QUANTICOL

A Quantitative Approach to Management and Design of Collective and Adaptive Behaviours

quanticol
http://www.quanticol.eu

D1.1

Multiscale modelling informed by smart grids

Revision: 1.0; March 26, 2014

Author(s): Nicolas Gast (EPFL/INRIA), Luca Bortolussi (CNR), Jane Hillston (UEDIN), Rytis Paškauskas (CNR), Mirco Tribastone (SOTON)

Due date of deliverable: Month 12 (March 2014)
Actual submission date: March 31, 2014
Nature: R. Dissemination level: PU

Funding Scheme: Small or medium scale focused research project (STREP)
Topic: ICT-2011 9.10: FET-Proactive 'Fundamentals of Collective Adaptive Systems' (FOCAS)
Project number: 600708

Coordinator: Jane Hillston (UEDIN)
e-mail: Jane.Hillston@ed.ac.uk
Fax: +44 131 651 1426

Part. no.	Participant organisation name	Acronym	Country
1 (Coord.)	University of Edinburgh	UEDIN	UK
2	Consiglio Nazionale delle Ricerche – Istituto di Scienza e Tecnologie della Informazione "A. Faedo"	CNR	Italy
3	Ludwig-Maximilians-Universität München	LMU	Germany
4	Ecole Polytechnique Fédérale de Lausanne	EPFL	Switzerland
5	IMT Lucca	IMT	Italy
6	University of Southampton	SOTON	UK
7	Institut National de Recherche en Informatique et en Automatique	INRIA	France

COOPERATION

Executive Summary

Multiple scales are inherent to collaborative adaptive systems (CAS). They can be temporal, for example, adjusting the production of a large power plant takes hours while appliances can switched on or off the instant a signal is received. They can also be organizational and represent a hierarchical structure, for example a smart building that contains rooms. Fluid approximations, and in particular mean-field limits, are a powerful tool to conduct a quantitative analysis of large stochastic systems. Yet existing techniques are not well adapted to study multiple-scale behaviour. This deliverable reports on Task 1.1 and 1.2 of Work Package 1. We mainly focus on two aspects: (1) how to define and construct mean-field limits in the presence of multiple scales and (2) what approximations are needed to build efficient control algorithms for smart grids. This deliverable presents results from published papers by the members of the project as well as the relevant literature. Our approach is to develop the theory in a way informed by examples. The results presented in this deliverable are motivated by their applicability to model smart grids and smart buildings. Hence, each section contains at least one concrete example of how the results apply to our case-studies.

We first present a review of model reduction techniques in the presence of multiple time scales. This occurs when the states of some objects evolve at a much faster time scale than others (for example, small electric appliances and big power plants). We compare existing reduction techniques for deterministic dynamical systems, a mature subject, with on-going work on stochastic systems. This review shows that a number of time-scale reduction techniques can be readily applied to mean-field models. It gives us tools to develop the analog for stochastic systems.

We then describe the situation when there are multiple population scales. Our basic example is when one centralized controller interacts with many appliances. We show that in such cases, the limit is naturally described by a stochastic hybrid system. We describe how to construct the limit and the limitations of the approach. This technique reduces greatly the complexity of the simulation while maintaining a good accuracy.

We develop a novel formalism to describe systems of systems. This can model systems that have a hierarchical organization. This formalism allows us to automatically reduce the complexity of the mean-field equations, by exploiting symmetries in the model. This method can be applied iteratively, to construct hierarchical abstractions of systems. We illustrate our method to describe the behaviour of a collection of smart buildings.

In the last section, we demonstrate the use of optimization tools for building control algorithms in electrical systems that have a large production of renewable energy. We model and treat two challenges: large forecast uncertainties and presence of delay due to multiple time scales. We study two directions based on centralized and distributed control. We develop storage and demand/response management policies, where a central controller sends signals to smart users to adapt the consumption to the production. These policies are more robust to forecast errors than existing strategies.

To conclude, this work package reports on progress that has been made on the fundamental aspects of multi-scale modeling as well as on advances in the smart-grid case study. This constitutes a first attempt to build generic tools that will be applicable to the analysis and the optimization of the other case studies. We will continue the development of these generic methods to incorporate spatial behaviour (Work Package 2) and apply these techniques to fluid model checking (Work Package 3).

Contents

1 Introduction

The QUANTICOL project will build a tool to analyze and improve the behavior of collective adaptive systems. This tool will rely on the use of deterministic approximation techniques, and in particular mean field techniques, to model the behavior and build control algorithms for complex stochastic systems. Our case-studies are characterized by the presence of multiple scales. These scales can be temporal – for example coexistence of fast and slow dynamics – or organizational – for example, a centralized controller and many individuals or a structured hierarchy of systems.

In this deliverable, we present the different research activities that have been carried out in the context of the project to model and understand multiple scale behavior. Within our contributions, we try to cover all aspects, from the most fundamental ones to the more applied ones. The organization of the deliverable reflects this. First, in Section 2, we present an quick introduction to mean-field approximation and its use in quantitative analysis. This section serves as an introduction for the main concepts developed in the rest of this deliverable as well as the Deliverables 2.1 and 3.1. Current mean field theory is focused mainly on ODE limits, which are not fully adequate to describe systems with multiple organizational scales or multiple temporal scales. Hence, in Sections 3, 4 and 5, we introduce a number of formalisms and methodologies to treat multiple scale behaviors. Finally, we present the applications to the control of smart grids in Section 6.

In Sections 3, 4 and 5, we present an overview of different model reduction techniques for multi-scale. The focus is on (1) **multiple time scales** and the existence of slow and fast behavior, (2) **multiple population scales**, when one population has a small number of individuals while the other one has a large number of individuals; and (3) **multiple organizational scales**, when the system can be viewed as a system of systems. More specifically, our goal has been to develop formalisms from which mean field equations can be automatically constructed. In particular, we study systems of systems and models with both stochastic and immediate transitions. This overview is a first step toward building generic approximation methods for multi-scale systems. In the near future, we want to strengthen the mathematical foundations of these methods and continue the development of their application to our case studies.

In Section 6, we explore **control problems in electricity networks**. We consider the joint problem of managing generation and storage in the presence of forecast uncertainties. This problem naturally leads to multi-scale behavior, in which the short-time scale (consumption and storage) has an influence on the slow time-scale (generation). We explore ways for an operator to stabilise a system by sending active signals to control users' consumption. These problems forced us to develop a novel methodology based on stochastic Lagrangian decomposition and trajectorial forecast. We think that this method will serve as a building block for building efficient control algorithms that are widely applicable. In particular, we want to apply these algorithms to other resource allocation problems linked with the smart grid and the bike-sharing case studies to build generic optimization tools.

This deliverable is based on a number of publications by members of our team. Section 2 is detailed in [BHLM13]. Section 3 summarizes [PB14]; Section 4 is based on [BP13]; Section 5 covers [BTH14] and Section 6 describe the results of [GTLB14, GLBT14, GLBPT13].

2 An Overview of Mean-Field Models used in Quantitative Analysis

Markovian stochastic processes are a classical tool used to describe the temporal behavior of a system. It is widely used for quantitative modeling in many domains like computer science, computational biology, electrical engineering, etc. One of the main limitations of Markovian models is the so-called *curse of dimensionality*: the state space grows exponentially with the number of elements constituting the system. This poses practical limitations on our ability to analyze these systems (like steady state computation, transient analysis or stochastic model-checking), as soon as the number of elements of the system exceeds a few tens. This problem is particularly prominent in collective adaptive systems that can have hundreds or thousands of agents.

To overcome this limitation, there has been a growing interest in approximation techniques based on differential equations. These techniques build a deterministic approximation of the system that is more and more accurate as the number of agents of the system goes to infinity. They are therefore particularly well adapted for very large systems. These techniques can automatically construct a set of differential equations that describe the emergent behavior of the system. In the remainder of the section, we will quickly review these models and their limitations. More details can be found in [BHLM13, BB08, DN08].

2.1 Stochastic Models of Population Dynamics

Our starting point is a population model, in which many agents, of heterogeneous types, interact together. Each agent is described as an automaton and changes state either spontaneously or due to the interaction with other agents. A classical example is the worm epidemic model of [BHLM13]. Each agent is described by a state (Susceptible, Exposed, Infected or Recovered) that changes as a results of an interaction with other agents (*e.g.*, infection by another infected agent) or by interaction with external sources. The behavior of a single agent can be described as in Figure 1(a). This model is fully described in [BHLM13].

A suitable representation of such systems is to specify the agent counts in each type. Each transition in the model can then be described in terms of how it affects such a state vector. When the time between two transitions is exponentially distributed, this model can be described by a continuous time Markov chain (CTMC). The state of the stochastic model is then described by the joint probability of having X_1 agents in state 1, X_2 agents in state 2, etc., at time t. The probability at some later time is found by solving a system of linear ordinary differential equations (ODEs), known as the 'Master equation' or 'Kolmogorov's equation'. The main issue with the Master equation is the state space explosion: A model that can accommodate up to N_1 agents in state 1, N_2 agents in state 2 etc., has dimension $N_1 \times N_2 \times \cdots \times N_n = \exp(\lambda n)$[1]. This explosion is a severe limitation for quantitative analysis.

2.2 Mean-Field Approximation

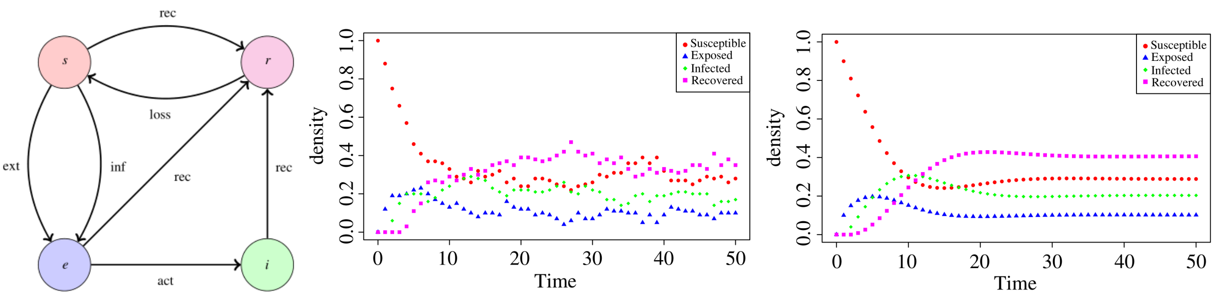
It can be shown that, for a large range of models, the asymptotic behavior (as the number of agents grows large) of the stochastic model is the solution of a set of differential equations. These equations are often referred to as the mean-field approximation or fluid approximation of the original system. Each ODE determines the temporal evolution of a species, denoted by x_1, x_2, ..., x_n. Using vector notation $\mathbf{x} = (x_1, \dots, x_n)$, the system of ODEs can be cast as a single vector ODE

$$\dot{\mathbf{x}} \equiv \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{f}(\mathbf{x}). \tag{1}$$

The temporal solution of this ODE is called a *trajectory* $\mathbf{x}(t) = (x_1(t), \dots, x_n(t))$.

An example of such approximation is shown in Figure 1. We simulate one trajectory of the stochastic model of a worm epidemic and report the results in Figure 1(b). This figure shows the evolution of the proportion of agents that are in one of the four states over time. The mean-field approximation of the model is represented in Figure 1(c). While there is a discrepancy between the stochastic system and the ODE, the latter is able to predict very well the trends of the dynamics of the original model. The mean-field approximation is a *nonlinear* ODE in n dimensions, the stochastic model is a *linear* ODE in $\exp(\lambda n)$ dimensions. This dimension causes more problems than any nonlinearity-related limitation of a mean-field model.

[1] λ is a logarithm of the geometric mean of the N_is, a number that varies, typically, between one and ten.



(a) Model of one agent. (b) Simulation for $N = 100$ agents. (c) Mean-field approximation (ODE).

Figure 1: The model of a worm epidemic from [BHLM13]. We plot the evolution of the number of agents in each state as a function of time. The behavior of the stochastic system with $N = 100$ agents is close to the one of the mean-field approximation (courtesy of [BHLM13]).

Current mean-field approximations are mainly limited to differential equation limits, which are not adapted to multiple scale dynamics. In the remainder of this work-package, we report on the different research activities that we conducted to extend these models to our case studies, in particular to describe multiple organizational and temporal scales. One of our focuses has been to try to build automatic processes for constructing and evaluating the limiting model.

The theoretical foundations developed in this workpackage are directly related to Workpackage 2, that deals with abstractions of spatial representation and Workpackage 3, which develops model checking techniques based on the fluid approximation. The work conducted in this workpackage is driven by the case studies. As such, it is very close to Workpackage 5. The main difference is that the emphasizes of this workpackage is on the theoretical foundation whereas Workpackage 5 focuses on the data gathering and modeling related to the case studies.

3 Model Reduction for Multiple Time Scales

An important aspect of many population models, including CAS scenarios, is the presence of multiple time scales, which in its simplest form means the coexistence of fast and slow events. An example in the context of smart grids is the presence of auxiliary generators that activate with various delays in case of a peak in the demand. In general, multiple time scales arise in CAS when we consider smart systems in which computational devices interact together and with humans: computations are much faster than communication between devices, which in turn act at a much faster time scale than the reaction time of humans.

Detection and treatment of such multiple scales is a non-trivial problem, especially in the context of quantitative models, like those considered in the QUANTICOL project. In this section, we present a review of several existing methodologies, developed in physics and applied mathematics, which could be suitable to apply to models. As we are in a preliminary stage of investigation, we will not discuss specific examples. Applications of these techniques to smart grids and other CAS systems will be a priority for future work. In particular, we consider model reductions both for stochastic models (CTMC) and for their mean-field limits, given by a set ODEs. The methodology for the fluid models is more mature, hence we will begin the discussion with it, turning afterwards to CTMCs.

3.1 Reduction Techniques for Fluid Models

In some models, multiple time scales are present. Their effects can be visualized by plotting the evolution of a population $x(t)$ versus time. An example is shown in Figure 2. The curve has two clearly identifiable legs: a brief, steep initial drop/rise, crossing over, in a knee-like manner, to a slowly varying plateau. A practical issue for QUANTICOL is the prohibitively high cost that these models impose for accuracy of a numerical solution. For the stability of numerical integrators, the fast time-scale imposes a very small time-step, which is computationally expensive and can jeopardize the accuracy of the solution due to numerous finite arithmetic operations.

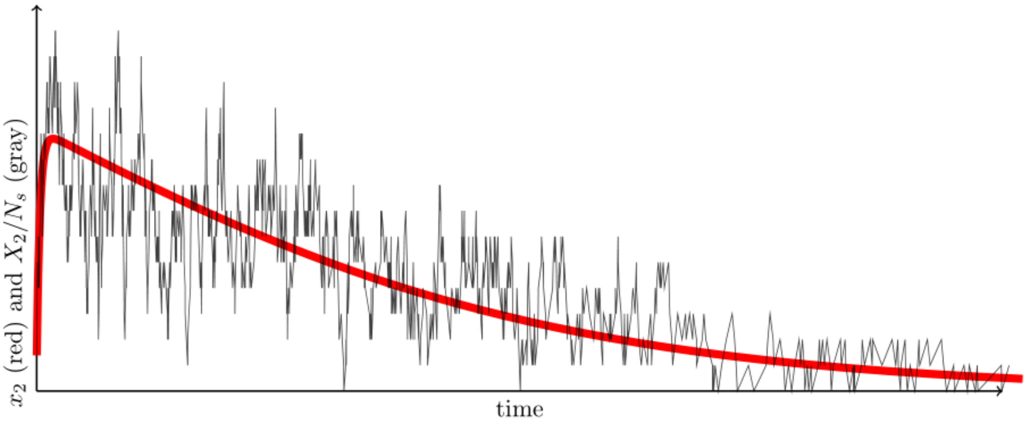


Figure 2: Evolution of the number of sensors that are transmitting data in the wireless sensor network model of [PB14]. We compare the fluid model (thick red curve) with the stochastic model.

It is long known that the so-called *model reductions*, developed by physicists and applied mathematicians, can sometimes be very effective in approximating the most relevant slow time-scale events. Several numerical methods exist in the literature, each based on some notion of *fast and slow* separation of some quantity of the model such as transition rates, species counts, probabilities, Jacobian eigenvalues. There is no silver bullet; each known methodology has its strengths and weaknesses, and limits of applicability. We believe that the winning strategy is to know all of them, but apply the most appropriate one for a given situation. We describe some of these methods here.

The singular perturbation theory [Ver05] is applicable when the mean-field ODE has a small parameter ϵ such that (1) assumes the Tikhonov form: $\mathbf{x}$ can be partitioned in two groups, $\mathbf{x} = (\mathbf{y}, \mathbf{z})$ such that : $\dot{\mathbf{y}} = \mathbf{g}(\mathbf{y}, \mathbf{z})$, $\dot{\mathbf{z}} = \frac{1}{\epsilon}\mathbf{h}(\mathbf{y}, \mathbf{z})$. $\mathbf{z}$ is called the *fast variable* and $\mathbf{y}$ is called the *slow variable*. In the limit $\epsilon = 0$ (i.e. the infinite separation of time scales), the solution is constrained by $\mathbf{h}(\mathbf{y}, \mathbf{z}) = 0$. This equation defines an embedded surface, called the *slow manifold*. For small $\epsilon > 0$, the slow time-scale solution remains close to the slow manifold.

The quasi-equilibrium (QE) approximation can be applied when the changes of some variables are slower than the changes of the others. It typically describes an equilibration process. When ϵ, the ratio between the slow and the fast time-scale, is small, the system is almost in equilibrium, but slowly diverge because of a *leak*, induced by $\epsilon > 0$. Sometimes, an intermediate category of transitions is added, distinguishing three groups: the *fast*, the *rate-controlling*, and the *dormant* transitions [LG94, GRZ10].

The quasi steady-state (QSS) approximation corresponds to a re-scaling where some transition are faster than others by a factor $1/\epsilon$ [GRZ10, SS89]. In that case, we can approximate the dynamics by considering that these variables are in steady-state. We emphasize that QSS is used in a more general context of *fast-slow* separation of species, even when a model cannot be brought to the Tikhonov form.

In simplest QSS or QE setups, the corresponding ODE can be brought to a Tikhonov form by a combination of a coordinate transformation and an appropriate scaling of variables and the time. A paradigmatic example of this situation is the mean-field Michaelis-Menten model [JG11]. Either QSS or QE approximation is applicable to this model in the appropriate limit. Excellent references to consult are Segel and Slemrod [SS89] for the singular perturbation theory, and Fraser [Fra88] for the relation between QE and QSS approximations. In more complex models, the Tikhonov form is not available, and one has to resort to numerical techniques. In our opinion, three techniques deserve special attention.

Fraser's method, developed by Fraser et al [RF90, NF89, Fra88] builds on a geometric picture: there exists a manifold that is strongly attractive transversally and weakly attractive internally. This means that all *external* trajectories $\mathbf{x}(t)$ tend to approach this surface and land on it, following a rapid transient. Afterwards, they transition along this surface slowly, towards the equilibrium. Fraser builds an iterative procedure that allows one to find a parameterization of this slow surface. Suppose that this so-called *center manifold* is computed. Then, to complete the approximate trajectory $\mathbf{x}(t)$ one only needs to find the landing point. Fraser's method is a numerical technique that does not require the Tikhonov form. Fraser's ideas have been recently extended in [NF13a, NF13b].

The **Intrinsic Low-Dimensional Manifold** method (ILDM) by Maas & Pope [MP92b, MP92a] is also a geometric method. The objective is, just like in Fraser's method, to parameterize a surface that contains all the slow dynamics. However, it is more literally a *reduction* technique because it allows the dimensionality d of the parameterization to be defined by the user and it allows the selection of which species are looked at (an example: in a network with 10 species we could ask for a two-dimensional manifold of species 3, parameterized by species 1 & 2). ILDM is more general than Fraser's method. It is one of the most sophisticated methods and is used heavily in chemistry and in controlled combustion in particular. Various case studies and extensions of ILDM exist. For ideas about algorithmic implementation, see [BM11, CK11]; for coupling with transport, diffusion equations and the temperature, see [BM07]; for additional techniques, such as the principal component analysis (PCA) and multivariate adaptive spline regression (MARS), used in conjunction with ILDM, see [YPC13]. For an overview of recent developments, see [KSR$^+$11]. For more on the related subject of diffusion-reaction equation reduction, see [Dav06a, Dav06b].

The **Computational Singular Perturbation** (CSP) method by Lam & Goussis [LG94, Lam93, LG91, LG89] was born in the numerical computation community, but has been used in a broader context as well (see [GN06] for an example of circadian rhythms in Drosophila). The authors themselves consider this method as a kind of *stiff ODE integrator* with an added bonus that a special structure of mean-field models is taken into account. As in the ILDM, CSP is based on the fast-slow separation in the eigenvalue spectrum of the Jacobian. In contrast to ILDM, CSP attempts to make a local coordinate transformation along the trajectory, to bring the system to a local Tikhonov form. In our opinion, ILDM and CSP are the two methods to be considered for complicated mean-field models. Although the original exposition of [LG94] was quite heuristic, [ZKK04b, ZKK04a] give a more formal presentation and error analysis. For an interesting discussion about the relation between manifolds and degeneracy problems, using the van der Pol oscillator as an example, see [Gou13]; for the hybridization of ILDM and CSP, see [ZAI13]. For an existing numerical toolkit integrated into the *Chemical Workbench* computational package, see [LOC$^+$12].

ILDM and CSP have always been the two main competiting algorithm for model reductions, [Gou12]. We believe that these two methods should be considered for the models of QUANTICOL's case studies.

3.2 Reduction Techniques for Stochastic Models

Time-scale decomposition for Markov chains on finite state spaces has been dominated by arguments based on near-complete decomposability of the infinitesimal generator matrix since the 60s [SA61]. The techniques are also known under the name of matrix perturbation theory, and require that the generator is a (block) diagonally dominant matrix [Ste90, BO91, GVL12, Ste01, Kat95]. The infinitesimal generator matrix of a CTMC is said to be amenable to time-scale decomposition if it has a block structure in which the diagonal blocks represent fast events or transitions, whilst slow events occur only in off-diagonal blocks. This means that the process will spend extended periods interacting within the *fast* sets of states represented by the diagonal blocks, and only occasionally move between these sets via a slow transition. In time-scale decomposition this alternating behaviour between sequences of fast transitions and intermittent slow transitions, leads to near-complete decomposability, where it is assumed that the fast sets of state reach quasi-steady state between the slow transition [Cou77]. Consequently, the fast submodels are solved separately, as well as an macro-model capturing the slow dynamics which represents each set as a single state and the transitions between them. There are known error bounds for the technique based on the magnitude of the largest element in an off-diagonal block. Time-scale decomposition has previously been characterised for both stochastic process algebras [Mer98, HM95] and stochastic Petri nets [AI89, BT93] meaning that models amenable to near-complete decomposability can be recognised by syntactic or structural conditions respectively. But these characterisations are based on models with only individual instances of components and not populations. In recent work, Pourranjbar has developed a characterisation for large scale stochastic process algebra models, which can avoid consideration of large populations, making the detection of a partition for time-scale decomposition more efficient [Pou14].

However, it may easily happen that a model has multi-scale characteristics but is does not have this block-diagonal structure. For these, more complicated models, we find inspiring the fast–slow species, rate, eigenvalues separation paradigm of mean-field reductions. As the survey in section 3.1 suggests, this approach could yield new techniques that could reach beyond the diagonal setting. Indirect modelling of the Master equation in terms of stochastic trajectories is a promising trend, thanks largely to the exact *stochastic simulation algorithm* (SSA). Because SSA computes stochastic trajectories, it is tempting to draw appropriate analogies with the mean-field. From section 3.1, one may suspect that QE/QSS are equivalent to making certain assumptions about the underlying Markov chain, which aren't restricted to mean-field modelling. A series of recent articles have established that such parallels between the mean-field QE/QSS scenarios and stochastic models could indeed be drawn. In what follows, we summarize what seem the most reasonable implications of QE and QSS scenarios to a stochastic model.

As with the Tikhonov form of the ODE in section 3.1, we assume that the variable $\mathbf{X} = (X_1, \ldots, X_n)$ can be partitioned into groups of slow and fast variables, $\mathbf{X} = (\mathbf{Y}, \mathbf{Z})$. The stochastic model analogy of the $\epsilon = 0$ limit is the marginal probability of the slow variable, $P(\mathbf{Y}; t) = \sum_{\mathbf{Z}} P(\mathbf{Y}, \mathbf{Z}; t)$. This probability satisfies a Master equation, with all the fast rates having been eliminated from the equation, but the slow rates remaining, having been averaged over, weigthed by the *fast conditional to slow* probability $P_c(\mathbf{Z}|\mathbf{Y}; t)$. The separation of time scales is implemented as the assumption of a Markovian property of P_c (contested in [CGP05b] and defended in [HR05]). Therefore P_c, too, satisfies a Master equation, usually of a *birth-death* type. From this point on, several options have been explored in the literature. Rao & Arkin [RA03] assume that P_c is in an instant, *local* equilibrium, parameterized by the current value of the slow variable. One may immediately recognize a QE scenario here. Haseltine & Rawlings [HR02] assume that the fast variable is too fast for the diffusion to be relevant. Instead, they use a deterministic approximation, but assume that the spread of a fast variable is quite small[2]. This is the QSS scenario. Alternatively, Cao et al. [CGP05a, CGP05b] use fictitious virtual transitions for the fast dynamics. We mention a stochastic simulation software with some model-reducing capabilities [WFLP12]. It uses weighted Petri nets for automatic lumping, and the stochastic simulation algorithm or tau-leaping.

To conclude, we could draw a historical parallel. The mean-field QE reduction for the famous and paradigmatic Michaelis-Menten model was first discussed in 1913 (see a review [JG11]) whereas the corresponding stochastic model reduction was pioneered between 2002 and 2005 [RA03, HR02, CGP05b, HR05] – *merely* 90 years later. The mean-field techniques, ILDM, CSP, and Fraser's methods reached their maturity somewhere between 1985 and 1995 [SS89, Fra88, Ver05, MP92a, LG94]. The community is still waiting for the analogous stochastic methodology to be developped.

3.3 Fluid Analysis of Models with Immediate Transitions

It is sometimes relatively easy to separate fast and slow transitions, for example, if the time-scale separation is so large that fast transitions are modelled directly as immediate. This is the modelling

[2] This discussion demonstrates that system size, moment expansions and other approximations are relevant to stochastic reductions. See [BB11a, BB11b] for news in moment expansions, or a classic [Gar85] for general background.

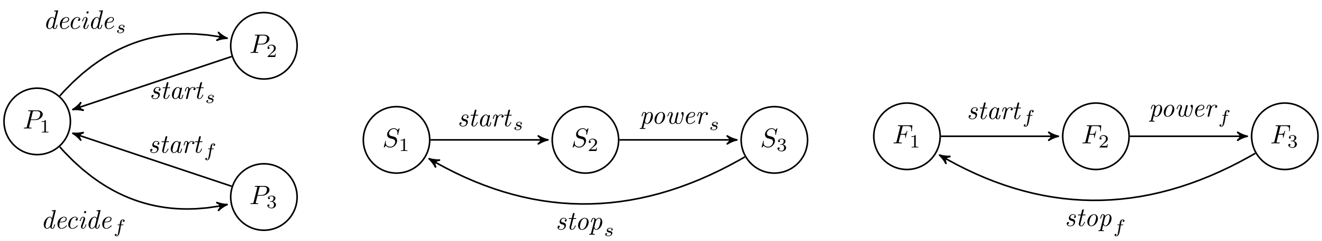


Figure 3: Fast and slow generator model in the interacting automata formalism of [BT13].

approach, for instance, of Generalised Stochastic Petri Nets [Haa02]. There are plenty of examples of (nearly) immediate transitions in performance analysis of computer systems: the spawning of a thread/process (compared to serving a dynamic web page), the forking/joining of threads (relative to the thread's own computation), and so on. Similar scenarios, can be found also in the context of smart grids. An example is the activation of power generators depending on the current offered load. There are two classes of generators: *fast* and *slow*, where the speed refers to the time they take to start producing electricity after they are powered on. We can model the activation time of the fast generators as immediate, compared to slow generators. Generators are controlled by an operator, who is interested in finding optimal policies to match the demand at a minimum cost/carbon footprint [GTLB14].

The most common modelling frameworks for stochastic processes with immediate actions are Generalised Stochastic Petri Nets [Haa02] and Interactive Markov Chains [Her02]. In both cases, to obtain a CTMC, one has to generate the state space of the model, identify the states having an immediate action active, called *vanishing states*, and remove them. However, this is computationally infeasible if the state space is very large, as we expect for CAS models, and it does not make it possible to obtain mean-field equations. In [BT13], we propose an alternative approach for the direct generation of fluid equations from a model specification in terms of interacting automata. Our models are composed of populations of interacting automata, each of which can participate in activities that trigger an immediate transitions or in activities that trigger an action after a random delay. We allow synchronisation of automata on immediate and delayed activities, currently with some restriction. A simple model of fast and slow generators in terms of interacting automata is shown in Figure 3. It can be explained as follows:

- Automaton P_1 models a provider. It decides which generator to use. After the decision is made, it starts the respective generator. Thus $decide_s$ and $decide_f$ are independent actions, while $start_s$ and $start_f$ are synchronized actions. These actions take effect after some delay.
- Automaton S_1 models a slow generator. Once it receives a $start_s$ message, it powers up. After some time, it stops. For this automaton we assume that all actions are delayed; actions $power_s$ and $start_s$ are independent actions.
- Automaton F_1 models a fast generator. It is analogous to S_1, except that $power_f$ is now an immediate independent action.

The major contribution in [BT13] is an algorithm for deriving a consistent set of fluid equations for such a class of models. This allows a fast analysis in scenarios with many generators and providers, and is most effective in optimisation problems, like trying to maximize electricity production, subject to environmental constraints, for instance on the pollution or carbon footprint. The interesting aspect is that the so-obtained equations present discontinuities in the vector field, which in some cases, but not all cases, can be eliminated by suitable coordinate transformations. We are currently extending this work by removing restrictions on the model and by better characterising the relationship between the stochastic and the fluid model.

3.4 Discussion and Future Work

In this section we have presented a survey of time-scale reduction techniques, which is the starting point to identify the most suitable approaches for CAS systems. Our next step will then be an investigation of different scenarios of CAS in which these methods can be beneficial. We will obviously start from the case studies of the project, focussing mainly on smart grids.

A picture that emerges from the survey in Section 3.2, is that several stochastic model approximations are possible, which are equivalent to a single mean-field QE or QSS scenario. The reason for this is a richer structure of distributions compared to deterministic fluid models. In all the examples looked at, a multiple scale reduction was followed by an additional layer of approximations, applied to distributions. Moment closures, equilibrium or deterministic approximations, mentioned in Section 3.2, yield possibly non-equivalent reduced models. A dedicated study of this issue is warranted, to develop a more complete understanding of the implications of these secondary approximations to the QE and QSS scenarios. In particular, we want to better clarify what are the formal relationship between a reduced stochastic and the corresponding reduced mean-field models. Specifically, we want to understand under which conditions mean-field convergence continues to hold.

Multiple time-scale reduction has to be approached within the QUANTICOL framework also from a practical side. Here we will investigate an automatic reduction scheme, combining the available techniques and selecting the most appropriate one for the scenario of interest. This can take the form of a decision tree that checks several conditions to select the reduction technique to be applied.

Another direction we want to investigate is how to bridge these techniques, which are defined in terms of *low-level* mathematical models, with the high level specification of systems in terms of stochastic process algebras (SPA). In particular, we want to understand what the connections between the syntax of a process and the possible reductions of its mathematical representation are, by extending the approach of [Pou14]. When this is possible, static analysis at the SPA level will be a much more effective method to construct reduced models.

4 Multiple Population Scales and Hybrid Models

Mean-field is an effective approach when we deal with systems composed of interacting agents, all present in large quantities. This assumption is usually verified for decentralised models, where the control is fully distributed. However, there are situations which diverge from such a scenario, where some populations of agents are present in a single or few copies. An example in smart grids is the use of appliances like water boilers to provide a service to the grid, as described in [GLBT14]. Boilers can be turned on or off randomly as a consequence of hot water usage, but are also connected to a central control system, which can switch them off synchronously when there is a shortage of available energy in the network, in order to decrease the demand. A single controller interacts with a population of boilers. This simple example is shown in Figure 4.

In those cases, we can still apply mean-field techniques to portions of the system, namely those with a large number of agents (in the example above, the water boilers). This results in (stochastic) hybrid models, in which some parts of the system are approximated continuously (and deterministically) and some others are kept discrete and stochastic [BP13, Pah09]. These models are usually easier to analyse than the fully stochastic ones. In the following, we will quickly introduce stochastic hybrid models and hybrid mean-field limits, concluding the section with a discussion on future perspectives and on the potentiality for CAS.

4.1 Hybrid Models: an Overview

Stochastic Hybrid Systems (SHS) are dynamical systems which are subject to a seamlessly interleaved discrete and continuous dynamics. In this project, in particular, we will focus on stochastic hybrid

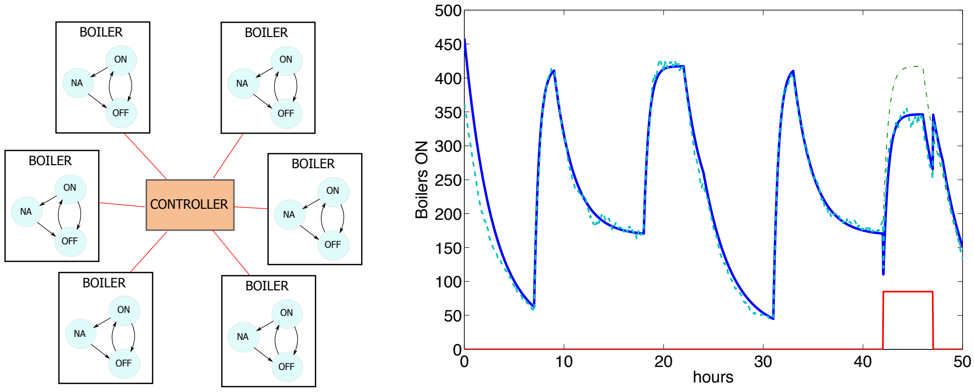


Figure 4: Simple model of a smart water boiler, which can be switched off by a central controller to act as a virtual energy reservoir. Each boiler (left picture) has three states: on, off, and not available. This last state corresponds to the boiler being switched off by the central controller. Boilers switch randomly between states on and off. The on rate, in particular, depends on the period of the day, and is higher during the morning and evening, when people usually take showers. The controller, in response to high energy demands, can turn off only boilers that are in the on state. We assume it turns off a fraction of active boilers, here 50%. Reactivation of boilers happens after a random time, connected with the duration of the energy demand peak. If the number of boilers is sufficiently high, then we can approximate it as continuous. The controller, instead, remains a discrete agent, giving rise to a hybrid model. In the right figure, we show a trajectory of the hybrid model (boilers on, solid blue line), compared with a trajectory of the fully stochastic model (dashed cyan line) for a small system with 500 boilers. Already in this case the accuracy is good. The red square line at the bottom shows the period in which the controller turned the boilers off. The thin blue dash-dotted line shows the number of active boilers as if the controller would have not acted, visually depicting the amount of energy saved. In this example, we have noticed a speed-up in the simulation of one order of magnitude with respect to a wholly discrete simulation.

systems [Buj12, Dav93], in which the continuous dynamics is deterministic, while the discrete dynamics is stochastic and Markovian. This class of models is also known as Piecewise Deterministic Markov Processes, PDMP [Dav93], and has been first investigated in the context of economic systems. More specifically, each PDMP is composed of a set of discrete states, called also modes, and a set of continuous variables. In each mode, continuous variables evolve deterministically, following the trajectories of a mode-specific vector field. Discrete states can change due to discrete transitions of two kinds. Stochastic Markovian events, firing spontaneously, can lead the system to a different mode, possibly resetting the value of continuous variables. Immediate transitions, instead, happen when the continuous trajectory hits the boundary of the activation region of the transition, usually specified by a guard predicate, and can also induce a change both in the discrete and in the continuous components of the state of the system. The dynamics that can be described within this framework are quite rich, encompassing, for instance, Generalised Semi-Markov Processes, in which stochastic transitions can fire at generally distributed times [Whi80], and Piecewise-Smooth Dynamical Systems, which are often employed in performance analysis [GG12]. We stress that SHS are different from more standard notions of hybrid automata [Hen96], in which discrete transitions are activated non-deterministically rather than stochastically. Moreover, more complex notions of SHS can also be considered [Buj12], for instance allowing continuous dynamics to be driven by stochastic differential equations.

4.2 Hybrid Mean-Field

Like ODEs emerge as the mean-field limits of sequences of Continuous Time Markov Chains when all populations grow with the system size, PDMP should intuitively appear as mean-field limits in which only a portion of the population variables increases, while other populations remain independent from the system size. This statement can be properly formalised, resulting in a set of theorems characterising hybrid mean-field convergence [Bor12, Bor10, CDMR12]. It turns out that convergence holds under some suitable conditions on rates and update vectors of population models, which are often satisfied in practice. We sketch some details below, referring the interested reader to [Bor12].

- The starting point is a population CTMC model, as defined, for example, in [BHLM13]. The dynamics is defined by a set of transitions, which happen at exponentially distributed random times, according to a state dependent rate, and modify system variables according to a specified update policy. Immediate transitions, happening in zero time when their guard is satisfied, can be considered as well.
- Stochastic transitions are partitioned into two classes: those to be approximated as continuous and those remaining discrete. System variables are partitioned accordingly: *continuous variables* are all and only those modified by continuous transitions. *Discrete variables* define the modes of the hybrid automaton: each possible tuple of their values is a discrete *mode*.
- Continuous transitions are then assembled together to construct a set of ODEs in each mode. The mode specificity is due to the dependency of rates on discrete variables. Discrete transitions remain stochastic, firing at exponentially distributed times (if their rate depends on system size, however, a different treatment is needed [CDMR12, BB08]).
- Instantaneous transitions become forced transitions in the hybrid automata jargon. Convergence results require further technical conditions on their guards [Bor12].
- Guards in stochastic transitions of the original population CTMC can be considered as well. However, guards in transitions that are approximated continuously introduce discontinuities in the ODEs, which require a special treatment in terms of piecewise smooth dynamical systems [Bor11] or differential inclusions [GG12].

Going back to the water boiler example, we can construct a hybrid mean-field limit in which the number of boilers being on and off is approximated by a continuous variable, while the centralised controller remains discrete. Specifically, it is capable of triggering a transition, depending on the global energy request, that can switch a fraction of boilers off. In the hybrid limit, this becomes a stochastic (or an instantaneous) transition with a reset inducing a discontinuous jump in the continuous variable. A more detailed discussion of the model, together with a visual comparison of the stochastic system with its hybrid limit, is shown in Figure 4.

The simulation of hybrid mean-field models is generally faster than their purely stochastic counterpart [BP13, Pah09]. Abstracting continuously some populations avoids keeping track of many events during the simulation, so that only a few key transitions have to be tracked explicitly. Moreover, one can use conditional moment closures [MLSH11], which provide approximate equations for the average behaviour of the system, conditional on being in a given mode, plus an approximation of the time-evolution of the probability of discrete states.

4.3 Discussion and Future Work

Hybrid mean-field is likely to play an important role in CAS modelling, whenever one considers a system where decentralised controllers interact with a centralised one. Regarding the case studies of the QUANTICOL project, at the moment this seems most relevant for smart energy scenarios.

Regarding future work, there are several directions towards which we can extend the current state of the art. One important point for the project is the automatic construction of the hybrid limit. We aim at producing a method to automatically partition transitions and variables into discrete and continuous. Another interesting aspect to consider is the integration of the hybrid approximation with the multiple time-scale reduction techniques that will be considered in the context of this project, similar to [Gal14] . Furthermore, we could also investigate the use of hybrid mean-field techniques to deal with systems of systems, in particular to properly understand the nature of the approximation discussed below in Section 5. As far as analysis techniques are involved, we will investigate the use of conditional moment closures in the context of CAS models, possibly looking at higher order methods, for instance taking into account conditional variances, or developing some kind of hybrid linear noise approximation.

5 Model Reduction for Multiple Organisational Scales

In this section we discuss a novel formalism to describe *systems of systems* (SoS), namely systems which are hierarchically structured in multiple levels, each with a complex dynamics [BTH14]. The formalism supports the automatic generation of a set of lumped mean-field equations, which greatly reduces the dimensionality of the equations. We will then discuss this formalism in the context of energy control for smart buildings.

5.1 Hierarchical Modelling of Systems of Systems

Modelling and analysis of nested (or hierarchical) structures occurs frequently in many domains, including CAS. For instance, a cloud environment may be thought of as a collection of many computers, each containing other components, such as processors and threads [KMT09]. Note that the hierarchical nesting in these systems is genuinely *structural* and not an abstraction used to hide detail as is done, for example, in statecharts [Har87]. The organisational complexity of these systems means that it is hard to predict their behaviour and it is imperative that performance and reliability characteristics are investigated prior to deployment.

When formal reasoning about these systems is performed with reactive models based on a discrete state-space representation, the problem size grows extremely quickly with the population of components, making the analysis infeasible in practice. For example, nested Markov models have previously been studied, but all the techniques put forward in order to tackle the heavy computational costs due to layering and large multiplicities are concerned with exploiting symmetries and still yield a Markov chain. For example in [Buc94, Buc99], Buchholz exploits Kronecker algebra with hierarchical models to express certain classes of stochastic Petri nets and queueing networks, but this work is limited to only two levels of nesting. Lanus, Yis, and Trivedi consider a class of hierarchical Markov models where the states of automata of arbitrary size can be partitioned in such a way that a reduced model can be constructed which preserves steady-state reward measures of availability and performance [LYT03]. However, the state space size of the CTMC is still dependent (exponentially in the worst case) on the number of the reduced automata.

In our SoS approach we construct a Markov model of nested automata, i.e., automata which are hierarchically organised in a tree. Each node of the tree is assigned a multiplicity, which indicates how many copies of the stochastic automaton are present in the SoS *within each copy* of its parent automaton. Thus the formalism is readily able to capture systems that contain elements at a number of different organisational scales.

5.2 Mean-Field Equations for Systems of Systems

The mean-field approximation is based on a system of ODEs which initially associates an equation with each element of the CTMC state descriptor of the SoS automata. These ODEs estimate the expectation of the model. Unfortunately the state descriptor grows exponentially with the number of levels, and polynomially with the branching level in the class tree, the number of automata in each class, and the number of states in each automata class. Consequently, even mean-field approximation could rapidly become infeasible. Fortunately it is possible to exploit a property of symmetry between

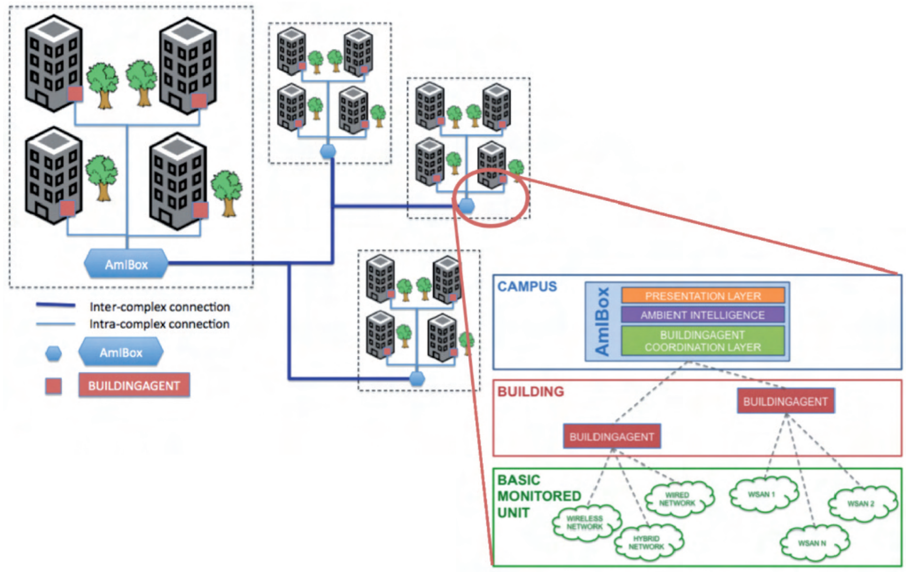


Figure 5: Hierarchical design of a smart building. Courtesy of [ROA13].

the ODEs. Informally, the approach is to establish that two distinct equations for any two automata copies of the same kind (i.e. belonging to the same node in the tree) yield the same ODE solution. This generalises earlier work [TT12], in which Tschaikowski and Tribastone established a basic result of ODE symmetry in the context of the Markovian process algebra PEPA. Following this approach a significant reduction of the ODE system size is possible by considering a representative set of equations for each node in the tree, irrespective of the multiplicities involved.

The dynamics of a SoS is the effect of two kinds of interactions between its components: *horizontal interaction* and *vertical interaction*. The horizontal interaction describes the dynamics of entities at the same level of the hierarchy. This dynamic is not independent of the context, but can be affected vertically by the state of the containing node and by the state of the automata contained in the interacting ones. Another form of vertical interaction occurs when a state change at a certain level in the tree can propagate its effects on its descendant nodes: think about the effect of a computer losing power on the processes running inside it. The first kind of interaction is described by *events*, which are specified by rules at system level, while the second kind of dynamics is described by a *causal map*.

In the ODE corresponding to a state of an automaton in the state descriptor, each term corresponds to a distinct event. Events may influence multiple automata (multiple equations in the set of ODEs) because each event has a synchronisation set specifying which and how many automata are involved in the event. An event also has an associated rate function defining the rate of the interaction. The rate function is a multiplier for the event term within the corresponding ODE, for each automata involved in the event. The additional part of dynamics is the casual map which defines how the transition of a parent automaton impacts on its child automata.

5.3 Smart-building Example

Let us illustrate our SoS approach using an example inspired by [ROA13], where the authors present *SmartBuildings*, a project which aims at improving energy efficiency in large complexes by means of a computer network. The architecture, shown in Figure 5, clearly indicates a hierarchical structure consisting of potentially many elements at each level. At the top level is an *AmiBox*, a smart device that coordinates different buildings; each building is managed by a *BuildingAgent*, which is in turn responsible for managing *basic monitored units*. Communication is possible between AmiBoxes, for instance to coordinate reactions to signals from energy providers; building agents may react to requests from the responsible AmiBox, and carry out control of their monitored units.

To formally describe this system using our model, we first consider the tree that captures the hierarchical structure, where the root represents the top element and each node is an element type. We use $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{U}$ to denote AmiBoxes, BuildingAgents, and monitored units, respectively. The tree is given by $\mathbf{A} \rightarrow \mathbf{B} \rightarrow \mathbf{U}$. Each node is associated with an automaton that describes the *local states* of each element. For instance, a very simple model could be given by two local states, $\mathbf{A} = \{A_{\text{on}}, A_{\text{off}}\}$, and a single transition $A_{\text{on}} \xrightarrow{\text{switch}} A_{\text{off}}$, describing a switch-off event for an AmiBox. With a transition is associated a function that describes the rate at which the event occurs for a single element. For instance, $F_{\text{switch}}(\mathbf{A}) = \lambda {A_{\text{on}}}^{3}$, with $\lambda > 0$, would model the fact that a single AmiBox is switched off after some time when it is in state A_{on}, according to an exponential distribution with mean delay $1/\lambda$. Let us assume that $\mathbf{B} = \{B_{\text{on}}, B_{\text{off}}\}$, i.e., also BuildingAgents have two simple states. The rule $A_{\text{on}} \xrightarrow{\text{switch}} A_{\text{off}} \implies B_{\text{on}} \rightarrow B_{\text{off}} :: 1$ models that whenever an AmiBox is switched off, every BuildingAgent is deactivated. This is expressed by the term " :: 1" which formally means that transitions $B_{\text{on}} \rightarrow B_{\text{off}}$ occur with probability 1 for each agent in state B_{on}. This is an example of vertical interaction; moreover, to model a vertical repercussion one level further down, it is sufficient to define a transition for the local state B_{off} and a rule of vertical interaction for it.

So far we have described the behaviours of individual entities. The description is completed by assigning the population of entities. For instance, $N_{\mathbf{A}} = 2$ and $N_{\mathbf{B}} = 4$ would mean that there are two AmiBoxes, each containing 4 building agents.

Using standard approaches, this system can be approximated using a system of ODEs, giving the time-course evolution of the number of automata in each local state. However, a direct translation into ODEs would require us to keep track of the identity of the parent where a population of automata resides. For instance, the ODE system would have equations for $b^1_{\text{off}}(t)$ and $b^2_{\text{off}}(t)$, denoting the number of BuildingAgents in the off state in the first and in the second AmiBox at time t, respectively. It is easy to see that the number of ODEs grows polynomially with the number of levels in the tree. Our main result in [BTH14] is to show that the ODEs enjoy symmetries that allow us to reduce the ODE system to a significantly (i.e., polynomially) smaller one having a single equation for each local state, *independently from the identity of the parent*. In this example we would have that $b^1_{\text{off}}(t) = b^2_{\text{off}}(t)$, $b^1_{\text{on}}(t) = b^2_{\text{on}}(t)$, and so on. In other words, the size of the ODE system becomes independent of all the multiplicities in the model. This extends our capability of studying quantitative models of hierarchical systems. Furthermore, we show in [BTH14] that while the computational gain is significant, the error induced by the ODE approximation is negligible and tends to decrease as the system size increases.

5.4 Discussion and Future Work

The model reduction approach of [BTH14], discussed in this section, can play an important role to simplify the treatment of multi-scale systems. In addition to an implementation of the current work, we need also to investigate in more detail the quality of the so-obtained approximation, controlling, if possible, the error from the Markov population model. We also plan to look in more detail at the smart building case study, and at other scenarios from smart grids.

6 Distributed Optimization for Storage and Demand-Response

The classical approach to the control of electricity networks involves a combination of both frequency and voltage controls at different time scales. To simplify the control scheme, the networks are decomposed in different layers, from distribution network to transmission network. Traditional systems are managed in a centralized way. They use good load prediction mechanisms with minimal storage and few monitoring points. Each layer guarantees its behavior to other layers, using small safety margins to recover from unexpected events. This approach is robust and efficient as long as the forecast errors are small.

[3] Here $A_{\text{on}} = 1$ if the state is A_{on} and 0 otherwise.

These networks are experiencing a tremendous transformation and their control scheme has therefore to adapt. Future energy systems will incorporate more distributed generation, which is stochastic by nature. In particular, renewable energy sources, such as wind and solar, are highly volatile and difficult to predict. Current distribution and transmission networks are not powerful enough to support a very high penetration of distributed generation. For example, the German transmission network cannot transport enough energy from the wind farms in the east to where it is consumed in the west[4]

Within our work, we try to rethink this layered control scheme and model the dependencies between layers. Our goal is to build control algorithms for these systems of systems. Specifically, we study virtual or real storage policies to deal with the stochastic nature of renewable energy. These systems naturally have several time-scales. The base load power production is provided by large generation units (for example, nuclear power plants) that are slow to react and have to be scheduled hours in advance. At a much faster time-scale, demand-response programs allow a grid operator to remotely switch on or off the appliances of users. This can occur at a time resolution of a minute or less. Our approach has been to consider the multi-scale problem of jointly scheduling generation and storage or demand-response. Our contributions to this problem have been published in [GTLB14, GLBT14, GLBPT13]. They are summarized in the rest of this section.

6.1 Centralized Control and Markov Decision Problem with Delays

The ability of an electric generator to modify its electricity production varies greatly for different types of generators. Large power plants, such as nuclear or coal power plants, are efficient but cannot adapt rapidly to changes. In contrast, fast-ramping generators, like gas turbines are more expensive but react rapidly. Renewable energy sources, such as wind, are characterized by non-dispatchability, high volatility, and non-perfect forecasts. These undesirable features can lead to energy loss and/or can necessitate a large reserve in the form of fast-ramping fuel-based generators. The use of energy storage can, to some extent, compensate for these negative effects, by playing the role of a buffer between demand and production.

In [GTLB12, GTLB14], we revisit a model of storage proposed by [BGK12]. We study the impact on the performance of wind prediction quality. Our problem has two time-scales: conventional generators are slow to adapt but are controllable. Storage is fast but has limited capacities. We cast our problem as a stochastic optimization problem with delay. This results in a Markov decision process with delays. The stochastic nature of the problem comes from the uncertainties in wind forecast. We provide a theoretical bound on the trade-off between energy loss and the use of reserves, given a certain irreducible stochasticity in the forecast errors of the renewable production. When the energy capacity of the storage is large, this bound is tight and we develop scheduling algorithms that attain this bound.

We use the statistics of forecast errors to construct a policy for small to medium capacity. This policy uses a mix of model predictive control and stochastic optimization. More precisely, at each time step, we compute the expected value of the storage level for the future. We use this deterministic prediction and the past statistics of forecast errors to decide how much energy has to be produced and stored in the future. The resulting policy beats all tested policies, in particular the policy that aims to keep the storage level balanced. The proposed policies are evaluated on real data collected in the UK made available by ELEXON[5]. This study can serve as an example for designing an efficient energy market that accounts for the available storage, more specifically, incentive and pricing mechanisms that match the socially-optimal performance.

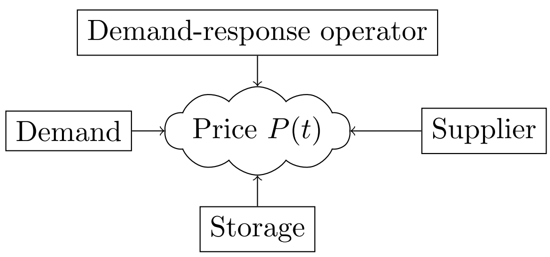
[4] `http://eurodialogue.org/Wind-energy-surplus-threatens-eastern-German-power-grid`, accessed 01/30/14.
[5] `https://www.elexonportal.co.uk/`

6.2 Decentralized Control, Mean-Field and ADMM

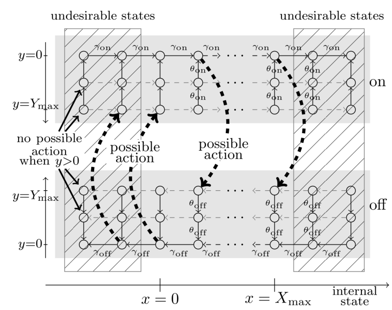
The electricity demand is traditionally considered as inflexible: the operator has no means to indicate to the user when to switch on or off an appliance. Hence, to guarantee the stability of the grid, the production has to adapt to consumption in real-time. Demand-response mechanisms have emerged in order to take advantage of loads that can be delayed or anticipated, like heating systems, air-conditioning (AC) or fridges. It can be implemented by having consumers react to prices or to congestion signals.

In [GLBPT13, GLBT14], we use a game-theoretic approach to study a two-stage market model composed of four players: (1) the consumer, who serves non-elastic loads, (2) the supplier, who provides base-load production, (3) the storage operator, who can buy or sell energy, and (4) the demand-response operator, who controls elastic loads. We mainly focus on real-time market over the period of one day. At each time slot t, each player buys electricity on the market at a price $P(t)$ to satisfy their need. A schematic representation of the market is depicted in Figure 6(a). The exact model is described in [GLBT14]. For our exposition in this document, it is sufficient to highlight the following characteristics of our model:

- The model takes into account the dynamical aspects of generation and demand-response. The uncertainties about wind or demand forecasting are represented by a non-stationary stochastic process.
- Each player has a selfish utility function that depends on its internal variables (for example the electricity demand or the generation cost) and the quantity of energy that is bought or sold at each time t. To take his decision, each player solves a stochastic optimal control problem.
- The elastic loads are composed of a set of appliances (fridges or boilers). The state of each appliance can be described by a Markov chain (represented in Figure 6(b)).
- The control is closed-loop. The players' decisions depend on their knowledge and observation up to time t and on the statistics about the future but they cannot observe the future.



(a) Schematic model of the real-time electricity market.



(b) Each flexible load is modeled by a Markov chain.

Figure 6: The real-time electricity market model and the demand-response model.

As in the previous section, the problem of maximizing the expected sum of the selfish utilities can be mapped as a Markov decision process, where the stochastic nature of the problem comes from the various possible forecasts. However, the state-space of this problem is huge. In [GLBT14], we show how to transform this problem into solvable problems by using the following technique:

- We first show that the natural pricing mechanism is efficient: there exists a price process such that all players agree on what is bought or sold at each time step. For this price, the players' selfish responses to the price coincide with a socially optimal policy.

- We then use a mean-field approximation to represent the state of the appliances and a trajectorial forecast to represent the noise process.
- This allows us to use a decentralized algorithm to solve numerically the control problem, by using the alternating direction method of multipliers (ADMM) [BPC+11].

This methodology is generic and scales linearly with the number of players. It can be implemented in a distributed way and be applied to other systems. We think that it is a good candidate to be later implemented in a tool that would provide automatic optimization of such systems.

In [GLBT14], we evaluate the optimal allocation and compare the social gain brought by demand-response with the gain for each player. The mechanism is evaluated using real wind generation data from the UK, obtained from ELEXON. We show that if all players have full information, then, as expected, the social welfare increases with the amount of elastic load available. However, if the day-ahead market cannot observe the state of the appliances, having too many elastic loads can be detrimental and leads to a decrease in the welfare. In that case, having a large player that controls many appliances introduces synchronization between devices, which makes the behavior of the system less predictable. This problem is probably hidden today as the installed capacity of demand-response is small but this is clearly a threat for the future. We also find that demand-response brings the same benefit as a storage system that would have a cycle-efficiency of 100%. Hence, it outperforms current storage technologies that have an efficiency of only 70 to 90%.

6.3 Discussion and Future Work

Our work on smart grids has mainly been focussed on how to solve the energy balance problem in a context where a large share of the electricity is produced by renewables. The key factor that defines our problem is the presence of multiple time-scales. For these scenarios, we developed centralized and distributed algorithms that will serve as building blocks for an implementation in the QUANTICOL tool.

There are still many open questions. The main question concerns network topology, which is a spatial component of the problem. In order to keep our models tractable and to develop generic methodologies, we did not consider the physical constraints of the network. These become important in a situation where the transmission network is unable to balance a local overproduction of renewable sources and an underproduction in a remote location. In this case, the placement of storage becomes crucial. We conjecture that our methodology can be used in a multiple-stage optimization problem: first by using local storage to solve local imbalance, and finally by aggregating these results.

This question is partially addressed in [CTLBP12] by other members of the team at EPFL. They study real-time voltage control in a local distribution network. They have developed an heuristic that uses a low-rate broadcast signal for controlling flexible loads or managing distributed storage. They numerically evaluated the performance of the scheme on the IEEE 34-node test feeder. They have shown that the same signal can be used to efficiently control supercapacitors or flexible loads whilst providing voltage control.

Other extensions of this work constitute the core of the research effort at EPFL on smart grid. The vision is to build a new control philosophy, which enables resources to directly communicate with each other and guarantee explicit real-time set-points for power absorption or injection. The goal is to build a recursive abstraction framework, that would make this method composable: subsystems can be aggregated into abstract models that hide the complexity from the upper layer.

7 Conclusion and Future Work

In this deliverable, we have reported on our early work on multi-scale modeling based on our expertise on smart grids. Even at this early stage of the project, there have been multiple outcomes, in terms of research articles or advances in the case studies. We made advances in the fundamental aspects of multi-scale modeling. We provided a literature review of techniques to reduce the complexity of the analysis of multiple time-scales models. We showed how stochastic hybrid systems can handle multiple population scales. Finally, we presented an abstraction technique to study multiple organizational scales. We also developed a new optimization method that we apply to various control problems in smart grids. This foundational work is a preliminary but necessary step for the implementation of software prototypes to be integrated in the tool chain of Work Package 5.

For the future of the project, we would like to build on the expertise gained in this work to develop new optimization tools. In particular, we will focus on the problem of decision making in the presence of multiple scale and imperfect forecasts. We think that the different areas of work presented in this deliverable, are the breakthrough that will allow us to achieve this goal. Our method, based on stochastic Lagrangian decomposition and trajectorial forecast, will allow people to go beyond heuristics and model predictive control and build more robust control policies.

A challenge will be to provide generic optimization tools based on these results, in order to make them as broadly applicable as possible. Our expertise gained in the decentralized control of smart grid systems will also be used for the other case studies. For example, we aim to study the effect of delays in the redistribution or incentives in bike-sharing systems, and planning of public transportation buses. We will also contribute to the more theoretical aspects of control. One direction would be, for example, to use second-order mean field approximations in our forecast model.

This work is directly related to Work Package 2, that focusses on spatial behavior. For example, hybrid approximation may be helpful when space is considered and mean-field approaches are crucial in patch models. We also aim to extend our method to the model checking problem, in collaboration with Work Package 3. The fluid model checking techniques developed in Work Package 3 are partially based on the construction of suitable hybrid approximations. A better understanding of the methods from this perspective may lead to efficient approximations of other classes of properties. Finally, we plan to use the expertise gained from smart grids to develop optimization algorithms for the other case studies (Work Package 5). Furthermore, hybrid approaches can be used in bus and bike sharing networks or when there is a centralized controller.

References

[AI89] H. Ammar and S. R. Islam. Time scale decomposition of a class of generalized stochastic Petri net models. *IEEE Transactions of Software Engineering*, 15(6):809–820, 1989.

[BB08] M. Benäim and J. L. Boudec. A class of mean field interaction models for computer and communication systems. *Performance Evaluation*, 2008.

[BB11a] B. Barzel and O. Biham. Binomial moment equations for stochastic reaction systems. *Physical Review Letters*, 106(15):150602, 2011.

[BB11b] B. Barzel and O. Biham. Stochastic moment equations - case closed. *Bulletin of the American Physical Society*, 56, 2011.

[BGK12] A. Bejan, R. Gibbens, and F. Kelly. Statistical aspects of storage systems modelling in energy networks. *46th Annual Conference on Information Sciences and Systems*, March 2012.

[BHLM13] L. Bortolussi, J. Hillston, D. Latella, and M. Massink. Continuous approximation of collective systems behaviour: a tutorial. *Performance Evaluation*, 2013.

[BM07] V. Bykov and U. Maas. Extension of the ILDM method to the domain of slow chemistry. *Proceedings of the Combustion Institute*, 31(1):465–472, January 2007.

[BM11] V. Bykov and U. Maas. Hierarchy analysis and reduction of reacting flow systems. In Krause, E and Shokin, Y and Resch, M and Kroner, D and Shokina, N, editor, *Computational science and high performance computing IV*, volume 115 of *Notes on Numerical Fluid Mechanics and Multidisciplinary Design*, pages 233–252. Springer, 2011.

[BO91] I. Babuška and J. Osborn. Eigenvalue problems. In *Handbook of Numerical Analysis*, volume 2, pages 641–787. Elsevier, 1991.

[Bor10] L. Bortolussi. Limit behavior of the hybrid approximation of stochastic process algebras. In *Proceedings of 17th International Conference on Analytical and Stochastic Modeling Techniques and Applications, ASMTA 2010*, volume 6148 of *Lecture Notes in Computer Science*, pages 367–381. Springer, 2010.

[Bor11] L. Bortolussi. Hybrid limits of continuous time Markov chains. In *Proceedings of Eighth International Conference on the Quantitative Evaluation of Systems, QEST 2011*, pages 3–12. IEEE Computer Society, 2011.

[Bor12] L. Bortolussi. Hybrid behaviour of Markov population models. *CoRR*, abs/1211.1643, 2012.

[BP13] L. Bortolussi and A. Policriti. (Hybrid) automata and (stochastic) programs. the hybrid automata lattice of a stochastic program. *Journal of Logic and Computation*, 23(4):761–798, 2013.

[BPC+11] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.

[BT93] A. Blakemore and S. Tripathi. Automated time scale decomposition of SPNs. In *Proceedings of 5th International Workshop on Petri Nets and Performance Modesl (PNPM'93)*, volume 6148, pages 248–257. IEEE, 1993.

[BT13] L. Bortolussi and M. Tribastone. Differential analysis of interacting automata with immediate actions. In *Proceedings of VALUETOOLS 2013*, 2013.

[BTH14] L. Bortolussi, M. Tribastone, and J. Hillston. Fluid performability analysis of nested automata models. 2014. Under review.

[Buc94] P. Buchholz. A class of hierarchical queueing networks and their analysis. *Queueing Systems*, 15:59–80, 1994.

[Buc99] P. Buchholz. An adaptive aggregation/disaggregation algorithm for hierarchical Markovian models. *European Journal of Operational Research*, 116(3):545–564, 1999.

[Buj12] L. M. Bujorianu. *Stochastic Reachability Analysis of Hybrid Systems.* Number 61 in Communications and Control Engineering. Springer Verlag London, 2012.

[CDMR12] A. Crudu, A. Debussche, A. Muller, and O. Radulescu. Convergence of stochastic gene networks to hybrid piecewise deterministic processes. *The Annals of Applied Probability*, 22(5):1822–1859, 2012.

[CGP05a] Y. Cao, D. Gillespie, and L. Petzold. Multiscale stochastic simulation algorithm with stochastic partial equilibrium assumption for chemically reacting systems. *Journal of Computational Physics*, 206(2):395–411, July 2005.

[CGP05b] Y. Cao, D. T. Gillespie, and L. R. Petzold. The slow-scale stochastic simulation algorithm. *The Journal of Chemical Physics*, 122(1):014116, 2005.

[CK11] E. Chiavazzo and I. Karlin. Adaptive simplification of complex multiscale systems. *Physical Review E*, 83(3), March 2011.

[Cou77] P. Courtois. *Decomposability: Queueing and Computer Systems Applications.* ACM series. Academic Press, 1977.

[CTLBP12] K. Christakou, D.-C. Tomozei, J.-Y. Le Boudec, and M. Paolone. GECN: Primary voltage control for active distribution networks via real-time demand-response. *IEEE Transaction on Smart Grid*, 2012.

[Dav93] M. H. A. Davis. *Markov Models and Optimization.* Chapman & Hall, 1993.

[Dav06a] M. J. Davis. Low-dimensional manifolds in reaction–diffusion equations. 1. Fundamental aspects. *The Journal of Physical Chemistry A*, 110(16):5235–5256, April 2006.

[Dav06b] M. J. Davis. Low-dimensional manifolds in reaction–diffusion equations. 2. Numerical analysis and method development. *The Journal of Physical Chemistry A*, 110(16):5257–5272, April 2006.

[DN08] R. Darling and J. R. Norris. Differential equation approximations for Markov chains. *Probability surveys*, 5:37–79, 2008.

[Fra88] S. J. Fraser. The steady state and equilibrium approximations: A geometrical picture. *The Journal of chemical physics*, 88:4732, 1988.

[Gal14] V. Galpin. Hybrid semantics for Bio-PEPA. *Information and Computation*, 2014.

[Gar85] C. W. Gardiner. *Handbook of stochastic methods*, volume 3. Springer Berlin, 1985.

[GG12] N. Gast and B. Gaujal. Markov chains with discontinuous drifts have differential inclusion limits. *Performance Evaluation*, 69(12):623–642, 2012.

[GLBPT13] N. Gast, J.-Y. Le Boudec, A. Proutière, and D.-C. Tomozei. Impact of storage on the efficiency and prices in real-time electricity markets. In *Proceedings of the fourth international conference on Future energy systems*, pages 15–26. ACM, 2013.

[GLBT14] N. Gast, J.-Y. Le Boudec, and D.-C. Tomozei. Impact of demand-response on the efficiency and prices in real-time electricity markets. *submitted for publication*, 2014.

[GN06] D. A. Goussis and H. N. Najm. Model reduction and physical understanding of slowly oscillating processes: The circadian cycle. *Multiscale Modeling & Simulation*, 5(4):1297–1332, January 2006.

[Gou12] D. A. Goussis. Quasi steady state and partial equilibrium approximations: their relation and their validity. *Combustion Theory and Modelling*, 16(5):869–926, October 2012.

[Gou13] D. A. Goussis. The role of slow system dynamics in predicting the degeneracy of slow invariant manifolds: The case of vdP relaxation-oscillations. *Physica D: Nonlinear Phenomena*, 248:16–32, April 2013.

[GRZ10] A. Gorban, O. Radulescu, and A. Zinovyev. Asymptotology of chemical reaction networks. *Chemical Engineering Science*, 65(7):2310–2324, April 2010.

[GTLB12] N. Gast, D. Tomozei, and J.-Y. Le Boudec. Optimal storage policies with wind forecast uncertainties. *ACM SIGMETRICS Performance Evaluation Review*, 40(3):28–32, 2012.

[GTLB14] N. Gast, D.-C. Tomozei, and J.-Y. Le Boudec. Optimal generation and storage scheduling in the presence of renewable forecast uncertainties. *accepted in Transaction on Smart Grid*, 2014.

[GVL12] G. H. Golub and C. F. Van Loan. *Matrix computations*, volume 3. JHU Press, 2012.

[Haa02] P. J. Haas. *Stochastic Petri nets modelling, stability, simulation*. Springer, New York, 2002.

[Har87] D. Harel. Statecharts: A visual formalism for complex systems. *Science of computer programming*, 8(3):231–274, 1987.

[Hen96] T. Henzinger. The theory of hybrid automata. pages 278–292. IEEE Comput. Soc. Press, 1996.

[Her02] H. Hermanns. *Interactive Markov chains: and the quest for quantified quality*. Springer-Verlag, 2002.

[HM95] J. Hillston and V. Mertsiotakis. A simple time scale decomposition technique for stochastic process algebras. *The Computer Journal*, 38(6), 1995.

[HR02] E. L. Haseltine and J. B. Rawlings. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *The Journal of Chemical Physics*, 117(15):6959, 2002.

[HR05] E. L. Haseltine and J. B. Rawlings. On the origins of approximations for stochastic chemical kinetics. *The Journal of chemical physics*, 123:164115, 2005.

[JG11] K. A. Johnson and R. S. Goody. The original Michaelis constant: Translation of the 1913 Michaelis–Menten paper. *Biochemistry*, 50(39):8264–8269, October 2011.

[Kat95] T. Kato. *Perturbation theory for linear operators*, volume 132. springer, 1995.

[KMT09] D. S. Kim, F. Machida, and K. S. Trivedi. Availability modeling and analysis of a virtualized system. In *PRDC*, pages 365–371, 2009.

[KSR+11] E. Krause, Y. Shokin, M. Resch, D. Kroner, and N. Shokina, editors. *The 4th Russian-German Advanced Research Workshop*. Springer, Berlin; Heidelberg, 2011.

[Lam93] S. H. Lam. Using CSP to understand complex chemical kinetics. *Combustion Science and Technology*, 89(5-6):375–404, 1993.

[LG89] S. H. Lam and D. A. Goussis. Understanding complex chemical kinetics with computational singular perturbation. In *Symposium (International) on Combustion*, volume 22, pages 931–941, 1989.

[LG91] S. H. Lam and D. A. Goussis. Conventional asymptotics and computational singular perturbation for simplified kinetics modeling. In M. D. Smooke, editor, *Reduced Kinetic Mechanisms and Asymptotic Approximations for Methane-Air Flames*, number 384 in Lecture Notes in Physics, pages 227—242. Springer-Verlag, 1991.

[LG94] S. H. Lam and D. A. Goussis. The CSP method for simplifying kinetics. *International Journal of Chemical Kinetics*, 26(4):461–486, 1994.

[LOC+12] A. V. Lebedev, M. V. Okun, V. A. Chorkov, P. M. Tokar, and M. Strelkova. Systematic procedure for reduction of kinetic mechanisms of complex chemical processes and its software implementation. *Journal of Mathematical Chemistry*, 51(1):73–107, August 2012.

[LYT03] M. Lanus, L. Yin, and K. S. Trivedi. Hierarchical composition and aggregation of state-based availability and performability models. *IEEE Transactions on Reliability*, 52(1):44–52, 2003.

[Mer98] V. Mertsiotakis. *Approximate Analysis Methods for Stochastic Process Algebras*. PhD thesis, IMMD V, Universität Erlangen-Nürnberg, 1998.

[MLSH11] S. Menz, J. C. Latorre, C. Schütte, and W. Huisinga. Hybrid stochastic–deterministic solution of the chemical master equation. *SIAM Interdisciplinary Journal Multiscale Modeling and Simulation (MMS)*, 2011.

[MP92a] U. Maas and S. B. Pope. Implementation of simplified chemical kinetics based on intrinsic low-dimensional manifolds. In *Symposium (International) on Combustion*, volume 24, pages 103–112, 1992.

[MP92b] U. Maas and S. B. Pope. Simplifying chemical kinetics: intrinsic low-dimensional manifolds in composition space. *Combustion and Flame*, 88(3):239–264, 1992.

[NF89] A. H. Nguyen and S. J. Fraser. Geometrical picture of reaction in enzyme kinetics. *The Journal of Chemical Physics*, 91(1):186—193, 1989.

[NF13a] P. Nicolini and D. Frezzato. Features in chemical kinetics. I. Signatures of self-emerging dimensional reduction from a general format of the evolution law. *The Journal of Chemical Physics*, 138(23):234101, 2013.

[NF13b] P. Nicolini and D. Frezzato. Features in chemical kinetics. II. A self-emerging definition of slow manifolds. *The Journal of Chemical Physics*, 138(23):234102, 2013.

[Pah09] J. Pahle. Biochemical simulations: stochastic, approximate stochastic and hybrid approaches. *Brief Bioinform.*, 10(1):53–64, 2009.

[PB14] R. Paskauskas and L. Bortolussi. Multi-scale reductions of mean-field and stochastic models tutorial and a survey of the literature. Technical Report TR-QC-4-2014, QUANTICOL, 2014.

[Pou14] A. Pourranjbar. *Modelling and Analysis of Large Scale Systems in PEPA*. PhD thesis, School of Informatics, University of Edinburgh, 2014. In preparation.

[RA03] C. V. Rao and A. P. Arkin. Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. *Journal of Chemical Physics*, 118(11):4999–5010, 2003.

[RF90] M. R. Roussel and S. J. Fraser. Geometry of the steady-state approximation: Perturbation and accelerated convergence methods. *The Journal of chemical physics*, 93(2):1072–1081, 1990.

[ROA13] G. L. Re, M. Ortolani, and G. Anastasi. Ambient intelligence for energy efficiency in a building complex, 2013. Retrieved from `http://ercim-news.ercim.eu/en92/special/ambient-intelligence-for-energy-efficiency-in-a-building-complex`.

[SA61] H. A. Simon and A. Ando. Aggregation of variables in dynamic systems. *Econometrica*, 29(2):111–138, 1961.

[SS89] L. A. Segel and M. Slemrod. The quasi-steady-state assumption: a case study in perturbation. *SIAM review*, 31(3):446–477, 1989.

[Ste90] G. W. Stewart. *Matrix perturbation theory*. Computer science and scientific computing. Academic Press, Boston, 1990.

[Ste01] G. W. Stewart. *Matrix algorithms*. SIAM, Philadelphia, 2001.

[TT12] M. Tschaikowski and M. Tribastone. Exact fluid lumpability for Markovian process algebra. In *CONCUR*, pages 380–394, 2012.

[Ver05] F. Verhulst. *Methods and applications of singular perturbations: boundary layers and multiple timescale dynamics*. Number 50 in Texts in applied mathematics. Springer, New York, 2005.

[WFLP12] S. Wu, J. Fu, H. Li, and L. Petzold. Automatic identification of model reductions for discrete stochastic simulation. *The Journal of Chemical Physics*, 137(3):034106, 2012.

[Whi80] W. Whitt. Continuity of generalized semi-Markov processes. *Mathematics of Operations Research*, 5(4):pp. 494–501, 1980.

[YPC13] Y. Yang, S. B. Pope, and J. H. Chen. Empirical low-dimensional manifolds in composition space. *Combustion and Flame*, 160(10):1967–1980, October 2013.

[ZAI13] S. Zhang, I. P. Androulakis, and M. G. Ierapetritou. A hybrid kinetic mechanism reduction scheme based on the on-the-fly reduction and quasi-steady-state approximation. *Chemical Engineering Science*, 93:150–162, April 2013.

[ZKK04a] A. Zagaris, H. G. Kaper, and T. J. Kaper. Analysis of the computational singular perturbation reduction method for chemical kinetics. *Journal of Nonlinear Science*, 14(1):59–91, 2004.

[ZKK04b] A. Zagaris, H. G. Kaper, and T. J. Kaper. Fast and slow dynamics for the computational singular perturbation method. *Multiscale Modeling & Simulation*, 2(4):613–638, 2004.