MolComML: The Molecular Communication Markup Language

Eduard Alarcon¹, Raúl G. Cid-Fuentes¹, Luca Felicetti², Mauro Femminella², Pietro Lio³, Gianluca Reali² and Josep Solé-Pareta¹

¹NaNoNetworking Center in Catalunya (N3Cat) Universitat Politècnica de Catalunya, Spain
²Department of Engineering, University of Perugia, CNIT RU via G. Duranti 93, 06125 Perugia, Italy.
³Computer Laboratory, University of Cambridge, Cambridge CB3 0FD, UK
eduard.alarcon@upc.edu, rgomez@ac.upc.edu, ing.luca.felicetti@gmail.com, mauro.femminella@unipg.it, pl219@cam.ac.uk, gianluca.reali@unipg.it, pareta@ac.upc.edu

ABSTRACT
Given the high multi-disciplinarity of Molecular Communications (MolCom), researchers often face significant difficulties to understand each other. This impairment not only affect researchers with different backgrounds, but it also affects the different software tools. This paper motivates the Molecular Communication Markup Language (MolComML). MolComML is proposed as an XML-based format to represent the considered elements, interactions, configuration and results of the experiments and simulations in the field of MolCom. MolComML is designed with the objective of converging all fields of research within MolCom to help the exchange of information. We overview its main functionality and define its basic composing elements.

Categories and Subject Descriptors
I.6.3 [Simulation and Modeling]: Simulation Languages

Keywords
Molecular communications, markup language, simulation, standardization

1. INTRODUCTION
Molecular Communication (MolCom) is a disruptive field of research that has been rapidly growing in the recent years [4]. Given its high multidisciplinary, bridging the disparate islands of knowledge stems as a pending challenge.

This issue does not only appear in the communication between researchers, but it also interferes the interaction between software and simulation tools, thus impairing reproducibility of the research outcomes. As an example, consider the existing simulation tools, which have been used to simulate diffusion-based channels. Among others, BiNS [12], N3Sim [19], Ns-3-based [7] or COMSOL-based [9] utilize different configuration and output files, making cross-validation and reproducibility often unfeasible, especially for complex communication architectures [14, 21].

This paper introduces the MolCom Markup Language (MolComML) as an XML-based language that promises to re-unite both numerical analysis and experimental synthesis by ensuring a flexible markup language [2]. This will allow cross-validation of experiments with the theoretical results, as well as to reduce significant researcher time in interfacing different software tools. As sketched in Fig. 1, the target of this activity can be compared with that of XML-based languages, which allow specifying network configuration in an unified way, such as NETCONF [11], or OF-CONF [1] for Software Defined Networking devices.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ACM Nanocom ’16 New York City, USA
Copyright 20XX ACM X-XXXX-XX-X/XX/XX ...$15.00.
the background and the related work. Section 3 defines the molecular communication markup language and overviews its properties, elements, and functionality. Finally, Section 4 concludes our work.

2. BACKGROUND AND RELATED WORK

In this section, we overview existing specification languages for other disciplines, as well as existing simulators and experimental test-beds.

2.1 Molecular Communication Standard

The model defined by the IEEE 1906.1 Working Group [7] stands as the first nanoscale and molecular communication standard and constitutes a recommended practice for the definition of a general framework for the nanoscale communications. This standard proposes an architecture based on the following blocks: NetDevice, Communication Interface, Medium, Motion, Field, Specificity, and Perturbation.

2.2 Specification languages

As for mark-up languages used in biology, one of the most important related language is SBML [18], that is nowadays the standard for representing computational models in system biology. More in details, it allows the communication and storing of computational models of biological processes and its success is due to the possibility to represent different classes of biological phenomena, such as cell signaling pathways, regulatory networks, and many others. The main purposes of SBML are essentially the sharing of the models on different software environments and the survival of such models beyond the lifetime of the software used to create them.

Another specification language used for the description of molecular biology is SBOL [16]. It is a standard for the specification and exchange of biological design information in synthetic biology that is able to go deeper in the description of the biological processes. In particular, it supports the explicit and unambiguous description of biological designs through a rigorous definition of rules and best practices on how to use the provided data models with a special focus on the design details. Hence, the SBOL standard is fully qualified to represent structural components of a biological design, such as DNA and RNA, proteins, small molecules, including also behavioral aspects.

The CellML language is an open standard based on the XML markup language [10]. The purpose of CellML is to store and exchange computer-based mathematical models. CellML allows scientists to share models even when they use different model-building software. It also enables them to reuse components from one model to another one, thus accelerating model building. CellML includes information about model structure, mathematics and metadata.

NeuroML is an XML-based model description language, which provides a powerful common data format for defining and exchanging models of neurons and neuronal networks [23]. The structure and behavior of ion channel, synapse, cell, and network model descriptions are based on underlying definitions provided in LEMS, a domain-independent language for expressing hierarchical mathematical models of physical entities. It includes two Application Programming Interfaces (APIs) written in Python to simplify the process of developing and modifying models expressed in NeuroML and LEMS.

2.3 MolCom Simulators and Experimentation

In this section, we list some of the main simulators that have been developed to characterize the broad set of MolCom systems:

- **BiNS - Biological and Nano-Scale communication simulator** is a simulation package for MolCom systems developed at the University of Perugia [12]. Its customizable design provides a set of tools for creating objects and for modeling the behavior of biological entities. The simulator contains generic type of software object, named Nano Objects. Nodes and carriers are specific implementations of the Nano Object, and, although they share its general features, they can exhibit very different functions.

- **CalComSim - Calcium Signaling-based Molecular Communication System Simulator** is a simulator designed for both synthetic and natural cellular communication found inside a 3D human tissues [6]. It implements three different types of tissues, namely non-excitable, excitable and hybrid, whereas cells are implemented as epithelial, smooth muscle cells and astrocytes. The simulator targets highly disparate fields of expertise, such as telecommunication engineers, biological scientists and pharmaceutical researcher.

- **N3Sim** is a simulation framework for diffusion-based molecular communications, which allows the evaluation of molecular networks with several transmitters and receivers [19]. The diffusion of particles through the medium is modeled as Brownian motion, taking into account particle inertia and collisions among particles. It implements a three-layer architecture, which are the user interface layer the data layer and the domain layer. The simulation parameters are determined by means of a text configuration file.

- **COMSOL Multiphysics** is a commercial multipurpose platform designed for simulating physics-based problems through a unified workflow for electrical, mechanical, fluid, and chemical applications [8]. It implements finite element analysis, for different physics and engineering applications. An example of the use of COMSOL Multiphysics for simulating a MolCom drug delivery system is presented in [8].

- **NS-3 Based Simulators**: NS-3 is a discrete-event network simulator, which was not originally developed for MolCom. Nevertheless, its flexible structure has allowed implementing some basic elements of MolCom, in particular a simulation tool has been developed in the framework of the IEEE P1906.1 working group [7]. NS-3 is organized in different software libraries that can work together. User programs, written in C++ or Python programming languages, can adapt these libraries or be linked with them.

- **NanoNS** is a tool developed on top of the NS-2 [17] network simulator. Its main objective is to provide a tool to create a better understanding of nanonetworks and facilitate the development of new communication techniques. It incorporates the simulation modules for various nanoscale communication paradigms based on a diffusive molecular communication channel.
based on Brownian motion in aqueous medium and a consequence of the constant thermal motion of atoms, molecules, and their chemical reactions.

- **BNSim** is a multithread java simulator for bacteria networks [25]. These networks interconnect engineered bacteria that communicate at nanoscales. BNSim integrates three simulation methods: (i) the Gillespie stochastic simulation algorithm; (ii) stochastic differential equations, used to model large-scale chemical system with a controlled level of approximation; (iii) a hybrid algorithm which integrates the above methods.

- **NCSim - Bacteria Nanonetworks** is a comprehensive simulation framework for molecular communications, utilizing flagellated bacteria for information delivery [5]. Its major focus is on different message encoding techniques. It can simulate several simultaneous links between the nanomachines. NCSim incorporates the stochastic model for bacteria mobility, and the plasmid/chromosome transfer between bacteria through the conjugation process. NCSim consists of three modules: (i) physical (PHY) layer of bacterial nanonetworks; (ii) scenarios generator and simulation monitor; and (iii) plotting tool, intended to post-process raw simulation data and to generate plots.

- **HLA Simulator.** In [3], the authors introduce a simulator design focusing on scalability, and adopting the high level architecture (HLA) model, which is standardized under IEEE 1516. It is used to design a distributed simulation tool for molecular communications, so that different scalability options can be used to include additional processing power to reduce the execution time. This model allows designing large systems, which could be difficult to do otherwise.

As for experimentation, significant examples for molecular communications are those illustrated in [13, 22]. In the former, real in-vitro communications between platelets and endothelial cells by means of CD40L molecules are tested. In the latter, the authors have combined cells with nanotechnology for an integrated molecular processing network. Specifically, engineered cell populations are triggered by a quorum sensing signal molecule to express surface-displayed fusions consisting of a fluorescent marker and an affinity peptide. The latter provides means for attaching magnetic nanoparticles in order to fluorescently activate subpopulations for coalescence into colour-indexed output. The resultant nano-guided cell network assesses quorum sensing activity and conveys molecular information as a 'bio-litmus' in a manner read by simple optical means.

### 3. MOLCOMML

A new MolCom Markup Language (MolComML) is necessary for bridging the areas of interest for molecular communications. Accordingly, this needs to be flexible and generic, such that it can be understood by any software and simulation platforms for molecular communications. It also needs to be capable of providing a large level of detail by providing all governing equations, parameter values and necessary conditions, such that it can entirely describe a MolCom environment in either a simulation or experimental test-bed.

In addition, due to the large growing rate of MolCom-related research and its high multi-disciplinarity, MolComML needs to be easily extensible to integrate the most recent and on-going advancements, compatible with existing languages of neighboring disciplines (e.g., SBML [18] for systems biology and NeuronML [23] for neurology). Finally, MolComML should be compliant with the IEEE P1906.1 recommended practice for nanoscale and molecular communication.

Its basic structure consists of several main blocks reflecting the main components of a general molecular communication case. Each block is composed of a set of required parameters and a set of custom parameters that could be defined each time according to the molecular communication needs. The UML diagram and the XML skeleton of the proposed MolComML are shown in Fig. 2 and Fig. 3, respectively. The proposed XML skeleton is based on the framework presented in [15], which is illustrated in Fig. 4.

#### 3.1 Objectives

The main objectives of the MolComML format are listed in what follows:

- **Represent the many different classes of molecular communication scenarios, in all levels of abstraction.**
- **Enable the use of multiple software tools without having to rewrite models to conform to different file formats.**
- **Ensure the survival of models beyond the lifetime of the software used to create them.**
- **Use a single language both to analyze the considered scenarios through software tools, as well as to synthesize actual experiments.** This ensures repeatability and cross-validation.
- **Enable models to be shared and published in a form that any researcher can use even by making use of different software environments.**
- **Enable future expandibility of the markup language.** Due to the rapid knowledge growth in this field, this language needs to be constantly updated. MolComML will be versioned to structure the integration of novel definitions and models.

#### 3.2 Elements and Functionality

This section overviews the composing elements of MolComML, as well as their basic functionality.

##### 3.2.1 Network Elements

Network elements are considered as the building blocks for users. They are created and interconnected to define the simulation or experimental set-up. Each element may be defined by different levels of abstraction. For instance, it can be an entirely conceptual entity; it can have some real physical interpretation, or both. Each network element has a set of standard attributes that could be extended by the introduction of custom ones. The most important attributes describe the shape and size of the element, its mass and time to live properties, the accepted and transmitted signals and, finally, the motion rules, if such element is equipped by autonomous propulsion system.

The main network elements are as follows:
3.2 Transmitter: These elements are in charge of encoding information in the form of a molecular communication. They need to include all the relevant parameters. Among others, rate of creation of molecules, rate of emission, and the molecule release mechanism.

3.2 Receiver: These elements are in charge of decoding the information by detecting the induced fluctuations in the molecular channel. They need to consider multiple configurations and types of receivers, such as absorbing receiver and receiver with absorbing receptors.

3.2 Signal: The transmitted signal carries the information towards the receiver. This can be based on DNA, proteins, ions, and others.

Derivations of these elements, as well as other elements (active or passive) in the communication channel that may affect the communication process can also be part of this list.

3.2.2 Communication Interface

Network Elements can be connected with other elements by using the Communication Interface element, that describes the external interfaces of each network element. Such interfaces have several properties that describe also the type of transmitted and received signals, their affinity and the direction of communication. Again, a subset of custom parameters can be defined in order to describe more in details the properties of each interface, as described in what follows.

3.2.3 Properties and parameters

It is often allowed the definition of a list of custom attributes, identified by the \(<param>\) tag. This approach allows a high customization capabilities, extending the predefined attributes or introducing completely new ones. The format is quite simple, you have just to define a set of new attributes after the \(<param>\) tag. This approach allows specifying both properties and attributes of each element described in the MolComML file.

3.2.4 Compartment Elements

The compartments are intended as a kind of well-stirred container of a particular type and finite size where species (e.g., chemical substances) may be located. A model may contain multiple compartments even of the same compartment type, and they can also be located inside each another, hierarchically. Connections between different compartments are handled by Gates, that define the rules for crossing. Note that each species (i.e. Network Element) in a model must be located within a compartment.

3.2.5 Gates

It is possible to define a list of Interconnection gates between a couple of adjacent compartments. Each gate is identified by a unique name, a position, size, shape and orientation, in order to create a sort of passing hole on the surface of both compartments.

3.2.6 Interaction Rules

A set of rules need to be defined. They restrict or specify the operation of the network elements and their connections. The collision behavior is a typical example: upon impact, molecules can join, merge, or absorb one another or they can bounce away from each other. Each rule could be a global rule valid everywhere or could be more specific, describing only a part of the communication environment or is valid only for a subset of network elements. In general, rules are described by mathematical expressions imported from an external model (e.g. MathML [24]). Here it is also
Figure 3: Example of MolComML code.

Figure 4: Graphical description of the MolComML to illustrate the framework in [15].

possible to initialize constants and variables of the imported equations.

3.2.7 Communication Channels

The transmitted signals are transferred to receivers through communication channels. There exist several channel types. Junction-based, diffusion, and diffusion-with-drift are examples of existing channel types. A channel element has to be defined and connected to each compartment placed in the communication environment. This ensures the description, with a high degree of accuracy, of the local environmental conditions, by means of specific mathematical rules defined in one or more external MathML files. It is also possible to initialize a set of the parameters used in the imported equations. Each channel definition could be shared between two or more Compartments. The association channel-compartment is defined in the compartment section.

3.2.8 Network Topology

Each Network Element described in section 3.2.1 defines only the main properties of such element. The Network Topology section is used to place the required elements in the proper Compartment. The Topology of the molecular communication network is described by defining the position and orientation of each element and also the communication protocol at the basis of the end-to-end communication. The initialization of any node parameters are declared here.

3.2.9 Protocol Stack

Taking inspiration from a previous work [20], in this section is defined the protocol stack that could be used for the communication needs. It is possible to define any number of protocol stacks and each one could have a custom structure composed of different depth and identification name. Each layer could map the well known protocol stacks of the traditional telecommunication field or define completely new layers. The layers are defined by a set of rule and each one is composed of two signals, the first is for the forward communication and the last is for the backward communication. For each signal, the type of carrier that will be transmitted
known rule
define a rule for each attribute of interest, through the well
tail, it is possible to define the list of Network Elements,
importing one, the list of imported elements is defined
</listOfUnitDefinitions> tags. The name chosen
external source with the name used in the MolComML file.
comprehensive or a partial import. For the lat-
replicability of research achievements.
This paper has introduced the Molecular Communication
Multi-hop conjugation based bacteria nanonetworks. IEEE
Comparative end-to-end analysis of
cellular communication in biological tissues. IEEE Transactions on
compartments and of their attributes, by specifying
rate control algorithms for each signal type defined above.
imported, by specifying its identification name on the
for each custom unit can be used in the other sections of
considered attribute.
3.2.12 External Sources
MolCom systems is expected to interact with elements
模范化 the system description in software, simula-
tion tools and experimental test-beds, targeting both mu-
ble to define the most common
units, and their multiples, in the International System. The
definition of custom units is based on the combination of the
previous ones, by setting their values, scales and exponents
through this format: customUnit = (value×10 scale ) exponent
inside the <listOfUnitDefinitions> tags. The name chosen
for each custom unit name in the external source with the name used in the MolComML file.
For each pairing, also the importing rule is specified; it could
communication, such as the synchronization and the
cdc2+-signaling-based molecular communication in
This work is supported by EU project H2020 FET Open
CIRCLE (Coordinating European Research on Molecular
communications in blood vessels. IEEE Transactions on NanoBioscience
Modeling CD40-based molecular communications in blood vessels. IEEE
4. CONCLUSIONS
This paper has introduced the Molecular Communication Markup Language (MolComML). It is an XML-based format to universalize the system description in software, simulation tools and experimental test-beds, targeting both mutual understanding among the disperse islands of knowledge in this highly multidisciplinary research field. It guarantees also reproducibility of research achievements.
5. ACKNOWLEDGMENTS
This work is supported by EU project H2020 FET Open
CIRCLE (Coordinating European Research on Molecular
Communications) No. 665564.
6. REFERENCES
cellular communication in biological tissues. IEEE Transactions on
Communications, 63(12):5128–5142, Dec 2015.


