HAL open science

# Concurrency and Composition in a Stochastic World

Christian Eisentraut, Holger Hermanns, Lijun Zhang

## To cite this version:

Christian Eisentraut, Holger Hermanns, Lijun Zhang. Concurrency and Composition in a Stochastic World. CONCUR 2010 - Concurrency Theory, 21th International Conference, Aug 2010, Paris, France. pp.21-39. hal-00650728

**HAL Id: hal-00650728**

**https://inria.hal.science/hal-00650728**

Submitted on 3 Jan 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Concurrency and Composition in a Stochastic World

Christian Eisentraut[1], Holger Hermanns[1,2], and Lijun Zhang[3]

[1] Saarland University, Saarbrücken, Germany
[2] INRIA Grenoble, Rhône-Alpes, France
[3] DTU Informatics, Technical University of Denmark, Kgs. Lyngby, Denmark

**Abstract.** We discuss conceptional and foundational aspects of Markov automata [22]. We place this model in the context of continuous- and discrete-time Markov chains, probabilistic automata and interactive Markov chains, and provide insight into the parallel execution of such models. We further give a detailled account of the concept of relations on distributions, and discuss how this can generalise known notions of weak simulation and bisimulation, such as to fuse sequences of internal transitions.

## 1 Introduction

Petri nets are a model of concurrency. Among the most successful and widespread variations we find a class of models tailored to performance and dependability evaluation, Generalised Stochastic Petri nets (GSPNs) [38,39]. GSPNs support stochastically timed behaviour and weighted immediate choices. A simple example GSPN is depicted in Figure 1. What we see are places and transitions, connected by directed arrows. There are two types of transitions, timed (drawn white) and immediate (drawn black) transitions. If enabled, the latter fire immediately, while the earlier fire after a delay that is distributed according to a negative exponential distribution. Immediate transitions have priority over timed transitions. Evaluation of a GSPN proceeds at the level of the

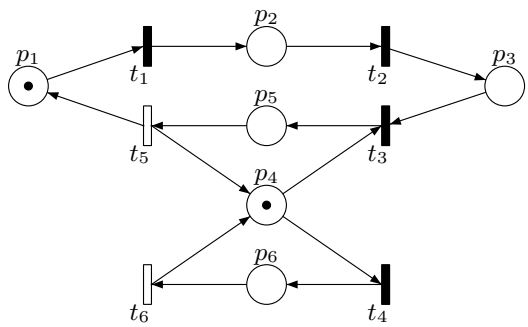


**Fig. 1.** A confused GSPN

reachability graph. That graph is reduced to a continuous-time Markov chain (CTMC), for which efficient steady-state and transient solvers are at hand [14,17]. Due to their formality, visual representation, and the availability of efficient evaluation support,

GSPNs have found widespread application in many diverse disciplines, including manufacturing, logistics, systems biology and so on [11,50,43,24].

Astonishingly, the above evaluation trajectory is incomplete. It is restricted to the class of *confusion-free* GSPNs. The net depicted in Figure 1 is confused. Confusion arises if a firing sequence admits the simultaneous enabling of multiple non-conflicting immediate transitions. Priorities and weights can eliminate confusion, but with strange effects. To shed some light on this phenomenon, assume that all weights of immediate transitions and also all rates of timed transitions are $1$ for the above example. In this case, the steady state probability of a token being present in place $p_5$ is $4/11$. This probability changes to $2/5$ if one replaces the immediate transition $t_2$ and place $p_2$ by a direct arc from $t_1$ to $p_3$. In other words, adding or removing an immediate 'stutter step' somewhere in a weighted confused net may change the performance figures obtained for such nets. The issue of confusion has been discussed in the literature [13,15,37,6], but even after 25 years of pragmatic use, the final word on the foundational semantics of GSPNs is yet to be spoken.

Petri nets are visual, but not compositional. Process calculi provide compositional theories for complex systems, especially those involving communicating, concurrently executing components [8]. This paper is not about Petri nets with stochastically timed behaviour and weighted immediate choices. It is not about Petri nets at all. It is about compositional theories in a setting with stochastically timed behaviour and weighted immediate choices. The paper revolves around Markov automata [22], a model that indeed solves the semantic challenges hinted at above. This is achieved by harvesting and intertwining results established independently for two process calculi that each extend classical concurrency models in simple yet conservative fashions: Probabilistic automata [44,7,12] (PA), and interactive Markov chains [25] (IMC). Though different in flavour, both are equipped with compositional theories for strong and weak bisimilarities and corresponding equational theories.

In probabilistic automata, there is no global notion of time. Concurrent processes may perform random experiments inside a transition. This is represented by transitions of the form $s \xrightarrow{a} \mu$, where $s$ is a state, $a$ is an action label, and $\mu$ is a probability distribution over states. Labelled transition systems are instances of this model family, obtained by restricting to Dirac distributions (assigning full probability to single states). Thus, foundational concepts and results of standard concurrency theory are retained in their full beauty, and extend smoothly to the model of probabilistic automata. By restricting to Dirac distributions (assigning full probability to single states), labelled transition systems arise, and standard concurrency theoretical concepts and results are retained in their full beauty, and these extend smoothly to the model of probabilistic automata. In Markov automata, probabilistic automata are employed to give a genuine semantics to weighted immediate choices and their sequential or concurrent execution.

Interactive Markov chains in turn arise from classical concurrency models by adding a second type of transitions $s \overset{\lambda}{\twoheadrightarrow} s'$, that can embody random delays governed by a negative exponential distribution with some parameter $\lambda$. This twists the model to one that is running on a continuous time line, and where executions of actions take no time and happens immediately – unless an action can be blocked by the environment. This is linked to the process algebraic notion of maximal progress for internal actions. By

dropping the second type of transitions, again, standard concurrency theory is regained in its entirety, and extends smoothly to the full IMC model. In Markov automata, IMC are used to represent stochastic timed behaviour and principal interaction possibilities.

Markov automata stand on the shoulders of PA and IMC. In a nutshell, the resulting model encompassed the expressiveness of GSPNs without semantic glitches, and with an entirely compositional theory. Due to the different time scales present in this model, this is a demanding endeavour. As in plain IMCs, internal immediate probabilistic transitions cannot be blocked and take no time to execute. Consequently, *MA* support fusing sequences of them. This implies that we need to partially ignore the branching structure of our probabilistic automata induced substructures when defining equalities, especially weak bisimilarity, on them. This is achieved by defining bisimulation as a relation on distributions over states, rather than as a relation on states, and by using the concept of distribution splitting, which is a concept of interest beyond the concrete Markov automata context. It allows for a surprisingly simple formulation of how transitions are fused, and allows to present various seemingly distinct preorders and equivalences in a unifying framework.

In this paper, we discuss conceptional and foundational aspects of Markov automata, partly rephrasing and complementing [22]. We place it in the context of some well-known and established models of concurrent computation that involve stochastically timed behaviour and weighted immediate choice, especially continuous- and discrete-time Markov chains, probabilistic automata and interactive Markov chains. We especially give insight into the parallel execution of such models, and discuss simulation and bisimulation relations on them. We show that the concept of relations on distributions generalises the respective standard relational notions. With this perspective on the different notions of bisimulation it becomes apparent how weak bisimulation for Markov automata appears as a natural generalisation of the constituent relations.

## 2 Preliminaries

*(Sub-)distributions.* A *subdistribution* $\mu$ over a set $S$ is a function $\mu : S \mapsto [0,1]$ such that $\sum_{s \in S} \mu(s) \leq 1$. We denote by $Supp(\mu) = \{s \in S \mid \mu(s) > 0\}$ the support of $\mu$ and define the probability of $S' \subseteq S$ with respect to $\mu$ as $\mu(S') := \sum_{s \in S'} \mu(s)$. Let $|\mu| := \mu(S)$ denote the *size* of the subdistribution $\mu$. We say $\mu$ is a *full distribution*, or distribution, if $|\mu| = 1$. Let $Dist(S)$ and $Subdist(S)$ be the set of distributions and subdistributions over $S$, respectively. For $s \in S$, we let $\Delta_s \in Dist(S)$ denote the *Dirac* distribution for $s$, i.e., $\Delta_s(s) = 1$. Let $\mu$ and $\mu'$ be two subdistributions. We define the subdistribution $\mu'' := \mu \oplus \mu'$ by $\mu''(s) = \mu(s) + \mu'(s)$, if $|\mu''| \leq 1$. Conversely, we say that $\mu''$ can be split into $\mu$ and $\mu'$. Or that $(\mu, \mu')$ is a splitting of $\mu''$. Moreover, if $x \cdot |\mu| \leq 1$, we let $x\mu$ denote the subdistribution defined by: $(x\mu)(s) = x \cdot \mu(s)$.

(Sub-)distributions can also be considered as sets over $S \times (0,1]$, where $(s_1, r_1)$, $(s_2, r_2) \in S \times (0,1] \wedge s_1 = s_2$ implies $r_1 = r_2$, and where the second components of the elements sum up to a number smaller or equal to 1. The set view on subdistribution will be widely used throughout the paper. For example, to denote the distribution $\mu$ with $\mu(s_1) = 0.75$ and $\mu(s_2) = 0.25$, we may write $\mu = [\![(s_1, 0.75), (s_2, 0.25)]\!]$. Let for an element $s \in S$ and a subdistribution $\mu$ over $S$ the expression $\mu - s$ denote

the subdistribution that is obtained from $\mu$ by removing the pair $(s, \mu(s))$ from $\mu$, if it exists. To make clear when we talk about sets representing subdistributions and when about general sets, we use $[\![$ and $]\!]$ for subdistributions, $\{$ and $\}$ for sets. Since $\oplus$ is associative and commutative, we may use the notation $\bigoplus_{i \in I}$ for arbitrary sums over a finite index set $I$.

*Labelled trees.* For $\sigma, \sigma' \in \mathbb{N}^*_{>0}$ we write $\sigma \leq \sigma'$ if there exists a (possibly empty) $\phi$ such that $\sigma\phi = \sigma'$. A partial function $\mathcal{T} : \mathbb{N}^*_{>0} \to L$, which satisfies

- if for $\sigma, \sigma' \in \mathbb{N}^*_{>0}$: $\sigma \leq \sigma'$ and $\sigma' \in \mathit{dom}(\mathcal{T})$ then $\sigma \in \mathit{dom}(\mathcal{T})$
- if $\sigma i \in \mathit{dom}(\mathcal{T})$ for $i > 1$, then also $\sigma(i-1) \in \mathit{dom}(\mathcal{T})$
- $\varepsilon \in \mathit{dom}(\mathcal{T})$

is called an *(infinite) $L$-labelled tree*. Let $\sigma \in \mathit{dom}(\mathcal{T})$: $\sigma$ is called a leaf of $\mathcal{T}$ if there is no $\sigma' \in \mathit{dom}(\mathcal{T})$ such that $\sigma < \sigma'$. The empty word $\varepsilon$ is called the root of $\mathcal{T}$. We denote the set of all leaves of $\mathcal{T}$ by $\mathit{Leaf}_{\mathcal{T}}$ and the set of all inner nodes by $\mathit{Inner}_{\mathcal{T}}$. If the tree has only one node, the root node, then this node is contained in both $\mathit{Inner}_{\mathcal{T}}$ and $\mathit{Leaf}_{\mathcal{T}}$. In any other case the two sets are disjoint. For a node $\sigma$ of a tree $\mathcal{T}$ let $\mathit{Children}(\sigma) = \{\sigma i \mid \sigma i \in \mathit{dom}(\mathcal{T})\}$. In this paper, we consider $L$-labelled trees with finite branching, i.e., $|\mathit{Children}(\sigma)| < \infty$ for all node $\sigma$.

## 3 Markov Automata

We integrate probabilistic automata and interactive Markov chains into one model, defined by means of a twofold transition relation $\longrightarrow$ and $\twoheadrightarrow$ [22]:

**Definition 1.** *A Markov automaton MA is a quintuple $(S, Act, \longrightarrow, \twoheadrightarrow, s_o)$, where*

- *$S$ is a nonempty finite set of states,*
- *$Act$ is a set of actions containing the internal action $\tau$,*
- *$\longrightarrow \subset S \times Act \times Dist(S)$ is a set of immediate transitions, and*
- *$\twoheadrightarrow \subset S \times \mathbb{R}_{\geq 0} \times S$ is a set of timed transitions, and*
- *$s_o \in S$ is the initial state.*

We let $s, u, v, t, E, F, G$ and their variants with indices range over $S$. For timed transitions, $\lambda, \mu \in \mathbb{R}_{\geq 0}$ denote rates of exponential distributions. For immediate transitions, $a$ ranges over $Act$, and $\mu$ ranges over $Dist(S)$. A immediate transition $(E, a, \mu) \in \longrightarrow$ is also denoted by $E \xrightarrow{a} \mu$, similarly we define $E \overset{\lambda}{\twoheadrightarrow} F$. We say an action $a \in Act$ is enabled in $E$, if there exists an immediate transition $E \xrightarrow{a} \mu$. A state $E \in S$ is called *stable* if $\tau$ is not enabled in $E$. If $E$ is stable, we use the shorthand notation $E{\downarrow}$. We employ the *maximal progress assumption*. This means that if a state is not stable, time is not allowed to progress, making timed transitions out of this state irrelevant [28]. As in IMC, this assumption is not evident in the model, but is part of the equivalences defined on it.

We define a (nonnegative) real-valued function $\mathit{rate}_{MA} : S \times S \mapsto \mathbb{R}_{\geq 0}$, that calculates the rate to reach a state $s'$ from a state $s$ by

$$\mathit{rate}_{MA}(s, s') = \sum \{\lambda \mid s \overset{\lambda}{\twoheadrightarrow} s'\}.$$

Moreover, we define $\mathit{rate}_{MA}(s) := \sum_{s'} \mathit{rate}_{MA}(s, s')$ as the *exit rate* of $s$. The index is omitted if clear from the context. The delay associated with a state $s$ that enables timed transitions is (negative) exponentially distributed with the exit rate $\mathit{rate}(s)$. In general, the probability to move from $s$ to the successor state in $s'$ equals the probability that the timed transitions that lead from $s$ to $'$ *wins the race*. Therefore, the *discrete branching probability* to move to $s'$ is given by $\mathbf{P}(s, s') := \frac{\mathit{rate}(s,s')}{\mathit{rate}(s)}$. For $s \in S$, we use $\mathbf{P}(s, \cdot)$ to denote this discrete branching distributions.

A *Dirac* distribution assigns full probability to a single outcome. We say that $\longrightarrow$ is *Dirac* if the distributions occurring as third components of $\longrightarrow$ are all Dirac, is *deterministic* if $E \xrightarrow{\alpha} \mu_1$ and $E \xrightarrow{\alpha} \mu_2$ implies that $\mu_1 = \mu_2$. Markov automata subsume many concurrent systems, which are discussed below.

1. *Labelled Transition Systems:* If $\Longrightarrow = \emptyset$ and $\longrightarrow$ is Dirac, we obtain labelled transition systems.
2. *Discrete-time Markov chains:* If $\Longrightarrow = \emptyset$ and $|Act| = 1$ and $\longrightarrow$ is deterministic, we obtain discrete-time Markov chains (DTMCs). In this case one usually ignores the single action, and writes it as a triple $(S, \mathbf{P}, s_0)$ where $\mathbf{P}$ is called the probability matrix, and is given by $\mathbf{P}(s, s') := \mu(s')$ provided $s \xrightarrow{\alpha} \mu$.
3. *Continuous-time Markov chains:* If $\longrightarrow = \emptyset$ we obtain continuous-time Markov chains (CTMCs). It is commonly represented as a triple $(S, \mathbf{Q}, s_0)$ where $\mathbf{Q}$ is called the infinitesimal generator matrix, and is given by $\mathbf{Q}(s, s') := \mathit{rate}(s, s')$ provided $s \neq s'$ and $\mathbf{Q}(s, s) = \mathit{rate}(s, s) - \mathit{rate}(s)$. The latter reflects that in the original mathematical formulation of CTMCs it is impossible to make a difference between staying in state $s$, and jumping back to $s$ from $s$.
4. *Probabilistic Automata:* If $\Longrightarrow = \emptyset$ we obtain probabilistic automata. If additionally $\longrightarrow$ is deterministic, we arrive at Markov decision processes (MDPs).
5. *Interactive Markov chains:* If $\longrightarrow$ is Dirac, we get interactive Markov chains (IMCs).

## 4 Parallel Composition

This section introduces parallel composition of *MA*s, and places it in the context of general operators for parallel composition, also motivating the rationale behind the semantic choices of the parallel operators in PA and IMC.

Assuming we are given two *MA*s $MA_1 = (S^1, Act^1, \longrightarrow^1, \Longrightarrow^1, s_o^1)$ and $MA_2 = (S^2, Act^2, \longrightarrow^2, \Longrightarrow^2, s_o^2)$, we consider a family of parallel operators $||_A$ indexed by some set $A \subseteq (Act^1 \cup Act^2) - \{\tau\}$. For a clear presentation we use these operators as syntactical means to denote some state $s_1 \;||_{\mathsf{A}}\; s_2$, which arises by the parallel composition of $s_1$ and $s_2$. As syntactic sugar, we lift them to subdistributions as follows: for subdistributions $\mu_1 \in \mathit{Subdist}(S_1)$ and $\mu_2 \in \mathit{Subdist}(S_2)$, $\mu_1 \;||_{\mathsf{A}}\; \mu_2$ denotes the subdistribution in $\mathit{Dist}(S_1 \times S_2)$ by distributing $||_{\mathsf{A}}$ element-wise. As an example, we have $(\mu_1 \;||_{\mathsf{A}}\; \mu_2)(s_1 \;||_{\mathsf{A}}\; s_2) := \mu_1(s_1) \cdot \mu_2(s_2)$.

**Definition 2.** *Let $MA_1$, $MA_2$ and $A$ be as discussed above. The parallel operator can be applied to the two MAs to form the parallel composition $MA_1 \;||_A\; MA_2 = (S, Act^1 \cup Act^2, \longrightarrow, \Longrightarrow, s_o)$ of processes where*

- $S = \{s_1 \,||_A\, s_2 \mid (s_1, s_2) \in S^1 \times S^2\}$,
- $(s_1 \,||_A\, s_2, a, \mu_1 \,||_A\, \mu_2) \in \rightarrow$ *iff either*
  - $a \in A$ *and* $(s_i, a, \mu_i) \in \rightarrow^i$ *for each* $i \in \{1, 2\}$, *or*
  - $a \notin A$ *and* $(s_1, a, \mu_1) \in \rightarrow^1 \wedge \mu_2 = \Delta_{s_2}$ *or* $(s_2, a, \mu_2) \in \rightarrow^2 \wedge \mu_1 = \Delta_{s_1}$
- $(s_1 \,||_A\, s_2, \lambda, s_1' \,||_A\, s_2') \in \twoheadrightarrow$ *iff either*
  - *if for each* $i \in \{1, 2\}$, $\text{rate}_{MA_i}(s_i, s_i') > 0 \wedge s_i = s_i'$ *then* $\lambda = \text{rate}_{MA_1}(s_1, s_1') + \text{rate}_{MA_2}(s_2, s_2')$, *otherwise*
  - $\lambda = \text{rate}_{MA_1}(s_1, s_1')$ *and* $s_2' = s_2$, *or* $\lambda = \text{rate}_{MA_2}(s_2, s_2')$ *and* $s_1' = s_1$,
- $s_o = s_o^1 \,||_{\mathsf{A}}\, s_o^2 \in S^1 \times S^2$ *is the initial state.*

In a process algebraic setting, the style of defining this operator can be made more elegant [34], but this is not the topic of this paper.

### 4.1 The Roots of *MA* Parallel Composition

It is illustrative to relate this operator to the ones it is derived from.

1. Whenever $MA_1$ and $MA_2$ are *labelled transitions systems*, $MA_1 \,||_A\, MA_2$ reduces to LOTOS-style parallel composition [9].
2. Whenever $MA_1$ and $MA_2$ are *discrete-time Markov chains* over the same singleton set $Act$, $MA_1 \,||_{Act}\, MA_2$ reduces to the synchronous product of the models, where both Markov chains proceed in lockstep.
3. Whenever $MA_1$ and $MA_2$ are *continuous-time Markov chains*, $MA_1 \,||_\emptyset\, MA_2$ reduces to the independent and concurrent execution of the two Markov chains, justified by the memoryless property.
4. Whenever $MA_1$ and $MA_2$ are *probabilistic automata*, $MA_1 \,||_A\, MA_2$ agrees with the essence of the parallel composition for PA [45] (neglecting minor differences in synchronisation set constructions).
5. Whenever $MA_1$ and $MA_2$ are *interactive Markov chains*, $MA_1 \,||_A\, MA_2$ reduces to the parallel composition for IMC [25].

A few further remarks are in order (despite they may seem obvious to many readers): *MA* takes the LOTOS-style parallel operator $||_A$ as a basis, but we could have equally well opted for CCS style [40], CSP style [30], asynchronous or synchronous $\pi$-calculus style [41], I/O style [36], etc. From a pragmatic perspective, the $||_A$-operator is a convenient 'Swiss army knife'. It can, as special cases, express asynchronous interleaving ($||_\emptyset$), synchronous product ($||_{Act}$ with $|Act| = 1$). It can also be used to encode shared variable communication, as well as asynchronous message passing communication. Shared variables can be modelled as separate *MA*, where states correspond to variable valuations, and transitions are put in place for reading and updating the state. Similarly, asynchronous message passing channels can be encoded as *MA* that keep memory of the messages in transit (see e.g. [4, Chapter 2] for details).

We mention this to make clear that a properly and well understood semantics for this one operator is the nucleus for a well understood semantics of essentially any prevailing communication paradigms found in the real world. Since the models developed with *MA* (just like GSPN, IMC, PA) are meant to be designed and evaluated in order to provide insight into performance and dependability of the system being modelled, a well understood semantics is essential.

### 4.2 A Connection between DTMCs and CTMCs

To shed some more light on the parallel behaviour of these models, we state an interesting observation relating the interleaving semantics for CTMCs to the synchronous semantic for DTMCs. They are both derived from our single parallel composition for *MA*s, but why do they make sense after all? This section establishes a probably astonishing connection between the two.

*Geometric and Exponential Distributions.* It is well known – and we thus here take it for granted – that for any state in a CTMC, the sojourn time of that state is exponentially distributed, and likewise, that for any state in a DTMC, the sojourn time of that state is geometrically distributed.

Furthermore, the exponential distribution can be considered as the limit of the geometric distribution, in the following sense: Let continuous random variable $X$ be exponentially distributed with parameter $\lambda > 0$, i.e. its density is $f_X(t) = \lambda e^{-\lambda t}$. Further, for $\Delta > 0$ with $p = \lambda\Delta < 1$, we consider the geometrically distributed discrete random variable $X_{\lambda\Delta}$ with parameter $\lambda\Delta$, i.e., with density function $f_{X_{\lambda\Delta}}(k) = \lambda\Delta(1-\lambda\Delta)^{k-1}$. Then the definition of Euler's number implies that the density of $X$ at time point $t > 0$ can be approximated by the density function of $X_{\lambda\Delta}$ at step $\lceil \frac{t}{\Delta} \rceil$, formally:

$$\forall t > 0.\ f_X(t) = \lim_{\Delta \to 0} f_{X_{\lambda\Delta}}(\lceil t/\Delta \rceil)/\Delta$$

One may consider $\Delta$ as a step size in continuous time (used to perform a single Bernoulli experiment, from which the geometric distribution is derived). This view is helpful in the discussion that follows.

*Discretised Markov Chain.* Let $M = (S, \mathbf{Q}, s_0)$ be a CTMC, and $\Delta$ be a sufficiently small step size. In the classical terminology [33], this CTMC is a family $\{C(t)\}$ of random variables each taking values in $S$, indexed by (continuous) time $t$, that obeys the Markov property. From this, we can derive a $\Delta$-*discretised* DTMC $M_\Delta = (S, \mathbf{P}_\Delta, s_0)$ by: $\mathbf{P}_\Delta(s, s') = Prob(C(\Delta) = s' \mid C(0) = s)$. It holds that:

- $\mathbf{P}_\Delta(s, s')$ equals $\mathbf{Q}(s, s')\Delta + o(\Delta)$, provided $s \neq s'$, and otherwise
- $\mathbf{P}_\Delta(s, s)$ equals $1 + \mathbf{Q}(s, s)\Delta + o(\Delta)$.

Here, $o(\Delta)$ subsumes the probabilities to pass through intermediate states between $s$ and $s'$ during the interval $\Delta$, and we have $\mathbf{P}_\Delta(s, s) \in (0, 1)$ – for sufficiently small $\Delta$.

Moreover, the rate between $s$ and $s'$ can be derived from the derivative:

- $\mathbf{Q}(s, s') = \lim_{\Delta \to 0} \mathbf{P}_\Delta(s, s')/\Delta$, provided $s \neq s'$, and otherwise
- $\mathbf{Q}(s, s) = \lim_{\Delta \to 0}(-\sum_{s' \neq s} \mathbf{P}_\Delta(s, s')/\Delta)$.

This observation justifies that the behaviour of a CTMC can be approximated arbitrarily closely by a $\Delta$-discretised DTMC, just by choosing $\Delta$ small enough, since in the limit $\mathbf{Q} = \lim_{\Delta \to 0}(\mathbf{P}_\Delta - \mathbf{I})/\Delta$, where $\mathbf{I}$ denotes the identity matrix. The limit is understood element-wise. All the above facts can, albeit usually stated in a different flavour, be found in many textbooks on Markov chains, for example in [47].

What is not found in textbooks is the question whether this approximation is compatible with parallel composition: For two CTMCs $M$ and $M'$, let $M_\Delta$ and $M'_\Delta$ denote the corresponding $\Delta$-discretised DTMCs respectively, which we assume labelled over the same singleton set $Act$. We now consider the synchronous product $M_\Delta \mid\mid_{Act} M'_\Delta$, where the two Markov chains evolve in lockstep with the step size – on a continuous time line – being $\Delta$. Now, how does this product relate to $M \mid\mid_\emptyset M'$, the parallel composition of the CTMCs $M$ and $M'$ under interleaving semantics? The following lemma answers this.

**Lemma 1.** *Let $M = (S, \mathbf{Q}, s_0)$ and $M' = (S', \mathbf{Q}', s'_0)$ be two CTMCs, let $M_\Delta, M'_\Delta$ denote the probability matrices in the discretised DTMCs, respectively. Moreover, let $M \mid\mid_\emptyset M' = (S \mid\mid_\emptyset S', \mathbf{Q}^{\mid\mid}, s_0)$ and $M_\Delta \mid\mid_{Act} M'_\Delta = (S \mid\mid_{Act} S', \mathbf{P}^{\mid\mid}_\Delta, s_0)$. Then,*

$$\mathbf{Q}^{\mid\mid} = \lim_{\Delta \to 0} (\mathbf{P}^{\mid\mid}_\Delta - \mathbf{I})/\Delta$$

*Proof.* Let $s, t \in S$ and $s', t' \in S'$. We consider a few cases:

1. Assume $s \neq t$ and $s' \neq t'$. By the synchronised product in the DTMCs, we have:

$$\mathbf{P}^{\mid\mid}_\Delta(s \mid\mid_{Act} s', t \mid\mid_{Act} t') = \mathbf{P}_\Delta(s,t) \cdot \mathbf{P}'_\Delta(s',t') = \mathbf{Q}(s,t)\mathbf{Q}'(s',t')\Delta^2 + o(\Delta)$$

   It holds now $\lim_{\Delta \to 0} \mathbf{P}^{\mid\mid}_\Delta(s \mid\mid_{Act} s', t \mid\mid_{Act} t')/\Delta = 0$. By definition of $\mid\mid_\emptyset$ we also have $\mathbf{Q}^{\mid\mid}(s \mid\mid_\emptyset s', t \mid\mid_\emptyset t') = 0$.
2. Now we consider the case $s = t$ and $s' \neq t'$. Under this assumption we have that $\mathbf{Q}^{\mid\mid}(s \mid\mid_\emptyset s', t \mid\mid_\emptyset t') = \mathbf{Q}'(s',t')$, and moreover, $\mathbf{P}_\Delta(s,s) = 1 + \mathbf{Q}(s,s)\Delta + o(\Delta)$. The rest can be shown similarly as previous case.
3. Finally we consider the case $s = t$ and $s' = t'$. In this case we have

$$\begin{aligned}\mathbf{P}^{\mid\mid}_\Delta(s \mid\mid_{Act} s', s \mid\mid_{Act} s') &= (1 + \mathbf{Q}(s,s)\Delta + o(\Delta)) \cdot (1 + \mathbf{Q}'(s',s')\Delta + o(\Delta)) \\ &= 1 + (\mathbf{Q}(s,s) + \mathbf{Q}'(s',s'))\Delta + o(\Delta)\end{aligned}$$

   Thus: $\lim_{\Delta \to 0}(\mathbf{P}^{\mid\mid}_\Delta(s \mid\mid_{Act} s', s \mid\mid_{Act} s') - 1)/\Delta = \mathbf{Q}(s,s) + \mathbf{Q}'(s',s')$, which is exactly $\mathbf{Q}^{\mid\mid}(s \mid\mid_\emptyset s', s \mid\mid_\emptyset s')$.

The above lemma derives the interleaving semantics for CTMCs through the limiting behaviour of their discretised DTMCs evolving synchronously.

## 5 Simulations and Bisimulations

We now discuss equivalences and preorders for *MA* and submodels thereof. We first introduce a notation that makes our further discussion more compact, at the price of mildly reduced readability. It enables a uniform treatment of immediate and timed transitions. In doing so, we introduce the special action $\chi(r)$ to denote the exit $r$ rate of a state. Moreover, we let $Act^\chi := Act \cup \{\chi(r) \mid r \in \mathbb{R}_{\geq 0}\}$, and $\alpha, \beta, \ldots$ range over this set.

**Definition 3.** *Let MA $= (S, Act, \rightarrow, \twoheadrightarrow, s_o)$ be an MA. Let $E \in S$ and $\alpha \in Act^\chi$. We write $E \xrightarrow{\alpha} \mu$ if*

- $E \xrightarrow{\alpha} \mu \wedge \alpha \in Act$ *or*
- $E\downarrow \wedge r = \mathrm{rate}(E) \wedge \alpha = \chi(r) \wedge \mu = \mathbf{P}(E, \cdot)$.

### 5.1 Strong Simulations and Bisimulation

*Strong Simulations.* Strong simulations for DTMCs were originally introduced [31] using the concept of *weight functions*. Weight function have since become a standard way of formalising specification and refinement relations between probabilistic processes. Given distributions $\mu, \mu'$ and a relation $R$, a weight function $\delta$ requires assigning weights in $[0, 1]$ for every pair of states in $R$, such that $\mu(s) = \sum_{s' \in S} \delta(s, s')$ and symmetrically $\mu'(s') = \sum_{s \in S} \delta(s, s')$. Owed to the need of assigning these weights, many proofs need human ingenuity and become certainly nontrivial. This is felt, for instance, in the proof of transitivity of strong similarity. As a consequence, a few equivalent reformulations of weight functions have emerged. One of them is based on computing the *maximum flow* on a particular network constructed out of $\mu$, $\mu'$ and $R$ [2]. This has been successfully exploited in decision procedures for various simulation preorders [2,52,21]. Below, we use another, rather recent, reformulation [51,21]: the existence of a weight function for distributions $\mu, \mu'$ with respect to $R$ is equivalent to the statement that, $\mu(A) \leq \mu(R(A))$ for every set $A$ of states. Here, we write $R(A) := \{s' \mid (s, s') \in R \ \wedge \ s \in A\}$. This alternative characterisation of weight function provides a very intuitive interpretation: for every set $A$, $\mu'$ assigns a higher probability to the related set $R(A)$ relative to $\mu$. The inspiration for this definition stems from [19], in which strong simulation relations are, in addition, required to be preorders.

**Definition 4 (Strong Simulations).** *Let MA $= (S, Act, \dashrightarrow, \twoheadrightarrow, s_o)$ be an MA. Let $\mathcal{R}$ be a binary relation on $S$. Then, $\mathcal{R}$ is a strong simulation iff $E\mathcal{R}F$ implies:*

1. *for all $\alpha \in Act$: $E \xrightarrow{\alpha} \mu$ implies $F \xrightarrow{\alpha} \mu'$ for some distribution $\mu'$ such that $\mu(A) \leq \mu'(R(A))$ for all $A \subseteq S$, and*
2. *for all $r \in \mathbb{R}_{\geq 0}$ : $E \xrightarrow{\chi(r)} \mu$ implies $F \xrightarrow{\chi(r')} \mu'$ for some distribution $\mu'$ and $r' \in \mathbb{R}_{\geq 0}$ such that $r \leq r'$ and $\mu(A) \leq \mu'(R(A))$ for all $A \subseteq S$.*

State $F$ strongly simulates $E$, written $E \precsim F$, if $(E, F)$ is contained in some strong simulation.

1. On *labelled transitions systems* $\precsim$ coincides with standard strong similarity [42,1].
2. On *discrete-time Markov chains* $\precsim$ coincides with strong similarity [31].
3. On *continuous-time Markov chains* $\precsim$ coincides with strong similarity [5].
4. On *probabilistic automata* $\precsim$ coincides with strong similarity [46].
5. On *interactive Markov chains* $\precsim$ coincides with strong similarity [32] if strengthening $r \leq r'$ to $r = r'$.

*Strong Bisimulations.* Strong bisimilarity in its diverse flavours is the most prominent equivalence relation for probabilistic models. For *MA*, the obvious combination of strong bisimilarity for IMC and strong bisimilarity for PA can be phrased as follows:

**Definition 5.** *Let MA $= (S, Act, \dashrightarrow, \twoheadrightarrow, s_o)$ be an MA. Let $\mathcal{R}$ be an equivalence relation on $S$. Then, $\mathcal{R}$ is a strong bisimulation iff $E\mathcal{R}F$ implies for all $\alpha \in Act^{\chi}$: $E \xrightarrow{\alpha} \mu$ implies $F \xrightarrow{\alpha} \mu'$ with $\mu(C) = \mu'(C)$ for all $C \in S/\mathcal{R}$.*

Two states $E$ and $F$ are strongly bisimilar, written $E \sim F$, if $(E, F)$ is contained in some strong bisimulation.

1. On *labelled transitions systems* $\sim$ is strong bisimilarity [42,40].
2. On *discrete-time Markov chains* $\sim$ coincides with strong bisimilarity [35] and corresponds to lumpability [33].
3. On *continuous-time Markov chains* $\sim$ coincides with lumping equivalence [26].
4. On *probabilistic automata* $\sim$ coincides with strong bisimilarity [46].
5. On *interactive Markov chains* $\sim$ coincides with strong bisimilarity [25].

For PA, coarser relations than strong bisimilarity and strong similarity – still treating internal transitions as ordinary transitions – are established based on the concept of combined transitions. The resulting relations are called *strong probabilistic (bi-)similarities* [46,45]. They can also be defined directly in our setting, by replacing $F \xrightarrow{a} \mu'$ by a convex combination of several $a$-labelled transitions starting in $F$. Details are left to the interested reader.

### 5.2 Weak Transitions

Weak transitions for probabilistic systems have been defined in the literature via probabilistic executions in [44], trees [20], or infinite sums [18]. We adopt the tree notation here. The material presented below concerning weak transitions provides no innovation over the classical treatment, it is included for the benefit of the reader.

We consider in the following $S \times \mathbb{R}_{\geq 0} \times Act^{\chi} \cup \{\epsilon\}$-labelled trees. Briefly, a node in such trees is labelled by the corresponding state, probability of reaching this node, and the chosen action (including the special action for timed transitions) to proceed. For a node $\sigma$ we write $\mathit{Sta}_t(\sigma)$ for the first component of $t(\sigma)$, $\mathit{Prob}_t(\sigma)$ for the second component of $t(\sigma)$ and $\mathit{Act}_t(\sigma)$ for the third component of $t(\sigma)$.

**Definition 6.** *Let MA* $= (S, Act, \rightarrowtriangle, \twoheadrightarrow, s_o)$ *be an MA. A* (weak) transition tree $\mathcal{T}$ *is a* $S \times \mathbb{R}_{\geq 0} \times Act^{\chi} \cup \{\epsilon\}$*-labelled tree that satisfies the following condition:*

1. $0 < \mathit{Prob}_{\mathcal{T}}(\varepsilon) \leq 1$
2. $\forall \sigma \in \mathit{Leaf}(\mathcal{T}) : \mathit{Act}_{\mathcal{T}}(\sigma) = \epsilon$.
3. $\forall \sigma \in \mathit{Inner}(\mathcal{T}) \setminus \mathit{Leaf}(\mathcal{T}) : \exists \mu : \mathit{Sta}_{\mathcal{T}}(\sigma) \xrightarrow{\mathit{Act}_{\mathcal{T}}(\sigma)} \mu$ *and*
$$\mathit{Prob}_{\mathcal{T}}(\sigma) \cdot \mu = [\![ (\mathit{Sta}_{\mathcal{T}}(\sigma'), \mathit{Prob}_{\mathcal{T}}(\sigma')) \mid \sigma' \in \mathit{Children}_{\mathcal{T}}(\sigma) ]\!]$$
4. $\sum_{\sigma \in \mathit{Leaf}} \mathit{Prob}_{\mathcal{T}}(\sigma) = \mathit{Prob}_{\mathcal{T}}(\varepsilon)$.

*We call the tree weak, if* $\mathit{Prob}_{\mathcal{T}}(\varepsilon) < 1$.

Restricting $Act^{\chi}$ to $Act$, a transition tree $\mathcal{T}$ corresponds to a probabilistic execution fragment: it starts from $\mathit{Sta}_{\mathcal{T}}(\epsilon)$, and resolves the non-deterministic choice by executing the action $\mathit{Act}_{\mathcal{T}}(\sigma)$ at the inner node $\sigma$. The second label of $\sigma$ is then the probability of reaching $\mathit{Sta}_{\mathcal{T}}(\sigma)$, starting from $\mathit{Sta}_{\mathcal{T}}(\epsilon)$ and following the selected actions. If in a node $\sigma$ the timed transition is chosen, the third label $\mathit{Act}_{\mathcal{T}}(\sigma) \in \mathbb{R}_{\geq 0}$ represents the

exit rate of $Sta_{\mathcal{T}}(\sigma)$. In this case, a child $\sigma'$ is reached with $Prob_{\mathcal{T}}(\sigma)$ times the discrete branching probability $\mathbf{P}(Sta_{\mathcal{T}}(\sigma), Sta_{\mathcal{T}}(\sigma'))$.

An *internal transition tree* $\mathcal{T}$ is a transition tree where each $Act_{\mathcal{T}}(\sigma)$ is either $\tau$ or $\epsilon$. Let $\mathcal{T}$ be a transition tree. Then the subdistribution associated with $\mathcal{T}$, denoted by $\mu_{\mathcal{T}}$, is defined as

$$\mu_{\mathcal{T}} \stackrel{\text{def}}{=} \bigoplus_{\sigma \in Leaf_{\mathcal{T}}} [\![(Sta_{\mathcal{T}}(\sigma), Prob_{\mathcal{T}}(\sigma))]\!] .$$

We say subdistribution $\mu_{\mathcal{T}}$ is *induced* by $\mathcal{T}$. Obviously, $\mu_{\mathcal{T}}$ is a full distribution if we have $Prob_{\mathcal{T}}(\varepsilon) = 1$. With the above definitions we are now able to express weak transitions:

**Definition 7.** *For $E \in S$ and $\mu$ a full distribution we write*

- $E \Longrightarrow \mu$ *if $\mu$ is induced by some internal transition tree $\mathcal{T}$ with $Sta_{\mathcal{T}}(\varepsilon) = E$.*
- $E \stackrel{\alpha}{\Longrightarrow} \mu$ *if $\mu$ is induced by some transition tree $\mathcal{T}$ with $Sta_{\mathcal{T}}(\varepsilon) = E$, where on every maximal path from the root at least one node $\sigma$ is labelled $Act_{\mathcal{T}}(\sigma) = \alpha$. In case that $\alpha \neq \tau$, then there must be exactly one such node on every maximal path. And all other inner nodes must be labelled by $\tau$.*
- $E \stackrel{\hat{\alpha}}{\Longrightarrow} \mu$ *if $\alpha = \tau$ and $E \Longrightarrow \mu$ or $E \stackrel{\alpha}{\Longrightarrow} \mu$.*

*For all three transition relations we say that the transition tree that induces $\mu$ also* induces the transition *to $\mu$.*

Note that $E \Longrightarrow \Delta_E$ and $E \stackrel{\hat{\tau}}{\Longrightarrow} \Delta_E$ holds independently of the actual transitions $E$ can perform, whereas $E \stackrel{\tau}{\Longrightarrow} \Delta_E$ only holds if $E \stackrel{\tau}{\rightarrowtriangle} \Delta_E$. For all $\alpha \neq \tau$, $E \stackrel{\hat{\alpha}}{\Longrightarrow} \mu$ is identical to $E \stackrel{\alpha}{\Longrightarrow} \mu$. Below we define the notion of *combined transitions* [44], which arise as convex combination of a set of transitions with the same label, including the label representing timed transitions.

**Definition 8.** *We write $E \stackrel{\alpha}{\Longrightarrow}_C \mu$, if $\alpha \in Act^{\chi}$ and there is a finite indexed set $\{(c_i, \mu_i)\}_{i \in I}$ of pairs of positive real valued weights and distributions such that $E \stackrel{\alpha}{\Longrightarrow} \mu_i$ for each $i \in I$ and $\sum_{i \in I} c_i = 1$ and $\mu = \bigoplus_{i \in I} c_i \mu_i$.*

We say that $E \stackrel{\alpha}{\Longrightarrow}_C \mu$ is *justified* by the set $\{(c_i, \mu_i)\}_{i \in I}$. Transitions relations from states to distributions can be generalised to take (sub)distributions $\mu$ to (sub)distributions, by weighting the result distribution of the transition of each element $E \in Supp(\mu)$ by $\mu(E)$.

**Definition 9.** *Let $\rightsquigarrow \in \left\{\stackrel{\hat{}}{\Longrightarrow}, \Longrightarrow, \rightarrowtriangle, \twoheadrightarrow, \longrightarrow\right\}$. Then, we write $\mu \rightsquigarrow \gamma$ if $\gamma = \bigoplus_{s_i \in Supp(\mu)} \mu(s_i)\mu_i$, where $s_i \rightsquigarrow \mu_i$ holds for all $s_i \in Supp(\mu)$.*

We say that $\mu \rightsquigarrow \gamma$ is *justified* by the transitions $s_i \rightsquigarrow \mu_i$.

### 5.3 Weak Simulations and Bisimulations over Subdistributions

Weak simulations and bisimulations are means to abstract from internal transitions and sequences thereof. In our setting this means that we intend to fuse distributions that arise from sequences of internal transitions in an *MA* into single, accumulated distributions.

*Relating distributions.* Bisimulations for PA and IMC have in the past been defined as relations on states, akin to bisimulations for LTS. The latter reside in a single state at every point in time during their execution, thus bisimulation relations on states are all that is needed to capture their behaviour adequately. In contrast, a stochastic system resides – for every point in time – in a certain state with a certain probability. The system behaviour is thus not correctly determined by the states occupied, but instead by probability distributions over states in which a system can reside at any point in time. It thus seems natural to define bisimulation relation as relations on distributions over states.

Several simulation relations for PA adopt this approach [46,44,48] in an asymmetric way, simulating a state by a distribution over states. Among these relations, probabilistic forward similarity [44] is the coarsest. We denote it by $\leq_{fwd}$. In Figure 2, state $u$ is forward simulated by state $v$ and vice versa. In an asymmetric way, $\leq_{fwd}$ achieves exactly what we intend to achieve for *MA*: we aim at fusing distribution along internal immediate transition sequences. It is however not obvious how to extend the definition of forward simulation, which relates states to distributions, to a *bisimulation* setting, which should then relate distributions to distributions. Even partially state-based approaches seem to fail, since in order to equate $u$ and $v$, the two distributions $[\![(v', 1)]\!]$ and $[\![(E, \frac{1}{3}), (F, \frac{2}{3})]\!]$ must be considered equivalent. However, from a state-based point of view, the two distributions must be different, assuming that $E$ and $F$ represent states of different equivalence classes, since neither $E$ nor $F$ alone can then be equivalent to state $v'$ .

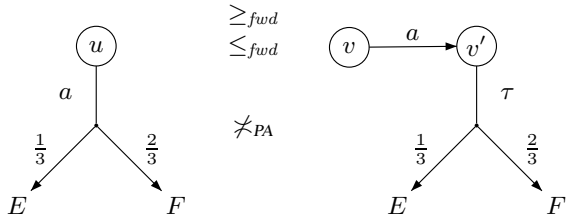


**Fig. 2.** Probabilistic forward simulation versus probabilistic weak bisimulation

We instead advocate a bisimulation-over-distribution approach to define a notion of weak bisimilarity that satisfies our demands [22]. In the sequel, we reiterate and rephrase several interesting aspects of this approach, and characterise the semantical relationship between weak bisimilarity and the standard bisimilarities of PA and IMC.

*Naïve Weak Bisimulation over States.* In the following, we will show that the standard bisimulations of PA and IMC can easily by cast as relations over distributions. In favour of a concise presentation, we will not consider the standard bisimulations for PAs and IMC separately, but only investigate a relation $\asymp$, which we define as a direct combination of IMC and PA weak bisimilarity, such that on the IMC submodel we obtain IMC weak bisimilarity [25], and on the PA submodel, we obtain (a divergence sensitive variation) of PA weak bisimilarity, namely stable weak bisimilarity. The variation in the latter case is owed to the maximal progress assumption, inherited from IMC and necessary for general *MA*s. This has, however, no influence on the technical development.

We have first introduced this relation in [22, Def. 10], as a naïve approach of defining a suitable weak equivalence for *MA*s. As we have argued there, this relation is, however, not suitable to achieve the intended effect of fusing internal transitions. We thus call this relation *naïve weak bisimulation*.

**Definition 10 (Naïve Weak Bisimulation).** *Let MA* $= (S, Act, \rightarrowtriangle, \twoheadrightarrow, s_o)$ *be a MA. For two states* $s, s' \in S$*,* $s \asymp s'$ *holds if* $(s, s') \in \mathcal{E}$ *for some equivalence relation* $\mathcal{E}$ *on* $S$ *for which* $E\mathcal{E}F$ *implies for all* $\alpha \in Act^{\chi}$ *and for all equivalence classes* $C$ *of* $\mathcal{E}$*,* $E \xrightarrow{\alpha} \mu$ *implies* $F \overset{\hat{\alpha}}{\Longrightarrow}_C \gamma$ *for some* $\gamma$ *and* $\forall C \in S/\mathcal{E} : \mu(C) = \gamma(C)$*.*

*Weak (Bi-)simulations over Subdistributions.* We will now introduce two weak bisimulation – and also two weak simulation – relations that relate distributions (or subdistributions, to be precise). One of them is *weak bisimulation* for Markov automata as introduced in [22, Def. 11], the other bisimulation appears new. We will show that this new bisimulation relation on distributions induces a bisimilarity on states that coincides with naïve weak bisimulation, which itself is defined directly over states. This strong connection bridges between the state-based and distribution-based approach and allows us to make precise their relationship. We call the new relation *semi-weak* bisimulation, since it is *weak*, meaning partially oblivious to internal transitions, but yet finer than weak bisimulation for Markov automata.

Both relations rely on the idea of equating subdistributions exactly when they can be split into subdistributions, such that for each component in one splitting there exists a related component of the other splitting that exhibits identical behaviour, and vice versa. Remarkably, the definitions only differ in one specific point. For *semi-weak* (bi)simulation, splittings need to be immediately related to match their behaviour. For *weak* (bi)simulation, we relax the conditions such that it suffices if equated distributions are able to reach distributions that can then be split suitably by internal transitions. To make explicit that the relations only differ in the way subdistributions are split, we define two sets of possible splittings of a subdistribution.

- For weak (bi)simulation, we use a set

$$\mathit{split}(\mu) = \{(\mu_1, \mu_2) \mid \exists \mu' : \mu \Longrightarrow_C \mu' \wedge \mu' = \mu_1 \oplus \mu_2\}$$

  where each splitting of an internal successor subdistribution of $\mu$ is a valid splitting.
- For semi-weak (bi)simulation, we use a more restricted set

$$\mathit{split}^{\circ}(\mu) = \{(\mu_1, \mu_2) \mid \mu = \mu_1 \oplus \mu_2\}$$

  where only direct splittings of $\mu$ are valid splittings.

Since weak and semi-weak bisimulation only differ in this one point, we will define them simultaneously in one definition. In what follows, the expression $\mathit{split}^{(\circ)}$ needs to be replaced by *split* in order to obtain weak bisimulation. Semi-weak bisimulation is obtained by replacing it by $\mathit{split}^{\circ}$.

**Definition 11 (Weak Bisimulations).** *A relation* $\mathcal{R}$ *on subdistributions over* $S$ *is called a* (semi-)weak bisimulation *iff whenever* $\mu_1 \mathcal{R} \mu_2$ *then for all* $\alpha \in Act^{\chi}$*:* $|\mu_1| = |\mu_2|$ *and*

1. $\forall E \in Supp(\mu_1)$: $\exists {\mu_2}^g, {\mu_2}^s$: $({\mu_2}^g, {\mu_2}^s) \in split^{(\circ)}(\mu_2)$ *and*
   (i) $[\![(E, \mu_1(E))]\!]\ \mathcal{R}\ {\mu_2}^g$ *and* $(\mu_1 - E)\ \mathcal{R}\ {\mu_2}^s$
   (ii) *whenever* $E \xrightarrow{\alpha} \mu_1'$ *for some* $\mu_1'$ *then* ${\mu_2}^g \overset{\hat{\alpha}}{\Longrightarrow}_C \mu''$ *and* $(\mu_1(E) \cdot \mu_1')\ \mathcal{R}\ \mu''$
2. $\forall F \in Supp(\mu_2)$: $\exists {\mu_1}^g, {\mu_1}^s$: $({\mu_1}^g, {\mu_1}^s) \in split^{(\circ)}(\mu_1)$ *and*
   (i) ${\mu_1}^g\ \mathcal{R}\ [\![(F, \mu_1(F))]\!]$ *and* ${\mu_1}^s\ \mathcal{R}\ (\mu_2 - F)$
   (ii) *whenever* $F \xrightarrow{\alpha} \mu_2'$ *for some* $\mu_2'$ *then* ${\mu_1}^g \overset{\hat{\alpha}}{\Longrightarrow}_C \mu''$ *and* $\mu''\ \mathcal{R}\ (\mu_2(F) \cdot \mu_2')$

*Two subdistributions $\mu$ and $\gamma$ are (semi-)weak bisimilar, denoted by $\mu \approx^{(\circ)} \gamma$, if the pair $(\mu, \gamma)$ is contained in some (semi-)weak bisimulation.*

It is worth noting that the weak bisimilarity $\approx$ in the above definition is identical to [22, Def. 11]. We immediately obtain simulation relations by removing Condition 2.

**Definition 12 (Weak Simulations).** *A relation $\mathcal{R}$ on subdistributions over $S$ is called a* (semi-)weak simulation *iff whenever $\mu_1 \mathcal{R} \mu_2$ then for all $\alpha \in Act^{\chi}$: $|\mu_1| = |\mu_2|$ and*

$\forall E \in Supp(\mu_1)$: $\exists {\mu_2}^g, {\mu_2}^s$: $({\mu_2}^g, {\mu_2}^s) \in split^{(\circ)}(\mu_2)$ *and*

(i) $[\![(E, \mu_1(E))]\!]\ \mathcal{R}\ {\mu_2}^g$ *and* $(\mu_1 - E)\ \mathcal{R}\ {\mu_2}^s$
(ii) *whenever* $E \xrightarrow{\alpha} \mu_1'$ *for some* $\mu_1'$ *then* ${\mu_2}^g \overset{\hat{\alpha}}{\Longrightarrow}_C \mu''$ *and* $(\mu_1(E) \cdot \mu_1')\ \mathcal{R}\ \mu''$

*Two subdistributions $\mu$ and $\gamma$ are (semi-)weak similar, denoted by $\mu \precapprox^{(\circ)} \gamma$, if the pair $(\mu, \gamma)$ is contained in some (semi-)weak simulation.*

It is not obvious that these relations are indeed equivalence relations and preorders, respectively. Reflexivity and symmetry is straightforward. The latter holds, because the union of two (semi-)weak bisimulations is again a (semi-)weak bisimulation. Since the pioneering work of Larsen and Skou [35], it has become a standard for bisimilarity notions defined in the stochastic setting, to presuppose the bisimulations to be equivalence relations on states. Because this property is not closed under union, several otherwise standard properties become difficult to establish. Owed to the distribution perspective on these relations illustrated above, the equivalence relation presupposition can be dropped, simplifying an easy exercise. Only transitivity needs a technical and involved proof. The proof for $\approx$ can be found in [22,23]. The crucial idea for this proof is that we can define (bi)simulation relations up-to-splitting. We refer the reader to [22] for further details. The proof for $\approx^{\circ}$ follows exactly the lines of that proof, but needs to distinguish fewer cases.

**Lemma 2.** *$\precapprox$ and $\precapprox^{\circ}$ are preorders, $\approx$ and $\approx^{\circ}$ are equivalence relations.*

It is apparent that $\approx$ and $\precapprox$ are weaker notions than $\approx^{\circ}$ and $\precapprox^{\circ}$ respectively:

**Theorem 1.** $\approx^{\circ} \subseteq \approx$ *and* $\precapprox^{\circ} \subseteq \precapprox$

The relations defined above relate subdistributions, but they induce relations on states in the obvious way: We call two states $E, F$ (semi-)weak bisimilar, denoted by $E \approx_{\Delta}^{(\circ)} F$, if $\Delta_E \approx^{(\circ)} \Delta_F$. Analogously, we call two states $E, F$ (semi-)weak similar, denoted by $E \precapprox_{\Delta}^{(\circ)} F$, if $\Delta_E \precapprox_{\Delta}^{(\circ)} \Delta_F$.

In the following we establish that $\asymp$ and $\approx_{\Delta}^{\circ}$ coincide. Since $\asymp$ is the naïve (state-based) integration of PA and IMC weak bisimulation, this fact provides insight into the twist achieved by moving from semi-weak to weak formulation.

**Theorem 2.** $\asymp \; = \; \approx^{\circ}_{\Delta}$

*Proof.* We first prove that the lifting of semi-weak bisimilarity to states, $\approx^{\circ}_{\Delta}$, is also a state-based bisimulation in the sense of Definition 10. The crucial point in this proof is the claim, that $\mu \approx^{\circ} \gamma$ implies $\forall C \in S/{\approx^{\circ}_{\Delta}} : \mu(C) = \gamma(C)$, since then the conditions of Definition 10 follow immediately. To see the claim, note that $\approx^{\circ}$ itself is a semi-weak bisimulation. Then, by repeated application of the left hand side of clause $(i)$ in the definition of $\approx^{\circ}$, we can split $\gamma$ into a family of subdistribution $\{\gamma_E\}_{E \in Supp(\mu)}$, such that every $E \in Supp(\mu)$ is matched by one of these distributions, and $\mu(E)\Delta_E \approx^{\circ} \gamma_E$ holds. In turn, we can split $\mu(E)\Delta_E$ into a family $\left\{\mu^E_F\right\}_{F \in Supp(\gamma_E)}$ accordingly, such that each state $F \in Supp(\gamma_E)$ is matched by the subdistribution $\mu^E_F$ , satisfying $\mu^E_F \approx^{\circ} \gamma_E(F)\Delta_F$. Every subdistributions $\mu^E_F$ must be of the form $\gamma_E(F)\Delta_E$. Hence, we know that $\gamma_E(F)\Delta_E \approx^{\circ} \gamma_E(F)\Delta_F$. In total, we have split $\mu$ and $\gamma$ into sets of subdistributions, such that there is a total matching of subdistribution of one set with subdistributions of the other set. Matched subdistributions have the same size and the single elements of their supports are equivalent up to $\approx^{\circ}_{\Delta}$. From here we can immediately conclude that $\forall C \in S/{\approx^{\circ}_{\Delta}} : \mu(C) = \gamma(C)$ holds.

For the other direction we show that the relation

$$\mathcal{R} = \{(\mu, \gamma) \mid \forall C \in S/{\asymp} : \mu(C) = \gamma(C)\}$$

is a semi-weak bisimulation. Then, whenever $E \asymp F$, the pair $(\Delta_E, \Delta_F)$ is contained in the semi-weak bisimulation $\mathcal{R}$, which implies $E \approx^{\circ}_{\Delta} F$. Let us consider an arbitrary pair $(\mu, \gamma) \in \mathcal{R}$. By symmetry it suffices to check the necessary conditions for an arbitrary $E \in Supp(\mu)$. Let $C$ be the equivalence class of $\asymp$ containing $E$. Since $\mu(C) = \gamma(C)$, there exists a splitting $\gamma^g \oplus \gamma^s$ of $\gamma$ with $Supp(\gamma^g) = \{F_1, \ldots, F_k\}$ and $F_i \asymp E$ for each $F_i$, and furthermore, $\forall C \in S/{\asymp} : \gamma^s(C) = (\mu - E)(C)$. Hence Condition $(i)$ is satisfied. Whenever $E \xrightarrow{\alpha} \mu'$, following Condition $(ii)$, then for each $F_i$ we immediately deduce from $E \asymp F_i$ that $F_i \overset{\alpha}{\Longrightarrow}_C \gamma_{F_i}$ and $\forall C \in S/{\asymp} : \mu'(C) = \gamma_{F_i}(C)$. Let us set $\rho := \bigoplus_{i=1\ldots k} \gamma_{F_i}$. It is then straightforward to show that in total $\gamma^g \overset{\alpha}{\Longrightarrow}_C \rho$ and that $\forall C \in S/{\asymp} : \mu(E) \cdot \mu'(C) = \rho(C)$. By the choice of $\mathcal{R}$ this immediately implies $(\mu(E)\mu', \rho) \in \mathcal{R}$, which suffices to establish Condition $(ii)$. □

Just like $\asymp$, most existing weak relations for systems with probabilistic or stochastic timed transitions can be recast as relations on distributions, which can be formulated as slight adaptations of $\approx^{\circ}$ and $\precsim^{\circ}$, respectively. So, with $\approx^{\circ}$ and $\precsim^{\circ}$ at hand, the exact extra power of the distribution-based perspective, combined with distribution splitting, becomes apparent: $\approx^{\circ}$ and $\approx$ only differ in their use of *split*$^{\circ}(\mu)$ and *split*$(\mu)$, respectively. The latter allows additional internal transition sequences, and is the key to fuse distributions along sequences thereof. It is thus a natural generalisation comparable to the classical passage from strong transitions to weak transitions.

*Discussion.* We will now summarise the relationship between the respective standard notions of weak (bi-)similarity on the submodels and weak (bi-)similarity for Markov automata. Since all of these relations are defined over states, we will compare them to $\approx_{\Delta}$ and $\approx^{\circ}_{\Delta}$.

1. On *labelled transitions systems*, both $\approx_\Delta$ and $\approx^\circ_\Delta$ coincide with stable weak bisimilarity [49]. They both coincide with standard weak bisimilarity [40] if no infinite sequences of internal transitions appear in the LTS. This difference is inherited from IMC, and owed to the maximal progress assumption [28]. The same applies to $\precsim_\Delta$, $\precsim^\circ_\Delta$, and weak similarity on LTS [40].
2. On *discrete-time Markov chains*, $\approx_\Delta$ and $\approx^\circ_\Delta$ coincide with weak bisimilarity [3,5]. We claim that $\precsim_\Delta$ and $\precsim^\circ_\Delta$ can be adapted such that they both coincide with weak similarity for state-labelled DTMC [5].
3. On *continuous-time Markov chains*, $\approx_\Delta$ and $\approx^\circ_\Delta$ coincide with lumping equivalence [26], due to the fact that our weak transitions do not affect timed transitions. For this reason we see no obvious way to adapt $\precsim_\Delta$ and $\precsim^\circ_\Delta$ such that they match weak similarity for state-labelled CTMC [5].
4. On *probabilistic automata*, $\approx^\circ_\Delta$ coincides with weak bisimilarity [44], if restricting to models without infinite sequences of internal transition. This slight restriction is again a legacy of the maximal progress assumption. This technical discrepancy carries over to all other relations defined on PA. If instead we adapt the definition and remove the stability condition, the adapted version of $\approx^\circ_\Delta$ and weak bisimilarity on PA [44] coincide. The same holds for $\precsim^\circ_\Delta$ and weak similarity for PA [44]. Remarkably, $\precsim_\Delta$ and probabilistic forward similarity $\leq_{fwd}$ [44] coincide.
5. On *interactive Markov chains* $\approx_\Delta$ and $\approx^\circ_\Delta$ coincide with weak bisimulation [25]. A weaker variant is found in [10]. To the best of our knowledge no weak similarity relations for IMC have been introduced in the literature so far, so the one jointly induced by $\precsim$ and $\precsim^\circ$ is new.

The fact that on PA $\precsim_\Delta$ and $\leq_{fwd}$ agree is especially interesting, since a bisimulation variant of this relation was not known to date, but is now at hand with $\approx_\Delta$. Furthermore, $\approx_\Delta$ has a selection of distinguishing properties. We refer the reader to [22] for details. We mention only briefly, that $\approx_\Delta$ is a congruence with respect to parallel composition. The congruence property can be established for other standard process algebraic operators – with the usual root condition being needed to arrive at a congruence for non-deterministic choice.

We finally want to correct our claim [22], that a reformulation of PA weak bisimilarity as a relation on distribution would not be compositional with respect to subdistribution composition, now turns out to be wrong. It is easy to show that $\approx^\circ$ is indeed compositional with respect to this operator, and since on PA, $\approx^\circ$ coincides with PA weak bisimilarity (except for divergence behaviour), this also holds for PA weak bisimilarity.

## 6 Conclusions

This paper has tried to provide insight into the foundational aspects of Markov automata, a model that integrates probabilistic automata and interactive Markov chains. We have laid out the principal ingredients of a compositional theory for *MA*, and have discussed how a lifting of relations to (sub)distributions, together with the notion of distribution splitting, enables us to cast a variety of existing simulations and bisimulations in a uniform setting, making subtle differences and semantic choices apparent.

Markov automata target a domain of concurrency modelling and evaluation, where designers find it adequate to work with durations that are memoryless, and need to represent branching probabilities as well as concurrency in a convenient manner. In this area, GSPNs have seen broad applicability, but, as we have highlighted, only with incomplete semantic understanding. The *MA* model changes the picture, it can serve as a semantic foundation for GSPN, and, since it is compositional in a strict sense, we think it is the nucleus for a fully compositional and usable language for this modelling domain. Noteworthy, the *MA* model is – just like IMC – rich enough to allow for non-exponential distributions, namely by approximating them as phase-type distributions [27], albeit at the price of a state space increase.

PA as well as IMC are supported by mature software tools, PRISM [29] and CADP [16]. We are eagerly exploring possibilities to arrive at tool support for the analysis of *MA*.

In this paper, we have restricted our attention to finite and finitely-branching models. It remains for further work to establish the results of this paper in a setting with (un)countably many states or transitions.

**Acknowledgements.** Christian Eisentraut and Holger Hermanns are supported by the DFG as part of the Transregional Collaborative Research Centre SFB/TR 14 *AVACS*, by the European Community's Seventh Framework Programme under grant agreement n$^{o}$ 214755 *QUASIMODO*, and by the NWO-DFG bilateral project *ROCKS*. The work of Lijun Zhang has received partial support from *MT-LAB*, a VKR Centre of Excellence.

## References

1. Abadi, M., Lamport, L.: The existence of refinement mappings. Theoretrical Computer Science 82(2), 253–284 (1991)
2. Baier, C., Engelen, B., Majster-Cederbaum, M.E.: Deciding bisimilarity and similarity for probabilistic processes. J. Comput. Syst. Sci. 60(1), 187–231 (2000)
3. Baier, C., Hermanns, H.: Weak bisimulation for fully probabilistic processes. In: Grumberg, O. (ed.) CAV 1997. LNCS, vol. 1254, pp. 119–130. Springer, Heidelberg (1997)
4. Baier, C., Katoen, J.-P.: Principles of Model Checking, 1st edn. MIT Press, Cambridge (2008)
5. Baier, C., Katoen, J.-P., Hermanns, H., Wolf, V.: Comparative branching-time semantics for Markov chains. Information and Computation 200(2), 149–214 (2005)
6. Balbo, G.: Introduction to generalized stochastic Petri nets. In: Bernardo, M., Hillston, J. (eds.) SFM 2007. LNCS, vol. 4486, pp. 83–131. Springer, Heidelberg (2007)
7. Bandini, E., Segala, R.: Axiomatizations for probabilistic bisimulation. In: Orejas, F., Spirakis, P.G., van Leeuwen, J. (eds.) ICALP 2001. LNCS, vol. 2076, pp. 370–381. Springer, Heidelberg (2001)
8. Bergstra, J.A., Ponse, A., Smolka, S.A.: Handbook of Process Algebra. Elsevier, Amsterdam (2001)
9. Bolognesi, T., Brinksma, E.: Introduction to the iso specification language lotos. Computer Networks 14, 25–59 (1987)
10. Bravetti, M.: Revisiting interactive markov chains. Electr. Notes Theor. Comput. Sci. 68(5) (2002)
11. Busch, H., Sandmann, W., Wolf, V.: A numerical aggregation algorithm for the enzyme-catalyzed substrate conversion. In: Priami, C. (ed.) CMSB 2006. LNCS (LNBI), vol. 4210, pp. 298–311. Springer, Heidelberg (2006)

12. Cheung, L., Stoelinga, M., Vaandrager, F.W.: A testing scenario for probabilistic processes. J. ACM 54(6) (2007)
13. Chiola, G., Donatelli, S., Franceschinis, G.: GSPNs versus SPNs: What is the actual role of immediate transitions? In: PNPM, pp. 20–31. IEEE, Los Alamitos (1991)
14. Chiola, G., Franceschinis, G., Gaeta, R., Ribaudo, M.: GreatSPN 1.7: Graphical editor and analyzer for timed and stochastic Petri nets. Perf. Eval. 24(1-2), 47–68 (1995)
15. Chiola, G., Marsan, M.A., Balbo, G., Conte, G.: Generalized stochastic Petri nets: A definition at the net level and its implications. IEEE TSE 19(2), 89–107 (1993)
16. Coste, N., Garavel, H., Hermanns, H., Hersemeule, R., Thonnart, Y., Zidouni, M.: Quantitative evaluation in embedded system design: Validation of multiprocessor multithreaded architectures. In: DATE, pp. 88–89 (2008)
17. Courtney, T., Daly, D., Derisavi, S., Gaonkar, S., Griffith, M., Lam, V.V., Sanders, W.H.: The Möbius modeling environment: Recent developments. In: QEST, pp. 328–329. IEEE, Los Alamitos (2004)
18. Deng, Y., van Glabbeek, R.J., Hennessy, M., Morgan, C.: Testing finitary probabilistic processes. In: CONCUR, pp. 274–288 (2009)
19. Desharnais, J., Gupta, V., Jagadeesan, R., Panangaden, P.: Approximating labelled Markov processes. Inf. Comput. 184(1), 160–200 (2003)
20. Desharnais, J., Gupta, V., Jagadeesan, R., Panangaden, P.: Weak bisimulation is sound and complete for PCTL*. Inf. Comput. 208(2), 203–219 (2010)
21. Desharnais, J., Laviolette, F., Tracol, M.: Approximate analysis of probabilistic processes: Logic, simulation and games. In: QEST, pp. 264–273 (2008)
22. Eisentraut, C., Hermanns, H., Zhang, L.: On probabilistic automata in continuos time. In: LICS. IEEE, Los Alamitos (to appear, 2010)
23. Eisentraut, C., Hermanns, H., Zhang, L.: Probabilistic automata in continuous time. Reports of SFB/TR 14 AVACS 62, Saarland University (April 2010), `http://www.avacs.org`
24. Florio, V.D., Donatelli, S., Dondossola, G.: Flexible development of dependability services: An experience derived from energy automation systems. In: ECBS, pp. 86–93. IEEE, Los Alamitos (2002)
25. Hermanns, H. (ed.): Interactive Markov Chains. LNCS, vol. 2428, p. 57. Springer, Heidelberg (2002)
26. Hermanns, H., Herzog, U., Katoen, J.-P.: Process algebra for performance evaluation. Theoretical Computer Science 274(1-2), 43–87 (2002)
27. Hermanns, H., Katoen, J.-P.: Automated compositional Markov chain generation for a plain-old telephone system. Science of Comp. Progr. 36(1), 97–127 (2000)
28. Hermanns, H., Lohrey, M.: Priority and maximal progress are completely axioatisable (extended abstract). In: Sangiorgi, D., de Simone, R. (eds.) CONCUR 1998. LNCS, vol. 1466, pp. 237–252. Springer, Heidelberg (1998)
29. Hinton, A., Kwiatkowska, M.Z., Norman, G., Parker, D.: PRISM: A tool for automatic verification of probabilistic systems. In: Hermanns, H., Palsberg, J. (eds.) TACAS 2006. LNCS, vol. 3920, pp. 441–444. Springer, Heidelberg (2006)
30. Hoare, C.A.R.: Communicating Sequential Processes. Prentice Hall, Englewood Cliffs (1985)
31. Jonsson, B., Larsen, K.G.: Specification and refinement of probabilistic processes. In: LICS, pp. 266–277. IEEE, Los Alamitos (1991)
32. Katoen, J.-P., Klink, D., Neuhäußer, M.R.: Compositional abstraction for stochastic systems. In: Ouaknine, J., Vaandrager, F.W. (eds.) FORMATS 2009. LNCS, vol. 5813, pp. 195–211. Springer, Heidelberg (2009)
33. Kemeny, J.G., Snell, J.L., Knapp, A.W.: Denumerable Markov Chains, 2nd edn. Springer, Heidelberg (1976)

34. Klin, B., Sassone, V.: Structural operational semantics for stochastic process calculi. In: Amadio, R.M. (ed.) FOSSACS 2008. LNCS, vol. 4962, pp. 428–442. Springer, Heidelberg (2008)
35. Larsen, K.G., Skou, A.: Bisimulation through probabilistic testing. Information and Computation 94(1), 1–28 (1991)
36. Lynch, N.A., Tuttle, M.R.: Hierarchical correctness proofs for distributed algorithms. In: PODC, pp. 137–151 (1987)
37. Marsan, M.A., Balbo, G., Chiola, G., Conte, G.: Generalized stochastic Petri nets revisited: Random switches and priorities. In: PNPM, pp. 44–53. IEEE, Los Alamitos (1987)
38. Marsan, M.A., Balbo, G., Chiola, G., Conte, G., Donatelli, S., Franceschinis, G.: An introduction to generalized stochastic Petri nets. Microel. and Rel. 31(4), 699–725 (1991)
39. Marsan, M.A., Balbo, G., Conte, G., Donatelli, S., Franceschinis, G.: Modelling with Generalized Stochastic Petri Nets. John Wiley & Sons, Chichester (1995)
40. Milner, R.: Communication and Concurrency. Prentice Hall, Englewood Cliffs (1989)
41. Palamidessi, C.: Comparing the expressive power of the synchronous and asynchronous pi-calculi. Mathematical Structures in Computer Science 13(5), 685–719 (2003)
42. Park, D.M.R.: Concurrency and automata on infinite sequences. Theoretical Computer Science, pp. 167–183 (1981)
43. Sandmann, W., Wolf, V.: Computational probability for systems biology. In: Fisher, J. (ed.) FMSB 2008. LNCS (LNBI), vol. 5054, pp. 33–47. Springer, Heidelberg (2008)
44. Segala, R.: Modeling and Verification of Randomized Distributed Real-Time Systems. PhD thesis, Laboratory for Computer Science, Massachusetts Institute of Technology (1995)
45. Segala, R.: Probability and nondeterminism in operational models of concurrency. In: Baier, C., Hermanns, H. (eds.) CONCUR 2006. LNCS, vol. 4137, pp. 64–78. Springer, Heidelberg (2006)
46. Segala, R., Lynch, N.: Probabilistic simulations for probabilistic processes. Nordic Journal of Computing 2(2), 250–273 (1995)
47. Stewart, W.J.: Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modeling. Princeton University Press, Princeton (2009)
48. Stoelinga, M., Vaandrager, F.W.: Root contention in IEEE 1394. In: Katoen, J.-P. (ed.) AMAST-ARTS 1999, ARTS 1999, and AMAST-WS 1999. LNCS, vol. 1601, pp. 53–74. Springer, Heidelberg (1999)
49. van Glabbeek, R.J.: The linear time - branching time spectrum II. In: Best, E. (ed.) CONCUR 1993. LNCS, vol. 715, pp. 66–81. Springer, Heidelberg (1993)
50. Wolf, V.: Modelling of biochemical reactions by stochastic automata networks. ENTCS 171(2), 197–208 (2007)
51. Zhang, L.: Decision Algorithms for Probabilistic Simulations. PhD thesis, Universität des Saarlandes (2008)
52. Zhang, L., Hermanns, H., Eisenbrand, F., Jansen, D.N.: Flow faster: Efficient decision algorithms for probabilistic simulations. Logical Methods in Computer Science 4(4) (2008)