A scalable runtime platform for multiagent-based simulation

Tobias Ahlbrecht, Jürgen Dix, Michael Köster,
Philipp Kraus, and Jörg P. Müller

Department of Informatics, Clausthal University of Technology, Julius-Albert-Str. 4
D-38678 Clausthal-Zellerfeld,Germany
`firstname.lastname@tu-clausthal.de`
`http://www.in.tu-clausthal.de`

Abstract. Using purely agent-based platforms for any kind of simulation requires to address the following challenges: (1) *scalability* (efficient scheduling of agent cycles is difficult), (2) *efficient memory management* (when and which data should be fetched, cached, or written to/from disk), and (3) *modelling* (no generally accepted meta-models exist: what are essential concepts, what just implementation details?). While dedicated professional simulation tools usually provide rich domain libraries and advanced visualisation techniques, and support the simulation of large scenarios, they do not allow for "agentization" of single components. We are trying to bridge this gap by developing a *distributed, scalable runtime platform for multiagent simulation, MASeRaTi*, addressing the three problems mentioned above. It allows to plug-in both dedicated simulation tools (for the *macro view*) as well as the agentization of certain components of the system (to allow a *micro view*). If no agent-related features are used, its performance should be as close as possible to the legacy system used.

Paper type: Technological or Methodological

1 Introduction

In this paper, we describe ongoing work on a distributed runtime platform for multiagent simulation, *MASeRaTi*, that we are currently developing in a joint project (`http://simzentrum.de/en/projects/desim`). The idea for *MASeRaTi* evolved out of two projects, Planets and MAPC. In both projects, we implemented, completely independently, running systems for different purposes. One to simulate urban traffic management, the other to simulate arbitrary agent systems in one single platform.

Agent-based traffic modelling and simulation: We developed *ATSim*, a simulation architecture that integrates the commercial traffic simulation framework *AIMSuN* with the multiagent programming system *JADE* (implemented in JAVA): *ATSim* was realized within Planets, a project on cooperative traffic management (`http://www.tu-c.de/planets`).

Agent-based simulation platform: We implemented, in JAVA, an agent-based
   platform, *MASSim*, which allows several simulation scenarios to be plugged
   in. Remotely running teams of agents can connect to it and *play against each
   other* on the chosen scenario. *MASSim* has been developed since 2006 and
   is used to realise the MAPC, an annual contest for multiagent systems.

   While the former system centers around a commercial traffic simulation plat-
form (*AIMSuN*), the latter platform is purely agent-based and had been devel-
oped from scratch. Such an agent-based approach allows for maximal freedom
in the implementation of arbitrary properties, preferences, and capabilities of
the entities. We call this the *micro-level*: each agent can behave differently and
interact with any other agent.
   The traffic simulation platform *AIMSuN*, which easily runs tens of thousands
of vehicles, however, does not support such a micro-level view. Often we can only
make assumptions about the *throughput* or other *macro*-features. Therefore, with
ATSim, we aimed at a hybrid approach to traffic modelling and integrated the
JADE agent platform in order to describe vehicles and vehicle-to-X (V2X) com-
munication within a multiagent-based paradigm. One of the lessons learned dur-
ing the project was that it is extremely difficult to *agentize*.[1] certain entities (by,
e.g. plugging in an agent platform) or to add agent-related features to *AIMSuN*
in a scalable and natural way.
   Before presenting the main idea in more details in Section 2, we point to
related work (Section 1.1) and comment about the overall structure of this paper.

1.1   Related work

In the past decade a multitude of simulation platforms for multiagent systems
have been developed. We describe some of them with their main features and
note why they are not the solution to our problem. The *Shell for Simulated Agent
Systems (SeSAm) [22]* is an IDE that supports visual programming and facili-
tates the simulation of multiagent models. *SeSAm*'s main focus is on modelling
and not on scalability.
   GALATEA [9] is a general simulation platform for multiagent systems de-
veloped in Java and based on the High Level Architecture [24]. *PlaSMA* [14]
was designed specifically for the logistics domain and builds upon JADE. *Any-
Logic* (http://www.anylogic.com/) is a commercial simulation platform writ-
ten in Java that allows to model and execute discrete event, system dynamics and
agent-based simulations, e.g. using the included graphical modelling language.
MATSim (http://www.matsim.org/) was developed for large-scale agent-based
simulations in the traffic and transport area. It is open-source and implemented
in Java. The open-source simulation platform *SUMO* [23] was designed to man-
age large-scale (city-sized) road networks. It is implemented in C++ and sup-
ports a microscopic view of the simulation while it is not especially agent-based.

---

[1]To agentize means *to transform given legacy code into an agent so that it belongs
to a particular multiagent system (MAS)* This term was coined in [29]. In [28], Shoham
used the term *agentification* for the same purpose.

Mason [26] is a general and flexible multiagent toolkit developed for simulations in Java. It allows for dynamically combining models, visualizers, and other mid-run modifications. It is open-source and runs as a single process. *NetLogo*[30] is a cross-platform multiagent modelling environment that is based on Java and employs a dialect of the Logo language for modelling. It is intended to be easily usable while maintaining the capability for complex modelling.

TerraME (`http://www.terrame.org/`) is a simulation and modelling framework for terrestrial systems which is based on finite, hybrid, cellular automata or *situated agents*. We are using a similar architecture (Section 3), but we add some features for parallelisation and try to define a more flexible model and architecture structure.

Most frameworks with IDE support are not separable, so the architecture cannot be split up into a simulation part (e.g., on a High Performance Computing (HPC) cluster) and a visualisation/modelling part for the UI. Therefore an enhancement with HPC structure produces a new design of large parts of the system. Known systems like *Repast HPC* (`http://repast.sourceforge.net/`) use the parallelisation structure of the message passing interface MPI(`http://www.mcs.anl.gov/research/projects/mpi/`), but the scenario source code must be compiled into platform specific code. Hence, the process of developing a working simulation requires a lot of knowledge about the system specifics.

Repast HPC represents a parallel agent simulation framework written in C++. It introduces local and non-local agents which can be distributed along with the environment among different processes. Technically, it uses Boost and Boost.MPI to create the communication between the processes. A dedicated scheduler defines the simulation cycle. A problem of *Repast HPC* is the "hard encoding" structure of the C++ classes, which requires good knowledge about the Repast interface structure. In our architecture, we separate the agent and scheduling structure into different parts, creating a better fit of the agent programming paradigm and the underlying scheduler algorithms.

Also, a number of meta models for multiagent-based simulation (MABS) have been developed so far. *AMASON* [21] represents a general meta-model that captures the basic structure and dynamics of a MABS model. It is an abstraction and does not provide any implementation. *MAIA* [15] takes a different approach by building the model on institutional concepts and analysis. The resulting meta-model is very detailed, focusing on social aspects of multiagent systems. *easyABMS* [13] provides an entire methodology to iteratively and visually develop models from which code for the *Repast Simphony* toolkit can be generated. The reference meta model for *easyABMS* is again very detailed making it possible to create models with minimal programming effort.

To summarize, we find that most platforms are either written in Java or are not scalable for other reasons. Many are only used in academia and simply not designed to run on a high performance computing cluster. Common challenges relate to agent runtime representation and communication performance.

1.2   Structure of the paper

In Section 2 we discuss our past research (*ATSim* and *MASSim*), draw conclusions and show how it led to the new idea of a highly scalable runtime platform for simulation purposes. We also give a more detailed description of the main features of *MASeRaTi* and how they are to be realized. The main part of this paper is Section 3, where we describe in some detail our simulation platform, including the system meta-model and the platform architecture. Section 4 presents a small example on which we are testing our ideas and the scalability of the system as compared to *MASSim*, a purely agent-based approach implemented in Java. We conclude with Section 5 and give an outlook for the next steps to be taken.

2   Essential concepts and features of *MASeRaTi*

In this section, we first present our own research in developing the platforms *ATSim* (Subsection 2.1) and *MASSim* (Subsection 2.2). We elaborate on lessons learned and show how this resulted in the new idea of the scalable runtime platform *MASeRaTi* (Subsection 2.3).

2.1   Traffic simulation (*ATSim*)

Most models for simulating today's traffic management policies and their effects are based on macroscopic physics-based paradigms, see e.g. [17]. These approaches are highly scalable and have proven their effectiveness in practice. However, they require the behaviour of traffic participants to be described in simple physical equations, which is not necessarily the case when considering urban traffic scenarios. Microscopic approaches have been successfully used for freeway traffic flow modelling and control [27], which is usually a simpler problem than urban traffic flow modelling and control, due to less dynamics and better predictability.

In [8], we presented the *ATSim* simulation architecture that integrates the commercial traffic simulation framework *AIMSuN* with the multiagent programming system *JADE*. *AIMSuN* is used to model and simulate traffic scenarios, whereas *JADE* is used to implement the informational and motivational states and the decisions of traffic participants (modelled as agents). Thus, all features of *AIMSuN* (e.g. rich GUI, tools for data collection and data analysis) are available in *ATSim*, while *ATSim* allows to simulate the overall behaviour of traffic, and traffic objects can be modelled as agents with goals, plans, and communication with others for local coordination and cooperation.

AIMSuN (Figure 1(a), left side) provides an API for external applications to access its traffic objects via Python or C/C++ programming languages. But the *JADE*-based MAS (right side of Figure 1(a)) is implemented in Java, which leads to problems with scalability. To enable *AIMSuN* and the MAS to work together in *ATSim*, we used CORBA as a middleware. Technically we implemented a CORBA service for the MAS and an external application using the
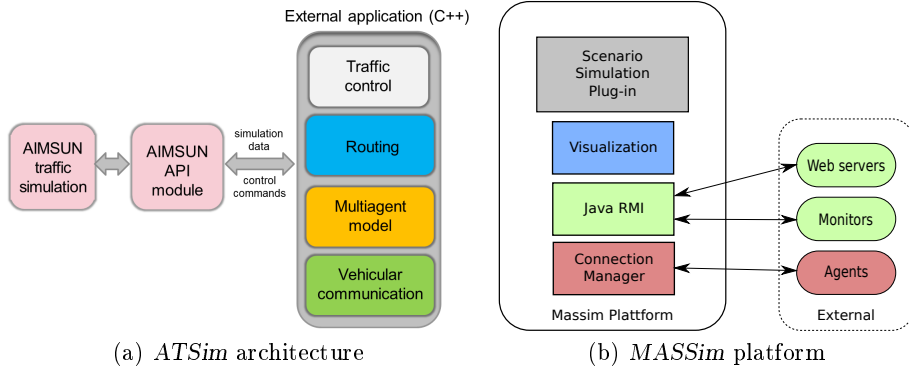
(a) *ATSim* architecture (b) *MASSim* platform

Fig. 1. Overview of the platforms

AIMSuN API to access the traffic objects simulated by *AIMSuN*. The CORBA service allows our external application to interact with the MAS directly via object references. For details on the integration architecture, we refer to [8]. Two application scenarios were modelled and evaluated on top of *ATSim*: The simulation of decentralized adaptive routing strategies, where vehicle agents learn local routing models based on traffic information [12], and cooperative routing based on vehicle group formation and platooning [16]. The overall system shown in Figure 1(a) was developed in a larger research project and contained additional components for realistic simulation of V2X communication (extending the OMNET++ simulator), and for formulating and deploying traffic control policies; see [11].

Our evaluation of the *ATSim* platform using a mid-sized scenario (rush hour traffic in Southern Hanover, one hour, approx. 30.000 routes, see [11]) showed that while the agent-based modelling approach is intuitive and suitable, our integration approach runs into scalability issues. Immediate causes identified for this were the computationally expensive representation of agents as Java threads in Jade and the XML-based inter-process communication between Jade and the *AIMSuN* simulator. In addition, system development and debugging proved difficult because two sets of models and runtime platforms needed to be maintained and synchronised.

2.2 Multi-Agent Programming Contest (MAPC)

The *MASSim* platform [5,4] is used as a simulation framework for the Multi-Agent Programming Contest (MAPC) [2](http://multiagentcontest.org). Agents are running remotely on different machines and are communicating in XML with the server over TCP/IP. The server computes the statistics, generates visual output and provides interfaces for the simulation data while the simulation is running.

A drawback of dividing the simulation in such a way is the latency of the
network that can cause serious delays. Network communication becomes a bot-
tleneck when scaling up; the slowest computer in the network is determining
the overall speed of the simulation. Running the simulation in one Java virtual
machine leads to a centralised approach that might impede an optimal run (in
terms of execution time) of a simulation.

Figure 1(b) depicts the basic components of the *MASSim* platform. *MAS-
Sim* will mainly serve us as a reference to compare scalability with *MASeRaTi*
right from the beginning (using the available scenarios). We want to ensure
that *MASeRaTi* outperforms *MASSim* in both computation time and number
of agents.

2.3   *MASeRaTi*: Challenges and requirements

Our new simulation platform, *MASeRaTi* (`http://tu-c.de/maserati`), aims
at combining the versatility of an agent-based approach (the *micro-view*) with
the efficiency and scalability of dedicated simulation platforms (the *macro-view*).
We reconsider the three challenges mentioned in the abstract for using a purely
agent-based approach.

Scalability: Efficient scheduling of agent cycles is a difficult problem. In agent
platforms, usually each agent has her own thread. Using e.g. Java, these
threads are realised in the underlying operating system which puts an upper
limit of approximate 5000 agents to the system. *These threads are handled
by the internal scheduler and are therefore not real parallel processes.* In the
MASeRaTi architecture we develop a micro-kernel where agents truly run
in parallel. In this way, we reduce the overhead that comes with each thread
significantly. We believe that this allows for a much better scalability than
agent systems based on (any) programming language, where all processes are
handled by the (black-box) operating system. Additionally, many simulation
platforms use a verbose communication language (e.g., XML or FIPA-ACL)
for the inter-agent communication that becomes a bottleneck when scaling
up. We exploit the efficient synchronisation features of MPI instead.

Efficient memory management: Which data should when be fetched from
disk (cached, written)? Most agent platforms are based on Java or simi-
lar interpreter languages. When using them we have no control over the
prefetching or caching of data (agents need to access and reason about their
belief state): this is done by the runtime mechanism of the language. We do
not know in advance which available agent is active (random access), but
we might be able to *learn* so during the simulation and thereby optimise
the caching mechanism[1]. This is the reason why we are using the scripting
language *Lua* in the way explained in the next section.

Modelling support: As of now, no generally accepted meta-model for multi-
agent-based simulations exists. We would like to distinguish between essen-
tial concepts and implementation details. What are the agents in the sim-
ulation? Which agent features are important? We also want the modelling

framework to assist a simulation developer in creating her scenario as well as hide the complexity of a parallelised simulation, while not being restrictive in terms of modelling capability.

So the main problem we are tackling is the following: *How can we develop a scalable simulation environment, where the individual agents can be suitably programmed and where one can abstract away from specific features?* We would like to reason about the macro view (usually supported by dedicated simulation tools) as well as zooming into the micro view when needed. The overhead for supporting the microview should not challenge overall system scalability:

(1) If no agents are needed (no micro-view), the performance of *MASeRaTi* should be as close to the legacy code (professional simulation tools) as possible.
(2) If no legacy code at all is used, *MASeRaTi* should still perform better or at least comparable to most of the existing agent platforms (and it should have similar functionality).

Due to general considerations (Amdahl's law[18]) and the fact that not all processes are completely parallelizable, it is not possible to achieve (1) perfectly (no agents: performance of *MASeRaTi* equals performance of legacy code).

In addition to a scalable platform we also provide a *meta-model* for multiagent-based simulations (MABS) in order to address the third challenge. The focus in this paper is on the first two challenges. The meta-model serves as a general starting point for the development of a MABS and ensures a certain structure of a simulation that is needed by the underlying platform in order to facilitate scalability. We have chosen *Lua* mainly because of its efficiency. It allows both object-oriented and functional programming styles and is implemented in native C. For details we refer to Section 3.2.

To conclude, we formulate these basic requirements for *MASeRaTi* that directly follow from the identified challenges: (1) the support of a macro and micro view of a simulation, (2) a scalable and efficient infrastructure, and (3) a multiagent-based simulation modelling framework that also supports non-agent components.

3   Overview of the system

The overall architecture of our framework is inspired by concepts from *game development*. The state of the art in developing massively multiplayer online role-playing games (MMORPGs) consists in using a client-server architecture where the clients are synchronised during game play [10] via a messaging system. Well-known games include Blizzards's *World of Warcraft* or EA's *SimCity 2013*, which supports multiplayer gaming with an "agent-based definition" in its own Glassbox engine(http://andrewwillmott.com/talks/inside-glassbox).

While a game architecture is a good starting point for our purposes, of course we cannot assume a server system with hundreds of hardware nodes, which is

powerful enough to handle a MMORPG system. Also, for developing purposes, we need a system running on a single node (think of a common desktop PC). The source code (i.e. scenario) developed there must then be transferable to a HPC system, where the real simulation is executed.

Our underlying agent-oriented meta-model uses the well established concept of a BDI agent [28,31] in a variant inspired by the agent programming language Jason [7] combined with the idea of an entity [3] that evolved out of experiences gathered in the MAPC. However, the used concepts are completely exchangeable, due to the flexibility of Lua, and developers are not forced to use them. In our agent model, agents connect to these `entities` in the simulation world. Agents consist of a *body* and a *mind*: While the mind (being responsible for the deliberation cycle, the mental state etc.) does not have to be physically grounded, the entity has to be located in an area of the simulation. Thus, an entity is an object with attributes that an agent can control and that might be influenced by the actions of other agents or the overall simulation. Intuitively, an agent can be viewed as a puppet master that directs one (or more) entities.

3.1   Architecture

Our system is composed of three layers (Fig. 2):

Micro-kernel (MK): The micro-kernel represents the technical backbone of the system. It is written in the C++ programming language to get the necessary performance and provides basic network parallelisation and scheduling algorithms. The layer defines the system's underlying structure containing interfaces e.g. for plug-ins, serialisation, Prolog for knowledge representation and reasoning, or statistical accumulation. In short, this bottom layer describes a meta-model for a generic parallel simulation (Section 3.2).

Agent-model layer (AML): The agent-model layer (Section 3.4) introduces agent-oriented concepts (e.g. agents, environments, artifacts etc.) to the system and thus describes a model of an agent-based simulation. It is implemented in the scripting language Lua(http://www.lua.org/) [20] to ensure maximum flexibility. Due to the multi-paradigm nature of Lua, pure object-oriented concepts are not supported by default. That is, Lua uses only simple data types and (meta-) tables. Fortunately, based on this, we can create an object-oriented structure in Lua itself. This allows us to work in a uniform fashion with UML models regarding the AML and the scenario layer.

Scenario layer (SL): The third and topmost layer represents the instantiation of the AML with a concrete scenario, e.g., a traffic setting or the MAPC cow scenario (to be introduced later in Section 4). It is illustrated by dotted boxes in Fig. 2 to emphasise the distinction from the AML layer. Section 4 provides an example of a concrete scenario fitting this layer.

An important aspect is the linkage between the three layers, and in particular the connections between the micro-kernel and the AML (illustrated in Fig. 2) which is discussed further in the following sections.
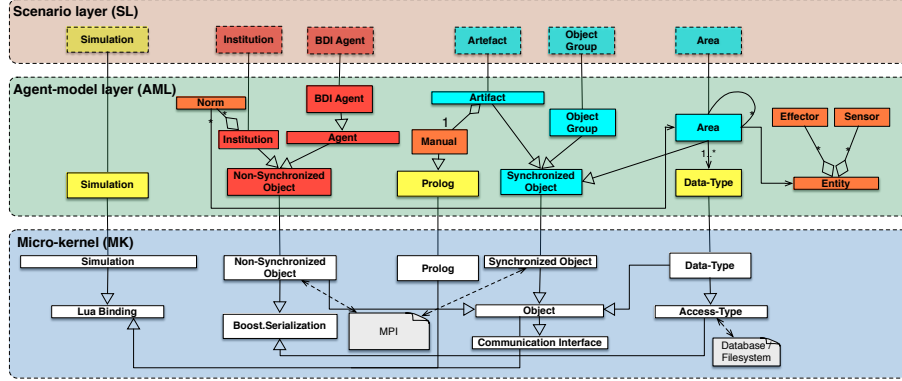
Fig. 2. *MASeRaTi* system architecture: UML class diagram

3.2 Micro-kernel

The micro-kernel represents the technical side of the system and is split up into two main structures (Fig. 3(b)). The core part (below) contains the scheduler algorithms, the core and memory management, the network and operating system layers and the plug-in API together with a Prolog interpreter. Above these core utilities, the Lua interpreter (top) is defined and each class structure on the core can be bound to "Lua objects". The Lua runtime is instantiated for each process once, so there is no elaborated bootstrapping.

The choice of Lua was mainly motivated by the scaling structure and the game developing viewpoint. Lua, a multi paradigm language, has been used for game development for many years ([25]). An advantage of Lua is the small size of its interpreter (around 100 kBytes) and the implementation in native C with the enhancement to append its own data structures into the runtime interpreter with the binding frameworks. The multiparadigm definition of Lua, especially object-oriented and functional [20], can help us to create a flexible metamodel for our simulation model. Lua can also be used with a just-in-time compiler.

The kernel defines basic data structures and algorithms (Fig. 3(a)):

Simulation: A global singleton simulation object, which stores all global operations in the simulation e.g. creating agents or artifacts. It defines the initialization of each simulation; the constructor of the Simulation object must create the world object, agent objects, etc.

Object: Defines the basic structure of each object within the simulation. All objects have got a UUID (Universally Unique Identifier), a statistical map for evaluating statistical object data, the (pre-/post-)tick methods to run the object and the running time data, which counts the CPU cycles during computation (for optimisation).

Prolog: An interface for using Prolog calls within the simulation.
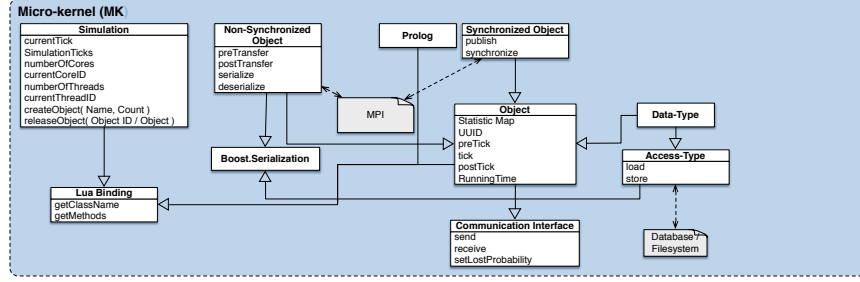
(a) Micro-kernel: UML class diagram
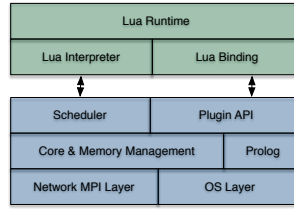
(b) Architecture

Fig. 3. Micro-kernel data model (a) and architecture (b)

Each class is derived from the *Lua Binding* class, so the objects will be mapped into the AML.

The mapping between the micro-kernel and the AML is defined using a *language binding concept*. The Lua interpreter is written in native C. Based on this structure, a C function can be "pushed" into the Lua runtime. The function will be stored into a global Lua table; the underlying C function is used with a script function call.

Our concept defines the micro-kernel in UML; instantiated C++ objects are mapped into the runtime environment by a Lua binding framework (e.g. Lua Bridge (`https://github.com/vinniefalco/LuaBridge`) or Luabind (`http://www.rasterbar.com/products/luabind.html`)). Classes and objects in Lua are not completely separate things, as a class is a table with anonymous functions and properties. If a Lua script creates an object, it calls the constructor, which is defined by a meta-table function, the underlying C++ object will be also created and allocated on the heap. The destructor call to an object deterministically removes the Lua object and its corresponding C++ object. All C++ objects are heap-allocated and encapsulated by a "smart pointer", as this helps avoiding memory leaks. This concept allows consistent binding between the different programming languages and the layer architecture.

Each *Object* comes from the *Communication interface*, which allows an object to send any structured data to another object. Three subclasses inherit from the
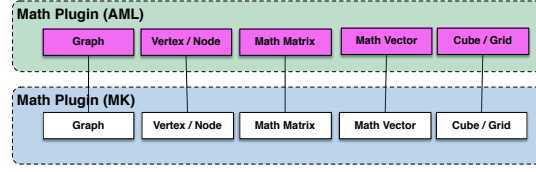
Fig. 4. Math plug-in architecture example

central *Object*. This structure is necessary for creating a distributed and scalable platform with optimisation possibility:

Synchronised Object: An object of this type is synchronised over all instances of the micro-kernel (thread and core synchronised). It exists also over all instances and needs a blocking communication. In the agent programming paradigm the world must be synchronised.

Non-Synchronised Object: This object exists only on one instance of the micro-kernel and can be transferred between different instances of the micro kernel. It should be used e.g. for agents and norms, because the evaluation is independent from other objects. Using the "execution time" of the tick (time complexity), we can group such objects together.

Data-Type: This object represents a data structure, e.g. a multigraph for the traffic scenario with routing algorithms (Dijkstra, $A^\star$ and $D^\star$). The data types will be pushed into the micro-kernel with the plug-in API. The *Access-Type* creates the connection to the storing devices.

Synchronised and *non-synchronized* objects are implemented via Boost.MPI[2] structure, and the *Access-Type* defines the interface to a database or the filesystem for storing / loading object data. The access via the data interface will be defined by the Boost.Serialization library[2], so we can use a generic interface. Based on the *Data-Type* we can use the defined plug-in API for math datatypes, which allows e.g. to create a (multi-) graph interface for our traffic scenario, based on Boost-Graph[2]. A plug-in is defined in a two-layer structure (Fig. 4). The plug-in is written in C++ (the algorithm part) and based on the Lua binding structure mapped into the higher layers (the function invoking part). This two layer strucutre enables us to use a differential equation solver like OdeInt(`http://www.odeint.com/`) to simulate the macroscopic view in the simulation (e.g. a highway traffic model can be simulated with a differential equation while employing a microscopic agent-based view for an urban traffic area. The "glue" between these two types can be defined by a "sink / source datatype").

The plug-in interface is based on a native C implementation to avoid problems with name managing in the compiler and linker definition. Plug-ins are stored in a dynamic link library; they are loaded upon start of the kernel.

---

[2]`http://boost.org/doc/libs/release/libs/`

Design tradeoffs Next, we discuss alternatives and trade-offs when designing
a runtime system to take a deeper look into that of *MASeRaTi*. During runtime
we propose to ask which objects need to be defined as synchronised or non-
synchronised datasets. The implementation of the FIPA-ACL definition, e.g., is
a blocking operation, because we can update the object only after we have pro-
cessed the input data, so each external data input creates a slower performance.
With the implementation we create additional workload, because parser, lexer
and interpreter must also process the data.
One *MASeRaTi* runtime instance implements a thread-pool (see Subsection 3.3)
which processes all objects. Scalability is obtained by looking at local instances
and taking the global view over all instances into account.

3.3   Optimisation

In [32], Wooldridge describes some pitfalls in agent developing:

1. *"You forget that you are developing multithreaded software"*.
2. *"Your design does not exploit concurrency"*.
3. *"You have too few agents"*.

    As discussed in Section 3.2 there are two disjoint sets of objects in our simu-
lation: non-synchronised and synchronised objects. Taking the above three state-
ments seriously, our aim is to design a scalable, multi-threaded and multi-core
system which can handle a large number of agents that act concurrently. With
the technical restrictions (memory and number of threads), we need another
approach, which is inspired by the technical view of an MMORPG:

- We create a scheduler on its own to handle the agents. It is based on a thread
  pool.
- We measure the average of the calculation time of each agent when it is
  active (counting the CPU cycles).
- Based on this result, we optimise the number of agents between the micro-
  kernels with a thread-/core-stealing algorithm (in future work we aim to
  describe this with a stochastic process).

After having defined one discrete simulation step, we denote this step "tick" and
the process definition of one step is as follows (Figure 5):
    Each simulation object owns a tick method, which is called by the scheduler
(pre/post tick calls are not shown here). There exist only two global blocking
operations for synchronisation over all kernel instances. Each micro-kernel pro-
cess runs the (global) synchronised objects first. After finishing, the simulation
environment is synchronised on each kernel. In the next step, the kernel runs
the non-synchronised objects. This second loop can be run in a fast way, e.g.
the agents do nothing, so the micro-kernel idles, then the while-loop sends "steal
requests" and gets objects from the other instances (Figure 6).
    Figure 7 shows the stealing process (bullets are the agents, with different
calculation times) over all running *MASeRaTi* instances
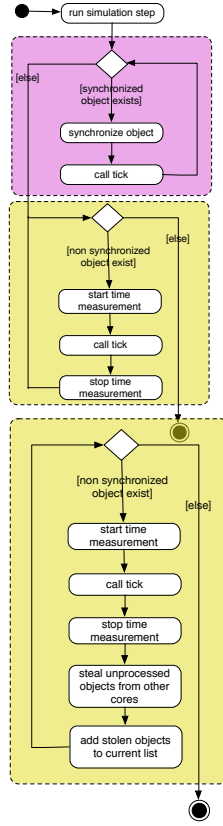
```
foreach simulation step
    foreach synchronized object
        synchronize object
        call tick
    wait for all threads are finished

    foreach non−synchronized object
        start object time measurement
        call tick
        stop object time measurement
    wait for all threads are finished
```

Fig. 5. Simulation tick

```
foreach unprocessed objects on each
    core

    foreach local non−sync. object
        start time measurement
        call tick
        stop time measurement

    steal unprocessed non−sync.
        objects from other cores
    push the stolen objects into the
        local working list

wait for all
```
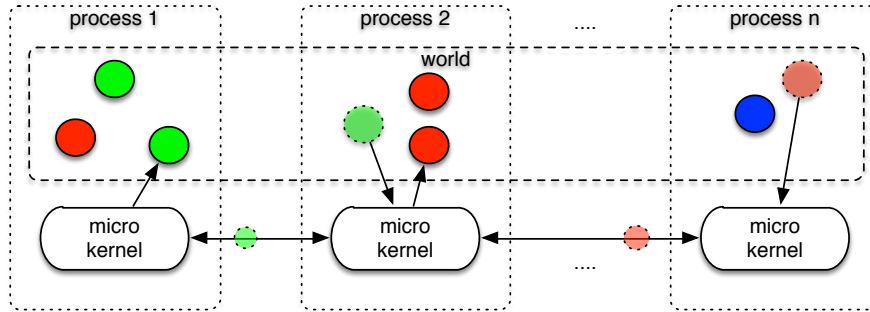
Fig. 6. Stealing process on an instance

Fig. 7. Stealing process over all instances

This idea allows the processing of a large number of agents with different (complex) plans and belief bases, because we can define the optimisation process
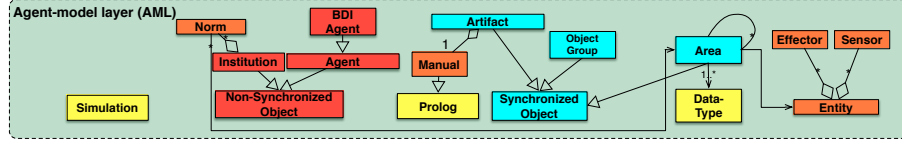
Fig. 8. Agent-model layer: UML class diagram

with different targets and methods. The simulation consists of a finite number of discrete steps and objects, so we can describe the process with a discrete stochastic approach.

3.4 Agent-model Layer

The agent-model layer (AML) (depicted in Fig. 8) defines a meta-model of an agent-based simulation. It provides the basic structure and serves as a starting point for an implementation. We start by explaining the structure, followed by the overall system behaviour; we end with a general description of the development process. Realization details (pseudo code) can be found in the appendix of [1].

Structure. The structure of the meta-model is heavily influenced by the goal of creating a simulation which can be distributed over several nodes or cores. In such a parallelised multiagent simulation, the developer has to determine for each object whether it has to be present on each and every core or if it is sufficient to have the object running independently on a single core only. We prefer the latter, since that implies *less execution time*. In contrast, an object like a global artifact has to be accessible by virtually any other object. Thus, it must be made available and therefore executed on each core.

The goal of the AML is to simplify the development of parallel multiagent simulations by defining a number of abstract objects or object categories that normally have to be synchronised and those that can usually run independently. Nevertheless, a developer can easily modify the AML to her needs, in particular redefining the synchronicity of objects.

Figure 8 illustrates the structure of the AML. Mainly, a simulation consists of a singleton `Simulation`, the non-synchronised object types `Agent`, `Norm`, and the synchronised classes `Area`, `Artifact`, `ObjectGroup`. While for the `Simulation` only one instance is allowed, the other objects can be instantiated several times. All instantiated objects are being executed in a step-based fashion and therefore implement a `tick` method that is called exactly once per simulation cycle.

Simulation: The `simulation` class in the AML is the Lua-based counterpart to the `simulation` class in the MK. It is responsible for the creation, initialisation and deletion of objects. Thus, it is in full control over the simulation.

Agent: As we aim to simulate as many agents as possible we have to ensure that this part of the model can run independently of the rest. Therefore we define two kinds of agents as non-synchronised objects: a generic agent based on [31] and a more sophisticated BDI agent [28] inspired by Jason [7]. The agent interacts with the environment through `entities` [3]. In general an agent can have random access to the simulation world. Therefore, we can only encapsulate some parts of the agent, namely the internal actions and functions like reasoning. But the effects on the environment have to be synchronised to make them known to all other objects. This is the reason for splitting the agent into two parts: the mind (the `agent`) and the body (the `entity`). The generic agent has three methods that are invoked in the following order: (1) `perceive`, (2) `think`, and (3) `act`. Inside these methods, those of the respective entity can be called directly while communication between objects has to be realised by a synchronised object (for instance by means of an `artifact`).

BDI Agent: The BDI agent is a little more sophisticated and consists of a `Belief Base` representing the current world view, a set of `Events` describing changes in the mental state, a set of plans `Plans`, and a set of `Intentions` describing the currently executed plans. Fig. 9 shows an overview of the agent cycle. Black (continuous) lines represent the activity flow while red (dashed) lines show the data flow. The agent cycle is executed from within the agent's `tick` method. In each `tick`, the agent first perceives the environment and checks for new messages. Based on this information, the belief base is updated and an event for each update is generated. From the set of events one particular event is selected and a plan that matches this event will be chosen and instantiated. During a simulation run this might result in multiple instantiated plans at the same time and allows the agent to pursue more than one goal in parallel. We decided to limit the agent to the execution of one external action (that affects the environment) but allow several internal actions per simulation `tick`. The next method selects the next action of an instantiated plan (i.e. the next action of an intention). In contrast to Jason, the agent cycle does not stop here if it was an internal action or a message, i.e., an action that does not affect the environment. Thus, the agent selects the next event (if possible) or next intention (if possible) until (1) it reaches a global timeout (set by the simulation) or (2) an external action is executed that forces a synchronisation, or (3) if the set of events and intentions are both empty.

Artifact: For all passive objects of a simulation we use the `artifact` methodology defined in [6]. Basically, each artifact has a `type`, a `manual` in Prolog (a description of the possible actions associated with it) and a `use` method that allows an agent to execute a particular action, i.e. make use of the artifact. Due to the generality of this approach the developer decides whether the actions are known by the agents beforehand or not. Additionally, since the artifact is defined as a synchronous object, one can consider a variation of this object that implements a method for each of its offered capabilities and allows for direct method invocation.
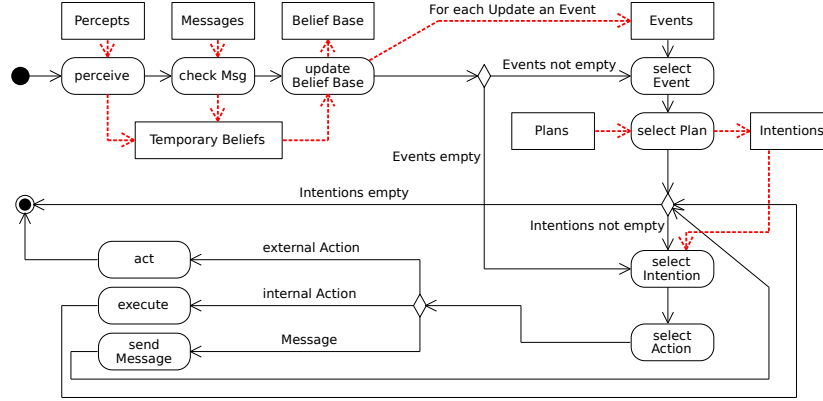
Fig. 9. BDI agent cycle: Activity diagram and data flow. Activity is red/dashed and data flow is black/solid

.

Area: So far, we defined the main actors of a simulation but how are they connected among each other? An artifact does not have to be located inside a real simulation, i.e., it does not need a physical position (in contrast, most objects do need one). Therefore, we define an `area` as a logical or physical space (similar to the term *locality* introduced by [19]). There can be several areas, subareas, and overlapping areas. In the general case, agents can perform a random access on the environment, so the areas have to be synchronised and thus be available on all cores of the simulation platform. Within an area, we define some basic data structures and algorithms for path finding, etc. The most important issue, the connection of the non-synchronised agents with the synchronised areas is realised by the use of entities. Agents perceive the environment and execute actions by using the entities' sensors and effectors.

Entity: An entity can be seen as the physical body of an agent located inside an area. An agent can register to it, get the sensor data, and execute actions that possibly change the environment. The entity has some effectors and sensors that are easily replaceable by the simulation developer. Since such an entity represents the physical body of an agent and is meant to connect an agent with the environment, it has to be synchronised over all cores.

Institutions & Norms: For now, we provided a rudimentary model as a placeholder for future extension: An institution is an object that checks for norm violations and compliance. It operates as a monitor and is also responsible for sanctioning. But a developer can also decide to separate these two tasks. For the future, we are planning to focus only on three kinds of norms: obligations, permissions, and prohibitions. Additionally, we will focus on exogenous norms (events that occur in at least one area) and not on rules that affect the agent's mind, plans etc.

ObjectGroup: Finally, an `ObjectGroup` – as the name implies – defines a group of objects. It can be used to group agents, artifacts or other objects. Method calls on an `ObjectGroup` are forwarded to all group members, i.e., with a single method call, all corresponding methods (with the same type signature) of the group members are invoked. In order to reduce overhead and to avoid circular dependencies we only allow a flat list of members at the moment. However, if a hierarchy is needed, it can be easily implemented.

Agent-model layer behaviour. So how does the overall behaviour look? Initially the simulation object creates a number of agents, areas, object groups, norms, etc., and changes the global properties in the three phases: `preTick`, `tick`, and `postTick`. It can delete and create new agents during runtime. But if the simulation developer decides to allow an agent to create another agent, this is consistent with the meta-model. The agent cycles are executed in each `tick` method. Also, the main procedures of artifacts, norms and areas are executed in this phase. The `preTick` is intended to be used as a preparation phase and the `postTick` phase for cleaning up.

Design tradeoffs. As we have seen, the AML tries to facilitate the modelling of a parallel multiagent simulation by helping the developer deciding whether objects have to be synchronised or not. Of course, our classification might not fit each and every possible use case. But because of the flexibility of this layer, it is possible to easily adapt the AML to the specific domain.

Also, the layer cannot provide all of the concepts related to the agent-oriented paradigm. We tried to identify those which are of utmost importance and thus form something like the least common denominator of all agent-based simulations. If further concepts are needed, they can be easily added on demand or might be readily available if already implemented for another use case.

We mentioned that our BDI agent is restricted to perform at most one external action per simulation cycle, while it is allowed to perfom internal actions until it runs out of time. It will be easy to change this behaviour if it proves to be disadvantageous both in terms of agent or platform perfomance.

We provided a BDI agent in order to (1) show how to transfer an agent concept to the platform at this level of implementation and (2) ensure that the platform is easily usable if no specific kind of agent is needed. While our platform is open to use any agent concepts, it does not have to.

This section contains some heavy technical machinery and describes even some low level features that are usually not mentioned. But our main aim is to ensure scalability in an agent-based simulation system. In order to achieve that, we came up with some ideas (using Lua and how to combine it with BDI-like agents) that can only be understood and appreciated on the technical level that we have introduced in this section.

4   Evaluation: Cow scenario

Scalability is an important requirement of the platform and therefore has to be evaluated early on. For that reason we chose the *cow scenario* from the MAPC as a first simulation that is realistic enough in the sense that it enforces the cooperation and coordination of agents. As it is already implemented for the *MASSim* platform, it can easily serve as a first benchmark.

In addition, we can test the suitability of the proposed meta-model and test a first implementation. Furthermore, the cow scenario contains already some elements of more complex scenarios (as in traffic simulation).

The cow scenario was used in MAPC from 2008 to 2010. The task for the agents is to herd cows to a corral. The simulated environment contains two corrals—one for each team—which serve as locations where cows should be directed to. It also contains fences that can be opened using switches. Agents only have a local view of their environment and can therefore only perceive the contents of the cells in a fixed vicinity around them. A screenshot of the visualisation as well as a short description of the scenario properties are depicted in Fig. 10. For a detailed description we refer to [4]. Using the proposed meta-model AML we can now implement the cow scenario in the following way.[3]

Fig. 11 shows how we derived the cow scenario classes from appropriate superclasses of the agent-model layer. The grid of the environment is implemented as an `Area`. Obstacles are defined by a matrix that blocks certain cells. The two corrals are subareas located inside the main area. Fences will become `Artifacts`. Similarly, we define a switch as an artifact that controls and changes the state (opened or closed) of a fence when getting activated. The cows are realised by a reactive agent that perceives the local environment and reacts upon it. For such a reactive agent the basic `Agent` definition together with an `entity` representing the cow are sufficient, while for the cowboy agents we need a more complex behaviour that facilitates coordination and cooperation. For this reason we use the `BDIAgent` (recall Fig. 9) class and create an `entity` for each cowboy agent. Furthermore, for each entity we create a simple `MoveEffector` that can be used by the entities to alter their position and a `ProximitySensor` providing the entities with their percepts. Additionally, we have to define the two teams by using the notion of an `ObjectGroup`. Finally, the `simulation` creates all agents and entities, assigns them to the two teams and creates the simulation world.

To conclude, this (very preliminary) evaluation shows that it is possible to express each aspect of the scenario using the predefined classes without the need to derive further ones from the synchronised or non-synchronised objects. (Nonetheless, doing so still remains a possibility). Regarding the suitability of Lua, it is an extremely flexible language that comes at the cost of a certain degree of usability: any newcomer needs some time to master it. But even then, having appropriate tools and methodologies that support the modelling process is a necessity to ensure an improved workflow and reduced error-proneness.

[3] The corresponding Lua code can be found in the appendix of [1].
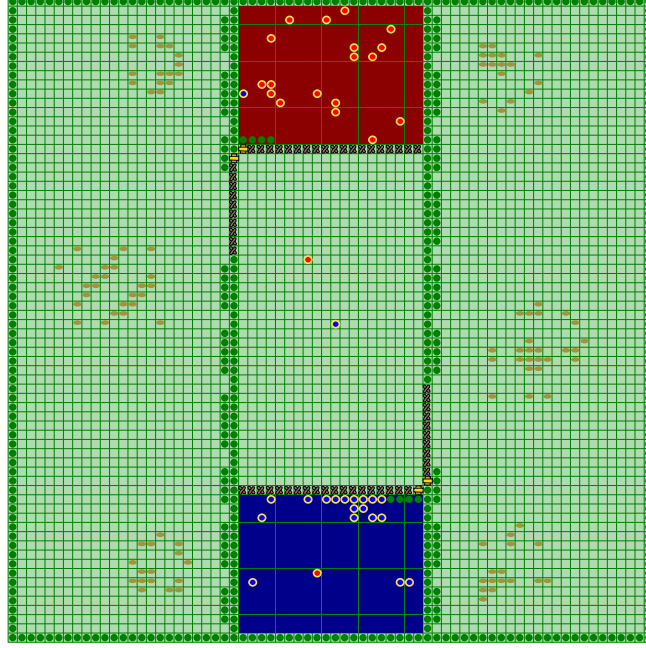
Fig. 10. The environment is a grid-like world. Agents (red (at top) and blue (at the bottom) circles) are steered by participants and can move between adjacent cells. Obstacles (green circles) block cells. Cows (brown circles) are steered by a flocking algorithm. Cows form herds on free areas, keeping distance to obstacles. If an agent approaches, cows get frightened and flee. Fences (x-shapes) can be opened by letting an agent stand on a reachable cell adjacent to the button (yellow rectangles). An agent cannot open a fence and then definitely go through it. Instead it needs help from an ally. Cows have to be pushed into the corrals (red and blue rectangles).

5   Conclusion and outlook

In this paper, we described ongoing work towards a distributed runtime platform for multiagent simulation. The main contributions of this paper are: (1) an analysis of the state of the art in agent-based simulation platforms, leading to a set of requirements to be imposed on a simulation platform, focusing on runtime scalability and efficient memory management; (2) the proposal of a novel architecture and design of the *MASeRaTi* simulation platform, bringing together a robust and highly efficient agent kernel (written in Lua) with a BDI agent interpreter including multiagent concepts such as communication and computational norms; and (3) an initial proof of concept realization featuring a simple application scenario.

The work presented in this paper provides the baseline for further research during which the *MASeRaTi* system will be extended and improved. Issues such as optimisation of the scheduler and the caching mechanisms sketched in the
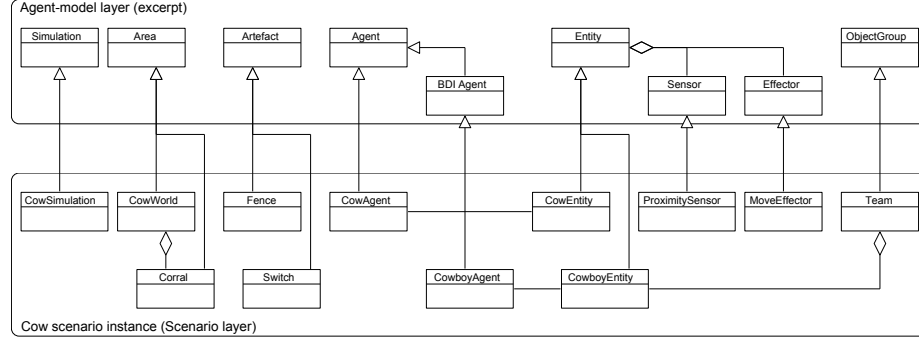
Fig. 11. Cow scenario: UML class diagram

appendix of [1] will be explored in more detail. Also, systematic experimental evaluation will be carried out using more sophisticated and much larger traffic simulation scenarios. As the *ATSim* platform introduced in Section 2.1 can deal with a few thousand (vehicle) agents, we aim *MASeRaTi* to scale up to one million agents of comparable complexity (corresponding to the micro-simulation of multimodal traffic in a large city, including public transport, cars, pedestrians, city logistics, and infrastructure).

Simulation results obtained this way can be compared to the performance of other simulation frameworks using benchmark data; scalability can also be described by varying certain parameters (e.g. number / complexity of agents) and investigating the resulting degradation of performance. An idea for evaluating our optimisation approach (and in particular the adaptive mechanism for allocating agents to nodes of the runtime system) is the following: By modelling agents' preferences, capabilities, and interactions, a certain structure is imposed on the resulting MAS. We intend to evaluate the degree to which this structure can be mapped to the allocation of agents to the nodes of the distributed runtime system, by a (domain-independent!) entropy measure. We are also planning to consider different communication technologies like Bittorrent(`http://www.libtorrent.org/`) for the inter-object communication.

Given the three objectives in the abstract, our focus in this paper has been on the first two: scalability and efficient memory management, whereas we only touched the third, modelling. Here, one avenue of research is to develop appropriate modelling tools to support the *MASeRaTi* architecture. Finally, methodologies for simulation development will be explored, starting from established methodologies such as GAIA, Tropos, or ASPECS.

References

1. Tobias Ahlbrecht, Jürgen Dix, Michael Köster, Philipp Kraus, and Jörg P. Müller. A scalable runtime platform for multiagent-based simulation. Technical Report IfI-14-02, TU Clausthal, February 2014.

2. Tobias Ahlbrecht, Jürgen Dix, Michael Köster, and Federico Schlesinger. Multi-Agent Programming Contest 2013. In Massimo Cossentino, Amal Fallah Seghrouchni, and Michael Winikoff, editors, *Engineering Multi-Agent Systems*, volume 8245 of *Lecture Notes in Computer Science*, pages 292–318. Springer Berlin Heidelberg, 2013.
3. Tristan Behrens. *Towards Building Blocks for Agent-Oriented Programming*. PhD thesis, Clausthal University of Technology, 2012.
4. Tristan Behrens, Mehdi Dastani, Jürgen Dix, Michael Köster, and Peter Novák. The Multi-Agent Programming Contest from 2005-2010. *Annals of Mathematics and Artificial Intelligence*, 59:277–311, 2010.
5. Tristan M. Behrens, Jürgen Dix, Mehdi Dastani, Michael Köster, and Peter Novák. Technical Foundations of the Agent Contest 2008. Technical Report IfI-08-05, Clausthal University of Technology, December 2008.
6. Rafael H. Bordini, Mehdi Dastani, Jürgen Dix, and Amal El Fallah-Seghrouchni, editors. *Multi-Agent Programming: Languages, Tools and Applications*. Springer, Berlin, 2009.
7. Rafael H. Bordini, Jomi F. Hübner, and Michael Wooldridge. *Programming multi-agent systems in AgentSpeak using Jason*. Wiley & Sons, 2007.
8. Viet-Hung Chu, Jana Görmer, and Jörg P. Müller. ATSim: Combining AIMSUN and Jade for agent-based traffic simulation. In *Proceedings of the 14th Conference of the Spanish Association for Artificial Intelligence (CAEPIA)*, volume 1. AEPIA, 2011. Electronic Proceedings.
9. Jacinto Dávila and Mayerlin Uzcátegui. Galatea: A multi-agent simulation platform. In *Proceedings of the International Conference on Modeling, Simulation and Neural Networks*, 2000.
10. Mattijs Driel, Jos Kraaijeveld, Zhi Kang Shao, and Remco van der Zon. A Survey on MMOG System Architectures, 2011.
11. Maksims Fiosins, Jelena Fiosina, Jörg P. Müller, and Jana Görmer. Reconciling Strategic and Tactical Decision Making in Agent-oriented Simulation of Vehicles in Urban Traffic. In *Proceedings of the 4th International ICST Conference on Simulation Tools and Techniques*, SIMUTools '11, pages 144–151, ICST, Brussels, Belgium, 2011. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).
12. Maksims Fiosins, Jelena Fiosina, Jörg P. Müller, and Jana Görmer. Agent-Based Integrated Decision Making for Autonomous Vehicles in Urban Traffic. In Yves Demazeau, Michal Pechoucek, Juan Corchado, and Javier Perez, editors, *Advances on Practical Applications of Agents and Multiagent Systems*, volume 88 of *Advances in Intelligent and Soft Computing*, pages 173–178. Springer Berlin / Heidelberg, 2011.
13. Alfredo Garro and Wilma Russo. easyABMS: A domain-expert oriented methodology for agent-based modeling and simulation. *Simulation Modelling Practice and Theory*, 18(10):1453–1467, 2010.
14. Jan D. Gehrke and Christian Ober-Blöbaum. Multiagent-based Logistics Simulation with PlaSMA. In *Informatik 2007 - Informatik trifft Logistik, Band 1. Beiträge der 37. Jahrestagung der Gesellschaft für Informatik*, pages 416–419. Technologie-Zentrum Informatik, 2007.
15. Amineh Ghorbani, Pieter W. G. Bots, Virginia Dignum, and Gerard P. J. Dijkema. MAIA: a Framework for Developing Agent-Based Social Simulations. *J. Artificial Societies and Social Simulation*, 16(2), 2013.

16. Jana Görmer and Jörg P. Müller. Group Coordination for Agent-Oriented Urban Traffic Management. In Yves Demazeau et al., editor, *Advances on Practical Applications of Agents and Multi-Agent Systems (Proc. of PAAMS 2012)*, volume 155 of *Advances in Soft Computing*, pages 245–248. Springer-Verlag, 2012.

17. Dirk Helbing, Ansgar Hennecke, Vladimir Shvetsov, and Martin Treiber. MASTER: macroscopic traffic simulation based on a gas-kinetic, non-local traffic model. *Transportation Research Part B: Methodological*, 35(2):183 – 211, 2001.

18. Mark D. Hill and Michael R. Marty. Amdahl's Law in the Multicore Era. *Computer*, 41(7):33–38, July 2008.

19. Michaela Huhn, Jörg P. Müller, Jana Görmer, Gianina Homoceanu, Nguyen-Thinh Le, Lukas Märtin, Christopher Mumme, Christian Schulz, Nils Pinkwart, and Christian Müller-Schloer. Autonomous agents in organized localities regulated by institutions. In *Digital Ecosystems and Technologies Conference (DEST), 2011 Proceedings of the 5th IEEE International Conference on Digital Ecosystems and Technologies*, pages 54–61, May 2011.

20. Roberto Ierusalimschy. Programming with multiple paradigms in lua. *Functional and Constraint Logic Programming*, pages 1–12, 2010.

21. Franziska Klügl and Paul Davidsson. AMASON: Abstract Meta-model for Agent-Based SimulatiON. In Matthias Klusch, Matthias Thimm, and Marcin Paprzycki, editors, *Multiagent System Technologies*, volume 8076 of *Lecture Notes in Computer Science*, pages 101–114. Springer Berlin Heidelberg, 2013.

22. Franziska Klügl and Frank Puppe. The Multi-Agent Simulation Environment SeSAm. In H. Kleine Büning, editor, *Proceedings of Simulation in Knowledge-based Systems*. Universität Paderborn, Reihe Informatik, Universität Paderborn, April 1998.

23. Daniel Krajzewicz, Georg Hertkorn, C. Rössel, and P. Wagner. Sumo (simulation of urban mobility). In *Proc. of the 4th Middle East Symposium on Simulation and Modelling*, pages 183–187, 2002.

24. Frederick Kuhl, Judith Dahmann, and Richard Weatherly. *Creating computer simulation systems: an introduction to the high level architecture*. Prentice Hall PTR Englewood Cliffs, Upper Saddle River, NJ, USA, 2000.

25. Roberto Ierusalimschy Luiz Henrique de Figueiredo, Waldemar Celes. *Lua Programming Gems*. Roberto Ierusalimschy, 2008.

26. Sean Luke, Claudio Cioffi-Revilla, Liviu Panait, and Keith Sullivan. Mason: A new multi-agent simulation toolkit. In *Proceedings of the 2004 SwarmFest Workshop*, volume 8, 2004.

27. Markos Papageorgiou, Jean-Marc Blosseville, and Habib Hadj-Salem. Modelling and real-time control of traffic flow on the southern part of Boulevard Peripherique in Paris: Part I: Modelling. *Transportation Research Part A: General*, 24(5):345 – 359, 1990.

28. Yoav Shoham. Agent-Oriented Programming. *Artificial Intelligence*, 60(1):51–92, 1993.

29. V. S. Subrahmanian, Piero Bonatti, Jürgen Dix, Thomas Eiter, Sarit Kraus, Fatma Özcan, and Robert Ross. *Heterogenous Active Agents*. MIT Press, 2000.

30. Seth Tisue and Uri Wilensky. NetLogo: A simple environment for modeling complexity. In *International Conference on Complex Systems*, pages 16–21, 2004.

31. Gerhard Weiss, editor. *Multiagent systems*. MIT-Press, 2013.

32. Michael J. Wooldridge. An Introduction to MultiAgent Systems, 2009.