The Benefits of Sometimes Not Being Discrete

Jane Hillston

LFCS, School of Informatics, University of Edinburgh, UK
jane.hillston@ed.ac.uk
http://www.quanticol.eu

Abstract. Discrete representations of systems are usual in theoretical computer science and they have many benefits. Unfortunately they also suffer from the problem of state space explosion, sometimes termed the *curse of dimensionality*. In recent years, research has shown that there are cases in which we can reap the benefits of discrete representation during system description but then gain from more efficient analysis by approximating the discrete system by a continuous one. This paper will motivate this approach, explaining the theoretical foundations and their practical benefits.

1 Introduction

Over the last twenty to thirty years, areas of quantitative modelling and analysis, such as performance, dependability and reliability modelling have embraced formal models [37]. This trend has been motivated by the increasing concurrency of the systems under consideration and the difficulties of constructing the underlying mathematical models, which are used for analysis, by hand. In particular concurrent modelling formalisms such as stochastic Petri nets and stochastic process algebras have been widely adopted as high-level modelling languages for generating underlying Markovian models. Moreover, there has been much work exploring how the properties of the high-level languages can be exploited to assist in the analysis of the underlying model through a variety of techniques (e.g. decomposition [23, 39], aggregation based on bisimulation [38], etc).

However, a combination of improved model construction techniques, and the increasing scale and complexity of the systems being developed, has led to ever larger models; and these models now frequently defy analysis even after model reduction techniques such as those mentioned above. The problem is the well-known curse of dimensionality: the state space of a discrete event system can grow exponentially with the number of components in the system.

Fortunately, over the last decade a new approach has emerged which offers a way to avoid this state space explosion problem, at least for one class of models. When the system under consideration can be presented as a *population* model and the populations involved are known to be *large*, then a good approximation of the discrete behaviour can be achieved through a continuous or fluid approximation. Moreover, this model is scale-free in the sense that the computational effort to solve it remains the same even as the populations involved grow larger. Of course, there is a cost, in the sense that some information is lost and it is no longer possible to analyse the system in terms of

P. Baldan and D. Gorla (Eds.): CONCUR 2014, LNCS 8704, pp. 7–22, 2014.

individual behaviours. But when average behaviours or expectations are required, for example in situations of collective behaviour, the fluid approach has substantial benefits.

The rest of this paper is organised as follows. Section 2 gives an intuitive explanation of how the fluid approximation approach has been widely used in biological modelling for many years, before presenting the mathematical foundations for the approach as provided by Kurtz's Theorem in Section 3. The attraction of combining the technique with the compositional models generated by process algebras is explained in Section 4, with discussion of how the mapping has been developed for a variety of process algebras. In Section 5 we give an overview of extending these results into the model checking arena, and in Section 6 we briefly summarise and conclude.

2 Biologists Just Do It!

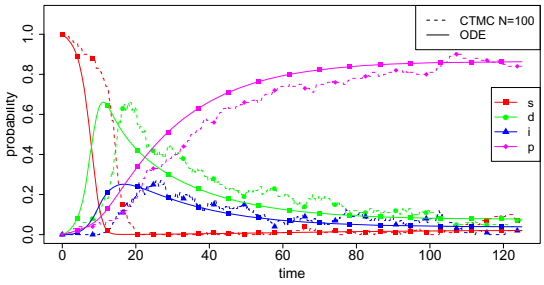
In several disciplines fluid approximations have long been used, often without concern for formal foundations. The most noticeable example of this is in biological modelling of intra-cellular processes. These processes result from the collisions of molecules within the cell, an inherently discrete process. Yet, the most common form of mathematical model for these processes is a system of ordinary differential equations (ODEs) which captures the collective behaviour in terms of concentrations of different molecular states, rather than the states of individual molecules. At heart, this is a fluid approximation, as highlighted by Kurtz [46] and Gillespie [32]. But it has been so widely adopted that many biologists no longer recognise that there is a fundamental shift in representation taking place.

That there was an implicit transformation taking place during model construction became more obvious when formal representations started to be used to describe intracellular biological processes [58]. In the early 2000s researchers recognised that the intracellular processes were highly concurrent systems, amenable to description formalisms used to describe concurrency in computer systems. This led to a plethora of adopted and developed process algebras for describing cellular processes e.g. [18, 56, 57, 24]. Whilst most focussed on the discrete representation and subsequent discrete event simulation of an underlying continuous time Markov chain (CTMC) using Gillespie's algorithm [32], work such as [17, 20] established that it was also possible to derive the systems of ODEs more familiar to biologists from process algebra descriptions.

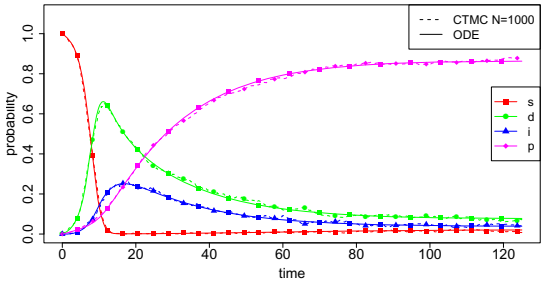
3 Kurtz's Theorem

At the foundations of fluid approximation is a fundamental result by Kurtz, dating back to the 1970s [45], which establishes that a sequence of CTMCs which satisfy some conditions and represent essentially the same system under growing populations, converges to a set of ODEs. At convergence the behaviour of the CTMC is indistinguishable from the behaviour of the set of ODEs. However, this theoretical limit is at an infinite population. Nevertheless in many practical cases we find empirically that sufficient convergence is often achieved at much lower populations, as illustrated in Fig. 1.

In order to explain this result in more detail we introduce a simple representation of Markov models of populations of interacting agents. Such models may be readily



(a) $N = 100$



(b) $N = 1000$

Fig. 1. Comparison between the limit fluid ODE and a single stochastic trajectory of a network epidemic example, for total populations $N = 100$ and $N = 1000$. This demonstrates how the accuracy of the approximation of behaviour captured by the fluid ODE improves as the population size grows.

derived from stochastic process algebras such as PEPA or EMPA [38, 7]. We consider the case of models of processes evolving in continuous time, although a similar theory can be considered for discrete-time models (see, for instance, [13]). In principle, we can have different classes of agents, and many agents for each class in the system. To keep notation simple, we assume here that the number of agents is constant and equal to N (making a closed world assumption) but analogous results can be derived for systems which include the birth and death of agents.

In particular, let us assume that each agent is a finite state machine, with internal states taken from a finite set S, and labelled by integers: $S = \{1, 2, \ldots, n\}$. We have a population of N agents, and denote the state of agent i at time t, for $i = 1, \ldots, N$, by $Y_i^{(N)}(t) \in S$. Note that we have made explicit the dependence on N, the total population size.

A configuration of a system is thus represented by the tuple $(Y_1^{(N)}, \ldots, Y_N^{(N)})$. This representation is based on treating each agent as a distinct individual with identity conferred by the position in the vector. However, when dealing with population models, it is customary to assume that single agents in the same internal state cannot be distinguished, hence we can move from the *individual representation* to the *collective*

representation by introducing n variables counting how many agents are in each state. This is sometimes termed a *counting abstraction*. Hence, we define

$$X_j^{(N)} = \sum_{i=1}^{N} \mathbf{1}\{Y_i^{(N)} = j\}, \tag{1}$$

where $\mathbf{1}\{Y_i^{(N)} = j\}$ is an indicator function with value 1 when $Y_i^{(N)} = j$ and zero, otherwise. Note that the vector $\mathbf{X}^{(N)} = (X_1^{(N)}, \ldots, X_n^{(N)})$ has a dimension independent of N; it is referred to as the collective, population, or counting vector. The domain of each variable $X_j^{(N)}$ is $\{0, \ldots, N\}$, and, by the closed world assumption, it holds that $\sum_{j=1}^{n} X_j^{(N)} = N$. Let us denote with $\mathscr{S}^{(N)}$ the subset of vectors of $\{1, \ldots, N\}^n$ that satisfy this constraint.

The dynamics of the population models is expressed in terms of a set of possible *events* or *transitions*. Events are stochastic, and take an exponentially distributed time to happen. Moreover their rate may depend on the current global state of the system. Hence, each event will be specified by a rate function, and by a set of update rules, specifying the impact of the event on the population vector.

In this model, the set of events, or transitions, $\mathscr{T}^{(N)}$, is made up of elements $\tau \in \mathscr{T}^{(N)}$, which are pairs $\tau = (\mathbf{v}_\tau, r_\tau^{(N)})$. Here $\mathbf{v}_\tau$ is the *update vector*; specifically $\mathbf{v}_{\tau,i}$ records the impact of event τ on the ith entry (ith population) in the population vector. The rate function, $r_\tau^{(N)} : \mathscr{S}^{(N)} \to \mathbb{R}_{\geq 0}$, depends on the current state of the system, and specifies the speed of the corresponding transition. It is assumed to be equal to zero if there are not enough agents available to perform a τ transition, and it is required to be *Lipschitz continuous* (when interpreted as a function on real numbers).

Thus we define a population model $\mathscr{X}^{(N)} = (\mathbf{X}^{(N)}, \mathscr{T}^{(N)}, \mathbf{x}_0^{(N)})$, where $\mathbf{x}_0^{(N)}$ is the initial state. Given such a model, it is straightforward to construct the CTMC $\mathbf{X}^{(N)}(t)$ associated with it; its state space is $\mathscr{S}^{(N)}$, while its infinitesimal generator matrix $Q^{(N)}$ is the $|\mathscr{S}^{(N)}| \times |\mathscr{S}^{(N)}|$ matrix defined by

$$q_{\mathbf{x},\mathbf{x}'} = \sum \{ r_\tau(\mathbf{x}) \mid \tau \in \mathscr{T},\ \mathbf{x}' = \mathbf{x} + \mathbf{v}_\tau \}.$$

As explained above, fluid approximation approximates a CTMC by a set of ODEs. These differential equations can be interpreted in two different ways: they can be seen as an approximation of the average of the system (usually a first order approximation, see [9, 68]). This is often termed a *mean field* approximation. Alternatively, they can be interpreted as an approximate description of system trajectories for large populations. We will focus on this second interpretation, which corresponds to a functional version of the law of large numbers. In this interpretation, instead of having a sequence of random variables, like the sample mean, converging to a deterministic value, like the true mean, in this case we have a sequence of CTMCs (which can be seen as random trajectories in $\mathbb{R}^n$) for increasing population size, which converge to a deterministic trajectory, the solution of the fluid ODE.

In order to consider the convergence, we must formally define the sequence of CTMCs to be considered. To allow models of different population sizes to be compared we normalise the populations by dividing each variable by the total population N. In this way,

the normalised population variables $\hat{\mathbf{X}}^{(N)} = \frac{\mathbf{X}^{(N)}}{N}$, or population densities, will always range between 0 and 1 (for the closed world models we consider here), and so the behaviour for different population sizes can be compared. In the case of a constant population, normalised variables are usually referred to as the *occupancy measures*, as they represent the fraction of agents which occupy each state.

After normalisation we must appropriately scale the update vectors, initial conditions, and rate functions [13]. Let $\mathscr{X}^{(N)} = (\mathbf{X}^{(N)}, \mathscr{T}^{(N)}, \mathbf{X_0}^{(N)})$ be the non-normalised model with total population N and $\hat{\mathscr{X}}^{(N)} = (\hat{\mathbf{X}}^{(N)}, \hat{\mathscr{T}}^{(N)}, \hat{\mathbf{X}}_{\mathbf{0}}^{(N)})$ the corresponding normalised model. We require that:

- initial conditions scale appropriately: $\hat{\mathbf{X}}_{\mathbf{0}}^{(N)} = \frac{\mathbf{X_0}^{(N)}}{N}$;
- for each transition $(\mathbf{v}_\tau, r_\tau^{(N)}(\mathbf{X}))$ of the non-normalised model, define $\hat{r}_\tau^{(N)}(\hat{\mathbf{X}})$ to be the rate function expressed in the normalised variables (obtained from $r_\tau^{(N)}$ by a change of variables). The corresponding transition in the normalised model is $(\mathbf{v}_\tau, \hat{r}_\tau^{(N)}(\hat{\mathbf{X}}))$, with update vector equal to $\frac{1}{N}\mathbf{v}_\tau$.

We further assume, for each transition τ, that there exists a bounded and Lipschitz continuous function $f_\tau(\hat{\mathbf{X}}) : E \to \mathbb{R}^n$ on normalised variables (where E contains all domains of all $\hat{\mathscr{X}}^{(N)}$), independent of N, such that $\frac{1}{N}\hat{r}_\tau^{(N)}(\mathbf{x}) \to f_\tau(\mathbf{x})$ *uniformly* on E. We denote the state of the CTMC of the N-th non-normalised (resp. normalised) model at time t as $\mathbf{X}^{(N)}(t)$ (resp. $\hat{\mathbf{X}}^{(N)}(t)$).

3.1 Deterministic Limit Theorem

In order to present the "classic" deterministic limit theorem, consider a sequence of normalised models $\hat{\mathscr{X}}^{(N)}$ and let $\mathbf{v}_\tau$ be the (non-normalised) update vectors. The *drift* $F^{(N)}(\hat{\mathbf{X}})$ of $\hat{\mathscr{X}}$, which is formally the mean instantaneous increment of model variables in state $\hat{\mathbf{X}}$, is defined as

$$F^{(N)}(\hat{\mathbf{X}}) = \sum_{\tau \in \hat{\mathscr{T}}} \frac{1}{N} \mathbf{v}_\tau \hat{r}_\tau^{(N)}(\hat{\mathbf{X}}) \tag{2}$$

Furthermore, let $f_\tau : E \to \mathbb{R}^n$, $\tau \in \hat{\mathscr{T}}$ be the limit rate functions of transitions of $\hat{\mathscr{X}}^{(N)}$. The *limit drift* of the model $\hat{\mathscr{X}}^{(N)}$ is therefore

$$F(\hat{\mathbf{X}}) = \sum_{\tau \in \hat{\mathscr{T}}} \mathbf{v}_\tau f_\tau(\hat{\mathbf{X}}), \tag{3}$$

and $F^{(N)}(\mathbf{x}) \to F(\mathbf{x})$ uniformly as $N \longrightarrow \infty$, as easily checked. The fluid ODE is

$$\frac{d\mathbf{x}}{dt} = F(\mathbf{x}), \quad with\, \mathbf{x}(0) = \mathbf{x_0} \in S.$$

Given that F is Lipschitz in E (since all f_τ are), this ODE has a unique solution $\mathbf{x}(t)$ in E starting from $\mathbf{x_0}$. Then, one can prove the following theorem:

Theorem 1 (Deterministic approximation [45, 27]). *Let the sequence $\hat{\mathbf{X}}^{(N)}(t)$ of Markov processes and $\mathbf{x}(t)$ be defined as above, and assume that there is some point*

$\mathbf{x_0} \in S$ *such that* $\hat{\mathbf{X}}^{(N)}(0) \to \mathbf{x_0}$ *in probability. Then, for any* finite *time horizon* $T < \infty$*, it holds that as* $N \longrightarrow \infty$*:*

$$\mathbb{P}\left\{ \sup_{0 \le t \le T} ||\hat{\mathbf{X}}^{(N)}(t) - \mathbf{x}(t)|| > \varepsilon \right\} \to 0.$$

Notice that the Theorem makes assertions about the trajectories of the population counts at all finite times, but nothing about what happens at steady state, i.e. when time goes to infinity.

3.2 Fast Simulation

Based on the Deterministic Approximation Theorem, we can consider the implications for a single individual in the population when the population size goes to infinity. Even as the collective behaviour tends to a deterministic process, each individual agent will still behave randomly. However, the Deterministic Approximation Theorem implies that the dynamics of a single agent, in the limit, becomes independent of other agents, and it will sense them only through the collective system state, or *mean field*, described by the fluid limit. This asymptotic decoupling allows us to find a simple, *time-inhomogenous*, Markov chain for the evolution of the single agent, a result often known as *fast simulation* [28, 30].

To see this decoupling we focus on a single individual $Y_h^{(N)}(t)$, which is a (Markov) process on the state space $S = \{1, \ldots, n\}$, conditional on the global state of the population $\hat{\mathbf{X}}^{(N)}(t)$. The evolution of this agent can be obtained by computing the rates q_{ij} at which its state changes from i to j, by projecting on a single agent the rate of global transitions that induce a change of state of at least one agent from i to j. Such a rate $q_{ij}(\hat{\mathbf{X}})$ still depends on the global system state, hence to track the evolution of agent $Y_h^{(N)}(t)$ we still need to know the global state of the system $\hat{\mathbf{X}}^{(N)}(t)$: e.g. solving any model checking problem on $Y_h^{(N)}(t)$ would requires us to work with the full Markov model $\hat{\mathbf{X}}^{(N)}(t)$.

However, as the size of the system increases, the deterministic limit theorem tells us the stochastic fluctuations of $\hat{\mathbf{X}}^{(N)}(t)$ tend to vanish, and this effect propagates to the stochastic behaviour of $Y_h^{(N)}(t)$, which can be approximated by making it dependent only on the fluid limit $\mathbf{x}(t)$. More precisely, we need to construct the time-inhomogeneous CTMC $z(t)$ with state space S and rates $q_{ij}(\mathbf{x}(t))$, computed along the fluid trajectory. The following theorem [28] guarantees that $z(t)$ is a good approximation of $Y_h^{(N)}(t)$:

Theorem 2 (Fast simulation theorem). *For any finite time horizon* $T < \infty$*,* $\mathbb{P}\{Y_h^{(N)}(t) \neq z(t)$*, for some* $t \le T\} \to 0$*, as* $N \to \infty$*.*

This theorem states that, in the limit of an infinite population, each agent will behave independently from all the others, sensing only the mean state of the global system, described by the fluid limit $\mathbf{x}(t)$. This *asymptotic decoupling* of the system, which can be generalised to any subset of $k \ge 1$ agents, is also known in the literature under the name of *propagation of chaos* [5].

Remark 1. For simplicity here we have considered a single class of agents without births or deaths. Nevertheless the same results hold for a model consisting of multiple classes of agents. In this case we construct a single agent class but partition the state space S into subsets, each of which represents the states of a distinct agent, and such that there are no transitions between subsets. The agents whose initial state is in each subset corresponds to agents of that class. Furthermore, events that capture birth and death can easily be included by allowing update vectors which are unbalanced in the sense that the total positive update is greater than or less than the total negative update. Such *open* systems can be handled in the same theory, see [12] for further details, but for clarity we will restrict to closed world models in this paper.

4 Stochastic Process Algebra with Fluid Interpretation

Kurtz's Theorem, or the Deterministic Approximation Theorem, has been established for many years. It has been widely used but when it is used directly from a CTMC model, it is the modeller's responsibility to prove that the model satisfies the necessary conditions for application of the theory, and moreover, to derive the corresponding ODEs. This must be done on a model-by-model basis. In recent years, the approach has been used for several performance and dependability models e.g. [3–5, 30].

This situation made it attractive to incorporate mean field or fluid approximation into the formal high-level language approaches which have developed over the last two decades for constructing CTMC models for quantitative analysis. From the perspective of the formal modelling community, this gives access to a scalable analysis technique which is immune to the problem of state space explosion; indeed, a technique which increases in accuracy as the size of the model grows. From the perspective of modellers already familiar with the mean field approach, it offers the possibility to establish the conditions for convergence at the language level via the semantics, once and for all, removing the need to fulfil the proof obligation on a model-by-model basis. Moreover the derivation of the ODEs can be automated in the implementation of the language.

Work has developed in both stochastic Petri nets, e.g. [66, 60, 61] and stochastic process algebras, e.g. [43, 40, 16]. Here we focus on the work in the process algebra context as it is more readily related to the agent-based CTMC model presented in the previous section. It is straightforward to see that components or agents within the process algebra description can be regarded as agents within the CTMC model, typically occupying different partitions within the notional complete state space for agents, as explained at the end of Section 3. When multiple instances of a component are present in the same context within the model, these constitute a *population*. In terms of the language the dynamic combinators are associated with the description of the behaviour of individual agents, essentially finite state machines, whereas static combinators, principally parallel composition, specify the structure of the system, which is now interpreted as the formation and interaction of populations.

The fluid approximation approach is only applicable to models where we have interactions of large populations (parallel compositions of large numbers of components with the same behaviour) within which each component has relatively simple behaviour rather than interactions between individuals each with complex behaviour. When this is

the case we need to make the shift from a state representation based on individuals, to one based on counting (analogous to the shift represented by equation (1)). How this is handled depends on the process algebra but is generally straightforward. For example, in PEPA models there is a simple procedure to reduce the syntactic representation to a state vector [40, 65], but in languages such as Bio-PEPA the mapping is more straightforward because the language was designed to support fluid approximation [24]. The actions of the algebra correspond to the events in the CTMC model, and the definition of the process and its continuation via an action is the basis for the definition of the update vector.

The first work relating process algebra and mean field models can be found in the thesis of Sumpter [62]. Sumpter developed models of social insects in the discrete synchronous process algebra WSCCS [63]. He then heuristically derived difference equations to capture the mean field representation of the model. This work inspired the work of Norman and Shankland [54], in which WSCCS is used to build models of the spread of infectious diseases and difference equation representations are derived. This led on to further work with ever more rigour introduced into the relationship between the difference equation/ODE models and the process algebra descriptions from which they were derived [52, 53, 51], but in later work the authors switched from using WSCCS to using PEPA and Bio-PEPA for their modelling of epidemics.

As previously mentioned, work in systems biology stimulated more widespread interest in the relationship between process algebra description and ODE models. The first work here was the mapping given from PEPA models constructed in a particular style, representing a reagent-centric view of biological signal transductions pathways, to equivalent ODE models, by Calder *et al.* [17]. This was subsequently generalised to more arbitrary PEPA models with large populations, where the mapping to the ODE was made completely systematic, based on an intermediate structure termed the *activity matrix* [40]. In the work of Bortolussi and Policriti the authors consider a different style of process algebra, stochastic Concurrent Constraint Programming (sCCP), and demonstrate a mapping, both from process algebra to ODEs and from ODEs to process algebra descriptions [16]. At around the same time Cardelli also constructed a systematic mapping from process algebra (in this case a variant of CCS) to ODEs, using a Chemical Parametric Form as an intermediary in this case [20]. The relationship between this interpretation of the process algebra model and the discrete-state stochastic semantics is explored in [19].

After these initial explorations of the possibilities to relate the inherently discrete representation of a process algebra model with a fluid approximation of the underlying Markov process, there came a sequence of papers establishing the mapping on a firmer foundation and considering the convergence properties which can be inferred from Kurtz's Theorem. For example in [31], Geisweiller *et al.*, working with a generalised form of PEPA models which allow two forms of synchronisation — both the usual PEPA synchronisation based on the bounded capacity, and the biological notion of mass action — show that the syntactically derived ODE models are indeed those which are obtained by the application of Kurtz's Theoreom, guaranteeing convergence in the limit. In [65], Tribastone *et al.* show how it is possible to fully formalise the derivation of the ODEs for PEPA models, via a structured operational semantics. In [16] Bortolussi

and Policriti construct a process algebra that matches a given set of ODEs in the limit. An alternative approach to the derivation of the fluid approximation model is taken in the work on Kappa [26], where the ODEs are derived as an abstract interpretation.

Some authors also considered how to make the derivation of ODEs from process algebra descriptions easier. As previously mentioned, the PEPA variant, Bio-PEPA [24] was explicitly constructed to maintain a counting abstraction, initially making the derivation of the activity matrix easier and later supporting a semantics in the style of [65]. Hayden and Bradley developed another variant of PEPA, termed Grouped PEPA, which makes clearer the population structures within models [34].

The system ODEs derived from a process algebra model are generally not amenable to algebraic solution, but instead are analysed by numerical simulation. This solution generates a trajectory, tracking the population counts of each local state over time, which can be interpreted as the expected population value over time. Such expected population counts are rarely the objective of quantitative modelling in computer science, although they are often the focus in biological systems. In computer systems derived measures such as throughput, response times, or first passage times are of more interest. In [64], Tribastone *et al.* establish when performance measures such as throughput and response time may legitimately be derived from a fluid approximation. Hayden *et al.* develop an approach to derive the more sophisticated first passage time distributions [36]. When the "passage" of interest relates to an individual component within the model the approach taken relies on the use of the fast simulation result. In further work [35], Hayden *et al.* show how response-time measures specified by stochastic probes can be readily calculated via the mean field approach.

5 Fluid Model Checking

Stochastic process algebra models have long been also analysed using quantitative model checking. In the case of stochastic model checking, there are some consolidated approaches, principally based on checking Continuous Stochastic Logic (CSL) formulae [2, 1, 59], and these are supported by software tools which are in widespread use such as PRISM [47, 48] and MRMC [41]. However these methods often depend on an explicit representation of the state space and consequently suffer from the state space explosion problem, which limits their applicability, particularly for population models. Even when statistical model checking is used, and the state space is only built on-the-fly, the size of population models may make adequate statistical sampling costly or even unattainable.

Thus it is natural to ask the question, to what extent can the fluid approximation techniques presented earlier in this paper be exploited to mitigate the problem of quantitative model checking of population CTMC-based models. The first work in this direction was presented in [11, 12], in which fluid approximation is used to carry out approximate model checking of behaviours of individual agents in large population models, specified as CSL formulae. This work builds on the Fast Simulation Theorem [30, 28], which characterises the limit behaviour of a single agent in terms of the solution of the fluid equation. Recall that the Fast Simulation Theorem states that a single agent senses the rest of the population only through its "average" evolution, as given by the

fluid equation. Thus if the modeller wishes to verify a property of an individual agent within a population of many interacting agents (possibly with a small set of different capabilities) the approach is to check the property in a limit model which consists of the discrete representation of the individual agent taking into account the average evolution of the rest of the system. In practice, for CTMC models, the discrete representation of the individual agent is a *time-inhomogeneous* CTMC (ICTMC), where the rates of transitions between states are determined by the fluid approximation of the rest of the system. Model checking of ICTMCs is far more complex than the homogeneous-time case, but this is compensated because only the local states of one agent need to be considered, so the state space is typically small. The authors termed this approach Fluid Model Checking. Preliminary ideas on using fluid approximation in continuous time for model checking population models, and in particular for an extension of the logic CSL, were informally sketched in [43], but no model checking algorithms were presented. Subsequently the work was more fully developed in [44], which relies substantially on [11].

In the Fluid Model Checking approach the technicalities come from the time-inhomogeneous nature of the process being checked. As in the CTMC case, model checking CSL formulas of ICTMC can be expressed in terms of reachability calculations on an ICTMC, typically with modified structure that makes some states absorbing. However, these calculations are more complex as rates are not constant, but changing over time as the state of the whole system evolves and influences the considered agent. This introduces discontinuities in the satisfaction probabilities as, for example, states in the ICTMC may change from being in the goal set to not, as time progresses. Thus the solution of the Kolmogorov equations to calculate the reachability must be conducted in a piecewise manner, between the time points at which the sets of goal states and unsafe states change over time. Convergence and quasi-decidability results are presented that guarantee the asymptotic consistency of the model checking [12].

Like all results from Kurtz's theorem, the Fluid Model Checking result pertains to models within a finite time horizon. However useful properties in CSL are sometimes expressed in terms of the steady state operator $\mathscr{S}$. Subsequently, Bortolussi and Hillston consolidated the Fluid Model Checking approach by incorporating the next state operator and the steady state operator [14]. This latter involved establishing when Kurtz's result can safely be extended to the infinite time horizon in this context.

A limitation of the Fluid Model Checking approach is that only properties of a single individual agent (or small set of agents) within a population can be checked. But for population models it is natural to wish to evaluate more global properties such as if a proportion of agents within a population have reached a particular state within a given time period. In [15], Bortolussi and Lanciani present an alternative approach which is able to deal with such properties. Their work is based on a second-order fluid approximation known as Linear Noise Approximation [68]. This can be regarded as a functional version of the Central Limit Approximation [45].

The basic idea of [15] is to lift local specifications to collective ones by means of the Central Limit Theorem. Thus the properties that they consider are first expressed as a property of an individual agent, specified by a deterministic timed automaton with a single clock. This clock is taken to be global — it is never reset and keeps track of

global passing of time. For an individual this will be a linear-time property. Such an individual property $\varphi(t)$ is then lifted to the population level to estimate the probability that a given number of agents within the system which satisfy $\varphi(t)$.

The method presented in [15] allows us to quickly estimate this probability by exploiting the Central Limit or Linear Noise Approximation (LNA). The key idea is to keep some estimation of the variability in the system. Rather than solely using the fluid approximation of average behaviour of the normalised behaviour $x(t)$, fluctuations in the form of Gaussian processes of the order of $\sqrt{N}$, where N is the population size, are included.

$$\mathbf{X}^{(N)}(t) \approx Nx(t) + \sqrt{N}Z(t),$$

where $Z(t)$ is a Gaussian stochastic process, i.e. a process whose finite dimensional projection (marginal distributions at any fixed and finite set of times) are Gaussian. $Z(t)$ has zero mean, and a covariance given by the solution of an additional set of $\mathcal{O}(N^2)$ ODEs. More details can be found in [15, 68].

For the purposes of model checking the authors combine the automaton-based property specification with the model of an individual agent, using a product construction (taking into account the clock constraints). This produces a population model with more variables, counting pairs of state-property configurations. The LNA is applied to this new model. The authors show that for a large class of individual properties, it is possible to introduce a variable $X_\varphi(t)$ in the extended model that counts how many individual agents satisfy the local property up to time t. From the Gaussian approximation of $X_\varphi(t)$, then one can easily compute the probabilities of interest. In [15], the authors discuss preliminary results, which are quite accurate and computationally efficient.

A further use of mean field approximation in model checking has recently been developed for discrete time, synchronous-clock population processes by Loreti *et al.* [49]. Although also derived from Kurtz's Theorem, this work takes a different approach as it is an *on-the-fly* model checker, only examining states as they are required for checking the property, rather than constructing the whole state space initially [25, 8, 33]. Similarly to Fluid Model Checking [11], in [49] the authors focus on a single individual or small set of individuals, with properties expressed in PCTL, and consider their evolution in the mean field created by the rest of the system. Again fast simulation provides the foundation for the approach, but for the discrete case, Loreti *et al.* follow the approach of [50] in which the behaviour of each agent is captured by a finite state discrete time Markov chain (DTMC).

As previously, the authors consider a system comprised of N agents, each with some initial state. A system *global state* $\mathbf{C}^{(N)} = \langle c_1, \ldots, c_N \rangle$ is the N-tuple of the current local states of its object instances. The dynamics of the system arise from all agents proceeding in discrete time, synchronously. A transition matrix $\mathbf{K}^{(N)}$ defines the state transitions of the object and their probabilities, and this may depend on the distribution of states of *all* agents in the system. More specifically, $\mathbf{K}^{(N)}$ is a function $\mathbf{K}^{(N)}(\mathbf{m})$ of the *occupancy measure* vector $\mathbf{m}$ of the current global state $\mathbf{C}^{(N)}$ (switching to the counting abstraction and normalising). State labels are associated with the states of an agent in its specification, and a global state is taken to assume the labels of the first component in the N-tuple. Further global system atomic properties can be expressed.

In [49] the authors develop a model checking algorithm which can applied in both the exact probabilistic case, and for the approximate mean-field semantics of the models. Here we focus on the latter approach. In this discrete case, for N large, the overall behaviour of the system in terms of its occupancy measure can be approximated by the (deterministic) solution of a mean-field *difference equation*. Loreti *et al.* show that the deterministic iterative procedure developed in [50] to compute the average overall behaviour of the system and behaviour of individual agents in that context, combines well with on-the-fly probabilistic model checking for bounded PCTL formulas on the selected agents. Just as in Fluid Model Checking [11], since the transition probabilities of individual agents may depend on the occupancy measure at a given time, the truth values of formulas may vary with time. The asymptotic correctness of the model checking procedure has been proven and a prototype implementation of the model checker, FlyFast, which has been applied to a variety of models [49].

One drawback of mean-field or fluid approximation is that the convergence results apply to infinite populations and currently there are not useful bounds on the errors introduced when smaller populations are considered. Some promising work in this direction was recently published by Bortolussi and Hayden [10]. In this paper the authors consider the transient dynamics and the steady state of certain classes of discrete-time population Markov processes. They combine stochastic bounds in terms of martingale inequalities and Chernoff inequalities, with control-theoretic methods to study the stability of a system perturbed by non-deterministic noise terms, and with algorithms to over-approximate the set of reachable states. The key idea is to abstract stochastic noise non-deterministically and apply techniques from control theory to examine the phase space of the mean field limit. This gives a more refined view of the dynamic behaviour allowing tighter bounds than the previously proposed bounds of Darling and Norris [28] which expand exponentially with time.

6 Conclusions and Future Perspectives

The fluid approximation technique is suitable for models comprised of interactions of populations of components, each component having relatively simple behaviour (few or moderate numbers of local states) but many components within the population. Moreover, in these cases the accuracy of the approximation increases as the size of the population grows. Building such models with a discrete formal description technique supports careful specification of the interactions between the components. This is in contrast to when mean field or fluid approximation is applied in fields such as epidemiology where predefined sets of ODEs are used, without consideration for the implicit assumptions about the interactions of individuals.

However, the population models amenable to fluid approximation are not the only systems which suffer from state space explosion and the technique is not suitable for models comprised of a small number of individual components, each of which has very complex behaviour resulting in a large number of local states. Moreover, recent work by Tschaikowski and Tribastone has shown that if the mapping to ODEs is carried out naively, there can be a problem of *fluid state space explosion* [67]. Nevertheless, the approach offers new possibilities for model analysis, tackling systems which would previously have been completely intractable and opening new arenas of research.

Acknowledgement. This work is partially supported by the EU project QUANTICOL, 600708. The author is very grateful to Luca Bortolussi and Stephen Gilmore for help in the preparation of this paper.

References

1. Aziz, A., Singhal, V., Balarin, F., Brayton, R., Sangiovanni-Vincentelli, A.: Verifying continuous time Markov chains. In: Alur, R., Henzinger, T.A. (eds.) CAV 1996. LNCS, vol. 1102, pp. 269–276. Springer, Heidelberg (1996)
2. Baier, C., Haverkort, B., Hermanns, H., Katoen, J.P.: Model checking continuous-time Markov chains by transient analysis. In: Emerson, E.A., Sistla, A.P. (eds.) CAV 2000. LNCS, vol. 1855, pp. 358–372. Springer, Heidelberg (2000)
3. Bakhshi, R., Cloth, L., Fokkink, W., Haverkort, B.R.: Mean-field analysis for the evaluation of gossip protocols. In: Proceedings of 6th Int. Conference on the Quantitative Evaluation of Systems (QEST 2009), pp. 247–256 (2009)
4. Bakhshi, R., Cloth, L., Fokkink, W., Haverkort, B.R.: Mean-field framework for performance evaluation of push-pull gossip protocols. Perform. Eval. 68(2), 157–179 (2011)
5. Benaïm, M., Le Boudec, J.: A class of mean field interaction models for computer and communication systems. In: Performance Evaluation (2008)
6. Benaïm, M., Le Boudec, J.Y.: On mean field convergence and stationary regime. CoRR, abs/1111.5710 (2011)
7. Bernardo, M., Gorrieri, R.: A Tutorial on EMPA: A Theory of Concurrent Processes with Nondeterminism, Priorities, Probabilities and Time. Theor. Comput. Sci. 202(1-2), 1–54 (1998)
8. Bhat, G., Cleaveland, R., Grumberg, O.: Efficient On-the-Fly Model Checking for CTL*. In: Logic in Computer Science (LICS 1995), pp. 388–397 (1995)
9. Bortolussi, L.: On the approximation of stochastic concurrent constraint programming by master equation, vol. 220, pp. 163–180 (2008)
10. Bortolussi, L., Hayden, R.A.: Bounds on the deviation of discrete-time Markov chains from their mean-field model. Perform. Eval. 70(10), 736–749 (2013)
11. Bortolussi, L., Hillston, J.: Fluid model checking. In: Koutny, M., Ulidowski, I. (eds.) CONCUR 2012. LNCS, vol. 7454, pp. 333–347. Springer, Heidelberg (2012)
12. Bortolussi, L., Hillston, J.: Fluid model checking. CoRR, 1203.0920 (2012)
13. Bortolussi, L., Hillston, J., Latella, D., Massink, M.: Continuous approximation of collective systems behaviour: A tutorial. In: Performance Evaluation (2013)
14. Bortolussi, L., Hillston, J.: Checking Individual Agent Behaviours in Markov Population Models by Fluid Approximation. In: Bernardo, M., de Vink, E., Di Pierro, A., Wiklicky, H. (eds.) SFM 2013. LNCS, vol. 7938, pp. 113–149. Springer, Heidelberg (2013)
15. Bortolussi, L., Lanciani, R.: Model Checking Markov Population Models by Central Limit Approximation. In: Joshi, K., Siegle, M., Stoelinga, M., D'Argenio, P.R. (eds.) QEST 2013. LNCS, vol. 8054, pp. 123–138. Springer, Heidelberg (2013)
16. Bortolussi, L., Policriti, A.: Dynamical systems and stochastic programming: To ordinary differential equations and back. In: Priami, C., Back, R.-J., Petre, I. (eds.) Transactions on Computational Systems Biology XI. LNCS, vol. 5750, pp. 216–267. Springer, Heidelberg (2009)
17. Calder, M., Gilmore, S., Hillston, J.: Automatically deriving ODEs from process algebra models of signalling pathways. In: Proceedings of Computational Methods in Systems Biology (CMSB 2005), pp. 204–215 (2005)

18. Cardelli, L.: Brane Calculi. In: Danos, V., Schachter, V. (eds.) CMSB 2004. LNCS (LNBI), vol. 3082, pp. 257–278. Springer, Heidelberg (2005)
19. Cardelli, L.: On process rate semantics. Theor. Comput. Sci. 391(3), 190–215 (2008)
20. Cardelli, L.: From Processes to ODEs by Chemistry. In: Ausiello, G., Karhumäki, J., Mauri, G., Ong, L. (eds.) 5th IFIP International Conference On Theoretical Computer Science - TCS 2008. IFIP, vol. 273, pp. 261–281. Springer, Boston (2008)
21. Chen, T., Han, T., Katoen, J.-P., Mereacre, A.: LTL model checking of time-inhomogeneous markov chains. In: Liu, Z., Ravn, A.P. (eds.) ATVA 2009. LNCS, vol. 5799, pp. 104–119. Springer, Heidelberg (2009)
22. Chen, T., Han, T., Katoen, J.P., Mereacre, A.: Model checking of continuous-time Markov chains against timed automata specifications. Logical Methods in Computer Science 7(1) (2011)
23. Ciardo, G., Trivedi, K.S.: A Decomposition Approach for Stochastic Reward Net Models. Perform. Eval. 18(1), 37–59 (1993)
24. Ciocchetta, F., Hillston, J.: Bio-PEPA: A framework for the modelling and analysis of biological systems. Theor. Comput. Sci. 410(33-34), 3065–3085 (2009)
25. Courcoubetis, C., Vardi, M., Wolper, P., Yannakakis, M.: Memory-efficient algorithms for the verification of temporal properties Form. Methods Syst. Des. 1(2-3), 275–288 (1992)
26. Danos, V., Feret, J., Fontana, W., Harmer, R., Krivine, J.: Abstracting the Differential Semantics of Rule-Based Models: Exact and Automated Model Reduction. In: Proceedings of Logic in Computer Science (LICS 2010), pp. 362–381 (2010)
27. Darling, R.W.R.: Fluid limits of pure jump Markov processes: A practical guide (2002), `http://arXiv.org`
28. Darling, R.W.R., Norris, J.R.: Differential equation approximations for Markov chains. Probability Surveys 5 (2008)
29. Donatelli, S., Haddad, S., Sproston, J.: Model checking timed and stochastic properties with CSL^{TA}. IEEE Trans. Software Eng. 35(2), 224–240 (2009)
30. Gast, N., Gaujal, B.: A mean field model of work stealing in large-scale systems. In: Proceedings of ACM SIGMETRICS 2010, pp. 13–24 (2010)
31. Geisweiller, N., Hillston, J., Stenico, M.: Relating continuous and discrete PEPA models of signalling pathways. Theor. Comput. Sci. 404(1-2), 97–111 (2008)
32. Gillespie, D., Petzold, L.: Numerical simulation for biochemical kinetics. In: System Modeling in Cellular Biology, pp. 331–353. MIT Press (2006)
33. Gnesi, S., Mazzanti, F.: An Abstract, on the Fly Framework for the Verification of Service-Oriented Systems. In: Results of the SENSORIA Project, pp. 390–407 (2011)
34. Hayden, R.A., Bradley, J.T.: A fluid analysis framework for a Markovian process algebra. Theor. Comput. Sci. 411(22-24), 2260–2297 (2010)
35. Hayden, R.A., Bradley, J.T., Clark, A.D.: Performance specification and evaluation with unified stochastic probes and fluid analysis. IEEE Trans. Software Eng. 39(1), 97–118 (2013)
36. Hayden, R.A., Stefanek, A., Bradley, J.T.: Fluid computation of passage-time distributions in large Markov models. Theor. Comput. Sci. 413(1), 106–141 (2012)
37. Hermanns, H., Herzog, U., Katoen, J.-P.: Process algebra for performance evaluation. Theor. Comput. Sci. 274(1-2), 43–87 (2002)
38. Hillston, J.: A Compositional Approach to Performance Modelling. Cambridge University Press (1995)
39. Hillston, J.: Exploiting Structure in Solution: Decomposing Compositional Models. In: Brinksma, E., Hermanns, H., Katoen, J.-P. (eds.) EEF School 2000 and FMPA 2000. LNCS, vol. 2090, pp. 278–314. Springer, Heidelberg (2001)
40. Hillston, J.: Fluid flow approximation of PEPA models. In: Proceedings of the Second International Conference on the Quantitative Evaluation of SysTems, QEST 2005, pp. 33 – 42 (2005)

41. Katoen, J.-P., Khattri, M., Zapreev, I.S.: A Markov Reward Model Checker. In: Proceedings of Quantitative Evaluation of Systems, QEST 2005, pp. 243–244 (2005)
42. Katoen, J.-P., Mereacre, A.: Model checking HML on piecewise-constant inhomogeneous markov chains. In: Cassez, F., Jard, C. (eds.) FORMATS 2008. LNCS, vol. 5215, pp. 203–217. Springer, Heidelberg (2008)
43. Kolesnichenko, A., Remke, A., de Boer, P.T., Haverkort, B.R.: Comparison of the mean-field approach and simulation in a peer-to-peer botnet case study. In: Thomas, N. (ed.) EPEW 2011. LNCS, vol. 6977, pp. 133–147. Springer, Heidelberg (2011)
44. Kolesnichenko, A., Remke, A., de Boer, P.-T., Haverkort, B.R.: A logic for model-checking of mean-field models. In: Proceedings of 43rd Int. Conference on Dependable Systems and Networks, DSN 2013 (2013)
45. Kurtz, T.G.: Solutions of ordinary differential equations as limits of pure jump Markov processes. Journal of Applied Probability 7, 49–58 (1970)
46. Kurtz, T.G.: Approximation of population processes. SIAM (1981)
47. Kwiatkowska, M., Norman, G., Parker, D.: Probabilistic symbolic model checking with PRISM: A hybrid approach. International Journal on Software Tools for Technology Transfer 6(2), 128–142 (2004)
48. Kwiatkowska, M., Norman, G., Parker, D.: PRISM 4.0: Verification of probabilistic real-time systems. In: Gopalakrishnan, G., Qadeer, S. (eds.) CAV 2011. LNCS, vol. 6806, pp. 585–591. Springer, Heidelberg (2011)
49. Latella, D., Loreti, M., Massink, M.: On-the-fly Fast Mean-Field Model-Checking. In: Abadi, M., Lluch Lafuente, A. (eds.) TGC 2013. LNCS, vol. 8358, pp. 297–314. Springer, Heidelberg (2014)
50. Le Boudec, J.-Y., McDonald, D., Mundinger, J.: A Generic Mean Field Convergence Result for Systems of Interacting Objects. In: Proceedings of Quantitative Evaluation of Systems (QEST 2007), pp. 3–18 (2007)
51. McCaig, C.: From individuals to populations: changing scale in process algebra models of biological systems. PhD thesis, University of Stirling (2008)
52. McCaig, C., Norman, R., Shankland, C.: Process Algebra Models of Population Dynamics. In: Horimoto, K., Regensburger, G., Rosenkranz, M., Yoshida, H. (eds.) AB 2008. LNCS, vol. 5147, pp. 139–155. Springer, Heidelberg (2008)
53. McCaig, C., Norman, R., Shankland, C.: From Individuals to Populations: A Symbolic Process Algebra Approach to Epidemiology. Mathematics in Computer Science 2(3), 535–556 (2009)
54. Norman, R., Shankland, C.: Developing the Use of Process Algebra in the Derivation and Analysis of Mathematical Models of Infectious Disease. In: Moreno-Díaz Jr., R., Pichler, F. (eds.) EUROCAST 2003. LNCS, vol. 2809, pp. 404–414. Springer, Heidelberg (2003)
55. Norris, J.R.: Markov Chains. Cambridge University Press (1997)
56. Priami, C.: Stochastic pi-Calculus. Comput. J. 38(7), 578–589 (1995)
57. Regev, A., Panina, E.M., Silverman, W., Cardelli, L., Shapiro, E.Y.: BioAmbients: an abstraction for biological compartments. Theor. Comput. Sci. 325(1), 141–167 (2004)
58. Regev, A., Shapiro, E.: Cellular Abstractions: Cells as computation. Nature 419(6905), 343–343 (2002)
59. Rutten, J., Kwiatkowska, M., Norman, G., Parker, D.: Mathematical Techniques for Analyzing Concurrent and Probabilistic Systems. CRM Monograph Series, vol. 23. American Mathematical Society (2004)
60. Silva, M., Recalde, L.: Petri nets and integrality relaxations: A view of continuous Petri net models. IEEE Trans. on Systems, Man, and Cybernetics, Part C 32(4), 314–327 (2002)
61. Silva, M., Recalde, L.: On fluidification of Petri Nets: from discrete to hybrid and continuous models. Annual Reviews in Control 28(2), 253–266 (2004)

62. Sumpter, D.T.J.: From Bee to Society: An Agent-based Investigation of Honey Bee Colonies. PhD thesis, University of Manchester (2000)
63. Tofts, C.M.N.: A Synchronous Calculus of Relative Frequency. In: Baeten, J.C.M., Klop, J.W. (eds.) CONCUR 1990. LNCS, vol. 458, pp. 467–480. Springer, Heidelberg (1990)
64. Tribastone, M., Ding, J., Gilmore, S., Hillston, J.: Fluid rewards for a stochastic process algebra. IEEE Trans. Software Eng. 38(4), 861–874 (2012)
65. Tribastone, M., Gilmore, S., Hillston, J.: Scalable differential analysis of process algebra models. IEEE Trans. Software Eng. 38(1), 205–219 (2012)
66. Trivedi, K.S., Kulkarni, V.G.: FSPNs: Fluid Stochastic Petri Nets. In: Ajmone Marsan, M. (ed.) ICATPN 1993. LNCS, vol. 691, pp. 24–31. Springer, Heidelberg (1993)
67. Tschaikowski, M., Tribastone, M.: Tackling continuous state-space explosion in a Markovian process algebra. Theor. Comput. Sci. 517, 1–33 (2014)
68. Van Kampen, N.G.: Stochastic Processes in Physics and Chemistry. Elsevier (1992)