International Journal of Informatics Society

12/15 Vol. 7 No.3 ISSN 1883-4566

Informatics Society

Aims and Scope

The purpose of this journal is to provide an open forum to publish high quality research papers in the areas of informatics and related fields to promote the exchange of research ideas, experiences and results.

Informatics is the systematic study of Information and the application of research methods to study Information systems and services. It deals primarily with human aspects of information, such as its quality and value as a resource. Informatics also referred to as Information science, studies the structure, algorithms, behavior, and interactions of natural and artificial systems that store, process, access and communicate information. It also develops its own conceptual and theoretical foundations and utilizes foundations developed in other fields. The advent of computers, its ubiquity and ease to use has led to the study of informatics that has computational, cognitive and social aspects, including study of the social impact of information technologies.

The characteristic of informatics' context is amalgamation of technologies. For creating an informatics product, it is necessary to integrate many technologies, such as mathematics, linguistics, engineering and other emerging new fields.

Guest Editor's Message

Masashi Saito

Guest Editor of Twenty-first Issue of International Journal of Informatics Society

We are delighted to have the twenty-first issue of the International Journal of Informatics Society (IJIS) published. This issue includes selected papers from the Eighth International Workshop on Informatics (IWIN2014), which was held at Prague, Czech Republic, Sep. 10-12, 2014. The workshop was the eighth event for the Informatics Society, and was intended to bring together researchers and practitioners to share and exchange their experiences, discuss challenges and present original ideas in all aspects of informatics and computer networks. In the workshop 24 papers were presented in five technical sessions. The workshop was successfully finished with precious experiences provided to the participants. It highlighted the latest research results in the area of networking, business systems, education systems, design methodology, groupware and social systems.

Each paper submitted IWIN2014 was reviewed in terms of technical content, scientific rigor, novelty, originality and quality of presentation by at least two reviewers. Through those reviews 15 papers were selected for publication candidates of IJIS Journal, and they were further reviewed as a Journal paper. This volume includes four papers among the accepted papers, which have been improved through the workshop discussion and the reviewers' comments.

We publish the journal in print as well as in an electronic form over the Internet. We hope that the issue would be of interest to many researchers as well as engineers and practitioners over the world.

Masashi Saito received B.E degree from Tokyo Institute of Technology in 1983, M. E. degree from Cornell University in 1992 and Ph.D degree from Osaka University in 2006. In 1983, he joined Mitsubishi Electric Corporation and has developed engineering workstations, Internet TVs, cellular phones, car navigation systems especially for operating system extension, Internet services and distributed processing. Since 2014, he has been a director of IPSJ (Information Processing Society of Japan) and a manager of SIG ITS (Intelligent Transportation Systems and Smart Community). He is a professor of Kanazawa Institute of Technology since 2015. His current research interests include ITS, smart city, distributed systems, smart systems. He is a member of IEEE, ACM and IPSJ.

[Invited Paper] Formal Methods for Mobile Robots: Current Results and Open Problems

Béatrice Bérard[†#], Pierre Courtieu[‡], Laure Millet[†#], Maria Potop-Butucaru[†#], Lionel Rieg[*], Nathalie Sznajder[†#], Sébastien Tixeuil[†#‖], and Xavier Urbain[⋆]

[†]UPMC Sorbonne Universités, LIP6-CNRS UMR 7606, France
[#] CNRS, LIP6-CNRS UMR 7606, France
[‡]CNAM, Cédric EA 4629, France
[*] Collège de France, France
[‖] Institut Universitaire de France, France
[⋆] ENSIIE, Cédric EA4629 and LRI CNRS UMR 8623, France

Abstract - Mobile robot networks emerged in the past few years as a promising distributed computing model. Existing work in the literature typically ensures the correctness of mobile robot protocols via *ad hoc* handwritten proofs, which are both cumbersome and error-prone.

This paper surveys state-of-the-art results about applying formal methods approaches (namely, model-checking, program synthesis, and proof assistants) to the context of mobile robot networks. Those methods already proved useful for bug-hunting in published literature, designing correct-by-design optimal protocols, and certifying impossibility results. We also present related open questions to further develop this path of research.

Keywords: Formal Methods, Mobile Robots, Distributed Algorithms, Model Checking, Program Synthesis, Proof Certification, Proof Assistant.

1 INTRODUCTION

The variety of tasks that can be performed by autonomous robots and their complexity are both increasing [1], [2]. Many applications envision groups of mobile robots that are self-organising and cooperating toward the resolution of common objectives, in the absence of any central coordinating authority.

The seminal model introduced by Suzuki and Yamashita [3] features a distributed system of k mobile robots that have limited capabilities: they are identical and *anonymous* (they execute the same algorithm and they cannot be distinguished using their appearance), they are *oblivious* (they have no memory of their past actions) and they have neither a common sense of direction, nor a common *handedness* (chirality). Furthermore these robots do not communicate by sending or receiving messages. However they have the ability to sense the environment and see all positions of the other robots.

Robots operate in cycles of three phases: *Look*, *Compute* and *Move*. During the *Look* phase robots take a snapshot of the positions of the other robots (in their own coordinate system). The collected information is used in the *Compute* phase where robots decide to move or to stay idle. In the *Move* phase, robots may move according to the computation of the previous phase.

In the original model [3], some non-empty subset of robots execute the three phases synchronously and atomically, giving rise to two variants: FSYNC, for the fully-synchronous model where all robots are scheduled at each step to execute a full cycle, and SSYNC, for the semi-synchronous model, where a strict subset of robots can be scheduled. This model had a huge impact on the community and was instrumental in deriving many new core problems in the area of distributed mobile entities. It was later generalised by Flocchini *et al.* [4] to handle full asynchrony and remove atomicity constraints (this model is called ASYNC [1], for asynchronous, in the sequel). One of the key differences between the fully- or semi-synchronous models, and the asynchronous model in the discrete setting is that in the ASYNC model, a robot can compute its next move based on an *outdated* view of the system. It is notorious that handwritten proofs for protocols operating in the ASYNC model are hard to write and read, due to many instances of case-based reasoning that is both cumbersome and error-prone.

Outline. The goal of the survey is to present recent advances in using formal methods for mobile robots following the model of Suzuki and Yamashita and derivatives. Formal methods are needed to certify that obtained results are correct, as previously published solutions were in fact incorrect.

We consider in this paper three main proposals in the domain of formal methods: model-checking, algorithm synthesis, and proof assistants.

The model itself may seem limited (robots have extremely few capabilities, compared to real life robots), but it permits to establish fundamental results (what are the tasks that are feasible, and what are those which are not). That is, it is a computability-centric model (as opposed to a efficiency-centric model).

In Section 2, we recall basic notions on model-checking, synthesis and games and proof assistants. We also briefly describe previous work on formal methods applied to robot algorithms. We present in Section 3.1 an overview of the various settings, as well as the formal models used in the sequel. In Section 4, we survey results in the three directions of model-checking, synthesis and proof-assistants. We conclude in Section 5 with several problems open for future research.

2 PRELIMINARIES

2.1 Model Cheking

Model-checking [5], [6] is an appealing technique that was developed for the verification of various models: finite ones but also in some cases infinite, parameterised, or even timed and probabilistic models. It has been successfully used for the verification of distributed systems from classical shared memory (consensus, transactional memory) to population protocols [7]–[12]. Unfortunately, it was proved in [13] that parameterised model checking is undecidable, and this general result was followed by several stronger ones for specific models, for instance in [14]. In such cases, a classical line of work consisted in combining model-checking with other techniques like abstraction, induction, etc., as first proposed in [15] or [16]. These techniques were largely used since, for instance in [17]–[20]. Although the problem is still open, we conjecture that parameterised model checking is undecidable for the robot model which leads to follow combined approaches.

2.2 Games and Protocols Synthesis

In the formal methods community, automatically synthesising programs that would be correct by design is a problem that raised interest early [21]–[24]. Actually, this problem goes back to Church [25], [26]. When the program to generate is intended to work in an open system, maintaining an on-going interaction with a (partially) unknown environment, it is known since [26] that seeing the problem as a *game* between the system and the environment is a successful approach. The system and its environment are considered as opposite players that play a game on some graph, the winning condition being the specification the system should fulfill whatever the environment behavior. Then, the classical problem in game theory of determining winning strategies for the players is equivalent to find how the system should act in any situation, in order to always satisfy its specification. The case of mobile autonomous robots that we focus on in this paper falls in this category of problems: the robots may evolve (possibly indefinitely) on a ring, making decisions based on some global state of the system at each time instant. The vertices of graph on which the players will play would then be some representation of the different global positions of the robots on the ring. The presence of an opposite player (or environment) is motivated by the absence of chirality of the robots: when a robot is on an axis of symmetry, it is unable to distinguish its two sides one from another, hence to choose exactly *where* it moves ; this decision is supposed to be taken by the opposite player.

2.3 Certification and Proof Assistants

Mechanical proof assistants are proof management systems in which a user can express data, programs, theorems and proofs. In sharp contrast with automated provers, they are mostly interactive, and thus require some kind of expertise from their users. Skeptical proof assistants provide an additional guarantee by checking mechanically the soundness of proofs *after* it has been interactively developped.

Various proof assistants emerged since the 60's, to name a few: Agda [27], NqThm [28] and its relative ACL2 [29], PVS [30], Mizar [31], Coq [32], Isabelle/HOL [33], etc.

In the context of program verification, Isabelle/HOL and Coq are amongst the most widely used; both are based on type theory. They have been successfully employed for various tasks such as the formalisation of programming language semantics [34], certification of an OS kernel [35], verification of cryptographic protocols [36], certification of RSA keys [37], mathematical developments as involved as the 4-colours threorem [38], the Feit-Thompson theorem [39], or the Kepler Conjecture [40].

During the last twenty years, the use of tool-assisted verification has extended to the validation of distributed processes.

In the context of process algebras, which can be used to describe and verify algorithms built from merge, sequential composition and encapsulation, Fokkink [41] and Bezem *et al.* [42] use a proof assistant to prove the equality between two processes, one of them being a specification.

TLA/TLAPS [43], [44] can enjoy an Isabelle back-end for its provers [45]. Gascard and Pierre [46] focus on interconnection networks that are symmetric: rings, tori, hypercubes. Based on a compositional approach of certified components, their work makes use of Nqthm.

Cansell and Méry's contribution to the RIMEL project [47] addresses the class of local computation (LC) algorithms. A catalogue of case studies like election algorithms, spanning tree construction and even Mazurkiewicz's enumeration algorithm have been developed in Event-B. The code of these algorithms is obtained by successive refinements starting from an abstract machine that translates directly to a specification. This code is *annotated* with logical formulas — mainly invariants on the state of the system — the proofs of which generate verification conditions through a calculus of weakest preconditions.

Küfner *et al.* [48] propose a methodology to develop (using Isabelle) proofs of properties of fault-tolerant distributed algorithms in an asynchronous message passing style setting. They focus on correctness proofs only.

Chou's methodology [49] is based on the HOL proof assistant. It aims at proving properties of concrete distributed algorithms through simulation with abstract ones. The methodology does not allow to prove impossibility results.

Castéran *et al.* [50] use Coq to state and prove invariants but also generic results about subclasses of LC systems, thanks to Castéran and Filou's library Loco [51]. Genericity is worth emphasising here as the approach is not limited to *particular instances* of algorithms. Castéran *et al.* actually propose proofs of negative results in Coq for some kinds of distributed algorithms in this graph relabelling setting.

Deng and Monin [52] use Coq to prove the correctness of distributed self-stabilising protocols in the population protocol model. This model permits to describe interactions of an arbitrary large size of mobile entities, however the considered entities lack movement control and geometric awareness that are characteristic of robot networks.

As a matter of fact, surprisingly few works consider using mechanised assistance for networks of *mobile entities*.

2.4 Previous Attempts for Mobile Robots

To our knowledge, in the context of mobile robots operating in discrete space, only two previous attempts, by Devismes *et al.* [53] and by Bonnet *et al.* [54], [55], investigate the possibility of automated verification of mobile robots protocols. The first paper uses LUSTRE [56] to describe and verify the problem of exploration with stop of a 3×3 grid by 3 robots in the SSYNC model, and to show by exhaustive searching that no such protocol can exist. The second paper considers the perpetual exclusive exploration by k robots of n-sized rings, and generates mechanically all *unambiguous* protocols for k and n in the SSYNC model (that is, all protocols that do *not* have symmetrical configurations). Those two works are restricted to the simpler SSYNC model rather than the more general and more complex ASYNC model. Second, they are either specific to a hard-coded topology (*e.g.*, a 3×3 grid [53]) that prevents easy reuse in more generic situations, or make additional assumptions about configurations and protocols to be verified (*e.g.* unambiguous protocols [54], [55]) that prevent combinatorial explosion but forbid reuse for proof-challenging protocols, which would most benefit from automatic verification.

3 FORMAL MODELLING

This section reviews the classical model for mobile robots that is due to Suzuki and Yamashita [3] (Section 3.1), then surveys formal modelling schemes that are tailored for model-cheking (Section 2.1), protocol synthesis (Section 3.3), and proof assistants (Section 3.4).

3.1 A Model for Mobile Robot Networks

Robots We consider a set of k mobile entities called robots, that are endowed with sensing, computing, and moving capabilities. They can observe (sense) the positions of other robots in the space they evolve in and based on these observations, they perform some local computations that can drive them to move to other locations.

Sensing Robots are usually endowed with visibility sensors that permit them to obtain the location of other robots. The obtained location is either *fine grained* (which usually denotes an arbitrary degree of precision in the such obtained location) or *coarse grained* (robots can only be observed at some specific discrete locations, each location being adjacent to at least one another). In the first case, the literature mostly refers to the *continuous space* model, while in the latter case, it is the *discrete space* model.

In some problem instances, robots may share the same position, which is called a *multiplicity point* or a *tower*. The ability for a robot to detect multiplicity is crucial to solve some particular tasks. We distinguish *weak* and *strong* multiplicity detection. The weak multiplicity detector detects whether there is zero, one or more than one robot at a particular location. The strong multiplicity detector senses the exact number of robots at a particular location. The multiplicity detector may be local or global. The *local* detector returns information only at the current position of the robot, while the *global* multiplicity detector return information about all observed positions.

A third characteristic of robot sensing capabilities is their visibility radius. It can be *infinite* (that is a robot is able to sense the position of all other robots) or finite. In the latter case, there exists a bound (that can be expressed either by a distance – in case of continuous space, or by a number of hops – in case of discrete space) beyond which a robot cannot sense anything.

It should be understood that all the sensing performed by robots are presented in the robot's own ego-centric coordinate system (that is called the local coordinate system in the sequel). The local coordinate systems of robots is not necessarily the same for all robots with respect to *origin* (the local coordinate systems are self-centered), *direction* (all robots need not agree on a common vertical north direction), and *chirality* (robots may have different sensing of left and right).

Computing As in classical distributed systems, robots are assumed to be able to perform the computing steps in negligible time.

Robots may be *oblivious* in the sense that they do not remember previously executed steps. Hence, volatile memory can be used to perform computing tasks in a single Look-Compute-Move loop, but the contents of the memory used in the computation are erased before the next loop occurs. By contrast, robots may have non-volatile memory: in this case they are *non-oblivious*.

Moving Robots may move only to the location computed in the computing phase of the current loop. In some instances, due to symmetry, the computed location may be ambiguous. To model this case, it is assumed that the actual move is decided by an adversary (also called demon, or scheduler). The demon can be viewed as an opponent in the game context. In the discrete space model, a robot may move only to a location that is adjacent to its current location. In the continuous space model, a robot moves toward its computed destination. With the *rigid* assumption, a move always perfoms to completion (that is, the robot is never interrupted). In the original model, a robot may be interrupted by an adversary before it finishes its move, but not before it has moved at least a minimum distance $\delta > 0$, where δ is a parameter of the model (unknown to the robots).

Atomicity There are two main models for atomicity. The historical model is the atomic model, where look-compute-move loops are executed in a lock-step fashion. In particular, in the atomic model, the robots that are selected for execution all sense at the same time, all compute at the same time, all move at the same time. In the current terminology [1], the atomic model is either referred to as the FSYNC (in the case where *all* robots execute at the same time) or as the SSYNC (in the case where a non-empty subset of the robots execute at the same time) model.

A less constrained model is the asynchronous and non-atomic model (or ASYNC in the current terminology [1]),

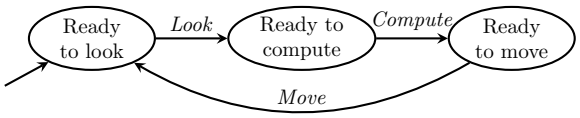


Figure 1: A generic automaton for the robot behaviour.

where robots look-compute-move loops are completely non-atomic and can each last an arbitrary period of time. In particular, in the ASYNC model, it is possible for a robot to observe another robot while it moves, or to perform the computing (and moving) phase with an observation that is long outdated. Of course, all executions in the atomic model are also valid in the ASYNC model. Thus, impossibility results for the atomic model extend in the ASYNC model, and protocols for the ASYNC model are also valid for the atomic model, but the converse is not true.

Demons Demons are an abstraction to characterise the degree of asynchrony in the robot network [57]. Demons can be seen as a predicate on system executions, that is, only executions matching the demon predicate can appear in a given context. The larger the set of executions in the predicate is, the more powerfull (and more general) the demon is. The most general demon in the context of mobile robots is the *fair* demon, which guarantes that in any configuration, any robot is activated within a finite number of steps. If the demon is k-*fair*, then between any two actions of a particular robot, any other robot is activated at most k times. Finally, the *synchronous* demon activates all robots all the time, always.

Faults Robots usually operate without failures (in which case they are said to be *correct*). Yet, some unexpected behaviours may occur. In the worst case, robots are *Byzantine*, meaning that they can behave arbitrarily. Note that to have an impact on the others, the only part of the misbehaviour to take into account is the move part. A less serious fault is the *crash* fault, where a robot unexpectedly stops moving forever.

3.2 A Formal Model for Robots on Graphs

In this section we describe the model proposed by Bérard *et al.* for the robots (in Section 3.2.1), the demons (in Section 3.2.2), and the system resulting from their interactions (in Section 3.2.3). This model encompasses all three FSYNC, SSYNC, and ASYNC operating modes, but assumes that individual robots can only operate in a discrete setting (that is, a graph).

3.2.1 Robot Modelling

All robots execute the same algorithm [1], hence the behaviour of each of them can be described by the finite automaton of Fig. 1. They operate in *Look*, *Compute*, and *Move cycles*.

To start a cycle, a robot takes a snapshot of its environment, which is represented by the *Look* transition. Then, it computes its future location, represented by the *Compute* transition. Finally the robot moves along an edge of the graph according to its previous computation, this effective movement is represented by the *Move* transition.

The algorithm is implemented in the *Compute* transition, hence the "Ready to move" state is divided into as many parts as there are possible movements according to the protocol under study.

Note that the original model [3] abstracts the precise time constraints (like the computational power or the locomotion speed of robots) and keeps only sequences of instantaneous actions, assuming that each robot completes each cycle in finite time. This model can be reduced by combining the *Look* and *Compute* phases to obtain the *LC* phase. This is simply done by merging the two states "Ready to look" and "Ready to compute" into a single state "Ready to Look-Compute".

3.2.2 Demon Modelling

Unlike robots that have the same behaviour regardless of the model, the demon is parameterised by the execution model and by the number of robots. It is also modelled by a finite automaton, one for each variant of the execution model. By synchronising one of these demons with robot automata, we obtain an automaton that represents the global behaviour of robots in the chosen model.

To describe these demon models, we consider a set $Rob = \{r_1, \ldots, r_k\}$ of robots. We denote by LC_i,$Move_i$ the respective *LC* and *Move* phases of robot r_i. Note that LC_i and $Move_i$ are actually sets of possible actions in the corresponding phases. For a subset $Sched \subseteq Rob$, we denote the synchronisation of all LC_i (resp. $Move_i$) actions of all robots in *Sched* by $\prod_{r_i \in Sched} LC_i$ (resp. $\prod_{r_i \in Sched} Move_i$).

In the SSYNC model, a non-empty subset of robots is scheduled for execution at every phase, and operations are executed synchronously. In this case, the automaton is a cycle, where a set $Sched \subseteq Rob$ is first chosen. In this cycle the *LC* and *Move* phases are synchronised for this set of robots. A generic automaton for SSYNC is described in Fig. 2(a). Actually, the "*Sched* chosen" state has to be divided into 2^k states, where k is the number of robots, in order to represent all possible sets *Sched*.

The FSYNC model is a particular instance of the SSYNC model, where all robots are scheduled for execution at every phase, and operate synchronously thereafter: In each global cycle, $Sched = Rob$, hence all global cycles are identical.

The ASYNC model is totally asynchronous: any finite delay may elapse between *LC* and *Move* phases. During each phase a set *Sched* is chosen, and all robots in this set execute an action: the action Act_i is either in LC_i or in $Move_i$ depending on the current state of robot r_i. Hence, a robot can move according to an outdated observation. The automaton for this demon is depicted in Fig. 2(b).

3.2.3 System Modelling

To describe the global model, we denote by $Pos = \{0, \ldots, n-1\} \subseteq \mathbb{N}$ the set of positions on the graph. A configuration of the system is a mapping $c : Rob \to Pos$ associating with each

(a) SSYNC model.

(b) ASYNC model.

Figure 2: The Demons automata.

robot r its position $c(r) \in \mathit{Pos}$. Hence, in a graph of n nodes with k robots, there are n^k possible configurations.

The model of the system is an automaton

$$M = (S, s_0, A, T)$$

obtained by the synchronised product of k robot automata and all the possible configurations, as defined above, the demon is used to define the synchronisation function. The alphabet of actions is $A = \prod_{r_i \in \mathit{Rob}} A_i$, with $A_i = \mathit{LC}_i \cup \mathit{Move}_i$ for each robot r_i. In this product, states are of the form $s = (s_1, \ldots, s_k, c)$ where s_i is the local state of robot r_i, and c the configuration. An initial state is of the form $s_0 = (s_{1,0}, \ldots, s_{k,0}, c)$ where $s_{i,0}$ is the initial local state of robot r_i and c is an arbitrary configuration.

A transition of the system is labelled by a tuple

$$a = (a_1, \ldots, a_k)$$

where $a_i \in A_i \cup \{\varepsilon, -\}$ for all $1 \leq i \leq k$ and

$$(s_1, \ldots, s_k, c) \xrightarrow{a} (s'_1, \ldots, s'_k, c')$$

if and only if for all i, $s_i \xrightarrow{a_i} s'_i$, and c' is obtained from c by updating the positions of all robots such that $a_i \in \mathit{Move}_i$. To represent the scheduling, we denote by $\prod_{r_i \in \mathit{Sched}} \mathit{Act}_i$ the action $(a_1, \ldots, a_k)$ such that $a_i = -$ if $r_i \notin \mathit{Sched}$ and $a_i \in \mathit{LC}_i \cup \mathit{Move}_i \cup \{\varepsilon\}$ otherwise.

3.3 Protocol Synthesis and Reachability Games

To enable robot protocol synthesis (that is, the automatic generation of robot protocols for a given problem in a given setting), the approach of Millet *et al.* [58] is to reuse the modelling presented in Section 3.1 for robots, schedulers, and their interactions, and to revisit reachability games in this context.

We now present classical notions on this subject. If A is a set of symbols, A^* is the set of finite sequences of elements of A (also called *words*), and A^ω the set of infinite such sequences, with ε the empty sequence. We note $A^+ = A^* \setminus \{\varepsilon\}$, and $A^\infty = A^* \cup A^\omega$. For a sequence $w \in A^\infty$, we denote its *length* by $|w|$. If $w \in A^*$, $|w|$ is equal to its number of elements. If $w \in A^\omega$, $|w| = \infty$. For all words $w = a_1 \cdots a_k \in A^*$, $w' = a'_1 \cdots \in A^\infty$, we define the *concatenation* of w and w' by the word noted $w \cdot w' = a_1 \cdots a_k a'_1 \cdots$. We sometimes omit the symbol and simply write ww'. If $L \subseteq A^*$ and $L' \subseteq A^\infty$, we define $L \cdot L' = \{w \cdot w' \mid w \in L, w' \in L'\}$.

A game is composed of an *arena* and *winning conditions*.

Arena An arena is a graph $\mathcal{A} = (V, E)$ in which the set of vertices $V = V_p \uplus V_o$ is partitioned into V_p, the vertices of the protagonist, and V_o the vertices of the opponent. The set of edges $E \subseteq V \times V$ allows to define the set of successors of some given vertex v, noted $vE = \{v' \in V \mid (v, v') \in E\}$. In the following, we only consider finite arenas.

Plays To play on an arena, a token is positioned on an initial vertex. Then the token is moved by the players from one vertex to one of its successors. Each player can move the token only if it is on one of her own vertices. Formally, a play is a path in the graph, i.e., a finite or infinite sequence of vertices $\pi = v_0 v_1 \cdots \in V^\infty$, where for all $0 < i < |\pi|$, $v_i \in v_{i-1}E$. Moreover, a play is finite only if the token has been taken to a position without any successor (where it is impossible to continue the game): if π is finite with $|\pi| = n$, then $v_{n-1}E = \emptyset$.

Strategies A strategy for the protagonist determines where she brings the token whenever it is her turn to play. To do so, the player takes into account the history of the play, and the current vertex. Formally, a strategy for the protagonist is a (partial) function $\sigma : V^* \cdot V_p \to V$ such that, for all sequence (representing the current history) $w \in V^*$, all $v \in V_p$, $\sigma(w \cdot v) \in vE$ (i.e. the move is possible with respect to the arena). A strategy σ is *memoryless* if it does not depend on the history. Formally, it means that for all $w, w' \in V^*$, for all $v \in V_p$, $\sigma(w \cdot v) = \sigma(w' \cdot v)$. In that case, we may simply see the strategy as a function $\sigma : V_p \to V$.

Given a strategy σ for the protagonist, a play $\pi = v_0 v_1 \cdots \in V^\infty$ is said to be *σ-consistent* if for all $0 < i < |\pi|$, if $v_{i-1} \in V_p$, then $v_i = \sigma(v_0 \cdots v_{i-1})$. Given an initial vertex v_0, the *outcome* of a strategy σ is the set of plays starting in v_0 that are σ-consistent. Formally, given an arena $A = (V, E)$, an intial vertex v_0 and a strategy $\sigma : V^* V_p \to V$, we let

$$\mathit{Outcome}(A, v_0, \sigma) = \left\{ v_0\pi \in V^\infty \mid v_0\pi \text{ is a play and is } \sigma\text{-consistent} \right\}$$

Winning conditions, winning plays, and winning strategies We define the *winning condition* for the protagonist as a subset of the plays $\mathit{Win} \subseteq V^\infty$. Then, a play π is *winning* for the protagonist if $\pi \in \mathit{Win}$. In this work, we focus on the simple case of reachability games: the winning condition is then expressed according to a subset of vertices $T \subseteq V$ by $\mathit{Reach}(T) = \{\pi = v_0 v_1 \cdots \in V^\infty \mid \exists 0 \leq i < |\pi| : v_i \in T\}$. This means that the protagonist wins a play whenever the token is brought on a vertex belonging to the set T. Once it has

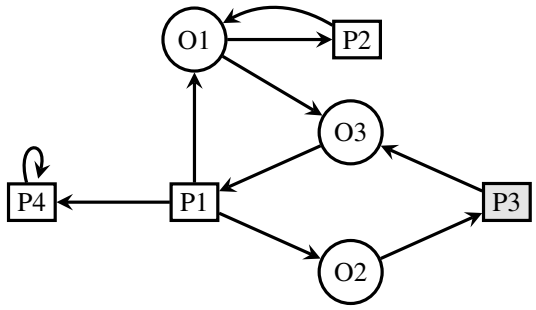


Figure 3: A two-player game. In this figure protagonist vertices are represented by rectangles and antagonist vertices by circles. The winning condition is $Reach(\{P3\})$. Any path in the graph is a play. From P2 the protagonist has no winning strategy. From P1 a (memoryless) winning strategy is to go to O2. Winning positions are $\{P1, P3\}$.

happened, the play is winning, regardless of the following actions of the players.

Given an arena $\mathcal{A} = (V, E)$, an initial vertex $v_0 \in V$ and a winning condition Win, a *winning strategy* σ for the protagonist is a strategy such that any σ-consistent play is winning. In other words, a strategy σ is winning if $Outcome(\mathcal{A}, v_0, \sigma) \subseteq Win$. The protagonist wins the game $(\mathcal{A}, v_0, Win)$ if she has a winning strategy for $(\mathcal{A}, v_0, Win)$. We say that σ is winning on a subset $U \subseteq V$ if it is winning starting from any vertex in U: if $Outcome(\mathcal{A}, v_0, \sigma) \subseteq Win$ for all $v_0 \in U$. A subset $U \subseteq V$ of the vertices is *winning* if there exists a strategy σ that is winning on U.

Solving a reachability game Given an arena $\mathcal{A} = (V, E)$, a subset $T \subseteq V$, one wants to determine the set $U \subseteq V$ of winning positions for the protagonist, and a strategy $\sigma : V^*V_p \to V$ for the protagonist, that is winning on U for $Reach(T)$.

Figure 3 represents a reachability 2-player game. We recall now a well-known result [59] on reachability games:

Theorem 1 *The set of winning positions for the protagonist in a reachability game can be computed in linear time in the size of the arena. Moreover, from any position, the protagonist has a winning strategy if and only if she has a memoryless winning strategy.*

3.4 A Formal Model with Coq for Robots in Continuous Spaces

In this section, we survey the modelling in COQ that was introduced by Auger *et al.* [60], [61] and by Courtieu *et al.* [62]. This model enables to deal with FSYNC and SSYNC execution models in a two-dimensional Euclidian space setting (where coordinates are modeled by real numbers), but assumes the *rigid* model of movement, where move phases always complete.

Since there is a wide variety of different assumptions, the model must be highly flexible. The higher-order expressiveness of proof assistants allows many aspect of the model to remain abstract. In a particular setting, one may instantiate carefully the abstract parts with concrete definitions corresponding to the assumptions under consideration. We provide such examples of particular instances in the following.

The formal framework is parameterised by the following: (1) The number of correct and Byzantine robots. (2) The topological space in which robots move, i.e. the type of *locations* (infinite line, discrete grid, discrete ring network, etc). (3) The observing capabilities of robots, i.e. what kind of *spectrum* do they receive from their sensors. This is where anonymity and multiplicity assumptions are specified for example. (4) The distributed protocol running on each non-Byzantine robot, which we call the *robogram*. (5) The execution model (FSYNC, etc.) and the degree of fairness under consideration. Proof of distributed systems are supposed to state properties for any execution, i.e. for any infinite sequence of successive activations of robots that obeys the assumptions under consideration (fairness, etc.). Traditionally, such an infinite sequence is called a *demon*. Characterising the authorised executions through the definition of a given demon is one of the crucial step of instantiating our framework on a particular setting.

3.4.1 Robots

We consider the union of two given disjoint finite sets of (robot) *identifiers*: G referring to robots that behave correctly, and B referring to the set of Byzantine ones. Note that at this level, in order to express any kind of properties about programs, all robots can be identified. The behaviour of correct and byzantine robots is defined later.

```
Variable nG nB: nat.
(* Number of good and byz. robots.
Left abstract *)
Definition G := Finite nG.
(* Type of good robots *)
Definition B := Finite nB.
(* Type of byzantine robots *)
Inductive ident := Good: G → ident
| Byz: B → ident.
(* Disjoint union *)
```

In some cases the assumptions require that local robograms cannot tell robots apart (anonymity), or detect whether they are correct or Byzantine. This restriction of the model can be ensured by the notion of *spectrum*, described below, which characterises what a robot can see of the global position.

Locations, Positions, Similarities Robots are distributed in space, at places called *locations*. *Positions* are functions from the set of identifiers to the space of locations. The space of location is left abstract in the model, it can be instantiated by any type: the infinite line $\mathbb{Q}$ [60], [61] or $\mathbb{R}$ [62], the plan $\mathbb{R} \times \mathbb{R}$, a ring network $\mathbb{Z}/n\mathbb{Z}$, a line network $[i, j]$, etc.

`gp` denotes a position for correct robots, and `bp` a position for Byzantine ones. The position of all robots is then given by the combination `gp` $\uplus$ `bp` defined by a record in COQ.

```
Variable location : Type.
(* Space occupied by robots. Left abstract. *)
```

```
Record position := {
gp: G → location ;
bp: B → location
}.
```

Spectrum Generally speaking, robots compute their target position from the configuration they perceive of their siblings in the considered space. Depending on assumptions (*e.g.* anonymity, multiplicity detection, etc.) the observation may be more or less accurate. To allow for different assumptions to be studied, we leave the type *spectrum*, together with the notion of spectrum of a position, abstract.

```
Variable spectrum : Type.
Variable spectrum_of : position → spectrum.
```

In the following we distinguish a *demon* position (resp. spectrum), that is expressed in the global frame of reference (viewed from nominal position, orientation and zoom), from a *robot* position (resp. spectrum), that is expressed in the robot's frame of reference. At each step of the distributed protocol (see definition of `round` below) the demon position and spectrum are transformed (i.e., recentered, rotated and scaled) into the considered robots ones before being given as parameters to robograms. Depending on the assumptions under consideration, the zoom and rotation factors may be fixed for each robot or chosen by the demon at each step. They may also be shared by all robots or not, etc.

Example 1 *In a framework where anonymity holds and where robots do not enjoy multiplicity detection, one can define a spectrum as a set of robot locations (each element of the set is a location occupied by at least one robot), and* `spectrum_of` *as a function returning the set of locations occurring in its parameter* `p`.

```
Definition location := R.
Definition spectrum := set location.
Definition spectrum_of p : spectrum
    := collect_locations p.
```

Notice that a spectrum being a set *in this example, it masks the number of robots occupying the same location, thus ensuring that multiplicity is undetected. To account for multiplicity, one may define another instance where spectra are* multisets, *and* `collect_locations` *keeps record of redundant locations.*

Robogram The robogram is a function computing a target location from a spectrum.

```
Definition robogram := spectrum → location.
```

More precisely it computes the target location from the robot spectrum, that is: expressed in the robot's own frame of reference.

3.4.2 Demonic action and round

Assuming the SSYNC model, at each round the demon selects the new location of byzantine robots, the set of correct robots to be activated, and a frame for each of them. More precisely the frame is a way to change the frame of reference. Depending on the space the robots move in, it can be for instance rotation and scale factors. The type of *demonic action* is left abstract in the model but it should provide all these operations.

Example 1 (continued) *We continue on the previous example where we suppose the set of locations to be the infinite real line. The frame can be expressed by a real number as follows: the absolute value denotes the scaling with reference to the demon's point of view, a negative number means that the position is rotated (in this case: swapped), and the special* 0 *value means that the robot is actually not activated at this round.*

```
Record demonic_action := {
  locate_byz  : B → location ;
  frame       : G → R }
```

From these definitions we can formalise what it is for the distributed algorithm to perform a *round*. In an SSYNC context, a round consists in the computation of the new position of correct robots (i.e. a function of type `G →location`) from a robogram, a demonic action and the previous position. The function `round` defined below is thus a function returning a function. For each robot `g` it computes its new location by feeding the robogram with the spectrum recentered and distorted by the demon.

```
Definition round (r : robogram)
(da : demonic_action) (gp : G → location) :
G → location :=
fun g:G ⇒
  let l := gp g in
  (* current location of g *)
  let k := da.(frame) g in
  (* scale and rotation factor for g *)
  if k = 0 then l
  (* g not activated, g stays at l *)
  else
    let pos := repos gp da.(locate_byz) k l in
    (* position viewed from g *)
    let newloc := r (spectrum_of pos) in
    (* apply r on g's spectrum *)
    l + /k * newloc.
    (* Uncenter, unscale, unrotate *)
```

Where `repos gp bp k l` returns the `l`-centered, `k`-zoomed and rotated version of position `{|gp;bp|}`.

Demon, Fairness An actual *demon* is simply an infinite sequence (stream) of demonic actions, that is a coinductive object [63]. Coinductive types are of invaluable help to express in a direct way infinite behaviours, infinite datatypes and properties on them. The Coq proof assistant provides means for the developer to define *and to quantify* over both inductive and coinductive types, so as to express inductive and coinductive properties. Roughly, coinduction is used for properties that hold forever, and induction for properties that hold eventually.

```
CoInductive demon :=
  NextDemon : demonic_action → demon → demon.
```

The set of authorised demons also depends on the assumptions under consideration. For example, we define below the well-known notion of being a *fair* demon by a coinductive property over demons, which state that at each step of the demon any robot is activated after a finite number of steps.

```
Inductive LocallyFairForOne g (d : demon) :
  Prop :=
  | ImmediatelyFair :
      frame (demon_head d) g ≠ 0
            → LocallyFairForOne g d
  | LaterFair :
      frame (demon_head d) g = 0
            → LocallyFairForOne g
                        (demon_tail d)
            → LocallyFairForOne g d.

CoInductive Fair (d : demon) :
  Prop :=
  AlwaysFair :
    (∀ g, LocallyFairForOne g d)
               → Fair (demon_tail d)
               → Fair d.
```

Some of those definitions may be shortened, but this is a rather direct and generic way to express that, at each point of an *infinite* execution, a property holds *eventually*.

4 SURVEY OF RESULTS

Making use of the formal modelling presented in the previous section, recent papers were able to use formal methods to verify existing algorihtms (Section 4.1), synthesise new algorithms that are correct by design (Section 4.2), and provide certified impossibility results (Section 4.3). In this section, we review the main contributions published so far.

4.1 Model Checking

The model checking approach of Bérard *et al.* was used for studying the *Min-Algorithm* presented by Blin *et al.* [64]. The followed approach was to outline the properties that need to be satisfied for the particular problem of perpetual exclusive exploration, using LTL logic.

Problem specification The *Exclusive Perpetual Exploration* problem in [64] is defined in the general asynchronous model as follows.

For any graph G of size n and any initial configuration where robots are located on different vertices, an algorithm solves the perpetual exclusive exploration problem if it guarantees two properties: the *exclusivity* property and the *liveness* property. The first one requires that no two robots visit the same vertex or traverse the same edge at the same time, whereas the *liveness* property requires that each robot visits each vertex infinitely often.

In the considered models an execution where no robot is ever scheduled can happen, as well as an execution where a particular robot is never scheduled. To prevent such executions a fairness assumption has to be added: All robots have to be scheduled infinitely often. Thus the *liveness* property is satisfied only on executions where the *fairness* assumption holds.

Min-Algorithm In [64] the authors proposed an algorithm called *Min-Algorithm*, for $k = 3$ robots in a ring of size $n \geq 10$, such that n is not a multiple of 3. Starting from tower-free configurations (where no two robots occupy the same position), this algorithm ensures exclusive and perpetual exploration. It is based on a classification of tower-free configurations and a specific action to be taken by the robot in any recognized configuration. An equivalence class of tower-free configurations on the ring is described by a sequence of symbols R and F, indexed by integers: R_i stands for i consecutive nodes occupied by a robot, and F_j stands for j consecutive nodes free of robots. The algorithm is presented in Tables 1 and 2.

Verification The previous algorithm was modeled then implemented into the DiVinE [65] model-checker, using a ring of size 10, the smallest advertized size for the algorithm to work. The algorithm was verified to work properly in the FSYNC and SSYNC model, but a counter-example was found when run using the ASYNC model, among the 13.10^6 possible movements. This counter-example ends up in two robots colliding (and thus breaking the exclusion property), as explicited in Fig. 4.

In this counter example every ring represents a configuration, a configuration change occurs when a robot moves, in each configuration a computation is represented by a full arrow, and outdated computation by a dotted arrow.

Following the verification, a simple fix on the rule

$$RC5 \ :: \ (R_2, F_1, R_1, F_z) \ \rightarrow \ (R_1, F_1, R_1, F_1, R_1, F_{z-1})$$

allowed to correct the algorithm.

4.2 Algorithm Synthesis

The algorithm synthesis approach of Millet *et al.* [58] was used to produce a mobile robot protocol for the gathering problem in a ring shaped network. The followed approach was to encode a particular arena for the gathering task, and later use the UPPAAL TIGA tool to generate a winning strategy that can be developed into an algorithm.

Problem specification The *Gathering* problem is defined in the general asynchronous model as follows.

For any graph G of size n and any initial configuration, an algorithm solves the gathering problem if it guarantees that in any execution, all robots meet at the same vertex (not known beforehand) and remain there infinitely thereafter. Similarly as in the previous section, all robots have to be scheduled infinitely often.

4.2.1 Arena Encoding for Gathering

The authors construct an arena so that the player has a winning strategy if and only if a mobile robot algorithm permits robots to gather at a particular node independentely of the

Table 1: Rules for the legitimate phase of *Min-Algorithm*.

Legitimate Phase: $z \neq \{0,1,2,3,4\}$			
$RL1$::	(R_2, F_2, R_1, F_z)	$\rightarrow$	$(R_1, F_1, R_1, F_2, R_1, F_{z-1})$
$RL2$::	$(R_1, F_1, R_1, F_2, R_1, F_z)$	$\rightarrow$	(R_2, F_3, R_1, F_z)
$RL3$::	(R_2, F_3, R_1, F_z)	$\rightarrow$	(R_2, F_2, R_1, F_{z+1})

Table 2: Rules for the convergence phase of *Min-Algorithm*.

Convergence Phase:				
$RC1$::	(R_2, F_y, R_1, F_z)	$\rightarrow$	$(R_2, F_{min(y,z)}, R_1, F_{max(y,z)+1})$	avec $y \neq z \neq \{1,2,3\}$
$RC2$::	$(R_1, F_x, R_1, F_y, R_1, F_y)$	$\rightarrow$	$(R_1, F_x, R_1, F_{y-1}, R_1, F_{y+1})$	avec $x \neq y \neq 0$
$RC3$::	$(R_1, F_x, R_1, F_y, R_1, F_z)$	$\rightarrow$	$(R_1, F_{x-1}, R_1, F_{y+1}, R_1, F_z)$	avec $x < y < z$
$RC4$::	(R_3, F_z)	$\rightarrow$	(R_2, F_1, R_1, F_{z-1})	if a single robot moves
		$\rightarrow$	$(R_1, F_1, R_1, F_1, R_1, F_{z-2})$	if two robots move
$RC5$::	(R_2, F_1, R_1, F_z)	$\rightarrow$	(R_2, F_2, R_1, F_{z-1})	

initial configuration. In each configuration, the robots can choose among the following actions: $\Delta = \{\curvearrowright, \curvearrowleft, \uparrow, ?\}$, which contains $\mathbb{M} = \{\curvearrowright, \curvearrowleft, \uparrow\}$, the set of possible movements, and "?", used by disoriented robots indicating their will to move, yet unability to decide the exact direction of movement (*e.g.* due to symmetry). We note $\overline{\curvearrowright} = \curvearrowleft$, $\overline{\curvearrowleft} = \curvearrowright$, $\overline{\uparrow} = \uparrow$ and $\overline{?} = ?$.

The arena is $\mathcal{A}_{\text{gather}} = (V_p \uplus V_o, E)$, with $V_p = (\mathcal{C}/\equiv)$ denoting the player states, and $V_o = \mathcal{C} \times (\Delta^k)$ denoting the environment states. The size of the arena is then linear in n and exponential in k. The arcs in the arena are defined by relation E as a strict alternating sequence of states between the two players: $E \subseteq (V_p \times V_o) \cup (V_o \times V_p)$.

From a player state, the player chooses for each robot a move. There is the additional constraint that in any equivalence class, two robots with the same view take the same decision (the robot algorithm is deterministic).

A *decision function* f is a function that proposes a move based on a robot view. It is defined by $f : \mathcal{V} \rightarrow \Delta$ such that, for any view $V \in \mathcal{V}$, if $|V| = 1$ then $f(V) \in \{\uparrow, ?\}$, and if $f(V) = ?$ then $|V| = 1$ (a disoriented robot can only choose to move or not to move). When a decision function is run, the robots moves must be coherent with a global sense of orientation. Since $C = (d_1, \ldots, d_k) \in \mathcal{C}$, and $f : \mathcal{V} \rightarrow \Delta$, for any $1 \leq i \leq k$, we define $f(C, i) = f(view_i(C))$ if $(d_i, \cdots, d_k, d_1, \cdots d_{i-1})$ is the smallest element of $view_i(C)$ (in lexicographic order), and $f(C, i) = \overline{f(view_i(C))}$ otherwise.

Then, for every $v \in V_p, v' \in V_o$, $(v, v') \in E$ *iff* there exists a decision function f such that $v' = \big(C, (a_1, \ldots, a_k)\big)$ with $C = rep(v) = (d_1, \ldots, d_k)$, and for every $1 \leq i \leq k$, $a_i = f(C, i)$.

The game then continues from an enviroment position where the previous choices of the player are remembered. If a disoriented robot has deciced to move, the environment chooses the move to be performed by the robot among $\{\curvearrowright, \curvearrowleft\}$.

In $v' = (C, (a_1, \ldots, a_k)) \in V_o$, a set of movements

$$(m_i)_{i \in \{1,\ldots,k\}} \in \mathbb{M}^k$$

is v'-*compatible* if: for every $1 \leq i \leq k$ such that $a_i \neq ?$, $m_i = a_i$, and for every $1 \leq i \leq k$ such that $a_i = ?$, $m_i \neq \uparrow$.

Getting from an environment state to a player state is then expressed as: for every $v \in V_p$, $v' = (C, (a_1, \ldots, a_k)) \in V_o$, $(v', v) \in E$ *iff* there exists a tuple that is v'-compatible $(m_i)_{i \in \{1,\ldots,k\}}$ and such that $v = [C \oplus (m_i)_{i \in \{1,\ldots,k\}}]_{\equiv}$.

Theorem 2 *The winning position for the player in the*

$$(\mathcal{A}_{gather}, W)$$

game corresponds exactly to the gatherable configurations.

4.2.2 Synthesis of a Gathering Algorithm for Three Robots

The aforementioned arena permits to synthesise a deterministic protocol for the gathering problem of k robots in a n-sized ring. Let $T = [(-1, \cdots, -1, n-1)]_{\equiv} \in V_p$ be the equivalence class of all configurations where all robots are gahtered at a single node. Millet *et al.* [58] implemented the arena for three robots and various ring sizes ($n \in [3, 15]$ et $n = 100$) using the game resolution tool UPPAAL TIGA [66]. It was possible to confirm the impossibility of gathering from a starting configuration that is periodic, and possibility of gathering otherwise (that is, there exists a winning strategy in those remaining cases).

To obtain optimal strategies (with respect to the overall number of movements), one can use weighted arcs in the arena depending on the number of moving robots on that arc.

Figure 5 presents classes of configurations (satisfying some constraint φ), and the strategy found in this class (in the "Strategy" column). The "Robot Algorithm" column presents the corresponding robot algorithm executed by Robot r when its view $view_r$ satisfies φ. For all other views, the robot algorithm is $\uparrow$. This algorithm is correct by construction for $n \in [3, 15]$ and $n = 100$. An induction proof is given in [58], extending the results to any ring size n.

4.3 Certification of Impossibility Results

So far the aforementioned formalism proved to be useful and with a (relative) ease of use to certify impossibility results regarding oblivious and anonymous mobile robots [62], even when one allows for byzantine behaviours [60], [61].

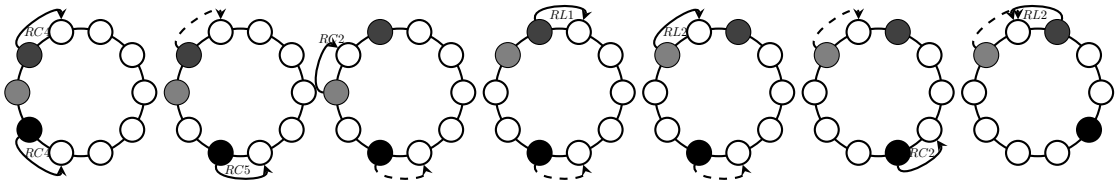


Figure 4: Counter example.

Strategy		
V_p	φ	V_o
$[(-1,-1,n-1)]_{\equiv}$		$((-1,-1,n-1),(\uparrow,\uparrow,\uparrow))$
$[(-1,d_1,d_2)]_{\equiv}$	$d_1 < d_2$	$((-1,d_1,d_2),(\uparrow,\uparrow,\curvearrowleft))$
$[(-1,\frac{n-1}{2},\frac{n-1}{2})]_{\equiv}$		$((-1,\frac{n-1}{2},\frac{n-1}{2}),(\uparrow,\uparrow,?))$
$[(d_1,d_1,d_2)]_{\equiv}$	$rep = (d_1,d_1,d_2)$	$((d_1,d_1,d_2),(\curvearrowright,\uparrow,\curvearrowleft))$
$[(d_1,d_1,d_2)]_{\equiv}$	$rep = (d_2,d_1,d_1)$	$((d_2,d_1,d_1),(\uparrow,\uparrow,?))$
$[(d_1,d_2,d_3)]_{\equiv}$	$d_1 < d_2 < d_3$	$(\curvearrowright,\uparrow,\uparrow))$

Robot Algorithm		
$view_r$	φ	$f : view_r$
$\{(d_1,-1,d_2),(d_2,-1,d_1)\}$	$d_1 < d_2$	$\curvearrowright$
$\{(\frac{n-1}{2},-1,\frac{n-1}{2})\}$		$?$
$\{(d_1,d_1,d_2),(d_2,d_1,d_1)\}$	$d_1 < d_2$	$\curvearrowright$
$\{(d_1,d_2,d_1)\}$	$d_2 < d_1$	$?$
$\{(d_2,d_1,d_3),(d_3,d_1,d_2)\}$	$d_1 < d_2 < d_3$	$\curvearrowright$

Figure 5: A Strategy and its corresponding algorithm.

From the point of view of the person who specifies the model and the properties, the theorems are stated in a natural way: mostly by quantifying over relevant demons, protocols (robograms), and concluding with a negation of the solution characterisation.

For instance, the impossibility of gathering for an even number of oblivious and anonymous mobile entities moving along $\mathbb{R}$ in [62] is simply expressed as follows:

```
Theorem noGathering :
  ∀ (G : finite) (r : robogram (G ⊎ G)),
  inhabited G
  → ∀ k : nat, (1 <= k)
    → ¬ (∀ d, kFair k d → solGathering r d)
```

It can be read as "for every finite set G that is non-empty, for every robogram r distributed over twice the cardinal of G robots (thus an even number), for every fairness constraint k, there is a k-fair demon d for which r fails to gather all robots".

Its proof amounts to showing that for a non-null even number of robots, any `k` and any robogram `r` there exists a k-fair demon that prevents r to gather all robots.

From the point view of the person with proof-assistant expertise who develops the actual (interactive) proof, the size of the development is reasonably short as it makes a fair use of the provided libraries. The size of the specialised development for the relevant notions and the aforementioned theorems (thus excluding for example the complete library for reals) is approximately 480 lines of specifications and 430 lines only of proofs. The file dedicated to the theorem itself is about 200 lines of specifications for 250 lines of proof scripts. This is a good indication on how adequate the framework is, as proofs are not too intricate and remain human readable.

Proving negative results has been emphasised here, yet it is worth noticing that this approach is not limited to impossibility results. Indeed, protocols can also be proved correct using this formal development, as it is easy to write an actual program within the language of Coq, a functional language. The statements are then of the form: for all demons, for any number of robots and initial positions that fulfill some constraint, the given robogram is a solution to some problem.

5 OPEN PROBLEMS

We surveyed recent results that make use of formal methods in the context of mobile robot networks. Model checking and algorithm synthesis were used in the discrete space model to find errors in existing literature (and possibly relieve protocol designers from the burden of manually checking small instances of the problem, thus permitting them to concentrate on abstract configurations where some global invariants hold) and general protocols that are correct by design in this context, while proof assistant was used to devise general impossibility results in the continuous space model. Many open challenges remain, we list a few of them in the sequel, hoping to pave the way for future research.

Arbitrary Sized Networks The main limitations implied by the model-checking and algorihtm synthesis approaches is that the space where robots evolve is *bounded*. That is, the number of robots k and the size of the ring n are given as parameter to generate the possible configuration. This permits to keep the modelling of the system simple, and to enumerate all possible situations. Getting generic results for any size n still requires a handmade approach, taking the mechanically verified instances as a base case for human generated induction. Mechanising the second part (*e.g.* with Coq or another similar tool) is a promising path.

Discrete *vs.* Continuous Space Going from the discrete space to the continuous space is another challenge (in the case of model-cheking and algorithm synthesis). Then, it becomes impossible to enumerate all possible configurations of robots, yet a completely different modelling of the configurations (*e.g.* based on some geometric invariants observed by the robots) could lead to limiting their classes to a tractable number. However, in this case, none of the presented approaches so far can be reused.

On the positive side, thanks to the abstraction level of the Pactole framework [67], setting the space to be $\mathbb{R}$, thus both unbounded and continuous, is not as complicated as one could imagine; it emphasises the relevance of a formal proof approach and how it is complementary to other formal verification techniques.

Atomic *vs.* Non-atomic Executions For the algorithm synthesis and proof assistant, we focused on the atomic FSYNC and SSYNC models. Breaking the atomicity of the individual Look-Compute-Move cycles (that is, considering automatic algorithm production for the ASYNC model [1], or writing impossibility results that are specific to that model) implies that robots cannot maintain a current global view of the system (their own view may be outdated), nor be aware of the view of other robots (that may be outdated as well). Then, the two-players game encoding of Millet *et al.* [58] is not feasible anymore. A natural approach would be to use distributed games, but they are generally undecidable as previously stated. So, a completely new approach is required for the automatic generation of non-atomic mobile robot algorithms.

The modelling of ASYNC is feasible in a proof assistant, and should not bring any additional difficulties in the specification of properties in that context. However, it would currently have a significant cost in terms of intricacy of the associated proofs. A really manageable formal development in an ASYNC model requires more automation at the proof level.

Toward Weaker Requirements A noteworthy added benefit of the Coq abstract framework is that keeping the abstractions as general as possible may lead to relaxing premises of theorems, thus potentially discovering new results (*e.g.* formalising weaker demons [57] and weaker forms of Byzantine behaviours could lead to stronger impossibility results).

Toward New Robotic Problems Solved While the modelling in the discrete space approaches is generic, the encoding of the problem has to be specific (LTL logic for model checking, identifying the winning configurations in the algorithm synthesis approach). The Coq approach remains generic with respect to the algorithm thank to its higher-order logic capabilities, however the suitability of the approach to obtain positive results (that is, certified algorihtms solving a praticular problem) has not been demonstrated yet on practical examples. This issue remains challenging as expertise is required to design the proper encoding in each formal model. Facilitating this step for algorithm designer is a long term research goal.

REFERENCES

[1] P. Flocchini, G. Prencipe, and N. Santoro, Distributed Computing by Oblivious Mobile Robots, Morgan & Claypool Publishers (2012).

[2] M. Potop-Butucaru, M. Raynal, and S. Tixeuil, "Distributed computing with mobile robots: An introductory survey," Network-Based Information Systems (NBiS), 2011 14th International Conference on, pp.318–324 (2011).

[3] I. Suzuki and M. Yamashita, "Distributed anonymous mobile robots: Formation of geometric patterns," SIAM Journal on Computing, pp.1347–1363 (1999).

[4] P. Flocchini, G. Prencipe, N. Santoro, and P. Widmayer, "Gathering of asynchronous robots with limited visibility," Theoretical Computer Science, pp.147–168 (2005).

[5] E. Clarke, O. Grumberg, and D. Peled, Model Checking, MIT Press (1999).

[6] C. Baier and J.P. Katoen, Principles of model checking, MIT press (2008).

[7] R. Guerraoui, T.A. Henzinger, and V. Singh, "Model checking transactional memories," Distributed Computing, pp.129–145 (2010).

[8] I. Chatzigiannakis, O. Michail, and P.G. Spirakis, "Algorithmic verification of population protocols," Stabilization, Safety, and Security of Distributed Systems, pp.221–235, Springer Berlin Heidelberg (2010).

[9] J. Clément, C. Delporte-Gallet, H. Fauconnier, and M. Sighireanu, "Guidelines for the verification of population protocols," Distributed Computing Systems, pp.215–224, IEEE (2011).

[10] T. Lu, S. Merz, and C. Weidenbach, "Towards verification of the pastry protocol using tla$^+$," Formal Techniques for Distributed Systems, pp.244–258, Springer Berlin Heidelberg (2011).

[11] T. Tsuchiya and A. Schiper, "Verification of consensus algorithms using satisfiability solving," Distributed Computing, pp.341–358 (2011).

[12] M.Z. Kwiatkowska, G. Norman, and D. Parker, "Probabilistic verification of herman's self-stabilisation algorithm," Formal Asp. Comput., Vol.24, No.4-6, pp.661–670 (2012).

[13] K.R. Apt and D. Kozen, "Limits for automatic verification of finite-state concurrent systems," Inf. Process. Lett., Vol.22, No.6, pp.307–309 (1986).

[14] J. Esparza, A. Finkel, and R. Mayr, "On the verification of broadcast protocols," 14th Annual Symp. on Logic in Computer Science, pp.352–359, IEEE (1999).

[15] Z. Manna and A. Pnueli, "Temporal verification diagrams," Proc.of Int. Conf. on Theoretical Aspects of Computer Software (TACS'94), LNCS, Vol.789, pp.726–765, Springer (1994).

[16] E.M. Clarke, O. Grumberg, and S. Jha, "Veryfying parameterized networks using abstraction and regular languages," Proc. of 6th Int. Conf. on Concurrency Theory (CONCUR'95), LNCS, Vol.962, pp.395–407, Springer (1995).

[17] N. Bjørner, A. Browne, E.Y. Chang, M. Colón, A. Kapur, Z. Manna, H. Sipma, and T.E. Uribe, "Step: Deductive-algorithmic verification of reactive and real-time systems," Proc. of 8th Int. Conf. on Computer Aided Verification (CAV'96), LNCS, Vol.1102, pp.415–418, Springer (1996).

[18] L. de Alfaro, Z. Manna, H.B. Sipma, and T.E. Uribe, "Visual verification of reactive systems," Proc. of 3d Int. Workshop on Tools and Algorithms for Construc-

tion and Analysis of Systems (TACAS'97), LNCS, Vol.1217, pp.334–350, Springer (1997).
[19] D. Cansell, D. Méry, and S. Merz, "Diagram refinements for the design of reactive systems," J. Univ. Comp. Sci., Vol.7, No.2, pp.159–174 (2001).
[20] T. Arons, A. Pnueli, S. Ruah, J. Xu, and L.D. Zuck, "Parameterized verification with automatically computed inductive assertions," Proc. of 13th Int. Conf. on Computer Aided Verification (CAV'01), LNCS, Vol.2102, pp.221–234, Springer (2001).
[21] E.M. Clarke and E.A. Emerson, "Design and synthesis of synchronization skeletons using branching time temporal logic," Proc. ofIBM Workshop on Logics of Programs (1981).
[22] Z. Manna and P. Wolper, "Synthesis of communicating processes from temporal logic specifications," ACM Trans. Program. Lang. Syst., Vol.6, No.1, pp.68–93 (1984).
[23] M. Abadi, L. Lamport, and P. Wolper, "Realizable and unrealizable specifications of reactive systems," Proc. ofICALP'89, LNCS, Vol.372, pp.1–17, Springer (1989).
[24] A. Pnueli and R. Rosner, "On the synthesis of a reactive module," Proc. ofPOPL'89, pp.179–190, ACM (1989).
[25] A. Church, "Logic, arithmetics, and automata," Proc. ofInt. Congr. of Mathematicians, pp.23–35 (1963).
[26] J.R. Büchi and L.H. Landweber, "Solving sequential conditions by finite-state strategies," Trans. Amer. Math. Soc., Vol.138, pp.295–311 (1969).
[27] Agda, http://wiki.portal.chalmers.se/agda/pmwiki.php.
[28] R.S. Boyer and J.S. Moore, A Computational Logic Handbook, Academic Press (1988).
[29] ACL2, http://www.cs.utexas.edu/users/moore/acl2/.
[30] PVS, http://pvs.csl.sri.com/.
[31] Mizar, http://mizar.uwb.edu.pl/.
[32] Coq, https://coq.inria.fr/.
[33] T. Nipkow, L.C. Paulson, and M. Wenzel, Isabelle/HOL — A Proof Assistant for Higher-Order Logic, Lecture Notes in Computer Science, Vol.2283, Springer-Verlag (2002).
[34] X. Leroy, "A Formally Verified Compiler Back-End," Journal of Automated Reasoning, Vol.43, No.4, pp.363–446 (2009).
[35] G. Klein, J. Andronick, K. Elphinstone, G. Heiser, D. Cock, P. Derrin, D. Elkaduwe, K. Engelhardt, R. Kolanski, M. Norrish, T. Sewell, H. Tuch, and S. Winwood, "seL4: Formal verification of an operating system kernel," Communications of the ACM, Vol.53, No.6, pp.107–115 (2010).
[36] J.B. Almeida, M. Barbosa, E. Bangerter, G. Barthe, S. Krenn, and S.Z. Béguelin, "Full Proof Cryptography: Verifiable Compilation of Efficient Zero-Knowledge Protocols," ACM Conference on Computer and Communications Security, ed. T. Yu, G. Danezis, and V.D. Gligor, pp.488–500, ACM (2012).
[37] L. Théry and G. Hanrot, "Primality Proving with Elliptic Curves," 20th International Conference on Theorem Proving in Higher Order Logics (TPHOLs 2007), ed. K. Schneider and J. Brandt, Lecture Notes in Computer Science, Vol.4732, Kaiserslautern, Germany, pp.319–333, Springer-Verlag, Sept. (2007).
[38] G. Gonthier, "Formal Proof—The Four-Color Theorem," in Notices of the AMS, Vol.55, p.1370, december (2008).
[39] G. Gonthier, "Engineering Mathematics: the Odd Order Theorem Proof," POPL, ed. R. Giacobazzi and R. Cousot, pp.1–2, ACM (2013).
[40] F.D. Team, "The Flyspeck Project." https://code.google.com/p/flyspeck/.
[41] W. Fokkink, Modelling Distributed Systems, EATCS Texts in Theoretical Computer Science, Springer-Verlag (2007).
[42] M. Bezem, R. Bol, and J.F. Groote, "Formalizing Process Algebraic Verifications in the Calculus of Constructions," Formal Aspects of Computing, Vol.9, pp.1–48 (1997).
[43] L. Lamport, R. Shostak, and M. Pease, "The Byzantine Generals Problem," ACM Transactions on Programming Languages and Systems, Vol.4, No.3, pp.382–401 (1982).
[44] L. Lamport, "Byzantizing Paxos by Refinement," DISC, ed. D. Peleg, Lecture Notes in Computer Science, Vol.6950, pp.211–224, Springer (2011).
[45] D. Cousineau, D. Doligez, L. Lamport, S. Merz, D. Ricketts, and H. Vanzetto, "TLA + Proofs," FM, ed. D. Giannakopoulou and D. Méry, Lecture Notes in Computer Science, Vol.7436, Paris, France, pp.147–154, Springer-Verlag, Aug. (2012).
[46] E. Gascard and L. Pierre, "Formal Proof of Applications Distributed in Symmetric Interconnexion Networks," Parallel Processing Letters, Vol.13, No.1, pp.3–18 (2003).
[47] D. Cansell and D. Méry, Logics of Specification Languages, ch. The Event-B Modelling Method: Concepts and Case Studies, pp.47–152, Springer-Verlag (2007).
[48] P. Küfner, U. Nestmann, and C. Rickmann, "Formal Verification of Distributed Algorithms - From Pseudo Code to Checked Proofs," IFIP TCS, ed. J.C.M. Baeten, T. Ball, and F.S. de Boer, Lecture Notes in Computer Science, Vol.7604, Amsterdam, The Netherlands, pp.209–224, Springer-Verlag, Sept. (2012).
[49] C.T. Chou, "Mechanical Verification of Distributed Algorithms in Higher-Order Logic," The Computer Journal, Vol.38, pp.158–176 (1995).
[50] P. Castéran, V. Filou, and M. Mosbah, "Certifying Distributed Algorithms by Embedding Local Computation Systems in the Coq Proof Assistant," Symbolic Computation in Software Science (SCSS'09), ed. A. Bouhoula and T. Ida (2009).
[51] Loco, http://www.labri.fr/~casteran/Loco.
[52] Y. Deng and J.F. Monin, "Verifying Self-stabilizing Population Protocols with Coq," Third IEEE International Symposium on Theoretical Aspects of Software Engi-

neering (TASE 2009), ed. W.N. Chin and S. Qin, Tianjin, China, pp.201–208, IEEE Computer Society, July (2009).
[53] S. Devismes, A. Lamani, F. Petit, P. Raymond, and S. Tixeuil, "Optimal grid exploration by asynchronous oblivious robots," Proc. of SSS, pp.64–76, Springer (2012).
[54] F. Bonnet, X. Défago, F. Petit, M. Potop-Butucaru, and S. Tixeuil, "Brief announcement: Discovering and assessing fine-grained metrics in robot networks protocols," Proc. of SSS 2012, LNCS, Vol.7596, pp.282–284, Springer (2012).
[55] F. Bonnet, X. Défago, F. Petit, M. Potop-Butucaru, and S. Tixeuil, "Discovering and assessing fine-grained metrics in robot networks protocols," 33rd IEEE International Symposium on Reliable Distributed Systems Workshops, SRDS Workshops 2014, Nara, Japan, October 6-9, 2014, pp.50–59, IEEE (2014).
[56] N. Halbwachs, P. Caspi, P. Raymond, and D. Pilaud, "The synchronous dataflow programming language lustre," Proceedings of the IEEE, Vol.79, No.9, pp.1305–1320, September (1991).
[57] S. Dubois and S. Tixeuil, "A Taxonomy of Daemons in Self-stabilization," Tech. Rep. 1110.0334, ArXiv eprint, October (2011).
[58] L. Millet, M. Potop-Butucaru, N. Sznajder, and S. Tixeuil, "On the synthesis of mobile robots algorithms: The case of ring gathering," Stabilization, Safety, and Security of Distributed Systems - 16th International Symposium, SSS 2014, Paderborn, Germany, September 28 - October 1, 2014. Proceedings, ed. P. Felber and V.K. Garg, Lecture Notes in Computer Science, Vol.8756, pp.237–251, Springer (2014).
[59] E. Grädel, W. Thomas, and T. Wilke, eds., Automata, Logics, and Infinite Games: A Guide to Current Research [outcome of a Dagstuhl seminar, February 2001], Lecture Notes in Computer Science, Vol.2500, Springer (2002).
[60] C. Auger, Z. Bouzid, P. Courtieu, S. Tixeuil, and X. Urbain, "Brief announcement: Certified impossibility results for byzantine-tolerant mobile robots," International Symposium on Distributed Computing (DISC), ed. Y. Afek, Lecture Notes in Computer Science, Vol.8205, Jerusalem, Israel, pp.577–578, Springer, October (2013).
[61] C. Auger, Z. Bouzid, P. Courtieu, S. Tixeuil, and X. Urbain, "Certified impossibility results for byzantine-tolerant mobile robots," SSS, ed. T. Higashino, Y. Katayama, T. Masuzawa, M. Potop-Butucaru, and M. Yamashita, Lecture Notes in Computer Science, Vol.8255, Osaka, Japan, pp.178–190, Springer, November (2013).
[62] P. Courtieu, L. Rieg, S. Tixeuil, and X. Urbain, "Impossibility of gathering, a certification," Information Processing Letters, Vol.115, pp.447–452 (2015).
[63] D. Sangiorgi, Introduction to Bisimulation and Coinduction, Cambridge University Press (2012).
[64] L. Blin, A. Milani, M. Potop-Butucaru, and S. Tixeuil, "Exclusive perpetual ring exploration without chirality," Distributed Computing, pp.312–327 (2010).
[65] J. Barnat, L. Brim, M. Češka, and P. Ročkai, "DiVinE: Parallel Distributed Model Checker (Tool paper)," Parallel and Distributed Methods in Verification and High Performance Computational Systems Biology, pp.4–7, IEEE (2010).
[66] Uppaal Tiga, `http://people.cs.aau.dk/~adavid/tiga/`.
[67] Pactole, `http://pactole.lri.fr`.

(Received May 11, 2015)
(Revised August 22, 2015)

Béatrice Bérard is full professor at University Pierre & Marie Curie (Paris 6, France) in the Modelling and Verification group. Her research interests cover formal verification and synthesis of reactive systems, possibly integrating quantitative features like time or probabilities.

Pierre Courtieu is associate professor at CNAM, Paris (France). He received his PhD from University Paris-Sud XI in 2001, and is a specialist of theorem proving and proof of programs. He is a member of the steering committee of the Coq proof assistant.

Laure Millet is currently a third year PHD student at University Pierre & Marie Curie - Paris 6 (France). She is working on formal methods and verification of distributed algorithms for autonomous and mobile robots.

Maria Potop-Butucaru is full professor at University Pierre & Marie Curie (Paris 6, France) in the Network and Performance Analysis group. She was previously, assistant professor in University Paris 11 and associate professor in University Rennes 1. Her research area includes distributed and fault tolerant algorithms and formal methods for static and dynamic systems with special focus on sensor and robot networks.

Lionel Rieg is currently *Attaché Temporaire d'Enseignement et de Recherche* (a temporary research assistant position) at Collège de France, Paris (France). Holder of the high academic competitive examination *Agrégation* in mathematics, he defended his PhD at École Normale Supérieure de Lyon in 2014. He is a specialist in logics, in particular regarding the Curry-Howard correspondance.

Nathalie Sznajder is assistant professor at University Pierre & Marie Curie - Paris 6 (France), LIP6. She received her PhD from ENS Cachan in 2009 and spent a year as a post-doctoral researcher at Université Libre in Bruxelles. Her research interests cover formal verification and control of (distributed, timed) systems.

Sébastien Tixeuil is full professor at University Pierre & Marie Curie - Paris 6 (France) and Institut Universitaire de France, where he leads the Networks and Systems department at LIP6. He received his Ph.D. from University of Paris Sud-XI in 2000. His research interests include fault and attack tolerance in dynamic networks and systems. He has co-authored more than 150 research papers in international journal and conferences.

Xavier Urbain is associate professor at ENSIIE, Évry (France). He received his PhD from University Paris-Sud XI in 2001, as well as his Habilitation in 2010. He is a specialist of automated deduction and certified proof, in particular aimed at exhibiting properties of programs.

Finding an Area with Close Phenotype Values to Predict Proteins That Control Phenotypes

Takatoshi Fujiki†, Empei Gaku†, and Takuya Yoshihiro‡

†Graduate School of Systems Engineering, Wakayama University, Japan
‡Faculty of Systems Engineering, Wakayama University, Japan
{s101044, s121010, tac}@sys.wakayama-u.ac.jp

Abstract - Because phenotypes of living creatures are expressed reflecting on interactions among genes and proteins, relations among phenotypes and proteins (or genes) have been regarded as a key issue to be clarified to understand the system of creatures. In this paper, we try to find the relation among two proteins A, B, and a phenotype P, where there is a group of samples *G*, whose expression levels of A and B are both close to one another, and they always have close values of P. In this paper, we propose a method to extract a pair of proteins that effect on a target phenotype, from a dataset that consists of protein expression profiles and phenotype values.

Keywords: Proteomic Analysis, Two-Dimensional Electrophoresis, Phenotype, Expression Profile, Data Mining

1 INTRODUCTION

After the entire human DNA sequence was made public, many post-genome researches started to investigate the systems of living creatures. Proteome analysis is a field of such a post-genome research. The proteome analysis is a research field to analyze comprehensively the entire protein sets, in which functions and interactions of proteins that maintain living creatures are actively investigated.

As a method in proteome analyses, there is a technique called 2D electrophoresis [1]. The 2D electrophoresis enables us to measure expression levels of thousands of proteins in a biological tissue simultaneously. From the protein expression profiles obtained by the technique, we can clarify the functions and the interactions of proteins.

In many researches, major goal of researchers is to identify proteins that effect on a certain phenotype. For this purpose, a method for discovering the relationship between one protein and one phenotype is often used. One of the most basic methods is to calculate the correlation coefficient between protein expression levels of a protein and values of a phenotype item. Relationship between two items can be revealed by a relatively simple statistical method. However, the correlation coefficient evaluates only the liner relationship between two items. In contrast, Qu, et al. proposed a method to discover the nonlinear relationship between a gene and a phenotype using orthogonal polynomials [2].

On the other hand, there are a few researches that try to discover relationships in which more than one proteins effect on one phenotype. Zhang, et al. studied the interaction among a triplet of genes by comparing the correlation coefficients of genes A and B between two cases where another gene C expresses and does not express [3]. As another method, Inoue, et al. developed an algorithm to predict interactions among three proteins A, B and C based on correlation coefficient [4], and Fujiki, et al. developed an algorithm to predict interactions among three proteins A, B and C based on conditional probability [5]. If we regard C as a phenotype, those methods can be used to investigate the relationship between proteins and phenotypes.

In this paper, we propose a new method to detect interactions from different approaches. Specifically, we try to find the relation among two proteins A, B, and a phenotype P, where there is a group of samples *G*, whose expression levels of A and B are close to one another, and they always have close values of P. We evaluate the proposed method by applying the proposed method to the real data set.

Note that, to the best of our knowledge, this study is the first study that tries to find a set of two proteins that effect on a phenotype by finding a group of samples *G* whose expression levels of proteins A and B are close to one another that also have close values of a phenotype P.

The remainder of this paper is organized as follows. In Section 2, we describe the relation among two proteins and a phenotype assumed in this paper. In Section 3, we describe the proposed algorithm in detail. In Section 4, we evaluate our method by applying it to a real protein expression profile and a data set of phenotype. Finally, in Section 5, we conclude our study.

2 THE RELATIONSHIP BETWEEN PROTEINS AND PHENOTYPE WE SUPPOSE

2.1 Phenotype of Creature

Phenotype is a character that a creature has. For example, phenotype is an individual's traits, such as a size of body, a color, a pattern, etc. It is generally said that phenotype is largely determined by genes, but also considerably depends on growth environment of individuals. Many researches try to investigate the system of creatures that determines phenotypes. Such kind of researches are especially valuable when they target on several economically important phenotypes. For example, beef marbling scores and carcass weight of Wagyu beef have direct impact on the economical price of beef.

2.2 Protein Expression Profile

Protein expression levels are the amount of each proteins included in a biological sample. The protein expression levels

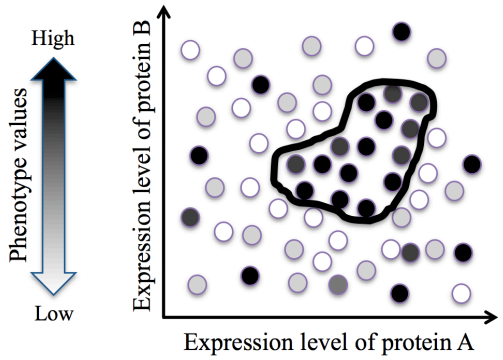


Figure 1: Example of Area That We Want to Demand.

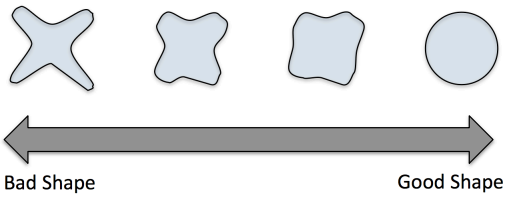


Figure 2: Good or Bad Shape of Area.

are typically measured by the 2D electrophoresis method [1]. This method is used for protein expression measurement widely.

The 2D electrophoresis is a method to separate proteins with 2-dimensions through two steps of electrophoresis. Generally, proteins are separated with isoelectric point in the first dimension, then they further are separated by molecular weight in the second dimension. Typically, the number of proteins included in a profile ranges from several hundred to thousands.

The expression profiles are the data that consists of expression levels of proteins included in a biological sample. The expression profiles are obtained by 2 steps. First, we obtain a 2D electrophoresis image through the 2D electrophoresis experiment. Second, we measure the areas of the islands revealed by the first step using image processing techniques.

2.3 Relationship of Two Proteins and A Phenotype We Suppose

We suppose two proteins that effect on a phenotype. In this paper, we try to find the relation among two proteins A, B, and a phenotype P, where there is a group of samples G, whose expression levels of A and B have close values a and b with each other, and they always have close value p of P.

Figure 1 shows an example of this relationship. We consider a 2-dimensional plane that has two axes of expression levels of proteins A and B. Each sample is plotted in this plane, and the deepness of the color of the samples represents phenotype values (i.e., samples with deep color represent high phenotype values and those with light color represent low phenotype values). Here, if there are no relationship among those two proteins and the phenotype, the distribution of the color of the samples would be uniform, i.e., the samples with various colors are plotted uniformly. In contrast,

Table 1: Data Format of Protein Expression Profiles.

Sample ID	Protein ID			
	A	B	C	⋯
1	0.000582	0.000107	0.000541	⋯
2	0.000563	0.000111	0.000458	⋯
3	0.000495	0.000126	0.000333	⋯
⋯	⋯	⋯	⋯	⋯

Table 2: Data Format of Phenotype Data Set.

Sample ID	Phenotype			
	Beef Marbling Standard	Carcass Weight	Rib-eye Area	⋯
1	4	422.7	44	⋯
2	9	470.7	53	⋯
3	7	433.5	50	⋯
⋯	⋯	⋯	⋯	⋯

if some relationships exist, it is thought that the distribution would not be uniform. In this paper, as shown in Fig. 1, we extract the area in which all the samples have close phenotype values. We consider that the existence of such areas indicates the relationship between proteins and phenotype. Namely, by extracting such areas, it is possible to estimate the combination of two proteins and the expression levels that control a phenotype.

3 EXTRACTION METHOD OF AREA WITH CLOSE PHENOTYPE VALUES

3.1 Format of Input Data

We use two sets of input data in the proposed method. One is a protein expression profile and the other is a set of phenotype data. We assume that the protein expression profile is obtained from the 2D electrophoresis experiment. The expression profile consists of the expression levels of each protein contained in each biological sample. We let $i(1 \leq i \leq I)$ be a sample, and let $j(1 \leq j \leq J)$ be a protein. Then, the expression level e_{ij} of a protein j included in a sample i is a real value. We show an example of the expression profile in Table 1.

We assume that a phenotype data set is represented by a table. Then the phenotype data set consists of the real values that represent the degree of phenotype (hereafter, we call them the *phenotype values*). We let $p(1 \leq p \leq P)$ be a phenotype, and the phenotype value p_i of a phenotype p included in a sample i is a real value. We show an example of the phenotype data set in Table 2. This example shows a case of brand cattle, in which we have BMS (Beef Marbling Standard), carcass weight, rib-eye area, etc. as phenotypes.

3.2 Areas That We Wish to Extract

In this paper, we extract a pair of proteins A and B that effect on a target phenotype p, by finding an area in which

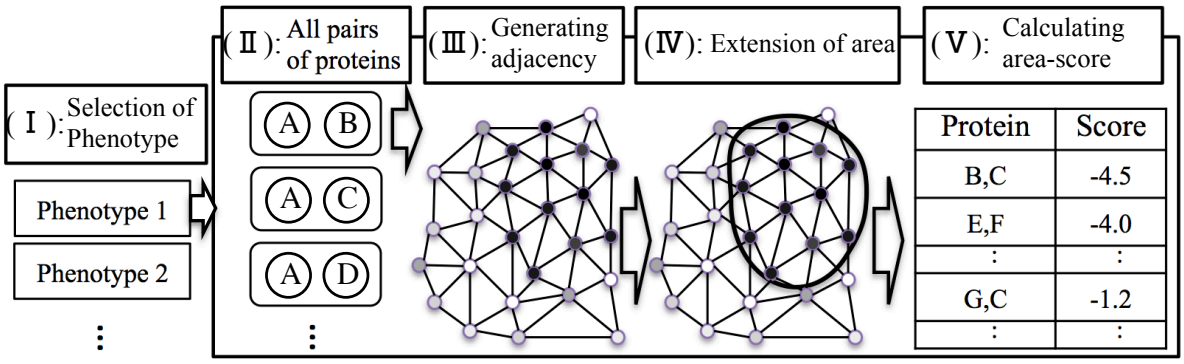


Figure 3: Overview of Our Proposed Method.

there is a set of samples whose expression levels of A and B are close to one another that also have close values of p on the 2-dimensional plane. In this section, we describe the criteria that the area should satisfy.

We consider two criteria with which we evaluate areas. Two criteria are on the phenotype values, and on the shape of the area, respectively. First, we describe the criterion on the phenotype values included in the area. It is required that the variance of the phenotype values included in the area is significantly smaller than those of all samples. Namely, it means that the samples that take a narrow range of phenotype values are included in it.

Next, we describe the criterion on the shape of the area. The criterion on shape is that the shape does not have big unevenness on a boundary line, i.e., the shape is "not warped" or "distorted". Namely, in this paper, we regard the circle as the best shape, whereas we regard the "warped" shape as bad shape. (See Fig. 2). Without this criterion, i.e., if an area is allowed to be any shape, we can extract areas of any distorted shape by choosing arbitrary samples freely that have close phenotype values. By limiting shape of areas, we can evaluate the area properly based on the sample distribution.

3.3 Overview of Proposed Method

We designed an algorithm to find the area that holds the criterion we described in Section 3.2. Note that the problem we treat is a combinatorial optimization problem whose search space is exponentially large so that we can hardly find the optimal solution. Thus, we designed our algorithm as a greedy one that explores areas from a small one by expanding it gradually with the best samples that forms the best areas at that time. We describe the overview of the proposed method as follows. (See Fig. 3 in parallel.)

(a) We select one phenotype to analyze (Fig. 3(Ⅰ)).

(b) We compose all the possible pairs of proteins for the phenotype selected in step (a) (Fig. 3(Ⅱ)).

(c) We generate an adjacency graph from the samples on a 2-dimensional plane whose two axes are the expression levels of proteins A and B (Fig. 3(Ⅲ)). We generate the adjacency graph as the Delaunay graph. We will give a short explanation of the Delaunay graph in the following Sections 3.4.

(d) We repeat extending the area using the graph that is generated in step (c) (Fig. 3(Ⅳ)). We start with the area that includes one arbitrarily sample (we call this sample *starting sample*). Then, we repeat extending the area with the most suitable samples until it comes to contain all samples. We perform this process from every starting sample. We describe this extending process in the following Section 3.5.

(e) We calculate the variance of the phenotype values for all the areas throughout the extending process i.e., we calculate the variance every time after extending the area with one sample. Then, for each individual area, we calculate its z-value (we call it the *area-score*) that indicates the statistical probability that the value of the variance occurs (Fig. 3(Ⅴ)). We extract the areas whose area-score is greater than the threshold. We describe about the calculating area-score in the following Section 3.6.

3.4 Step(c): Generating the Adjacency Graph from Samples

In this Section, we explain the algorithm to generate the adjacency graph from the samples on the 2-dimensional plane.

First, we generate a Voronoi diagram [6] on the 2-dimensional plane. A Voronoi diagram (Fig. 4) is a diagram obtained by dividing space into a number of areas. The boundary lines (dotted lines) between samples are composed of perpendicular bisectors between two samples. The plane is divided into areas (called Voronoi area) corresponding to each sample by the boundary line.

By connecting every pair of samples corresponding to two adjacent Voronoi areas, the Delaunay diagram (Fig. 5) that represents the adjacency among samples is generated. Then, we let $N(i)$ be the sample set adjacent to sample i.

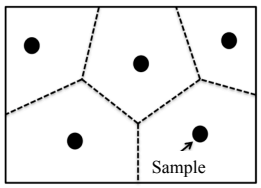


Figure 4: Volonoi Diagram.

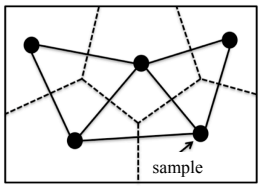


Figure 5: Delaunay Diagram.

3.5 Step(d): Extension of Areas

3.5.1 Overview of Extension Algorithm

We describe an algorithm to extend the area we wish to extract. We show the overview of the process as follows.

(1) An initial area consisting of one sample is determined by selecting a starting sample arbitrarily.

(2) We select a set of *extension candidate samples* from the samples that are adjacent to the current area so as not to make the shape of the extended area distorted.

(3) We select an *extension sample* from the set of *extension candidate samples* and extend the area by adding this extension sample.

(4) If the area does not include all samples, we return to (2).

Our algorithm searches for good areas through the process that expands an area by adding samples one by one greedily. Thus, the result largely depends on selection of starting samples. So, as for (1), the strategy to select starting samples should be determined according to the practical requirements. For instance, if users are interested in retrieving areas in which high-value samples are collected, it is recommended to start with high-value samples. Similarly, users may benefit from starting with low-value or middle-value samples for some cases. The strategy should be determined according to the situation. As for (2) and (3), details are described in the following sections 3.5.2 and 3.5.3, respectively.

3.5.2 Method to Select Extension Candidate Samples

In this Section, we explain the method to select the set of *extension candidate samples* mentioned in Section 3.5.1 (2). We let C be a set of extension candidate samples, and let D be the current area. C is a set of samples that satisfy conditions (i) and (ii) among the samples that is adjacent to D. We prevent extensions from creating donut-shape by setting these conditions as follows.

(i) Candidate sample must be adjacent to more than one samples that are included in D.

(ii) Samples on the boundary of D adjacent to the candidate sample must be continuous on the boundary.

We explain that the area does not become donut-shape using an example. In the area shown in Fig. 6, the samples that are surrounded by a black square are the samples that satisfies condition (i). Among them, the X-marked sample does not satisfy the condition (ii) because the three samples on the boundary line of D adjacent to this X-marked sample are not continuous on the boundary line. If we add this X-marked sample to D, the area that is extended becomes the donut-shape.

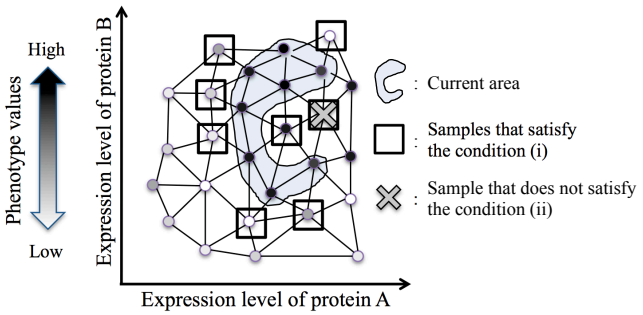


Figure 6: Conditions of Extension Candidate Sample.

3.5.3 Method to Select Extension Sample

We explain the algorithm to select a *extension sample* mentioned in Section 3.5.1. An extension sample is the sample that is the most desirable to be added to D from a set of extension candidate sample. As described in Section 3.2, the sample that is the most desirable is the sample that satisfies the two criteria on the phenotype values and on the shape of the area, respectively.

We describe the steps to select the extension sample from the extension candidate sample set C. First, we calculate the *shape-cost* $T(x)$ for every *extension candidate sample* $x \in C$. $T(x)$ evaluates the shape of the area that is created by adding the sample x to the current area D. Now, we let D_x be the area that is created by extending D with x. The less $T(x)$ is, the more the shape of D_x is distorted.

Next, we calculate the *phenotype-cost* $Z(x)$ to evaluate the phenotype value of the samples included in D_x. If $Z(x)$ is small, the phenotype value of x take a value close to the samples in D, and x would not increase the variance of the phenotype values of samples included in D. Therefore, we select the extension sample x such that $Z(x)$ is the smallest in C, and satisfy $T(x) \geq T_{thresh}$, where T_{thresh} is the threshold to the $T(x)$. Here, we consider the shape of the area that satisfies $T(x) \geq T_{thresh}$ to ensure the area has a good shape. If there is no x to satisfy $T(x) \geq T_{thresh}$ we regard x such that $Z(x)$ is the smallest in C as the extension sample.

We explain the method to calculate the shape-cost $T(x)$. We calculate $T(x)$ based on the ratio between the boundary length of D_x and the area of D_x. In general, if the area is the same, the boundary length is shorter when the shape is close to circle. We calculate $T(x)$ using this property. First, we let L_x and S_x be the length of the boundary line and the area of D_x, respectively. Here, the radius r_x of the circle whose circumference is just L_x is written as $r_x = \dfrac{L_x}{2\pi}$. Similarly, the radius r'_x of the circle whose area is just S_x is written as $r'_x = \sqrt{\dfrac{S_x}{\pi}}$. Finally, We define $T(x)$ as the ratio of r_x and r'_x

as follows:

$$T(x) = \frac{r'_x}{r_x} = \frac{2\sqrt{\pi S_x}}{L_x}.$$

Note that, $T(x)$ takes a value between 0 and 1, and the larger $T(x)$ is, the closer the shape is to a circle.

The phenotype-cost $Z(x)$ evaluates the variance of phenotype values included in D_x, which is created by extending D with x. We calculate $Z(x)$ as the z-value, which is known as a kind of statistic. $Z(x)$ is defined as $Z(x) = \left|\frac{p_x - \mu_x}{\sigma_x}\right|$, where p_x is the phenotype value of x, and μ_x and σ_x are the average and the standard deviation of the phenotype values of samples in D_x. Note that, $Z(x)$ represents the amount of difference between the phenotype value p_x and the average μ_x of phenotype values in D_x, which is measured as the number of the unit value σ_x. If the absolute value of $Z(x)$ is small, it means that p_x is close to the phenotype values of samples in D_x. Namely, if we add such p_x to D, the variance of D would not be increased.

3.6 Step(e) Calculating Area Score

In this Section, we explain the retrieval of the areas we wish to extract.

In this paper, we wish to extract the area with small variance of the phenotype values of the samples in the area. However, in general, if the number of samples in the current area is low, the variance is small. Therefore, in this paper, in order to retrieve the good area under variation of the number of samples, we calculate the variance of every area throughout our process to extend areas, and aggregate the variance of all the areas. Then, we judge whether each area is the one we wish to extract, by calculating the *area-score* of each area as a relative "position" of its variance in the distribution of variances of all areas that includes the same number of samples.

Now, we let (i_1, i_2) $(1 \leq i_1 < i_2 \leq J)$ be the pair of proteins, and let n $(1 \leq n \leq I)$ be the number of samples in the area. We suppose the extending process of the area with a starting sample m on the plane whose axes are two proteins (i_1, i_2). Here, we define the area where number of samples in the area is n as $D^{(n)}_{m,(i_1,i_2)}$. Note that, $D^{(n)}_{m,(i_1,i_2)}$ is determined uniquely by n, m and (i_1, i_2). Now, we let p_i be the phenotype value of sample i $(i \in D^{(n)}_{m,(i_1,i_2)})$, and $E[D^{(n)}_{m,(i_1,i_2)}]$ and $V[D^{(n)}_{m,(i_1,i_2)}]$ be the average and the variance of the phenotype values of samples in $D^{(n)}_{m,(i_1,i_2)}$.

We explain the method to calculate the *area-score*. First, we calculate $E[D^{(n)}_{m,(i_1,i_2)}]$ corresponding to combination of n, m and (i_1, i_2) as follows:

$$E[D^{(n)}_{m,(i_1,i_2)}] = \frac{1}{n} \sum_{i \in D^{(n)}_{m,(i_1,i_2)}} p_i.$$

Similarly, we calculate $V[D^{(n)}_{m,(i_1,i_2)}]$ as follows:

$$V[D^{(n)}_{m,(i_1,i_2)}] = \frac{1}{n-1} \sum_{i \in D^{(n)}_{m,(i_1,i_2)}} (p_i - E[D^{(n)}_{m,(i_1,i_2)}])^2.$$

Next, we calculate the average μ_n and the standard deviation σ_n of $V[D^{(n)}_{m,(i_1,i_2)}]$ with all areas whose number of samples in the area is n as follows:

$$\mu_n = \frac{1}{|M| \times J(J-1)/2} \sum_{m \in M} \sum_{1 \leq i_1 < i_2 \leq J} V[D^{(n)}_{m,(i_1,i_2)}],$$

$$\sigma_n = \sqrt{\frac{\sum_{m \in M} \sum_{1 \leq i_1 < i_2 \leq J} (V[D^{(n)}_{m,(i_1,i_2)}] - \mu_n)^2}{|M| \times \frac{J(J-1)}{2} - 1}}.$$

Finally, we calculate the z-value for the variance $V[D^{(n)}_{m,(i_1,i_2)}]$ of each area using μ_n and σ_n as the area-score $R^{(n)}_{m,(i_1,i_2)}$. The area-score $R^{(n)}_{m,(i_1,i_2)}$ is defined as follows:

$$R^{(n)}_{m,(i_1,i_2)} = \frac{V[D^{(n)}_{m,(i_1,i_2)}] - \mu_n}{\sigma_n}.$$

If $R^{(n)}_{m,(i_1,i_2)}$ is small, it means that the area rarely appears statistically. Therefore, we expect the area $D^{(n)}_{m,(i_1,i_2)}$ whose are-score $R^{(n)}_{m,(i_1,i_2)}$ is small enough for the output of the proposed method. For such areas D, we suppose there would be an interaction among two proteins i_1, i_2, and the phenotype.

4 EVALUATION AND DISCUSSION

4.1 Evaluation Method

We evaluate the proposed method by applying it to real protein expression profiles and a phenotype data set obtained by the author's collaborative work in Wakayama [7]. The protein expression profiles that we use in our evaluation are obtained by a 2D electrophoresis-based experiment [8].

A measurement error occurs in the measurement of the protein expression levels. Therefore, we performed 2D electrophoresis twice for each sample to confirm the accuracy of each electrophoresis experiment. From the result of the duplicated measurement, we removed the values considered to be low reliability from expression profiles. Specifically, we measured two expression values for each pair of a protein and a sample. If the larger expression level is larger than 1.3 times the value of the smaller expression level, we consider the expression level for the protein and the sample to be a null value as they are not reliable. Otherwise, the average of the two expression levels is used for each sample-protein pair. As a result, the expression profiles used for our evaluation consist of 90 samples and 47 proteins. In addition, the expression profiles are standardized in advance so that the average and the standard deviation of the expression levels with each sample are 0 and 1, respectively.

We performed an evaluation using "Carcass weight" as an important phenotype among many items included in the phenotype data set of beef cattle. As a pre-processing, we also standardized the phenotype data.

In order to evaluate the performance of the proposed method, we implemented a *simple method* to extend areas to be compared with the proposed method. The simple method is the

method that replaces the extension algorithm explained in Sections 3.4 and 3.5. The simple method adds a sample that is close in the Euclidean distance to the start sample m to the current area $D^{(n)}_{m,(i_1,i_2)}$. Consequently, the shape of the area $D^{(n)}_{m,(i_1,i_2)}$ that is obtained by the simple method is nearly a circle centered on the start sample m. Thus, the simple algorithm is equivalent to the algorithm that retrieves the best circular areas in the plane.

We evaluate the performance of the proposed method by comparing it with the simple method by calculating the variance $V[D^{(n)}_{m,(i_1,i_2)}]$ and the average $E[D^{(n)}_{m,(i_1,i_2)}]$.

Here, we describe the parameters in the evaluation experiment. We determined the threshold of the shape-cost $T(x)$ as $T_{thresh} = 0.7$ through a careful preliminary experiments to find a balancing point under the trade-off between the shape-cost and the area-score, and we set the number of samples in the area between 20 and 40 in order to ensure the reliability of the variance of the phenotype values in $D^{(n)}_{m,(i_1,i_2)}$. In addition, as the starting sample m, we use the sample whose phenotype value is within the bottom 10% among all samples. As actual requirements, because it is expected to extract the areas whose samples have low phenotype values, we confirm that the proposed method extracts the area whose phenotype value is low.

4.2 Result and Discussion

Tables 3 and 4 show the results of the ranking of top 10 combinations of proteins with respect to the area-scores. Table 3 is the result of the case where we applied the proposed method to the expression profiles and the phenotype data. On the other hand, Table 4 is the result of the simple method. These tables include the columns of protein ID of proteins A and B, the number of samples in the area, the area-score, $V[D^{(n)}_{m,(i_1,i_2)}]$ and $E[D^{(n)}_{m,(i_1,i_2)}]$. Note that, in Table 3 and Table 4, we leave only the best area out of the same protein pairs.

These results show that both $V[D^{(n)}_{m,(i_1,i_2)}]$ and $E[D^{(n)}_{m,(i_1,i_2)}]$ in the proposed method are smaller than those in the simple method. It was found from the result that the proposed method could extract areas better than the simple method. In order to confirm it in detail, Fig. 7 shows the scatter plots of the ranking of the top 50 areas extracted by the proposed method and the simple method. The vertical axis represents $E[D^{(n)}_{m,(i_1,i_2)}]$ and the horizontal axis represents $V[D^{(n)}_{m,(i_1,i_2)}]$. As is apparent from Fig. 7, both $E[D^{(n)}_{m,(i_1,i_2)}]$ and $V[D^{(n)}_{m,(i_1,i_2)}]$ extracted by the proposed method is found to be lower values than those of the simple method. From these results, we confirmed that the phenotype values of the areas extracted by the proposed method are lower than those extracted by the simple method, and the samples included in the area have close phenotype value each other. In other words, it can be said that the proposed method can extract "good area," compared with the simple method.

Next, we confirm whether the shape of the area extracted by the proposed method is "good shape" or not. As a typical example of the extracted areas, we show the shape of the rank-1 area in Fig. 8.

Table 3: Ranking of Areas with Proposed Method.

Ranking	Protein A	Protein B	Number of samples	Area score	Variance in area	Average in area	Shape score
1	3899	4491	39	-2.7545	0.2510	-0.5539	0.7003
2	5639	5735	31	-2.5615	0.1862	-0.6034	0.7012
3	3648	4491	38	-2.4033	0.3012	-0.4405	0.7057
4	828	5733	36	-2.3852	0.2832	-0.3981	0.7002
5	3648	5727	40	-2.3596	0.3283	-0.3444	0.7010
6	3899	3598	30	-2.3408	0.2153	-0.5549	0.7175
7	4491	5727	29	-2.3014	0.2090	-0.7281	0.7058
8	5636	5654	38	-2.3002	0.3193	-0.4944	0.7001
9	3648	5726	38	-2.2910	0.3209	-0.4495	0.7276
10	4491	5730	40	-2.2879	0.3406	-0.3662	0.7060

Table 4: Ranking of Areas with Simple Method.

Ranking	Protein A	Protein B	Number of samples	Area score	Variance in area	Average in area
1	3648	4491	31	-2.9546	0.3939	-0.3227
2	4491	5657	40	-2.8999	0.5544	-0.2364
3	4491	5688	40	-2.8186	0.5688	-0.2780
4	4491	5686	39	-2.8077	0.5571	-0.2671
5	4491	5721	26	-2.8066	0.3203	-0.4939
6	828	5660	38	-2.8003	0.5436	-0.1725
7	4491	5724	39	-2.6507	0.5856	-0.3966
8	4491	5734	36	-2.6493	0.5437	-0.2875
9	828	4991	40	-2.6394	0.6005	-0.2203
10	5637	5644	25	-2.6194	0.3477	-0.4936

Figure 8 shows the scatter diagram of the rank-1 area in Table 3. The horizontal axis and the vertical axis represent the standardized expression levels of protein A and B, respectively. The shape-cost of the area is 0.7003, which is the value close to threshold $T_{thresh} = 0.7$. We found that this area is close to a circular shape to same extent and is allowable as an area. That is, the shape of this area extracted by the proposed method is "good shape."

Then, we see whether this area is a "good area" or not by examining the phenotype value of the samples in the rank-1 area in Table 3. Figure 9 shows the histogram of the phenotype values included in the area, and a histogram of the phenotype values of all samples. The vertical axis represents the number of samples and the horizontal axis represents the carcass weight. Since the carcass weight has been standardized, the average of the carcass weight of all samples is 0, and the variance is 1. The phenotype values in the extended area are distributed in a relatively narrow range between -1.5 and 0.5, and the distribution is unimodal. We find that $V[D^{(n)}_{m,(i_1,i_2)}] = 0.2510$ is considerably lower than the whole variance 1. Moreover, the $E[D^{(n)}_{m,(i_1,i_2)}] = -0.5539$ is sufficiently smaller than the whole average 0.

From the above reasons, we found that the area extracted by the proposed method is the area that we want to find because both $V[D^{(n)}_{m,(i_1,i_2)}]$ and $V[D^{(n)}_{m,(i_1,i_2)}]$ are small enough.

One of the essential future tasks is to explore how to utilize the proposed method in practice. We do notice that the approach of three-way interactions (i.e., interactions among three proteins, or two proteins and a phenotype) generally

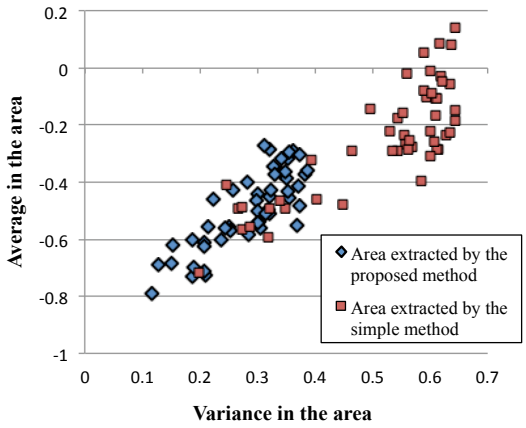


Figure 7: Distribution of Areas Extracted by Proposed Method and Simple Extension Method.

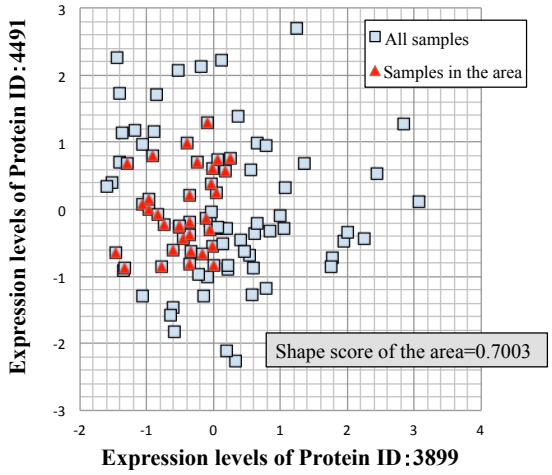


Figure 8: Distribution of Areas of Rank 1 in Table 3.

involves a difficulty in clarifying the performance of methods because specific physical interactions or phenomena are rarely connected directly to them. Thus, to reveal the practical capability of the analysis on three-way interactions, we require to design the processes in which these analytical methods effectively work. This is, in fact, a challenge that requires a considerable deal. For example, we can try to connect some physical interactions or phenomena to our analysis to clarify the direct meaning of our analysis. Also, we can try to show that our analytical result can support to explore biomarkers that control a target phenotype, or accelerate to find proteins included in some pathways or related to some biological functionality. Anyway, these are generally a part of important future work for the approach of three-way interaction analysis.

5 CONCLUSION

In this paper, we proposed a method to find areas with close phenotype values to predict proteins that control phenotypes. By extracting areas including samples with close phenotype

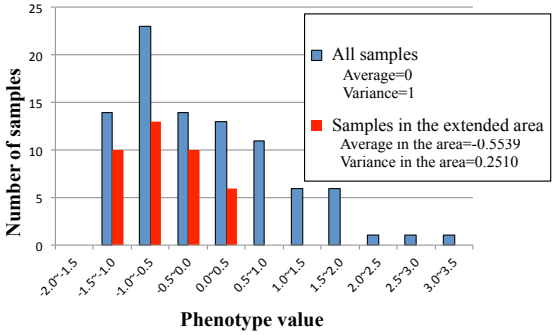


Figure 9: Histogram of Areas of Rank 1 in Table 3.

values, which rarely occur statistically, it is possible to estimate the relationship among two proteins and a phenotype.

We performed the evaluation experiment using real data set obtained by the author's collaborative work in Wakayama [7]. In order to evaluate the performance of the proposed method, we implemented a simple method to be compared with the proposed method. As a result, we found that the proposed method extracted the area better than the simple method. That is, the proposed method is able to extract the area that the variance of the phenotype values in the area is small.

ACKNOWLEDGEMENT

This work was partly supported by "the Program for Promotion of Basic and Applied Researches for Innovations in Bio-oriented Industry" of NARO (National Agriculture and Food Research Organization), and "the Program for Promotion of Stockbreeding" of JRA (Japan Racing Association).

REFERENCES

[1] A. Malcolm Campbell, Laurie J. Heyer, Discovering Genomics, Proteomics and Bioinformatics, Benjamin Cummings (2006).

[2] Y Qu, S Xu, "Quantitative trait associated microarray gene expression data analysis," Molecular Biology and Evolution, Vol. 23, No. 8, pp. 1558-1573 (2006).

[3] J. Zhang, Y. Ji, L. Zhang, "Extracting Three-way Gene Interactions from Microarray Data," Bioinformatics, Vol. 23, No. 21, pp. 2903–2909 (2007).

[4] E. Inoue, S. Murakami, T. Fujiki, T. Yoshihiro, A. Takemoto, H. Ikegami, K. Matsumoto, and M. Nakagawa, "Predicting Three-way Interactions of Proteins from Expression Profiles Based on Correlation Coefficient," IPSJ Transactions on Bioinformatics, Vol. 5, pp. 34–43 (2012).

[5] T. Fujiki, E. Inoue, T. Yoshihiro, M. Nakagawa, "Prediction of Combinatorial Protein-Protein Interaction from Expression Data Based on Conditional Probability," In: Protein-Protein Interactions - Computational and Experimental Tools, InTech Web Press, pp. 131–146 (2012).

[6] M. de Berg, O. Cheong, M. van Kreveld, M. Overmars, Computational Geometry: Algorithms and Applications 3rd ed, Springer (2008).

[7] Collaboration of Regional Entities for the Advancement of Technological Excellence in Wakayama, http://www.yarukiouendan.jp/techno/kessyu/
[8] K. Nagai, T. Yoshihiro, E. Inoue, H. Ikegami, Y. Sono, H. Kawaji, N. Kobayashi, T. Matsuhashi, T. Otani, K Morimoto, M. Nakagawa, A. Iritani and K. Matsumoto, "Developing an Integrated Database System for the Large-scale Proteomic Analysis of Japanese Black Cattle," Animal Science Journal, Vol. 79, No. 4 (2008) (in Japanese).

(Received November 22, 2014)
(Revised May 22, 2015)

Takatoshi Fujiki recieved his B.E. and M.E. degrees from Wakayama University in 2010 and 2012, respectively. He is currently a doctoral course student in Wakayama University. He is interested in data mining, machine learning, and bioinformatics. He is a student member of IPSJ.

Empei Gaku received his B.E. and M.E. degrees from Wakayama University in 2011 and 2013, respectively. He is currently working with Azbil Corporation.

Takuya Yoshihiro received his B.E., M.I. and Ph.D. degree from Kyoto University in 1998, 2000 and 2003, respectively. He was an Assistant Professor in Wakayama University in 2003-2009, and from 2009 he is an Associate Professor in Wakayama University. He is interested in computer networks, graph theory, bioinformatics, medical systems, and so on. He is a member of IEEE, IEICE and IPSJ.

Measurement of Olfaction in Children with Autism by Olfactory Display Using Pulse Ejection

Eri Matsuura*, Risa Suzuki*, Shutaro Homma*, and Ken-ichi Okada**

* Graduate School of Science of Technology, Keio University, Japan
** Faculty of Science and Technology, Keio University, Japan
{matsuura, risa, honma, okada}@mos.ics.keio.ac.jp

Abstract - Autism Spectrum Disorders (ASD) is considered as one of the developmental difficulties caused by dysfunctions of the brain. There are a variety of symptoms with ASD, but these can be significantly improved by appropriate treatment and education. Thus, it is important to find the patients with ASD while young. Recently, research on the olfactory senses of patients with ASD is being done. It is reported that there are differences of odor detection and identification abilities between people with ASD and controls. Our aim was to develop the screening examination by olfaction, hence we developed an application to assess the odor detection and identification abilities in children by using olfactory display which uses pulse ejection. We also investigated the olfactory abilities of the children with ASD using the application. However, we found some problems with application, and we made improvements. After the improvements, we assess the olfactory abilities of the typically developing children. As a result, we saw similar tendency as in the related works on olfactory abilities in children with ASD by the olfactory display.

Keywords: olfactory display, pulse ejection, autism spectrum disorders, children, interface application

1 INTRODUCTION

Autism Spectrum Disorders (ASD) is considered as one of the developmental difficulties caused by dysfunctions of the brain [1]. People with ASD have difficulties in three main areas referred to as the "Triad of Impairments" coined by Lorna Wing: impairment of social communication, social imagination, and social relation. Some examples are sudden start of conversations with strangers, difficulty in meeting their eyes, and parroting the question back. There are a variety of other symptoms with ASD, and it is not a uniform state. However, it is known that such difficulties can be significantly improved by appropriate treatment and education [2]. Therefore, it is important to detect a person with ASD while young. On the other hand, the examinations of ASD have many problems, and it is not easy to carry out an examination. Multiple staff spends time to make preparations for an examination, and it takes a long time to do an examination. For example, one examination takes a maximum of four hours in Japan, but it may take longer in other countries. Furthermore, the staffs are required to have a discussion to select the appropriate method to suit each subject's individuality. Because the examinations of ASD require such a preparation for all subjects, they are often needed to wait over one year, and the examinations cost too much money to take it lightly. In other words, no matter how much you want, it is not easy to have the examinations of ASD. There are also screening examinations of ASD, but they have some problems as well. For example, the Autism-Spectrum Quotient is unsuitable for detecting ASD while young, because it is developed for adults [3]. There is also the Modified Checklist for Autism in Toddlers (M-CHAT) as a screening examination for children [4], but the Japanese version of M-CHAT [5] has some problems likewise. It is reported that Japanese version of M-CHAT cannot detect 15% of children with ASD.

Recently, research on the olfactory senses of patients with ASD is being done. Because one of the symptoms with ASD is characteristic behavior towards odor, for example, they have a fetish for certain odors. It is reported that there are differences of odor detection and identification abilities between people with ASD and controls, hence checking olfactory abilities may make it possible to screen ASD. In this study, we assess olfactory abilities in children with ASD for developing a screening examination by olfaction.

2 OLFACTION IN PEOPLE WITH ASD

Many children with ASD have sensory difficulty [6]. For example, they don't want to get on the buses because of its odor, and they keep on sniffing certain odors. For these abnormal responses to sensory stimuli, E.Gal et al. claim that there is a change in symptoms during puberty, and in responses to sensory stimuli as well [7]. The studies on changes in symptoms with age have not been done much, but there have been reports that the symptoms were improved or more sever as people with ASD get older. Thus, it is thought that people with ASD do not show a similar trend, but the symptoms may vary according to age.

There are a variety of the symptoms of sensory difficulty, and it is reported that abnormal responses for olfaction and gustation are stronger than for vision and audition. Moreover, it is reported that abnormal responses for olfaction, gustation, and touch are stronger in the case of girls than boys, and the difficulties in olfaction of girls with ASD are more serious. Thus, it is believed that many of the patients with ASD respond characteristically to olfactory stimuli, and research on the olfactory senses is being done. There are several olfactory abilities such as the odor detection ability and the identification ability. The detection ability is the ability to detect an odor. If you cannot recognize what odor it is, you only have to be able to detect

This paper is recommended papers in IWIN 2014.

a certain odor. The odor identification ability is the ability to select the correct odor from several choices. You have to not only detect a certain odor but match the odor and the name correctly from several choices. In the previous studies, the detection and identification abilities were mainly measured. Suzuki et al. assessed olfaction of the adults with Asperger Syndrome (AS) by using University of Pennsylvania Smell Identification Test (UPSIT) [8] in 2003 [9]. 24 participants took part: 12 men with AS and 12 control participants. Odor identification ability of participants with AS were impaired, but there was no difference in odor detection ability. Bennetto et al. focused on odor identification ability in 2007 [10]. 48 participants aged 10 to 18 took part, 21 patients with ASD and 27 control participants. They assessed the odor identification ability using Sniffin' Sticks [11], and reported impaired ability of patients with ASD. There is also research focusing on pleasantness of odors. Hydlicka et al. assessed pleasantness of odors for 70 children aged around 10 years old (35 children with AS or high functioning autism (HFA), 35 control participants) using Sniffin' Sticks [12]. They reported that children with AS or HFA, compared to controls, perceived the odor of cinnamon and pineapple as less pleasant, the same was true of cloves. Dudova et al. assessed odor detection and identification abilities using Sniffin' Sticks, and relation between preference for odors and identification ability [13]. 70 children aged around 10 years old took part, where 35 were children with AS or HFA, and 35 were control participants. They reported impaired detection ability of children with AS or HFA, and compared to controls, they were better at correctly identifying the odor of orange and worse at correctly identifying the odor of cloves. As it was found that the odor of orange was favorite odor and the odor of cloves was least favorite odor for children with AS or HFA by Hydlicka et al. , they argued that the odor identification ability of children with ASD was related to pleasantness of odors. The patients with ASD have such characteristics regarding olfaction. However, the characteristics differed among age groups. As there is a possibility that the differences were caused by the change of symptoms associated with aging, it is believed that olfaction may be a physiological indicator of ASD by the age limit. Thus, by assessing olfaction in children, screening patients with ASD may be possible.

3 MESUREMENT OF OLFACTION BY OLFACTORY DISPLAY

It is important to find patients with ASD at while young. But the examinations of ASD have a variety of problems, and it is difficult to have an examination. Although there are some screening tests for ASD, they also have some problems. Therefore, it is necessary to develop an easy method to screen ASD. Because there are several characteristics of olfaction in people with ASD, we think measurement of olfaction in children will lead to screening examination of ASD. However, the current method of olfaction test in Japan requires great care; there are many tasks and takes too much time. There are also the problems of scent lingering in the air and filling the room. In addition, the current method may be tiring for children because it consists of repetition of simple tasks for a long time, and it has a possibility of measurement failure. Moreover, the odors used in the identification test are not always familiar to children. Thus, it is important to make a simpler method of olfactory test targeted at children. We therefore use olfactory display in the examination to solve these problems, and develop a screening application for children with ASD.

The screening examination we proposed is composed of the odor detection and identification tests. In the examination, we use the olfactory display using pulse ejection, which means scent is presented for a short duration. The person conducting the examination uses a personal computer (PC) to control the olfactory display, and the scents are presented simply by operating the PC. The usage of PC reduces the time and effort involved to present a scent. Since part of the operation is automatic, the person can still conduct the examination by only simple operations even if the person conducting the examination doesn't have enough knowledge about the olfaction tests. In addition, there are two application screens in our system, one for doctors to check the status and the other for children. Odor detection and identification test are composed of games manipulated by children so as not to bore them. The operations are simple, and children can manipulate intuitively using a touch panel monitor. Since the operations that doctors must do are made minimum, doctors can observe patients. We propose a screening application for children with ASD based on these concepts.

We hypothesized that odor detection ability in children with ASD is impaired and odor identification ability is not different from it in typically developing children, and assessed the olfaction in children.

4 OLFACTORY MESUREMENT METHOD

4.1 Olfactory Display for Medical Purposes

We developed an olfactory display as shown in Fig. 1. We call this display "Fragrance Jet for Medical Checkup (FJMC)." FJMC uses the technique used in ink-jet printers in order to produce a jet, which is broken into droplets from the small hole in the ink tank. This device can create pulse ejection for scent presentation: thus the problem of scent lingering in the air can be minimized. The 2D diagram of FJMC is as shown in Fig. 2. This display has one large tank and three small tanks, and one fragrance is stored in each tank. There are 255 minute holes in the large tank, and refer to the average ejection quantity per minute holes as "the unit average ejection quantity (UAEQ)", and the number of minute holes that emit scent at the same time as "the number of simultaneous ejection (NSE)." The device can change the ejection time at 667 μs intervals so the measurement can be controlled precisely. UAEQ from one minute hole on the large tank is 7.3 pL, and on the small tanks is 4.7 pL. As this display can emit a fragrance from multiple holes at the same time, the range of NSE is 0 to 127 for small tanks, and 0 to 255 for the large tank. These values determine the ejection quantity per unit time. Now, we refer to the average

Figure 1: Olfactory Display.

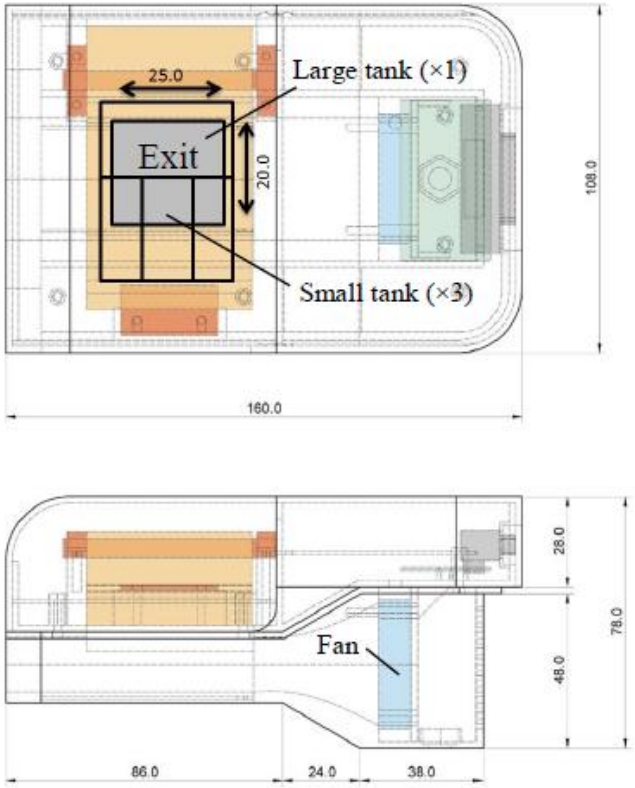


Figure 2: Plane and Side View of the Olfactory Display.

of ejection quantity from one minute hole per unit time as "the average ejection quantity per unit time (AEQUT)." Thus, it is possible to control the intensity of scent by the ejection quantity per unit time and the ejection time (ET), and the actual ejection quantity (EQ) can be calculated as follows.

$$EQ(pL) = AEQUT(pL/\mu s) \times NSE \times ET(\mu s) \quad (1)$$

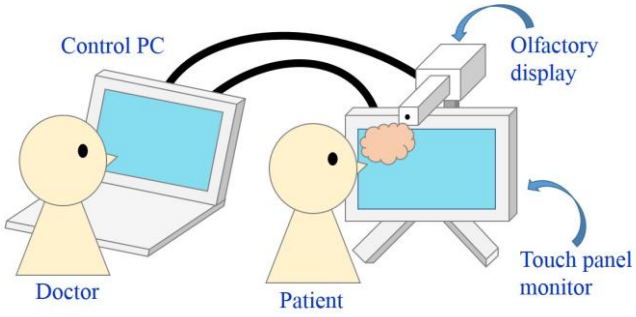


Figure 3: Conceptual Diagram.

4.2 Application for Measurement of Olfaction

We developed the screening application for children with ASD. The purpose of developing the application was to reduce the time and effort for the olfactory test and not to bore the children. We used FJMC, the PC, and the touch panel monitor in this system. Figure 3 shows a schema of the system. As FJMC uses pulse ejection, scent is emitted just short periods of time, and presented odor can be changed quickly by changing some parameters using the PC. The reason why we used touch panel monitor was to make children move their own hands. In addition, we incorporated gaming element in the olfactory test, and we also made the examination simple. The examination is composed of the odor detection and identification test. In the next section, we describe the method of use.

The odor detection test was designed as a treasure hunting game where patients try to find a smelling box out of three boxes. We used the odors of banana and pineapple, because banana and pineapple are, we thought, familiar to children. As regards the odor of pineapple, there was an additional reason. It was reported that children with AS or HFA perceived the odor of cinnamon and pineapple as less pleasant in the research of Hydlicka et al. Thus, we thought children with ASD had some characteristics regarding detection the odor of pineapple. In addition, there were four levels of intensity of the odors: 10, 20, 40, and 80. In this time, we set ET at 200ms, and the intensity was expressed by NSE. Moreover, we used only the small tanks. Detection threshold, which is the lowest intensity you can smell, was determined by using the raising method (the first intensity was 10) and three alternative forced choice procedures. There were three boxes on the screen for patients, one of them with a smell and the others odorless, and the subject select the smelling box from the three boxes. Detection threshold was determined by the intensity which the subject answered correctly twice in a row. If the subject selected the wrong answer, the intensity level was raised one level. In addition, the test was carried out firstly with the odor of banana and then the odor of pineapple. Figure 4 shows the screen for subjects. The subject pushed the arrow buttons to move the dog in front of the box the subject wanted to sniff. By pushing the button of the dog, the scent was presented and a sound rang at the same time. When the subject found the smelling box, he or she moved the dog in front of the box and pushed the "this" arrow button.

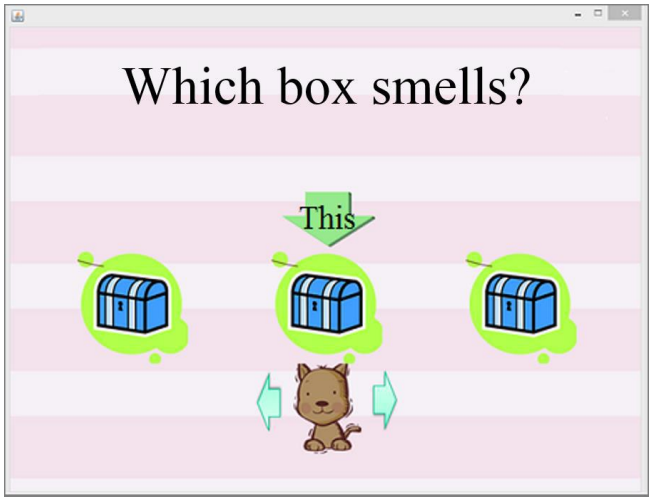


Figure 4: Odor Detection Test.

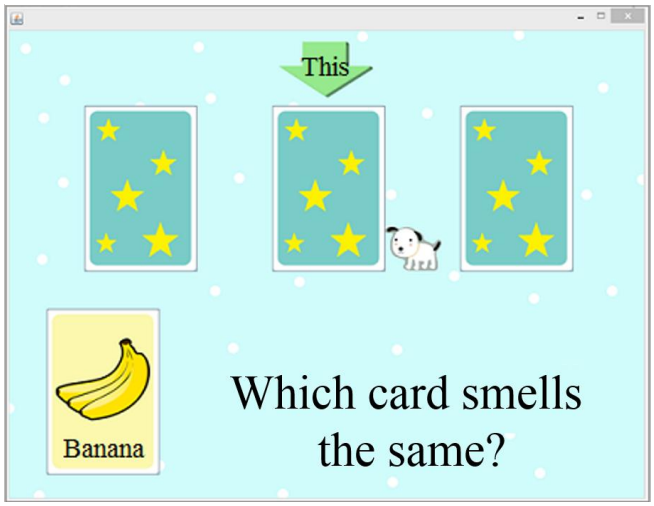


Figure 5: Odor Identification Test.

The odor identification test was designed as a card game where subjects try to find the same card by using odor as a clue. In this test, we used odors of banana, rose, and lavender. The intensity was decided by preparatory experiment, and was strong enough to sniff. Two tests were conducted in the odor identification test. Firstly, the subjects selected a card which the odor matched the illustration (Trial1). Secondly, the subject sniffed the odor of the target card, and selected a card with the same odor (Trial2). The reason why we carried out two tests was that we wanted to discuss whether the cause of impaired odor identification was the brain (the ability of associating odors and the image of odors) or simply olfaction. Both tests were carried out first with the banana and rose. In this examination, the number of correct answers was evaluated. Before the actual test, the subject checked the odor of banana and rose. The first two of the four tests were Trial1. Figure 5 shows the application screen for subjects for Trial1. The odor of banana, rose, and lavender was assigned in a random order on the three cards above. The odor was emitted by pressing the card, and the subjects sniffed the three cards to find the card with the odor matching the illustration on the bottom left. Since the goal of this test was the identification of odor and illustration, the odor was not emitted when touching the bottom left card. When they found the card with the odor

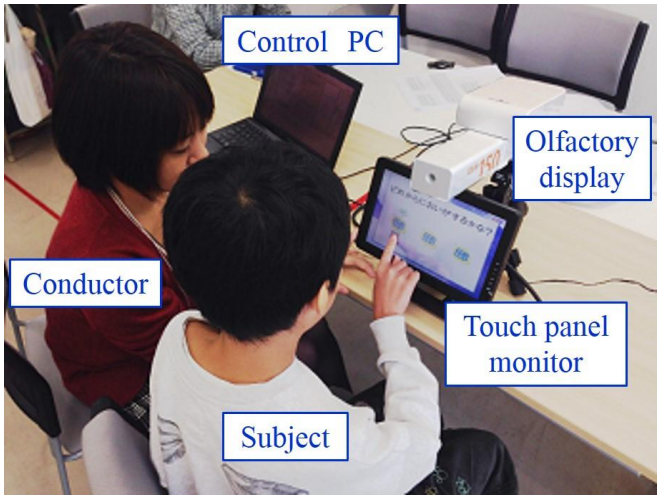


Figure 6: The State of the Experiments.

that matched the illustration, they touched the card again and the “this” arrow button to answer. After the two tests, Trial2 was conducted. Compared to the application screen for Trial1, only the bottom left card was different. Since there was no illustration on the card, first the subject had to touch the bottom left card to sniff the target odor. After sniffing the each odor for the three cards, the subject selected the card with the target odor.

5 EXPERIMENTS AND EVALUATION

5.1 Experiments for Children with ASD

The subjects were 15 patients with ASD (11 boys, 4 girls). Their ages were from 10 to 16 (mean age of 14.3 years, SD 1.74 years). We used the PC, the olfactory display, and the touch panel monitor in the experiment. The interface for doctors was displayed on the PC, and that for subjects was displayed on the touch panel monitor. The touch panel monitor was placed in front of the olfactory display so that sniffing the odor can be done while watching or touching the monitor. The experiment was conducted from 10:00 to 17:00 in a quiet room, and the subjects sat on the chair during the experiment. Figure 6 shows the experimental environment. We conducted the odor detection and identification tests, in that order.

In the odor detection test, the average of odor detection threshold for banana was 65.7 ± 55.0 and that for pineapple was 46.9 ± 44.3. However, two subjects had to do the test twice and other two subjects could not finish the test because of operation errors, so the average was calculated by excluding the two immeasurable subjects. Furthermore, the detection thresholds of the subjects who could not detect in 80 was determined as 160. In the odor identification test, we calculated the accuracy rate for each trial. The accuracy rate of banana in Trial 1 and 2 were 73.3%, and that of rose in Trial 1 was 73.3%, in Trial 2 was 60.0%. In addition, we could conduct the whole experiments within 15 minutes for all subjects. Then, we examined the olfactory senses of the subjects with ASD. The calculation was carried out by excluding the immeasurable subjects. There were no differences between the detection threshold of banana and

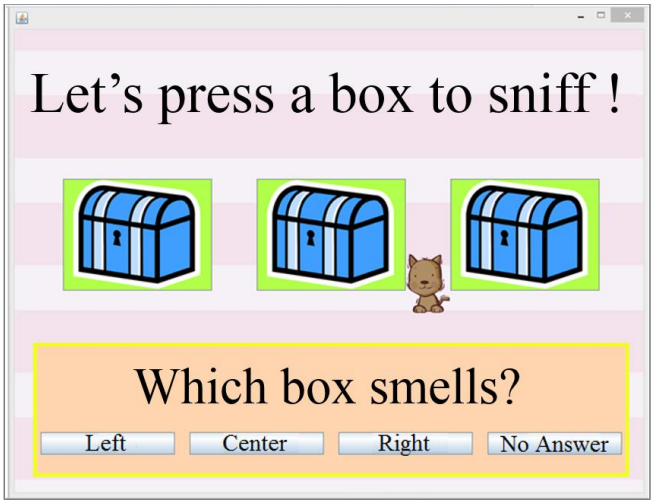


Figure 7: Odor Detection Test after the Improvements.

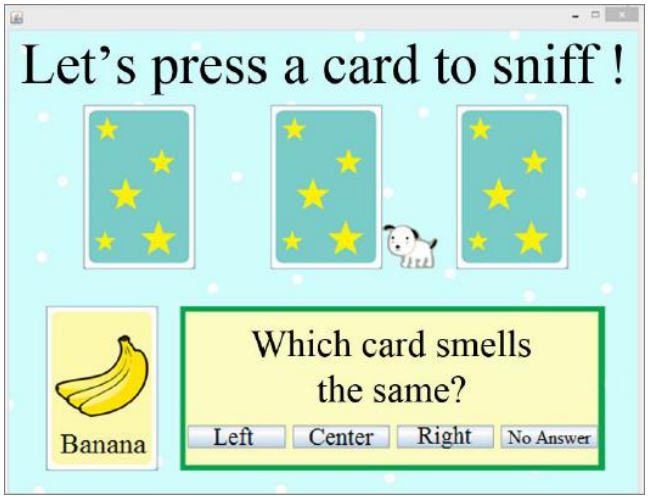


Figure 8: Odor Identification Test after the Improvements.

pineapple by the t-test ($p = 0.27 > 0.05$). Thus, it is thought that there are no differences between when using the odor of banana and when using the odor of pineapple. Next, we examined the differences between the results of identification in Trial1 (odor and illustration) and Trial2 (odor and odor) by the z-test, and we also saw no differences (banana: $p = 1.00 > 0.05$, rose: $p = 0.40 > 0.05$). Therefore, it is thought that there are no differences between the methods of two trials. We also compared the kinds of odor in Trial1 and Trial2, and there were no differences (Trial1: $p = 1.00 > 0.05$, Trial2: $p = 0.52 > 0.05$). From this results, it is thought that there are no differences between the kinds of odor as well.

5.2 Improvements of Interfaces

During the experiment, the subjects took a variety of characteristic behaviors. For example, there were many children who had shown a large interest in the touch panel monitor and the olfactory display. Some of them looked into the ejection hole of the olfactory display and stuck their nose in. In addition, they asked about the equipment very much but not so much about the operations, and the operation error was sometimes occurred in the odor detection test. Thus, we thought that it is necessary to improve the application to allow more intuitive manipulation. We also thought that it is essential that the application screen for subjects and the operation were made simpler in order to do the experiment easily for the children.

Because there were some operation errors in the odor detection test, we made improvements in the interfaces and the measurement method of the odor detection test. As regards the improvements of the interfaces, we made mainly three improvements: 1) automation of screen transition, 2) display method of the correct answer and the operation time on the application screen for doctors, and 3) adding a practice mode. Firstly, the screen control was not automatic in this application, and the doctor had to click on the "next" button to show the next screen for subjects. However, the doctor was talking with the subjects and helping them during the examination, hence the screen transition was troublesome. Therefore, the screen control was made almost automatic without the cases that the doctor should ask the subject whether the next test can be carried out. Secondly, since only the results of the tests were displayed on the application screen for doctors during the examination, they could not know which box was the correct answer until the subject selects an answer. In addition, some of the subjects took a long time to answer. We thought that there was a possibility that the operation time for the ASD group differs from that for the control group. Therefore, the correct answer and the operation time were displayed on the application screen for doctors. Finally, when describing the operation method verbally, there was no problem for the adults, but some children could not understand well without seeing the actual operations. Thus, in order to have a good understanding the operations, a practice mode was added in the odor detection and identification tests. However, the practice mode is not always conducted if the subject knows the operations. In consideration of this, we made it enabled to select the test mode, normal mode or practice mode, at the beginning of the test by doctors. The subject can run the test only once in the practice mode, and then run the test in the normal mode even if the patient selects a wrong answer.

In the odor detection test, the subjects often pushed the box or the "this" arrow button, and not the dog when they tried to sniff. So that children could make a more intuitive operation, we made the size of box bigger and the odors were emitted by touching the box. From this improvement, the time and effort for moving the dog to the box subjects want to sniff was not required. In addition, there was no choice available when the patient couldn't detect the odor, but had to choose one box. In order to solve these problems, the "this" arrow button was eliminated, and four choices (left, center, right, no answer) were placed at the bottom of the application screen for subjects. Figure 7 shows the application screen of the odor detection test after the improvements were made. There was also an improvement of the algorithm. Since the test started from the weakest odor intensity, children who could not detect the odor often appeared to have lost interest in the examination. It was most important to conduct the whole examination with interested, hence the examination method was changed to the descending method (the first intensity was 80) in order to enhance the motivation of the children. The timing of changing the level of odor intensity was also changed. If

the answer was correct twice in a row, the intensity was lowered one level. Therefore, if the answer was wrong once, the examination was finished and the last intensity which they detected twice became the detection threshold.

During the odor identification test, operation errors were not occurred. However, there was no choice available for when the subject could not find an answer as well. Thus, the "this" arrow button was also eliminated, and four choices were placed at the bottom of the application screen for subjects. Before the test, the subjects could check the odor of banana and rose but not lavender, so we modified the application so that the odor of lavender can be checked in the same section. Figure 8 shows the screen of the odor identification test after the improvement.

5.3 Experiments for Controls

The experiments were also conducted for the typically developing children using the application after the improvements. The subjects were nine children (five boys, four girls). Their ages were from 9 to 12 (mean age of 10.8 years, SD 0.916 years). The experiment was conducted from 17:00 to 21:00 in a quiet room. The instruments we used were the same as the experiments for children with ASD, and we conducted the odor detection and identification tests, in that order as well.

In the odor detection test, the average of odor detection threshold for banana was 15.0 ± 10.0 and that for pineapple was 12.2 ± 4.16. In this experiment, operation error was occurred once in the odor detection test, but there were no subjects who could not be assessed. There were also no subjects who could not detect in 80. In the odor identification test, the accuracy rate of banana in Trial 1 100%, in Trial 2 was 88.9%, and that of rose in Trial 1 and 2 were 66.7%. Although we added "no answer" to choices in the improvements, the subjects did not select the choice. Thus, the adding the choice did not affect the results. In addition, we could conduct the whole experiments for all subjects within 11 minutes including the time for description. Except for the time for description, the odor detection test could be conducted in about five minutes, and the odor identification test could be conducted in about one and a half minutes. In other words, measurement time per subject was about six and a half minutes. Then, we examined the olfactory senses of controls as well. There were no differences between the detection thresholds of banana and pineapple by the t-test ($p = 0.51 > 0.05$). Therefore, it is thought that the kinds of odor did not influence the result of detection thresholds. Next, we examined the differences between the results of identification Trial1 (odor and illustration) and Trial2 (odor and odor) by the z-test, and there were no differences (banana: $p = 0.29 > 0.05$, rose: $p = 1.00 > 0.05$). As with the children with ASD, significant differences could not be found in the odor detection ability between banana and pineapple, and in the odor identification ability between Trial1 and Trial2. In short, there were also no differences between the two trials in the case of controls. However, there were significant differences between the kinds of odor in Trial1 ($p = 0.03 < 0.05$). Although there were no differences between the kinds of odor, there was also the tendency that the accuracy rate of banana was higher than that of rose in Trial2 of the ASD group and the control group. We thought the reason for this was the subjects could remember the odor of banana easier than that of rose because banana was more familiar to human life than rose. Moreover, both odors of rose and lavender were the odor of flower. There was a possibility that the choices of answer affected the accuracy rate of rose.

Table 1: The results of detection threshold (M ± SD).

Group	Banana	Pineapple
ASD	65.7 ± 55.0	46.9 ± 44.3
Control	15.0 ± 10.0	12.2 ± 4.16

Table 2: The accuracy rate in odor identification test [%].

Group	Trial1		Trial2	
	Banana	Rose	Banana	Rose
ASD	73.3	73.3	73.3	60.0
Control	100	66.7	88.9	66.7

We compared the result of the children with ASD and the typically developing children (Table1, Table2). The detection thresholds of banana differed significantly between the ASD group and the control group ($p = 0.006 < 0.01$), and pineapple differed significantly as well ($p = 0.02 < 0.05$) by the t-test. Thus, it can be said that the odor detection ability of children with ASD is impaired. Next, there were no differences between the identification ability of Trial 1 of the ASD group and that of the control group (banana: $p = 0.09 > 0.05$, rose: $p = 0.73 > 0.05$) by the z-test, as well as the identification ability of Trial 2 (banana: $p = 0.36 > 0.05$, rose: $p = 0.74 > 0.05$). In other words, it can be said that there are no differences in the odor identification ability between ASD group and control group. In summary, we obtained the results that the odor detection ability of children with ASD was worse compared to typically developing children, but there were no differences in the odor identification ability. From these results, we thought that the ASD group could detect the odors in the odor identification test. In other words, the ASD group could perceive odors in olfactory function during the test. We thought that the part of the brain that processes perception was immature when they were young. We also thought that the reason why there were no differences between the identification abilities of the ASD group and the control group was that the subjects were young. As the symptoms of ASD may change with age [7], there were no differences in the case of children. We thought that the odor identification ability of children with ASD would become impaired at subsequent developmental stage from the related work [8].

The current results were in line with our hypothesis, but the experimental conditions were different. To compare the olfactory abilities accurately, we should reassess children with ASD using the improved application. However, we found on previous study that there was a tendency of the odor detection threshold to be lower using the raising method than the descending method. Therefore, we expect there is the same tendency when we conduct the experiments for the children with ASD again. We could also assess the olfactory abilities of the children with ASD

precisely by our olfactory display, and we obtained the similar results as in the related works [13]. As the subjects of the related work and our work were close in age, there is a high possibility that the odor detection ability is impaired in the children with ASD. However, we should assess the olfactory abilities by using other kinds of odor in order to examine in more detail if subjects can have the longer experiments. In such case, we should select the combination of odors more firmly in especially the odor identification test. Moreover, it is necessary to increase the sample size.

6 CONCLUSION

In this study, we developed the screening application for children with ASD. Since it is known that the patients with ASD have characteristics related to olfaction, we introduced the odor detection and identification test for screening ASD. The application targeted children because it is important to find the patients with ASD while young. Therefore, we incorporated gaming element in the olfactory test so as not to bore the children. Moreover, we used the olfactory display in the examination in order to minimize the problems in the current olfactory measurement method used in Japan. We conducted the experiments to assess the olfactory abilities of children with ASD using our application so as to develop the way of screening ASD. During the experiments, we found some problems with the application, and we improved the application. After the improvements, we conducted the experiments for typically developing children, and compared to the results of both groups, the children with ASD had less ability of odor detection. However, there were no differences in the odor identification ability between the ASD group and the control group. In short, we were able to see similar results as in the related works on olfactory abilities in children with ASD by the olfactory display using pulse ejection, and the result was in line with our hypothesis. In addition, we could conduct the experiments for all subjects within 15 minutes. This is very shorter than the experiment time for the existing tests. In the future, we will conduct the experiments for children with ASD and typically developing children under the same experimental conditions, and hope to contribute to the construction of the screening tests for ASD by increasing sample size.

ACKNOWLEDGMENTS

This work is supported in part by a KAKENHI, 2014, No. 26330229 from the Ministry of Education, Culture, Sport, Science and Technology Japan, Dr. Kumazaki at Research Center for Child Mental Development, University of Fukui, and Takasago International Corporation.

REFERENCES

[1] R. Stoner et al., “Patches of Disorganization in the Neocortex of Children with Autism,” The New England Journal of Medicine, Vol.370, No.13, pp.1209-1219 (2014).

[2] M. Nishiwaki, “Early Intervention for an Improvement of ASD Children’s Development,” Journal of Aichi University of Education Center for Clinical Practice in Education, Vol.3, pp.47-54 (2013) (in Japanese).

[3] S. Baron-Cohen, S. Wheelwright, R. Skinner, J. Martin, and E. Clubley, “The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Males and Females, Scientists and Mathematicians,” Journal of Autism and Developmental Disorder, Vol.31, No.1, pp.5-17 (2001).

[4] D.L. Robins, D. Fein, M.L. Barton, and J.A. Green, “The Modified Checklist for Autism in Toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders,” Journal of Autism and Developmental Disorders, Vol.31, No.2, pp.131-44 (2001).

[5] N. Inada, T. Koyama, E. Inokuchi, M. Kuroda, and Y. Kamio, “Reliability and validity of the Japanese version of the Modified Checklist for autism in toddlers (M-CHAT),” Reseach in Autism Spectrum Disorders, Vol.5, No.1, pp.330-336 (2011).

[6] S.D. Tomchek, W. Dunn, “Sensory Processing in Children With and Without Autism: A Comparative Study Using the Short Sensory Profile,” The American Journal of Occupational Therapy, Vol.61, No.2, pp.190-200 (2007).

[7] E.Gal, SA.Cermak, A.Ben-Sasson, “Sensory Processing Disorders in Children with Autism: Nature, Assessment, and Intervention,” Growing Up with Autism: Working with School-Age Children and Adolescents, The Guilford Publishers, pp.95-123 (2007).

[8] R.L. Doty, R.E. Frye, U. Agrawal, “Internal consistency reliability of the fractionated and whole University of Pennsylvania Smell Identification Test,” Perception & Psychophysics, Vol. 45, No.5, pp.381-384 (1989).

[9] Y. Suzuki, H. Critchley, A. Rowe, P. Howlin, D.G.M. Murphy, “Impaired Olfactory Identification in Asperger’s Syndrome,” Journal of Neuropsychiatry and Clinical Neuroscience, Vol.15, pp.105-107 (2003).

[10] L. Bennetto, E.S. Kuschner, S.L. Hyman, “Olfaction and Taste Processing in Autism,” Biological Psychiatry, Vol62. No.9, pp.1015-1021 (2007).

[11] T. Hummel, B. Sekinger, S.R. Wolf, E. Paul, and G. Kobal, “‘Sniffin’ Sticks’: Olfactory Performance Assessed by the Combined Testing of Odor Identification, Odor Discrimination and Olfactory Threshold,” Chemical Senses, vol.22 No.1, pp.39-52 (1997).

[12] M. Hrdlicka et al., “Brief Report: Significant Differences in Perceived Odor Pleasantness Found in Children with ASD,” Journal of Autism and Developmental Disorders, Vol.41, No.4, pp.524-527 (2011).

[13] I. Dudova et al. “Odor detection thresholds, but not odor identification, is impaired in children with

autism," European Child & Adolescent Psychiatry, Vol.20, No.7, pp.333-340 (2011).

(Received November 20, 2014)
(Revised April 6, 2015)

Eri Matsuura B.S. degree from the Department of Information and Computer Science of Keio University, Japan in 2014. She is currently working toward an M.S. degree in Open and Environment Systems at Keio University. Her research interests include fragrance information processing work.

Risa Suzuki B.S. degree from the Department of Information and Computer Science of Keio University, Japan in 2013. She is currently working toward an M.S. degree in Open and Environment Systems at Keio University. She received Young Researcher Awards from DICOMO 2013. Her research interests include fragrance information processing work.

Shutaro Homma B.S. degree from the Department of Information and Computer Science of Keio University, Japan in 2014. He is currently working toward an M.S. degree in Open and Environment Systems at Keio University. He received Excellent Presentation Awards from DICOMO 2015. His research interests include fragrance information processing work.

Ken-ichi Okada received his B.S., M.S. and ph. D. in instrumentation engineering from Keio University, in 1973, 1975, and 1982, respectively. He is currently a chief professor in the Department of Information and Computer Science at Keio University. His research interests include CSCW, groupware, human computer interaction, and ubiquitous computing. He has published 140 journal papers, 170 international conference papers, and 18 books entitled "Collaboration and Communication", "Designing Communication and Collaboration Support Systems", "Introduction to Groupware" and so on. He is a member of IEEE, ACM, IEICE and IPSJ. He was a chair of SIGGW, a chief editor of IPSJ Magazine, and a chief editor of IPSJ Journal. Dr. Okada received the IPSJ Best Paper Award in 1995, 2000, and 2008, the IPSJ 40 the Anniversary Paper Award in 2000, IPSJ Fellow in 2002 and the IEEE SAINT Best Paper Award in 2004.

A Joint Uplink/Downlink Resource Allocation Scheme in Wireless OFDMA Networks

Bingfeng Liu[†], Huifang Chen[†‡*], and Lei Xie[†‡]

[†]Dept. of Information Science and Electronic Engineering, Zhejiang University
[‡]Zhejiang Provincial Key Laboratory of Information Network Technology
No. 38, Zheda Road, Hangzhou 310027, P. R. China
{21231086, chenhf, xiel}@zju.edu.cn
[*]Corrsponding author: Huifang Chen, +86-571-87951820-217

Abstract- With the increasing demand for interactive services, it is important to investigate joint uplink/downlink resource allocation scheme in wireless networks, because the performance of the interactive service is determined by uplink and downlink jointly. In this paper, we propose a joint uplink/downlink resource allocation (RA) scheme in wireless orthogonal frequency division multiple access (OFDMA) networks with interactive service. The proposed RA scheme consists of a data interaction mechanism and a joint uplink/downlink RA algorithm. In the data interaction mechanism for interactive service, the base station encodes the received data from the interactive users over individual uplink channels using network coding technique, and multicasts the processed information over the same downlink channels. In the proposed RA algorithm, the uplink and downlink resources are allocated jointly for the interactive service. Simulation results show that the proposed joint uplink/downlink RA scheme allocates the resources efficiently in wireless OFDMA networks.

Keywords: Joint uplink/downlink resource allocation; network coding; wireless OFDMA networks

1 INTRODUCTION

Along with the development of mobile service, interactive services with coupled quality requirements between the uplink and downlink, such as video conference, mobile gaming, and so on, have being become more and more popular.

For interactive service, the overall end user satisfaction depends on good quality in both directions of uplink and downlink. Hence, only improving the transmission quality of one direction is not enough. For example, when a user makes a call, the overall user satisfaction is low if the quality of service (QoS) of downlink session is good, while the QoS of uplink session is bad [1]. For interactive communication between two users via the base station (BS) in wireless networks, the close interaction between the uplink and downlink exists such that improving the QoS of uplink or downlink solely is meaningless [2]. Therefore, it is important to address the joint uplink/downlink communication scheme to improve the overall end user satisfaction in wireless networks.

The joint uplink/downlink resource allocation issue in wireless networks has been concerned in some literatures. In [1], based on the states of uplink and downlink, the concept of the overall user satisfaction is introduced. The resource allocation is formulated as an optimization problem, which is solved using the Lagrangian dual approach. However, the analysis is quite simple, and the wireless channel features are not considered. A joint uplink/downlink resource allocation scheme based on the link quality and delay is proposed in [2]. By jointly considering the time varying channel conditions in the uplink and downlink, the proposed algorithm prevents the resource from wasting. In [3], the joint uplink/downlink resource allocation is formulated as an optimization problem with the coupled constraint on the requirements of uplink and downlink rates. The proposed algorithm maximizes the system throughput, as well as minimizes the gap between the allocated uplink and downlink rates. Based on the work in [3], a resource allocation scheme modeled as a two-sided stable matching game is proposed in [4]. Authors in [5] proposed a joint uplink/downlink resource allocation algorithm, where the scheduling process in one direction is controlled by the delay information of the other direction. This algorithm achieves an constrained delay behavior of BS.

On the other hand, Ahlswede, *et al.*, first proposed the concept of network coding in [6], and proved theoretically that the routers mixed information in different messages to achieve multicast capacity. Owing to the broadcast feature of the medium, the network coding technique in wireless networks has broad application prospects and has been received wide attention in recent years. The coded bi-directional relaying is first proposed in [7]. By analyzing the channel capacity, authors proposed a resource allocation scheme, and proved that the proposed scheme can improve the system performance significantly. The resource allocation scheme with network coding for downlink two-user orthogonal frequency division multiple access (OFDMA) system was proposed in [8]. It is concluded that the network coding mechanism can reduce power consumption effectively.

In order to exert the advantage of joint uplink/downlink resource allocation and network coding, we propose a joint uplink/downlink resource allocation scheme in wireless OFDMA networks with the interactive service. In this scheme, the network coding technique is used to improve the downlink resource utilization for the interactive service. Considering interactive users within a cell exchanging data via BS, a bi-directional data interaction model based on network coding is set up. The joint uplink/downlink resource allocation is formulated as an optimization problem with the objective to maximize the system throughput. The formulated problem is solved by the dual decomposition

method. Finally, we evaluate the performance of the proposed joint uplink/downlink resource allocation scheme by simulations.

The remainder of the paper is organized as follows. Section 2 introduces the system model and formulates the problem we aim to resolve. Section 3 presents the proposed solution for the problem formulated in section 2. The performance of the proposed scheme is evaluated and analyzed in section 4. Finally, we conclude the paper in section 5.

2 SYSTEM MODEL AND PROBLEM FORMULATION

2.1 System Model

We consider one cell in wireless FDD-OFDMA network, as shown in Fig. 1. There are M^{U} users using uplink, M^{D} users using downlink, where 2M users using both uplink and downlink form M pairs of interactive users via interactive links. If a user communicates over both a uplink and downlink with a uncoupling mode, we assume that one user only uses the uplink and another only uses the downlink in the system model. In other words, there are three types of users in the network, namely the users using uplink only, the users using downlink only and the interactive users using both uplink and downlink.

K^{U} and K^{D} subcarriers are allocated among users in the uplink with bandwidth B^{U} and the downlink with bandwidth B^{D}, respectively. The bandwidth of a subcarrier, $B^{\mathrm{U}}/K^{\mathrm{U}}$ or $B^{\mathrm{D}}/K^{\mathrm{D}}$, is small enough so that each subcarrier is assumed to experience flat-fading [9].

Furthermore, it is assumed that the BS has perfect knowledge of the channel state information (CSI) between BS and users, which enables BS to allocate available resources dynamically according to the service requirements and channel conditions. In literatures, the resource allocation schemes in the wireless OFDMA network can be classified into two categories, one is based on full CSI and another is based on partial CSI. In the resource allocation algorithms with full CSI, researchers focus on the resource allocation mechanisms and the theoretical performance. And in the resource allocation algorithms with partial CSI, researchers pay attention to the performance of these algorithms in the actual networks. In our work, we focus on the mechanism of the joint uplink/downlink resource allocation scheme in the wireless OFDMA network with interactive service. Hence, we assume that the BS has perfect knowledge of CSI and the difficulty of obtaining the CSI is not considered [10].

It is assumed that the noise is additive white Gaussian noise (AWGN) with the power spectral density N_0. Moreover, the impact of bit error rate (BER) and modulation coded scheme in the physical layer is assumed to be ignored.

2.2 Wireless OFDMA Channel Model

In the uplink, for user m and uplink subcarrier k, the channel gain is denoted as $h^{\mathrm{U}}_{m,k}$. The channel-to-noise ratio (CNR) of user m in uplink subcarrier k is $H^{\mathrm{U}}_{m,k}=\left|h^{\mathrm{U}}_{m,k}\right|^2\Big/\left(N_0B^{\mathrm{U}}/K^{\mathrm{U}}\right)$.

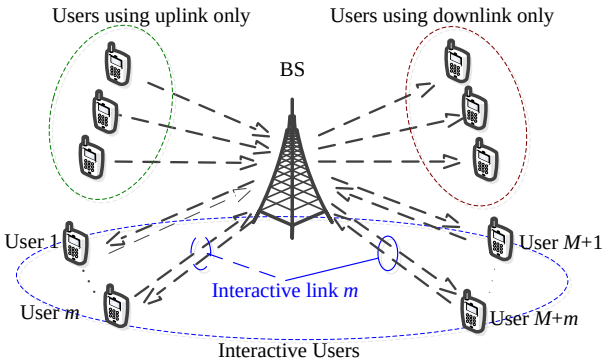


Figure 1: The structure diagram of a wireless network.

Let $\alpha^{\mathrm{U}}_{m,k}$ be the subcarrier allocation factor of user m in uplink subcarrier k. If uplink subcarrier k is allocated to user m, $\alpha^{\mathrm{U}}_{m,k}=1$; otherwise, $\alpha^{\mathrm{U}}_{m,k}=0$.

Let $p^{\mathrm{U}}_{m,k}$ be the power allocated to uplink subcarrier k by user m. Since $p^{\mathrm{U}}_{m,k}$ and $\alpha^{\mathrm{U}}_{m,k}$ are coupled, a new variable, $s^{\mathrm{U}}_{m,k}$ is introduced, and $s^{\mathrm{U}}_{m,k}=\alpha^{\mathrm{U}}_{m,k}p^{\mathrm{U}}_{m,k}$. Thereby, $s^{\mathrm{U}}_{m,k}$ and $\alpha^{\mathrm{U}}_{m,k}$ are decoupled [11].

The allocated rate is defined as the maximum achieved rate using the resource allocated by system. The allocated rate of user m in uplink subcarrier k, $C^{\mathrm{U}}_{m,k}$, is

$$C^{\mathrm{U}}_{m,k}=\frac{B^{\mathrm{U}}}{K^{\mathrm{U}}}\log_2\left(1+\frac{s^{\mathrm{U}}_{m,k}H^{\mathrm{U}}_{m,k}}{\alpha^{\mathrm{U}}_{m,k}}\right),\ m=1,2,...,M^{\mathrm{U}},k=1,2,...,K^{\mathrm{U}}. \quad (1)$$

And the allocated uplink rate of user m, C^{U}_{m}, is

$$C^{\mathrm{U}}_{m}=\sum_{k=1}^{K^{\mathrm{U}}}\alpha^{\mathrm{U}}_{m,k}C^{\mathrm{U}}_{m,k},\ m=1,2,...,M^{\mathrm{U}}. \quad (2)$$

In the downlink, for user m and downlink subcarrier k, the channel gain is denoted as $h^{\mathrm{D}}_{m,k}$. The CNR for user m in downlink subcarrier k is $H^{\mathrm{D}}_{m,k}=\left|h^{\mathrm{D}}_{m,k}\right|^2\Big/\left(N_0B^{\mathrm{D}}/K^{\mathrm{D}}\right)$.

Let $\alpha^{\mathrm{D}}_{m,k}$ be the subcarrier allocation factor of user m in downlink subcarrier k. If downlink subcarrier k is allocated to user m, $\alpha^{\mathrm{D}}_{m,k}=1$; otherwise, $\alpha^{\mathrm{D}}_{m,k}=0$.

Let $p^{\mathrm{D}}_{m,k}$ be the power allocated to user m in downlink subcarrier k by BS. Similarly, we introduce a new variable, and $s^{\mathrm{D}}_{m,k}=\alpha^{\mathrm{D}}_{m,k}p^{\mathrm{D}}_{m,k}$.

The allocated rate of user m in downlink subcarrier k, $C^{\mathrm{D}}_{m,k}$, is

$$C^{\mathrm{D}}_{m,k}=\frac{B^{\mathrm{D}}}{K^{\mathrm{D}}}\log_2\left(1+\frac{s^{\mathrm{D}}_{m,k}H^{\mathrm{D}}_{m,k}}{\alpha^{\mathrm{D}}_{m,k}}\right),m=1,2,...,M^{\mathrm{D}},k=1,2,...,K^{\mathrm{D}}. \quad (3)$$

And the allocated downlink rate of user m, C_m^{D}, is

$$C_m^{\mathrm{D}} = \sum_{k=1}^{K^{\mathrm{D}}} \alpha_{m,k}^{\mathrm{D}} C_{m,k}^{\mathrm{D}}, m = 1,2,...,M^{\mathrm{D}}. \quad (4)$$

2.3 Network-coding-based Bi-directional Data Interaction Model for Interactive Service

In network-coding-based bi-directional communication for interactive service, a pair of interactive users exchange data via BS when they locate within a cell.

The flow diagram of network-coding-based bi-directional data interaction between users A and B is shown in Fig. 2. The interactive data exchange process works as follows.

Step I: Two interactive users transmit data X and Y to the BS through the uplink channels separately.

Step II: Receiving X and Y, the BS encodes X and Y, e.g., with an XOR bitwise operation, $X \oplus Y$.

Step III: The BS multicasts encoded data, $X \oplus Y$, to interactive users. For decoding at user A, a bitwise XOR operation of $X \oplus Y$ and X is performed to obtain Y. Another user performs the corresponding operation to obtain X.

The design of the network-coding-based bi-directional communication is based on two key principles.

(1). The broadcast property of radio is exploited to implement a one-to-many multicast communication.

(2). Network coding makes it possible to implement multicast communication in the downlink for interactive service. That is, the data from two interactive users are mixed at the BS before forwarding them. Users decode the mixed data independently.

As shown in Fig. 1, user m and user $M+m$ are a pair of interactive users of the interactive link m. That is, user $M+m$ is the destination of user m, and vice versa. In this work, we only consider the interactive service with symmetric rate. However, the proposed approach here can be extended for generalized interactive service easily.

Due to the feature of symmetric interactive service, the maximum uplink rate of the interactive link m, R_m^{U}, is determined by the user with less allocated uplink rate. That is,

$$R_m^{\mathrm{U}} = \min\left\{C_m^{\mathrm{U}}, C_{M+m}^{\mathrm{U}}\right\}, m = 1,2,...,M. \quad (5)$$

In the downlink, the BS transmits encoded data to the pair of users of each interactive link by multicast. The downlink rate of interactive link m in downlink subcarrier k is determined by the user with less allocated downlink rate in this subcarrier. Hence, the maximum downlink rate of the interactive link m, R_m^{D}, is

$$R_m^{\mathrm{D}} = \sum_{k=1}^{K^{\mathrm{D}}} \min\left\{C_{m,k}^{\mathrm{D}}, C_{M+m,k}^{\mathrm{D}}\right\}, m = 1,2,...,M. \quad (6)$$

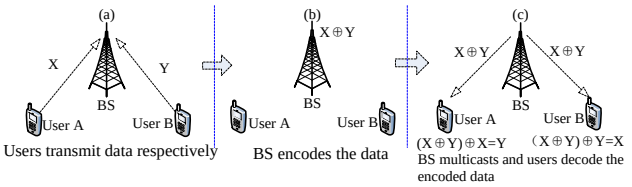


Figure 2: Network-coding-based bi-directional communication between interactive users.

When the BS encodes the data from two interactive users, massage padding of the shorter data packet is needed if two uplink data packets are unequal length. The length of encoded data packet equals to the length of longer data packet [12].

In addition, we assume that all users continue to send to or receive data and the data flows are continuous as similar as in [12]. In other words, data always exist in the cache of users and BS waiting for sending. And the impact of the flow dynamics on the resource allocation is not considered. To ensure the interactive users receiving and decoding data, the uplink allocated rate of interactive users should be less than the maximum downlink rate. Therefore, the interactive rate of the interactive link m, R_m, is determined by the minimum of the maximum uplink and downlink rate. That is,

$$R_m = \min\left\{R_m^{\mathrm{U}}, R_m^{\mathrm{D}}\right\} = \min\left\{C_m^{\mathrm{U}}, C_{M+m}^{\mathrm{U}}, R_m^{\mathrm{D}}\right\}, m = 1,2,...,M. \quad (7)$$

For interactive users, the actual uplink and downlink rates equal to the interactive rate of interactive link. If the uplink and downlink allocated rates exceed to the interactive rate, resources will be wasted.

For non-interactive users, the actual uplink and downlink rates equal to the uplink and downlink allocated rates, respectively.

2.4 Problem Formulation

Each uplink subcarrier is exclusively allocated to no more than one user. Hence,

$$\sum_{m=1}^{M^{\mathrm{U}}} \alpha_{m,k}^{\mathrm{U}} \le 1, \ \alpha_{m,k}^{\mathrm{U}} \in \{0,1\}, k = 1,2,...,K^{\mathrm{U}}. \quad (8)$$

Let P_m be the maximum power budget of user m. We have

$$\sum_{k=1}^{K^{\mathrm{U}}} s_{m,k}^{\mathrm{U}} \le P_m, m=1,2,...,M^{\mathrm{U}}. \quad (9)$$

For interactive service, BS transmits the encoded data to interactive users by multicast. Therefore, the downlink subcarriers allocated to the users of each interactive link are identical. That is,

$$\alpha_{m,k}^{\mathrm{D}} = \alpha_{M+m,k}^{\mathrm{D}}, m=1,2,...,M, \ k = 1,2,...,K^{\mathrm{D}}. \quad (10)$$

In the downlink of traditional wireless OFDMA networks, each downlink subcarrier is exclusively allocated to no more

than one user. However, in our proposed scheme, the downlink subcarriers can be are allocated to the two users of one interactive link simultaneously. Hence,

$$\sum_{m=M+1}^{M^{\mathrm{D}}} \alpha_{m,k}^{\mathrm{D}} \le 1,\ \alpha_{m,k}^{\mathrm{D}} \in \{0,1\},\ k=1,2,...,K^{\mathrm{D}}. \quad (11)$$

Let P_{BS} be the maximum power of BS. We have

$$\sum_{m=M+1}^{M^{\mathrm{D}}} \sum_{k=1}^{K^{\mathrm{D}}} s_{m,k}^{\mathrm{D}} \le P_{\mathrm{BS}}. \quad (12)$$

Define the system throughput be the sum of allocated uplink and downlink rates of all users. Considering the interactive service, the uplink and downlink resources should be allocated jointly to maximize the system throughput.

Therefore, the joint uplink/downlink resource allocation problem in the wireless OFDMA network with the interactive service can be formulated as

$$\begin{aligned} \max_{\{s_{m,k}^{\mathrm{U}}, s_{m,k}^{\mathrm{D}}, \alpha_{m,k}^{\mathrm{U}}, \alpha_{m,k}^{\mathrm{D}}\}} & \left(4\sum_{m=1}^{M} R_m + \sum_{m=2M+1}^{M^{\mathrm{U}}} C_m^{\mathrm{U}} + \sum_{m=2M+1}^{M^{\mathrm{D}}} C_m^{\mathrm{D}} \right) \\ \text{s.t. C1: } & \sum_{m=1}^{M^{\mathrm{U}}} \alpha_{m,k}^{\mathrm{U}} \le 1,\ \alpha_{m,k}^{\mathrm{U}} \in \{0,1\},\ k=1,2,...,K^{\mathrm{U}} \\ \text{C2: } & \sum_{k=1}^{K^{\mathrm{U}}} s_{m,k}^{\mathrm{U}} \le P_m,\ m=1,2,...,M^{\mathrm{U}} \\ \text{C3: } & \sum_{m=M+1}^{M^{\mathrm{D}}} \alpha_{m,k}^{\mathrm{D}} \le 1,\ \alpha_{m,k}^{\mathrm{D}} \in \{0,1\},\ k=1,2,...,K^{\mathrm{D}} \\ \text{C4: } & \alpha_{m,k}^{\mathrm{D}} = \alpha_{M+m,k}^{\mathrm{D}},\ m=1,2,...,M,\ k=1,2,...,K^{\mathrm{D}} \\ \text{C5: } & \sum_{m=M+1}^{M^{\mathrm{D}}} \sum_{k=1}^{K^{\mathrm{D}}} s_{m,k}^{\mathrm{D}} \le P_{\mathrm{BS}} \end{aligned}, \quad (13)$$

where C1 and C2 are the subcarriers and power allocation constraints of the uplink, respectively. C3 and C5 are the subcarriers and power constraints of the downlink, respectively. And C4 is the subcarriers allocation constraint of downlink for the interactive users.

3 PROBLEM SOLUTION

Obviously, the problem formulated in (13) is a mixed integer nonlinear programming problem. By introducing new variables, the problem can be formulated into a convex optimization problem [13]. That is

$$\begin{aligned} \max_{\{s_{m,k}^{\mathrm{U}}, s_{m,k}^{\mathrm{D}}, \alpha_{m,k}^{\mathrm{U}}, \alpha_{m,k}^{\mathrm{D}}\}} & \left(4\sum_{m=1}^{M} t_m + \sum_{m=2M+1}^{M^{\mathrm{U}}} C_m^{\mathrm{U}} + \sum_{m=2M+1}^{M^{\mathrm{D}}} C_m^{\mathrm{D}} \right) \\ \text{s.t. } & \text{C1~C5 in eq. (13)} \\ & \text{C6: } t_m \le C_m^{\mathrm{U}},\ m=1,2,...,M \\ & \text{C7: } t_m \le C_{M+m}^{\mathrm{U}},\ m=1,2,...,M \\ & \text{C8: } t_m \le R_m^{\mathrm{D}},\ m=1,2,...,M \end{aligned}. \quad (14)$$

It is shown in [14] that the dual gap of resource allocation problem in multicarrier system is nearly zero if the number of subcarriers is sufficiently large. And the dual problem is convex regardless of the convexity of the primal problem, and it can be solved easily [15]. Therefore, the problem formulated in (14) can be solved by the dual decomposition method.

The Lagrange function of the problem is given as

$$\begin{aligned} L = & \sum_{m=1}^{M} t_m \left(4 - \lambda_m^{\mathrm{U}} - \lambda_{M+m}^{\mathrm{U}} - \lambda_m^{\mathrm{D}}\right) + \sum_{m=1}^{M^{\mathrm{U}}} \sum_{k=1}^{K^{\mathrm{U}}} \alpha_{m,k}^{\mathrm{U}} f^{\mathrm{U}}\left(s_{m,k}^{\mathrm{U}}\right) \\ & + \sum_{m=M+1}^{M^{\mathrm{D}}} \alpha_{m,k}^{\mathrm{D}} f^{\mathrm{D}}\left(s_{m,k}^{\mathrm{D}}\right) + \sum_{m=1}^{M^{\mathrm{U}}} \mu_m^{\mathrm{U}} P_m + \mu^{\mathrm{D}} P_{\mathrm{BS}} \end{aligned}, \quad (15)$$

where λ_m^{U}, λ_m^{D}, μ_m^{U}, μ^{D} are the Lagrange multipliers. And

$$f^{\mathrm{U}}\left(s_{m,k}^{\mathrm{U}}\right) = \begin{cases} \lambda_m^{\mathrm{U}} C_{m,k}^{\mathrm{U}} - \mu_m^{\mathrm{U}} \dfrac{s_{m,k}^{\mathrm{U}}}{\alpha_{m,k}^{\mathrm{U}}}, m=1,2,...,2M \\ C_{m,k}^{\mathrm{U}} - \mu_m^{\mathrm{U}} \dfrac{s_{m,k}^{\mathrm{U}}}{\alpha_{m,k}^{\mathrm{U}}}, m=2M+1,...,M^{\mathrm{U}} \end{cases}, \quad (16)$$

and

$$f^{\mathrm{D}}\left(s_{m,k}^{\mathrm{D}}\right) = \begin{cases} \lambda_{m-M}^{\mathrm{D}} \min\left\{C_{m,k}^{\mathrm{D}}, C_{m-M,k}^{\mathrm{D}}\right\} - \mu^{\mathrm{D}} \dfrac{s_{m-M,k}^{\mathrm{D}}}{\alpha_{m-M,k}^{\mathrm{D}}}, m=M+1,...,2M \\ C_{m,k}^{\mathrm{D}} - \mu^{\mathrm{D}} \dfrac{s_{m,k}^{\mathrm{D}}}{\alpha_{m,k}^{\mathrm{D}}}, m=2M+1,...,M^{\mathrm{D}} \end{cases}. \quad (17)$$

The dual problem of (14) is given by

$$\min_{\{\lambda_m^{\mathrm{U}}, \lambda_m^{\mathrm{D}}, \mu_m^{\mathrm{U}}, \mu^{\mathrm{D}}\}} D, \quad (18)$$

where is D the Lagrange dual function as

$$D = \max_{\{s_{m,k}^{\mathrm{U}}, s_{m,k}^{\mathrm{D}}, \alpha_{m,k}^{\mathrm{U}}, \alpha_{m,k}^{\mathrm{D}}, t_m\}} L. \quad (19)$$

Taking the derivative of L with respect to t_m and setting it to be zero yields

$$4 - \lambda_m^{\mathrm{U}} - \lambda_{M+m}^{\mathrm{U}} - \lambda_m^{\mathrm{D}} = 0,\ m=1,2,...,M. \quad (20)$$

Hence, the Lagrange dual function, D, is equivalent to

$$D = \max_{\{s_{m,k}^{\mathrm{U}}, \alpha_{m,k}^{\mathrm{U}}\}} \sum_{m=1}^{M^{\mathrm{U}}} \sum_{k=1}^{K^{\mathrm{U}}} \alpha_{m,k}^{\mathrm{U}} f^{\mathrm{U}}\left(s_{m,k}^{\mathrm{U}}\right) + \max_{\{s_{m,k}^{\mathrm{D}}, \alpha_{m,k}^{\mathrm{D}}\}} \sum_{m=M+1}^{M^{\mathrm{D}}} \sum_{k=1}^{K^{\mathrm{D}}} \alpha_{m,k}^{\mathrm{D}} f^{\mathrm{D}}\left(s_{m,k}^{\mathrm{D}}\right). \quad (21)$$

The problem (21) can be decomposed into the uplink sub-problem and the downlink sub-problem. As $\alpha_{m,k}^{\mathrm{U}}$ and $s_{m,k}^{\mathrm{U}}$ are decoupled, the uplink sub-problem can be formulated as

$$\max_{\{s_{m,k}^{\mathrm{U}}, \alpha_{m,k}^{\mathrm{U}}\}} \sum_{m=1}^{M^{\mathrm{U}}} \sum_{k=1}^{K^{\mathrm{U}}} \alpha_{m,k}^{\mathrm{U}} f^{\mathrm{U}}\left(s_{m,k}^{\mathrm{U}}\right) = \max_{\{\alpha_{m,k}^{\mathrm{U}}\}} \sum_{m=1}^{M^{\mathrm{U}}} \sum_{k=1}^{K^{\mathrm{U}}} \max_{\{s_{m,k}^{\mathrm{U}}\}} \alpha_{m,k}^{\mathrm{U}} f^{\mathrm{U}}\left(s_{m,k}^{\mathrm{U}}\right). \quad (22)$$

The inner maximization of (22) is over the set of allocated power. Hence, we obtain the optimal value as

$$p_{m,k}^{*\mathrm{U}} = \max_{\left\{s_{m,k}^{\mathrm{U}} \ge 0\right\}} \alpha_{m,k}^{\mathrm{U}} f^{\mathrm{U}}\left(s_{m,k}^{\mathrm{U}}\right). \quad (23)$$

Taking the derivative of $f^{\mathrm{U}}\left(s_{m,k}^{\mathrm{U}}\right)$ with respect to $s_{m,k}^{\mathrm{U}}$, and setting them to be zero yields

$$p_{m,k}^{*\mathrm{U}} = \frac{s_{m,k}^{*\mathrm{U}}}{\alpha_{m,k}^{\mathrm{U}}} = \begin{cases} \left[\frac{\lambda_m^{\mathrm{U}}}{\mu_m^{\mathrm{U}} \ln 2} - \frac{1}{H_{m,k}^{\mathrm{U}}}\right]^+, m = 1,2,...,2M \\ \left[\frac{1}{\mu_m^{\mathrm{U}} \ln 2} - \frac{1}{H_{m,k}^{\mathrm{U}}}\right]^+, m = 2M+1,...,M^{\mathrm{U}} \end{cases}, \quad (24)$$

where $[x]^+ = \max\{0, x\}$.

The subcarrier allocation solution is obtained by considering the outer maximization of (22). That is,

$$\alpha_{m,k}^{*\mathrm{U}} = \max_{\left\{\alpha_{m,k}^{\mathrm{U}}\right\}} \sum_{m=1}^{M^{\mathrm{U}}} \sum_{k=1}^{K^{\mathrm{U}}} \alpha_{m,k}^{\mathrm{U}} f^{\mathrm{U}}\left(s_{m,k}^{*\mathrm{U}}\right). \quad (25)$$

Note that each uplink subcarrier is exclusively allocated to no more than one user anytime. Hence,

$$\sum_{m=1}^{M^{\mathrm{U}}} \sum_{k=1}^{K^{\mathrm{U}}} \alpha_{m,k}^{\mathrm{U}} f^{\mathrm{U}}\left(s_{m,k}^{*\mathrm{U}}\right) \le \sum_{k=1}^{K^{\mathrm{U}}} \arg\max_m f^{\mathrm{U}}\left(s_{m,k}^{*\mathrm{U}}\right). \quad (26)$$

Therefore, $\alpha_{m,k}^{\mathrm{U}}$ is obtained for each subcarrier k by finding the user m to maximize $f^{\mathrm{U}}\left(s_{m,k}^{*\mathrm{U}}\right)$. That is,

$$\alpha_{m,k}^{*\mathrm{U}} = \begin{cases} 1, \text{ if } k = \arg\max_m f^{\mathrm{U}}\left(s_{m,k}^{*\mathrm{U}}\right) \\ 0, \text{ otherwise} \end{cases}. \quad (27)$$

Similarly, the optimal power and subcarriers allocation in the downlink are obtained as

$$p_{m,k}^{*\mathrm{D}} = \begin{cases} \left[\frac{\lambda_{m-M}^{\mathrm{D}}}{\mu^{\mathrm{D}} \ln 2} - \frac{1}{\min\left\{H_{m,k}^{\mathrm{D}}, H_{m-M,k}^{\mathrm{D}}\right\}}\right]^+, m = M+1,...,2M \\ \left[\frac{1}{\mu^{\mathrm{D}} \ln 2} - \frac{1}{H_{m,k}^{\mathrm{D}}}\right]^+, m = 2M+1,...,M^{\mathrm{D}} \end{cases}, \quad (28)$$

and

$$\alpha_{m,k}^{*\mathrm{D}} = \begin{cases} 1, \text{ if } k = \arg\max_m f^{\mathrm{D}}\left(s_{m,k}^{*\mathrm{D}}\right) \\ 0, \text{ otherwise} \end{cases}. \quad (29)$$

Substituting (24), (27), (28) and (29) into (18), we obtain the dual problem of (14) is

$$\min_{\left\{\lambda_m^{\mathrm{U}}, \lambda_m^{\mathrm{D}}, \mu_m^{\mathrm{U}}, \mu^{\mathrm{D}}\right\}} \left[\sum_{m=1}^{M^{\mathrm{U}}} \sum_{k=1}^{K^{\mathrm{U}}} \alpha_{m,k}^{*\mathrm{U}} f^{\mathrm{U}}\left(s_{m,k}^{*\mathrm{U}}\right) + \sum_{m=M+1}^{M^{\mathrm{D}}} \sum_{k=1}^{K^{\mathrm{D}}} \alpha_{m,k}^{*\mathrm{D}} f^{\mathrm{D}}\left(s_{m,k}^{*\mathrm{D}}\right) + \sum_{m=1}^{M^{\mathrm{U}}} \mu_m^{\mathrm{U}} P_m + \mu^{\mathrm{D}} P_{\mathrm{BS}}\right] \quad .(30)$$

s.t. $\lambda_m^{\mathrm{U}} + \lambda_{M+m}^{\mathrm{U}} + \lambda_m^{\mathrm{D}} = 4,\ m=1,2,...,M$

$\lambda_m^{\mathrm{U}}, \lambda_m^{\mathrm{D}}, \mu_m^{\mathrm{U}}, \mu^{\mathrm{D}} \ge 0,\ m=1,2,...,M$

The problem formulated in (30) can be solved by sub-gradient algorithm by updating the Lagrange multipliers simultaneously along the sub-gradients of them. Because the problem is convex, the sub-gradient update algorithm is convergent.

The sub-gradients of the Lagrange multipliers are

$$\begin{aligned} \Delta\lambda_m^{\mathrm{U}} &= C_m^{\mathrm{U}}(\alpha_{m,k}^{*\mathrm{U}}, s_{m,k}^{*\mathrm{U}}) - R_m^{\mathrm{D}}(\alpha_{m,k}^{*\mathrm{D}}, s_{m,k}^{*\mathrm{D}}),\ m=1,2,...,M \\ \Delta\lambda_{M+m}^{\mathrm{U}} &= C_{M+m}^{\mathrm{U}}(\alpha_{M+m,k}^{*\mathrm{U}}, s_{M+m,k}^{*\mathrm{U}}) - R_m^{\mathrm{D}}(\alpha_{m,k}^{*\mathrm{D}}, s_{m,k}^{*\mathrm{D}}),\ m=1,2,...,M \\ \Delta\mu_m^{\mathrm{U}} &= P_m - \sum_{k=1}^{K^{\mathrm{U}}} s_{m,k}^{*\mathrm{U}},\ m = 1,2,...,M^{\mathrm{U}} \\ \Delta\mu^{\mathrm{D}} &= P_{\mathrm{BS}} - \sum_{m=M+1}^{M^{\mathrm{D}}} \sum_{k=1}^{K^{\mathrm{D}}} s_{m,k}^{*\mathrm{D}} \end{aligned}, \quad (31)$$

where $C_m^{\mathrm{U}}(\alpha_{m,k}^{*\mathrm{U}}, s_{m,k}^{*\mathrm{U}})$, $C_m^{\mathrm{U}}(\alpha_{M+m,k}^{*\mathrm{U}}, s_{M+m,k}^{*\mathrm{U}})$ and $R_m^{\mathrm{D}}(\alpha_{m,k}^{*\mathrm{D}}, s_{m,k}^{*\mathrm{D}})$ are obtained by substituting (24), (27), (28) and (29) into (2), (4) and (6), respectively. And the Lagrange multipliers are updated as

$$\begin{aligned} \lambda_m^{\mathrm{U}(l+1)} &= \left[\lambda_m^{\mathrm{U}(l)} - \varsigma_1^{(l)} \Delta\lambda_m^{\mathrm{U}(l)}\right]^+,\ m=1,2,...,M \\ \lambda_{M+m}^{\mathrm{U}(l+1)} &= \left[\lambda_{M+m}^{\mathrm{U}(l)} - \varsigma_2^{(l)} \Delta\lambda_{M+m}^{\mathrm{U}(l)}\right]^+,\ m=1,2,...,M \\ \mu_m^{\mathrm{U}(l+1)} &= \left[\mu_m^{\mathrm{U}(l)} - \varsigma_3^{(l)} \Delta\mu_m^{\mathrm{U}(l)}\right]^+,\ m = 1,2,...,M^{\mathrm{U}} \\ \mu^{\mathrm{D}(l+1)} &= \left[\mu^{\mathrm{D}(l)} - \varsigma_4^{(l)} \Delta\mu^{\mathrm{D}(l)}\right]^+ \end{aligned}, \quad (32)$$

where, l is the iteration number, $\varsigma_1^{(l)}$, $\varsigma_2^{(l)}$, $\varsigma_3^{(l)}$ and $\varsigma_4^{(l)}$ are the step size in the lth iteration.

Substituting the obtained optimal solution of Lagrange multipliers to (24), (27) and (28), we solve the optimal subcarriers and power allocation of original problem.

By decomposing at each iteration, we can decouple the design in uplink and downlink. In the uplink, there are K^{U} subcarriers to be allocated M^{U} users, and the computational complexity is O($K^{\mathrm{U}}M^{\mathrm{U}}$) per iteration. In the downlink, there are K^{D} subcarriers to be allocated to M pairs of interactive users and (M^{D}-2M) users, and the computational complexity is O(K^{D}(M^{D}-M)) per iteration. Hence, the total computational complex at each iteration is O($K^{\mathrm{U}}M^{\mathrm{U}}$+($M^{\mathrm{D}}$-$M$)$K^{\mathrm{D}}$). Therefore, the computational complexity of the proposed scheme is comparatively low.

4 SIMULATION RESULTS AND DISCUSSIONS

In this section, we will evaluate the proposed joint uplink/downlink resource allocation scheme with Monte Carlo simulations.

Table 1. Simulation Parameters

Parameters	value
Bandwidth of uplink/downlink, B^{U}/B^{D}	1 MHz
Number of uplink/downlink subcarriers, K^{U}/K^{D}	128
Power spectral density of AWGN, N_0	-174 dBm/Hz
Maximum power of users, P_m	0.125 W
Maximum power of BS, P_{BS}	2 W

It is assumed that the channel fading of each subcarrier follows an independent Rayleigh distribution. The channel gain is exponentially distributed, and the propagation loss is represented as $\kappa d_m^{-\chi}$ [16], where κ is a constant chosen to be −128.1dB, χ is the path loss exponent set to be 3.76, and d_m is the distance between user m to the BS. The other simulation parameters are listed in Table 1.

For comparison, the performances of the following three algorithms are given.

(1). The resource allocation algorithm is to maximize the throughput in the uplink and downlink directions independently, proposed in [10] and [17]. It is denoted as 'Independent RA algorithm'.

(2). The resource allocation algorithm is to maximize the throughput in the uplink and downlink directions jointly. It is denoted as 'Joint RA algorithm'.

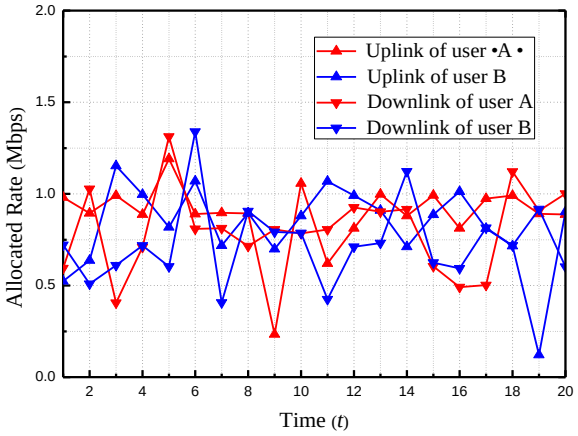
(3). The proposed resource allocation scheme is to maximize the throughput in the uplink and downlink directions jointly, where the network coding (NC) is used in the downlink multicast transmission for interactive service. It is denoted as 'Joint RA algorithm with NC'.

From Fig. 3 to Fig. 6, it is assumed that 24 users are distributed in a circular cell. The first 8 users in uplink and downlink form 4 pairs of interactive users via 4 interactive links. And other 16 users are 8 users using uplink only and 8 users using downlink only, respectively.

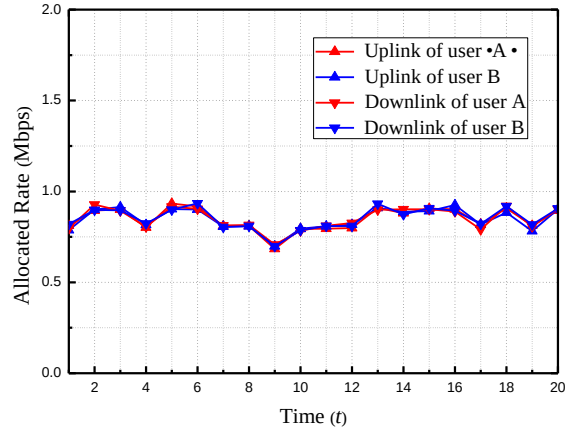
Figure 3 shows the comparison of the allocated uplink and downlink rates of a selected pair of interactive users using three resource allocation algorithms. The distances of all users to BS are 0.4 km, and the selected interactive users are denoted as user A and user B, respectively.

From Fig. 3(a), we observe that the allocated uplink and downlink rates of interactive users are independent each other. The system allocates resources according to the channel conditions for maximizing the uplink and downlink throughput. From Fig. 3(b) and 3(c), we observe that the allocated uplink and downlink rates of interactive users are fully coupled. The reason for this phenomenon is that Joint RA algorithm and Joint RA algorithm with NC minimize the gap between the uplink and downlink allocated rates of interactive users to save resources.

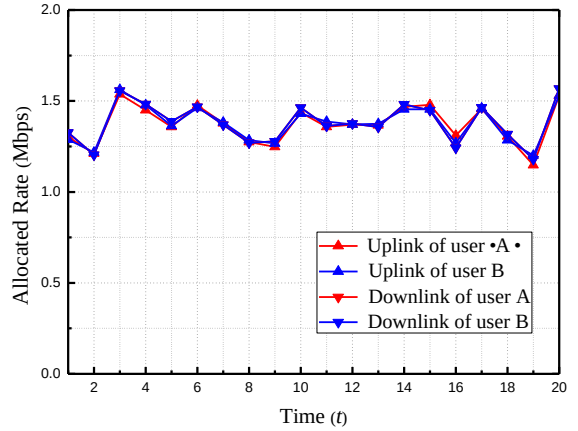
Moreover, comparing the results in Fig. 3(b) and 3(c), we observe that the allocated uplink and downlink rates of Joint RA algorithm with NC outperform Joint RA algorithm. The reason for this phenomenon is that Joint RA algorithm with NC improves the resource efficiency of downlink by network coding technique.



(a) Independent RA algorithm



(b) Joint RA algorithm



(c) Joint RA algorithm with NC

Figure 3: The comparison of the allocated uplink and downlink rates of a selected pair of interactive users.

Figure 4 shows the comparison of system throughput for three different algorithms, where the cell radius varies from 0.2 to 1.2 km, and all users are distributed uniformly.

From Fig. 4, we observe that, as the cell radius increases, the average channel condition of all users will deteriorate, and then the system throughput of three algorithms decreases. Moreover, due to fully coupling of the allocated uplink and downlink rates of each pair of interactive users, Joint RA algorithm and Joint RA algorithm with NC reduce the waste of the resource, and their system throughput

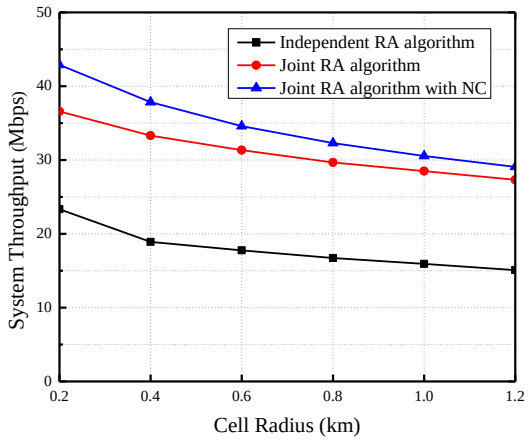


Figure 4: The comparison of the system throughput with different cell radius.

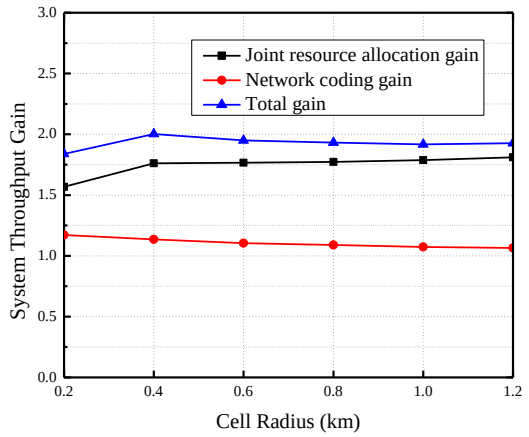


Figure 5: The comparison of the system throughput gain with different cell radius.

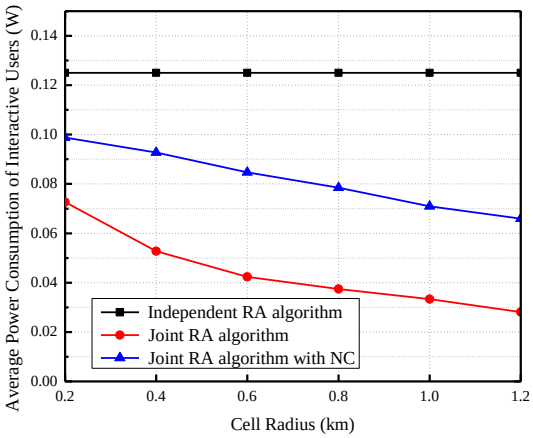


Figure 6: The comparison of average power consumption of interactive users.

outperforms that of Independent RA algorithm. However, owing to the network coding processing at BS, Joint RA algorithm with NC makes the resource utilization be more efficient, and achieves the largest system throughput among three RA algorithms.

In order to analyze the system throughput gain, Fig. 5 shows the comparison of the system throughput gain for three different algorithms, where the cell radius varies from 0.2 to 1.2 km.

The joint resource allocation gain is defined as the system throughput gain of Joint RA algorithm over Independent RA algorithm. The network coding gain is defined as the system throughput gain of Joint RA algorithm with NC over Joint RA algorithm. The total gain, which is the combined effect of both joint resource allocation gain and network coding gain, is defined as the system throughput gain of Joint RA algorithm with NC over Independent RA algorithm.

From Fig. 5, we observe that the joint resource allocation gain gradually increases as the cell radius increases. The reason for this phenomenon is that, as the cell radius increases, the average gap of channel conditions of each pair of interactive users will gradually increase and thereby, the advantage of joint uplink/downlink resource allocation schemes increases. To exploit the network coding advantage, it is preferable to encode the downlinks of each pair of interactive users with similar channel gains. For this reason, as the cell radius increases, the coding gain will gradually decreases, as shown in Fig. 5. Obviously, a tradeoff, between the joint resource allocation gain and network coding gain, is achieved. Hence, the total gain first increases, and then decreases along with the increase of cell radius, as shown in Fig. 5.

The average power consumption of interactive users is defined as the mean value of power consumption of all interactive users.

Figure 6 shows the comparison of the average power consumption of interactive users for three different RA algorithms, where the cell radius varies from 0.2 to 1.2 km.

From Fig. 6, we observe that, the average power consumption of interactive users of Joint RA algorithm and Joint RA algorithm with NC is lower than that of Independent RA algorithm. The reason for this phenomenon is that as the allocated uplink and downlink rates of each pair of interactive users are coupled in Joint RA algorithm and Joint RA algorithm with NC, the interactive users need not to allocate all power budget to subcarriers when the gap of the uplink and downlink channel conditions of each pair of interactive users are big enough.

Moreover, we also observe that the average power consumption of interactive users for Joint RA algorithm and Joint RA algorithm with NC decreases gradually as the cell radius increases. The reason for this phenomenon is that the system has more chance to save the power of interactive users since the average gap of channel conditions of each pair of interactive users gradually increases along with the increase of the cell radius.

In addition, due to the uplink rate of interactive users for Joint RA algorithm with NC outperforms that of Joint RA algorithm as shown in Fig. 3, the average power consumption of interactive users for Joint RA algorithm with NC also outperforms that of Joint RA algorithm, as shown in Fig. 6.

Figure 7 shows the comparison of the system throughput for three different RA algorithms as the number of interactive links, *M*, varies from 0 to 7, where the cell radius

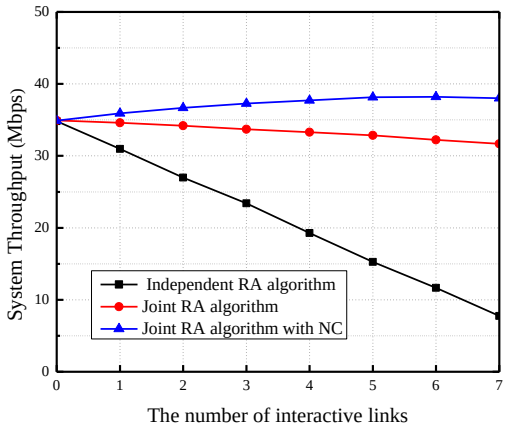


Figure 7: The comparison of the system throughput with different number of interactive links.

is 0.4 km. As the number of interactive links increases from 0 to 7, the number of users using uplink or downlink only decreases from 16 to 2.

As the number of interactive links increases, more and more users are influenced by the coupling constraint of the allocated uplink and downlink rates of each pair of interactive users. Hence, the system throughput of Independent RA algorithm and Joint RA algorithm decreases simultaneously. However, owing to the network coding processing at BS and the advantage of multicast in the downlink for the interactive users, the system throughput of Joint RA algorithm with NC gradually increases.

5 CONCLUSIONS

In this paper, we propose a joint uplink/downlink resource allocation scheme in wireless OFDMA networks with the interactive service. The proposed scheme consists of a bi-directional data interaction mechanism and a joint uplink/downlink resource allocation algorithm. In the data interaction mechanism, BS encodes the received data packets from two interactive users in individual uplink using the network coding technique, and multicasts the encoded data packet in the same downlink. In the proposed resource allocation algorithm for the interactive service, the uplink and downlink resources are allocated jointly to improve the system throughput, as well as considering the characteristic of interactive service.

Simulation results show that the performance of the proposed joint uplink/downlink resource allocation scheme, in terms of the system throughput and the power consumption of interactive users, outperforms other two resource allocation schemes.

In the future, we will investigate the joint uplink/downlink resource allocation problem for delay-sensitive interactive service in wireless networks.

ACKNOWLEDGEMENTS

This work was partly supported by National Natural Science Foundation of China (No. 61071129, No. 61171087), Ministry of Industry and Information Technology of P. R. China (No. 2012ZX03001035-004), and Science and Technology Department of Zhejiang Province (No. 2011R10035, No. 2012C01036-1).

REFERENCES

[1] S. Kim, and J.-W. Lee, “Joint resource allocation for uplink and downlink in wireless networks: a case study with user-level utility functions,” Proc. of IEEE VTC-Spring 2009, pp. 1-5 (2009).

[2] W. Saad, and Z. Dawy, “A utility-based algorithm for joint uplink/downlink scheduling in wireless cellular networks,” Collaborative Computing and Applications, Vol. 35, No. 1, pp. 348-356 (2012).

[3] A. M. El-Hajj, and Z. Dawy, “On optimized joint uplink/downlink resource allocation in OFDMA networks,” Proc. of IEEE ISCC 2011, pp. 248-253 (2011).

[4] A. M. El-Hajj, Z. Dawy, and W. Saad, “A stable matching game for joint uplink/downlink resource allocation in OFDMA wireless networks,” Proc. of IEEE ICC 2012, pp. 5354-5359 (2012).

[5] A. M. El-Hajj, and Z. Dawy, “On delay-aware joint uplink/downlink resource allocation in OFDMA Networks,” Proc. of IEEE ISCC 2013, pp. 257-262 (2013).

[6] R. Ahlswede, N. Cai, and S. Li, “Network information flow,” IEEE Trans. Information Theory, Vol. 46, No. 4, pp. 1204-1216 (2000).

[7] P. Larsson, N. Johansson, and K.-E. Sunell, “Coded bi–directional relaying,” Proc. of IEEE VTC-Spring 2006, pp. 851-855 (2006).

[8] P. Tarasak, and S. Sun, “Resource allocation for downlink two-user OFDMA systems with wireless network coding,” Proc. of IEEE SPAWC 2011, pp. 261-265 (2011).

[9] K. Seong, M. Mohseni, and J. M. Cioffi, “Optimal resource allocation for OFDMA downlink systems,” Proc. of IEEE ISIT 2006, pp. 1394-1398 (2006).

[10] E. Yaacoub, and Z. Dawy, “A survey on uplink resource allocation in OFDMA wireless networks,” IEEE Communications surveys & tutorials, Vol. 14, No. 2, pp. 322-337 (2012).

[11] M. Tao, Y. Liang, and F. Zhang, “Resource allocation for delay differentiated traffic in multiuser OFDM systems,” IEEE Trans. Wireless Commnumications, Vol. 7, No. 6, pp. 2190-2201 (2008).

[12] S. Katti, H. Rahuld, and D. Katabi, “XORs in the air: practical wireless network coding,” IEEE/ACM Trans. Networking, Vol. 16, No. 3, pp. 497-510 (2008).

[13] W. Rhee, and J. M. Cioffic, “Increase in capacity of multiuser OFDM system using dynamic subchannel allocation,” Proc. of IEEE VTC-Spring 2000, pp. 1086-1089 (2000).

[14] W.Yu, and R. Lui, “Dual methods for nonconvex spectrum optimization of multicarrier systems,” IEEE Trans. Communications, Vol. 54, No. 7, pp. 1310-1322 (2006).

[15] J. Tian, H. F. Chong, and Y.-C. Liang, "Network coding for intra-cell communications in OFDMA networks," IEEE Wireless communications letter, Vol 4, No.1, pp. 70-73 (2015).
[16] A. M. El-Hajj, and Z. Dawy, "On probabilistic queue length based joint uplink/downlink resource allocation in OFDMA networks," Proc. of IEEE ICT 2012, pp. 1-6 (2012).
[17] K. Kim, Y. Han, and S. Kim, "Joint subcarrier and power allocation in uplink OFDMA systems," IEEE Communications letters, Vol. 9, No. 6, pp. 526-528 (2005).

(Received April 4, 2015)
(Revised September 29, 2015)

Bingfeng Liu, graduate student in the Department of Information Science and Electronic Engineering, Zhejiang University, China. He received his B.S. degree at the college of Information Technical Science, Nankai University, in 2008. His current research interest is resource management in wireless networks.

Huifang Chen, Associate Professor with the Department of Information Science and Electronic Engineering, Zhejiang University, China. She received her B.S. degree in electronic engineering, M.S. and Ph.D. degrees in communications and information systems from Zhejiang University in 1994, 1997 and 2000, respectively. From October 2005 to September 2007, she was a post-doctoral research fellow with the Department of Computer Science, Shizuoka University, Japan. Her current research interests include information theoretic security and its applications, and wireless networks.

Lei Xie, Associate Professor with the Department of Information Science and Electronic Engineering, Zhejiang University, China. He received his B.S. degree in electronic engineering, M.S. and Ph.D. degrees in communications and information systems from Zhejiang University in 1994, 1997 and 2002, respectively. His current research interests include information theory and coding, coded modulation system, and security issue in wireless networks.

Submission Guidance

About IJIS

International Journal of Informatics Society (ISSN 1883-4566) is published in one volume of three issues a year. One should be a member of Informatics Society for the submission of the article at least. A submission article is reviewed at least two reviewer. The online version of the journal is available at the following site: http://www.infsoc.org.

Aims and Scope of Informatics Society

The evolution of informatics heralds a new information society. It provides more convenience to our life. Informatics and technologies have been integrated by various fields. For example, mathematics, linguistics, logics, engineering, and new fields will join it. Especially, we are continuing to maintain an awareness of informatics and communication convergence. Informatics Society is the organization that tries to develop informatics and technologies with this convergence. International Journal of Informatics Society (IJIS) is the journal of Informatics Society.

Areas of interest include, but are not limited to:

Computer supported cooperative work and groupware
Intelligent transport system
Distributed Computing
Multi-media communication
Information systems
Mobile computing
Ubiquitous computing

Instruction to Authors

For detailed instructions please refer to the Authors Corner on our Web site, http://www.infsoc.org/.

Submission of manuscripts: There is no limitation of page count as full papers, each of which will be subject to a full review process. An electronic, PDF-based submission of papers is mandatory. Download and use the LaTeX2e or Microsoft Word sample IJIS formats.

http://www.infsoc.org/IJIS-Format.pdf

LaTeX2e

LaTeX2e files (ZIP) http://www.infsoc.org/template_IJIS.zip

Microsoft Word™

Sample document http://www.infsoc.org/sample_IJIS.doc

Please send the PDF file of your paper to secretariat@infsoc.org with the following information:

Title, Author: Name (Affiliation), Name (Affiliation), Corresponding Author. Address, Tel, Fax, E-mail:

Copyright

Publisher

Address: Informatics Laboratory, 3-41 Tsujimachi, Kitaku, Nagoya 462-0032, Japan
E-mail: secretariat@infsoc.org

CONTENTS