculate the doer effect ratio for the unit tests: 3.4. While this is less than the value of 6 typically quoted as the representative doer effect size, Koedinger et al. [11] found doer effect ratios ranging from 2.2 to  $\infty$ . However, when we look at the final exam model, within-unit reading is no longer significant. This could be indicating that the within-unit reading is not beneficial for the far transfer of the final exam assessment when compared to the near transfer of the unit test. Because it is possible within-reading is not statistically different than zero, the doer effect ratio for the final exam is, once again, reported as  $\infty$ .

An interesting note is that the outside-unit reading coefficient was significant but negative on both the unit test and final exam models, showing an overall negative relationship between the amount of outside-unit reading and assessment performance. One possible explanation for this negative result is suggested from prior anecdotal observations of engagement behaviors of students with poor learning outcomes. Many of these students tended to read the same section(s) of text repeatedly, indicating they were struggling. This pattern of rereading without obtaining a good outcome may have contributed to this negative relationship. These struggling students also often did not meaningfully engage in practice, which is regrettable since the body of doer effect research would recommend that investing that study time in practice instead of rereading would have been more beneficial. Note particularly that within-unit reading was not significant, meaning no special relationship to outcomes beyond outside-unit reading was discernible. This negative relationship between reading behavior and outcomes should be a subject of additional future study.

#### IV. CONCLUSION AND FUTURE WORK

It is increasingly critical to utilize methods proven to benefit learners in online learning environments. In this paper, we used the same Macroeconomics course to do four doer effect analyses: replicating the correlational and causal models of Koedinger et al. [11] using both courseware unit tests and final exam scores. By studying the doer effect in this comprehensive manner on a single course, we can confirm correlational and causal findings using different learning outcomes and directly compare results from an assessment that serves as a near transfer of learning with an assessment that serves as an intermediate/far transfer of learning.

Our research question, “Can the correlational doer effect be replicated using both courseware unit tests and final exam scores?” was affirmed. Both linear regression models found that doing was positively significant, while reading was either not significant (unit tests) or negatively significant (final exam). Therefore, the doer effect ratio for both models was  $\infty$ .

Confirming the doer effect correlational analysis, it was also reasonable to ask our second research question: “Can the doer effect be accounted for by student characteristics?” Koedinger et al. [10] also checked to ensure that usage of the OLI courseware was still significant when accounting for student characteristics and found that it was. In this work, using a linear regression model that controlled for student characteristics—gender, underrepresented status, first-generation status, Pell eligible status, and age—we found that doing was still significant. Doing formative practice is still more effective than reading no matter student demographics.

Our research question—“Can the causal doer effect be replicated using both courseware unit tests and final exam scores?”—was positively answered. The courseware unit tests and final exam data produced results consistent with those of the original study. Replicating the findings of Koedinger et al. [11] using courseware designed with the same learning science principles but in a different domain and at a different higher education institution extends the generalizable nature of the doer effect findings. By engaging with a learning by doing design—formative practice questions integrated into the learning material—students activate the doer effect and increase their learning gains. This analysis confirms that even when controlling for an outside variable, doing the formative practice within the courseware caused better performance on both in-course unit tests and an external final exam. Doing practice *causes* better learning.

Some interesting observations can be drawn by comparing results across the different regression models. The standardized doing coefficient was always positive and highly significant, whereas the standardized reading coefficient was not always significant or positive. The doing coefficient also was always much larger than the reading coefficient. Although reading was positively significant and thus allowed computation of a numerical doer effect in only one of the six regressions performed in this work, we can still compare the magnitudes of the doing coefficients themselves across studies.

The doing coefficient was larger for the unit tests than the final exam in the correlational analyses, 0.4472 vs. 0.3655.

This was also the case for the within-doing coefficients in the causal analyses, 0.1629 vs. 0.1146. Although only a single course was studied, this is qualitatively consistent with a priori expectations; the unit tests are more proximal to the formative practice and hence less likely to be affected by learning decay, and alignment with the practice would be expected to be better with in-course summative assessments than an independently developed final exam.

Not only was it found that the doer effect could not be accounted for by student characteristics, inclusion of student characteristics as covariates had minimal impact on the value of the doing coefficient. Controlling for student characteristics changed the doing coefficient by -2.6 % and -5.7% in the unit test and final exam correlational models, respectively.

This work is to our knowledge the first to compare correlational and causal doer effect models on the same course. While not possible to make generalizable observations from a single course, especially quantitatively, it is interesting to note that significance of doing in the correlational model corresponded to significance in the causal model for both the unit tests and final exam. Should it turn out that observing a correlational doer effect generally tends to go along with a causal doer effect, this may be of practical interest because the data needed for a correlational study is simpler and available much more often than the data needed for a causal study. All these trends will be the subject of future investigation planned using a larger sample of courses.

The data available through courseware enable analysis and evaluation of learning principles, such as this one. Through large-scale data collected in a natural learning environment, learning analytics can broaden support for learning science concepts and strategies and provide generalizable results for additional learning contexts. In this particular case, the Macroeconomics courseware provided a comprehensive learning environment for students, but the final exam was what determined the course grade and final student outcome. This use-case may be similar to other higher education institutions where a high-stakes course assessment would take place as a proctored event outside of the learning environment. Identifying the doer effect using a final exam is encouraging because the potential for learning decay is greater than on a more proximal assessment, such as a unit test. What's more, separate development of the learning content and formative practice from the final exam could have made the doer effect more difficult to identify, but that was not the case. The use of a final exam for analysis may also be more typical of a college course where the content and exam are from different authors.

Learning engineering will continue to require not only collaboration of organizations and team members to engage in the LEP, but also the combination of different data sources to investigate learning principles in applied contexts. This study highlights the value of combining data from institutions and educational technology that collects large volumes of raw student data. Analysis for causality required both engagement data from the formative practice in the courseware as well as student learning outcomes from a final exam. As more data become available, combining data from different sources can accomplish valuable analysis of learning methods and principles. The doer effect research was critical to the design

of the courseware environment during the LEP, and this process is furthered by sharing this replication research.

The significance of causal doer effect findings suggests at least two main avenues for future work. The first is to bring the learning by doing method to learning environments at scale, to provide as many students as possible with the learning benefits possible through the doer effect [20]. Doing causes learning, and these findings have been replicated in a variety of subject domains, using learning resources created by different organizations, and implemented at different institutions. The second goal of future work is to use these findings for iterative improvement in the LEP by identifying ways of increasing the amount of practice students do. While variation in the amount of practice students did in the progression of the course was necessary for the statistical models, it would be ideal if every student did effectively all the formative practice available. If doing causes learning, students should engage in as much formative practice as possible to leverage the causal doer effect and maximize its contribution to their learning outcomes. Future work can focus on the role of instructor implementation practice [22] and student motivation in increasing engagement.

#### ACKNOWLEDGMENT

We gratefully acknowledge Bill Jerome and Ken Koedinger for helpful discussions of this work. We also thank Margaret Hsaio for assisting in the preparation for this project.

#### REFERENCES

- [1] R. Van Campenhout, B. G. Johnson, and J. A. Olsen, The Doer Effect: Replicating Findings that Doing Causes Learning. Presented at eLmL 2021 : The Thirteenth International Conference on Mobile, Hybrid, and On-line Learning. ISSN 2308-4367, pp. 1-6, 2021. Retrieved from: [https://www.thinkmind.org/index.php?view=article&articleid=elml\\_2021\\_1\\_10\\_58001](https://www.thinkmind.org/index.php?view=article&articleid=elml_2021_1_10_58001)
- [2] J. R. Anderson, A. T. Corbett, and F. Conrad, "Skill acquisition and the LISP tutor," *Cognitive Science*, vol. 13, pp. 467-506, 1989.
- [3] L. W. Anderson et al. A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives (Complete edition). New York: Longman. (2001).
- [4] P. Black, and D. William, "Inside the black box: raising standards through classroom assessment." *Phi Delta Kappan*, vol. 92(1), pp. 81-90, 2010. <https://doi.org/10.1177/003172171009200119>
- [5] J. Dunlosky, K. Rawson, E. Marsh, M. Nathan, and D. Willingham, "Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology." *Psychological Science in the Public Interest*, vol. 14(1), pp. 4-58, 2013. <https://doi.org/10.1177/1529100612453266>
- [6] J. Goodell, M. Lee, and J. Lis, "What we discovered at the roots of learning engineering." In *IEEE ICICLE Proceedings of the 2019 Conference on Learning Engineering*, Arlington, VA, May 2019.
- [7] IEEE ICICLE. "What is Learning Engineering?" Retrieved 01/11/2021 from: <https://sagroups.ieee.org/icicle/>
- [8] A. Kessler and Design SIG colleagues. *Learning Engineering Process Strong Person*, 2020. Retrieved 01/11/2021 from <https://sagroups.ieee.org/icicle/learning-engineering-process/>
- [9] P. A. Kirschner, J. Sweller, and R. E. Clark, "Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-

- based teaching.” *Educational Psychologist*, vol. 41, pp. 75–86, 2006. [http://doi:10.1207/s15326985ep4102\\_1](http://doi:10.1207/s15326985ep4102_1)
- [10] K. Koedinger, J. Kim, J. Jia, E. McLaughlin, and N. Bier, “Learning is not a spectator sport: doing is better than watching for learning from a MOOC.” In: *Learning at Scale*, pp. 111–120, 2015. Vancouver, Canada. <http://dx.doi.org/10.1145/2724660.2724681>
- [11] K. Koedinger, E. McLaughlin, J. Jia, and N. Bier, “Is the doer effect a causal relationship? How can we tell and why it’s important.” *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge, LAK 2016*, pp. 388–397. <http://dx.doi.org/10.1145/2883851.2883957>
- [12] K. R. Koedinger, R. Scheines, and P. Schaldenbrand, “Is the doer effect robust across multiple data sets?” *Proceedings of the 11th International Conference on Educational Data Mining, EDM 2018*, pp. 369–375.
- [13] M. Lovett, O. Meyer, and C. Thille, “The Open Learning Initiative: Measuring the effectiveness of the OLI statistics course in accelerating student learning,” *Journal of Interactive Media in Education*, vol. 2008(1), pp. 1–16. <http://doi.org/10.5334/2008-14>
- [14] Olsen, J., & Shackelford, S. (2021). Intersectionality and Incremental Value: What combination(s) of student attributes lead to the most effective adaptations of the learning environment? In: Sottolare R., Schwarz J. (eds) *Adaptive Instructional Systems. HCII 2021*. LNCS, vol. 12792. Springer. pp. 577–591. [https://doi.org/10.1007/978-3-030-77857-6\\_41](https://doi.org/10.1007/978-3-030-77857-6_41)
- [15] A. Renkl, R. Stark, H. Gruber, and H. Mandl, “Learning from worked-out examples: the effects of example variability and elicited self-explanations,” *Contemporary Educational Psychology*, vol. 23, pp. 90–108, 1998. <https://doi:10/1006/ceps.1997.0959>
- [16] M. Serra-Garcia, and U. Gneezy, “Nonreplicable publications are cited more than replicable ones,” *In Science Advances*, vol. 7, pp. 1–7, 2021. <http://doi.org/10.1126/sciadv.abd1705>
- [17] H. A. Simon, “The job of a college president,” *Educational Record*, vol. 48, pp. 68–78, 1967.
- [18] J. Sweller, “The worked example effect and human cognition,” *Learning and Instruction*, vol. 16(2), pp. 165–169, 2006. <https://doi.org/10.1016/j.learninstruc.2006.02.005>
- [19] R. Van Campenhout, “Learning engineering as an ethical framework: A case study of adaptive courseware,” In: R. Sottolare, J. Schwarz (eds) *Adaptive Instructional Systems, HCII 2021*, Lecture Notes in Computer Science, vol 12792, pp. 105–119, 2021. Springer, Cham. [https://doi.org/10.1007/978-3-030-77857-6\\_7](https://doi.org/10.1007/978-3-030-77857-6_7)
- [20] R. Van Campenhout, J. S. Dittel, B. Jerome, and B. G. Johnson, “Transforming textbooks into learning by doing environments: an evaluation of textbook-based automatic question generation.” In: *Third Workshop on Intelligent Textbooks at the 22nd International Conference on Artificial Intelligence in Education, 2021*. Retrieved from: [https://intextbooks.science.uu.nl/workshop2021/files/iTextbooks\\_2021\\_paper\\_6.pdf](https://intextbooks.science.uu.nl/workshop2021/files/iTextbooks_2021_paper_6.pdf)
- [21] R. Van Campenhout, B. Jerome, and B. G. Johnson, “The impact of adaptive activities in Acrobatiq courseware: Investigating the efficacy of formative adaptive activities on learning estimates and summative assessment scores,” In: R. Sottolare, J. Schwarz (eds) *Adaptive Instructional Systems, HCII 2020*, LNCS, vol. 12214, 2020. Springer. pp. 543–554. [https://doi.org/10.1007/978-3-030-50788-6\\_40](https://doi.org/10.1007/978-3-030-50788-6_40)
- [22] R. Van Campenhout and M. Kimball, “At the intersection of technology and teaching: The critical role of educators in implementing technology solutions. IICE 2021: The 6<sup>th</sup> IAFOR International Conference on Education.” Retrieved from: <https://papers.iafor.org/submission59028/>