CONVERGENCE OF STOCHASTIC GENE NETWORKS TO HYBRID PIECEWISE DETERMINISTIC PROCESSES

BY A. CRUDU, A. DEBUSSCHE, A. MULLER AND O. RADULESCU

Université de Rennes 1, ENS Cachan Bretagne, Université de Lorraine and Université de Montpellier 2

We study the asymptotic behavior of multiscale stochastic gene networks using weak limits of Markov jump processes. Depending on the time and concentration scales of the system, we distinguish four types of limits: continuous piecewise deterministic processes (PDP) with switching, PDP with jumps in the continuous variables, averaged PDP, and PDP with singular switching. We justify rigorously the convergence for the four types of limits. The convergence results can be used to simplify the stochastic dynamics of gene network models arising in molecular biology.

1. Introduction. Modern molecular biology emphasizes the important role of the gene regulatory networks in the functioning of living organisms. Recent experimental advances in molecular biology show that many gene products do not follow deterministic dynamics and should be modeled as random variables [Kepler and Elston (2001), Kaufmann and van Oudenaarden (2007)].

Markov processes approaches to gene networks dynamics, originating from the pioneering ideas of Delbrück [Delbrück (1940)], capture diverse features of the experimentally observed expression variability, such as bursting [Cai, Friedman and Xie (2006)], various types of steady-state distributions of RNA and protein numbers [Kaern et al. (2005)], noise amplification or reduction by network propagation [Paulsson (2004), Warren, Tănase-Nicola and ten Wolde (2006)], clock de-synchronization [Barkai and Leibler (2000)], stochastic transitions in cellular memory storage circuits [Kaufmann et al. (2007)].

However, the study of the full Markov dynamics of biochemical networks is a difficult task. Even the simplest Markovian model, such as the production module of a single protein, involves tens of variables and biochemical reactions and an equivalent number of parameters [Kierzek, Zaim and Zielenkiewicz (2001), Krishna et al. (2005)]. The direct simulation of such models by the Stochastic Simulation Algorithm (SSA) [Gillespie (1976)] is extremely time consuming.

In order to increase computational efficiency, several accelerated simulation algorithms are hybrid and treat fast biochemical reactions as continuous variables

[Haseltine and Rawlings (2002), Alfonsi et al. (2005), Alfonsi et al. (2004)]. Similar approaches reducing fast reactions can be justified by diffusion approximations for Markov processes [Ball et al. (2006)].

A different hybrid approach is to distinguish between molecular species according to their abundances. Species in small amounts can be treated as discrete variables, whereas species in large amounts can be considered continuous. It has been proposed that the dynamics of gene networks with well-separated abundances can be well approximated by piecewise deterministic Markov processes [Radulescu, Muller and Crudu (2007), Crudu, Debussche and Radulescu (2009)]. Piecewise deterministic processes (PDP) are used in operational research in relation with optimal control and various technological applications [Boxma et al. (2005), Ghosh and Bagchi (2005), Bujorianu and Lygeros (2004)]. Their popularity in physical, chemical and biological sciences is also steadily increasing as they provide a natural framework to deal with intermittent phenomena in many contexts [Radulescu, Muller and Crudu (2007), Zeiser et al. (2008)].

By looking for the best PDP approximation of a stochastic network of biochemical reactions, and depending on the time scales of the reaction mechanism, we can distinguish several cases [Crudu, Debussche and Radulescu (2009)]:

- Continuous PDP with switching: continuous variables evolve according to ordinary differential equations. The trajectories of the continuous variables are continuous, but the differential equations depend on one or several discrete variables. These discrete variables are involved in Markov jumps.
- PDP with jumps in the continuous variables: the same as the previous case, but the continuous variables can jump as well as the discrete variables.
- Averaged PDP: some discrete variables have rapid transitions and can be averaged. The resulting approximation is an averaged PDP.
- Discontinuous PDP with singular switching: the continuous variable has two time scalings. The switch between the two regimes is commanded by a discrete variable. The rapid parts of the trajectory of the continuous variable can be approximated by discontinuities.

In this paper, we justify rigorously these approximations that were illustrated by models of stochastic gene expression in Crudu, Debussche and Radulescu (2009). More precisely, we present several theorems on the weak convergence of biochemical reactions processes toward piecewise deterministic processes of the types specified above. The resulting piecewise deterministic processes can be used for more efficient simulation algorithms, and also, in certain cases, can lead to analytic results for the stochastic behavior of gene networks. Higher-order approximations of multiscale stochastic chemical kinetics, corresponding to stochastic differential equations with jumps, are not discussed in this paper. These extensions of our results will be studied in a forthcoming article.

The structure of this article is as follows. In Section 2, we present the PDP, a useful theorem on the uniqueness of the solution of a martingale problem and

the Markov jump model for stochastic regulatory networks. The four remaining sections discuss the asymptotic behaviors of the models, corresponding to the four cases presented above.

2. Piecewise deterministic processes.

We begin with a brief description of Piecewise Deterministic Processes (PDP) and collect useful results on these. We do not consider PDPs in their full generality. The reader is referred to Davis (1993) for further results.

Standard conditions. In this article, a PDP taking values in $E = \mathbb{R}^n \times \mathbb{N}^d$ is a process $x_t = (y_t, \nu_t)$, determined by its three characteristics:

(1) For all $\nu \in \mathbb{N}^d$, a C^1 vector field in $\mathbb{R}^n$, denoted by F_ν, which determines a unique global flow $\phi_\nu(t, y)$ in $\mathbb{R}^n$ such that, for $t > 0$,

$$\frac{d}{dt}\phi_\nu(t, y) = F_\nu(\phi_\nu(t, y)), \qquad \phi_\nu(0, y) = y \qquad \forall y \in \mathbb{R}^n.$$

We also use the notation: $F(y, \nu) = F_\nu(y)$.

(2) A jump rate $\lambda : E \to \mathbb{R}^+$ such that, for each $x = (y, \nu) \in E$, there exists $\epsilon(x) > 0$ such that

$$\int_0^{\epsilon(x)} \lambda(\phi_\nu(t, y), \nu)\, dt < \infty.$$

(3) A transition measure $Q : E \to \mathcal{P}(E)$, $x \mapsto Q(\cdot; x)$, where $\mathcal{P}(E)$ denotes the set of probability measures on E. We assume that $Q(\{x\}; x) = 0$ for each $x \in E$.

From these standard conditions, a right-continuous sample path $\{x_t : t > 0\}$ starting at $x = (y, \nu) \in E$ may be constructed as follows. Define

$$x_t(\omega) := (\phi_\nu(t, y), \nu) \qquad \text{for } 0 \le t < T_1(\omega),$$

where $T_1(\omega)$ is the realization of the first jump time T_1, with the following distribution:

$$\mathbb{P}_x(T_1 > t) = \exp\left(-\int_0^t \lambda(\phi_\nu(s, y), \nu)\, ds\right) =: H(t, x), \qquad t \in \mathbb{R}^+.$$

We have then $x_{T_1^-(\omega)}(\omega) = (\phi_\nu(T_1(\omega), y), \nu)$, and the post-jump state $x_{T_1(\omega)}(\omega)$ has the distribution given by

$$\mathbb{P}_x(x_{T_1} \in A | T_1 = t) = Q(A; (\phi_\nu(t, y), \nu))$$

on the Borel sets A of E.

We then restart the process at $x_{T_1(\omega)}(\omega)$ and proceed recursively according to the same procedure to obtain a sequence of jump-time realizations $T_1(\omega), T_2(\omega), \ldots$. Between each of two consecutive jumps, $x_t(\omega)$ follows a deterministic motion, given by the flow corresponding to the vector field F.

Such a process x_t is called a PDP. The number of jumps that occur between the times 0 and t is denoted by

$$N_t(\omega) = \sum_k \mathbb{1}_{t \geq T_k}(\omega).$$

It can be shown that x_t is a strong Markov process with right-continuous, left-limited sample paths [see Davis (1993)]. The generator $\mathcal{A}$ of the process is formally given by

$$(1) \qquad \mathcal{A}f(x) = F_\nu(x) \cdot \nabla_y f(x) + \lambda(x) \int_E (f(z) - f(x)) Q(dz; x);$$

for each $x = (y, \nu) \in E$, we have denoted by ∇_y the gradient with respect to the variable $y \in \mathbb{R}^n$. We do not need a precise description of the domain of $\mathcal{A}$ and just note that $\mathcal{A}$ is well defined for $f \in \mathcal{E}$, the set of functions $f : E \to \mathbb{R}$ such that:

(E1) f is bounded,
(E2) for all $\nu \in \mathbb{N}^d$, $f(\cdot, \nu) \in C^1(\mathbb{R}^n)$,
(E3) its derivatives are bounded uniformly in E.

We restrict to such test functions which are sufficient for our analysis. However, the generator has a much larger domain as shown in Davis (1993), Section 26. In particular, in all the situations considered here, it can be applied to locally Lipschitz functions with polynomial growth.

For $f \in \mathcal{E}$, we denote by

$$(2) \qquad L_f = \sup_{y \in \mathbb{R}^n} \|D_y f\|_\infty = \sup_{(y,\nu) \in E} \|D_y f(y, \nu)\|$$

the Lipschitz constant of f with respect to the variable y.

The space $\mathcal{E}$ is a Banach space when endowed with the norm

$$(3) \qquad \|f\|_\mathcal{E} = \|f\|_\infty + L_f.$$

If Z is a Banach space, $\mathcal{B}_b(Z)$ is the set of bounded Borel-measurable functions on Z; $C_b^k(Z)$ is the set of C^k-differentiable functions on Z, such that the derivatives, until the kth order, are bounded; $\mathcal{C}_b(Z)$ is the set of bounded continuous functions on Z. Also $D(\mathbb{R}^+; Z)$ is the set of processes defined on $\mathbb{R}^+$ with right-continuous, left-limited sample paths defined on $\mathbb{R}^+$ and taking values in Z and $C(\mathbb{R}^+; Z)$ is the set of continuous processes defined on $\mathbb{R}^+$ and taking values in Z.

The PDPs considered in this paper will always satisfy the following property:

HYPOTHESIS 2.1. *The three characteristics of the PDP satisfy the standard conditions given above. The jump rate λ is C^1-differentiable with respect to the variable $y \in \mathbb{R}^n$. For every starting point $x = (y, v) \in E$ and $t \in \mathbb{R}^+$, we suppose $\mathbb{E}(N_t) < \infty$.*

REMARK 2.2. $\mathbb{E}(N_t) < \infty$ implies in particular that $T_k(\omega) \to \infty$ almost surely. This assumption is usually quite easy to check in applications, but it is hard to formulate general conditions under which it holds, because of the complicated interaction between F, λ and Q. It can be shown, for instance, that if λ is bounded, then $\mathbb{E}(N_t) < \infty$ [cf. Davis (1993)].

For some results, we need the following stronger property:

HYPOTHESIS 2.3. *The functions F, λ and $x \mapsto \lambda(x) \int_E f(z) Q(dz; x)$, with $f \in \mathcal{E}$, are bounded on E, C^1-differentiable with respect to the variable $y \in \mathbb{R}^n$ and their derivatives with respect to y are also bounded.*

REMARK 2.4. In Hypothesis 2.1, we assume that the jump rates and the vector fields F_v are C^1. Our result extends to the more general assumption that these functions are locally Lipschitz. This extension requires only technical complications and notations. We chose the framework of C^1 objects to lighten the presentation. We believe that it allows the reader to see more clearly the main ideas of our work. For instance, our choice allows to consider a set of test functions $\mathcal{E}$ consisting of C^1 functions. This simplifies the various convergence arguments. Also, if one wishes to work with locally Lipschitz functions, one has to modify in particular Hypothesis 2.3 and the proof of Theorem 2.5 in the Appendix.

When Hypothesis 2.3 is satisfied, we set

$$M_F = \|F\|_\infty, \qquad L_F = \sup_{y \in \mathbb{R}^n} \|D_y F\|_\infty,$$

$$M_\lambda = \|\lambda\|_\infty, \qquad L_\lambda = \sup_{y \in \mathbb{R}^n} \|D_y \lambda\|_\infty,$$

and L_Q a constant such that, for all $f \in \mathcal{E}$ and $x = (y, v) \in E$:

$$\left\| D_y \left(\lambda(x) \int_E f(z) Q(dz; x) \right) \right\|_\infty \leq L_Q \|f\|_\mathcal{E}.$$

Denote by $(P_t)_{t \geq 0}$ the transition semigroup associated to the PDP constructed above and by P_x the law of the PDP starting from $x \in E$. Then, P_x is a solution of the martingale problem associated to $\mathcal{A}$ in the following sense:

$$f(x_t) - f(x) - \int_0^t \mathcal{A} f(x_s) \, ds$$

is a local martingale for any $f \in \mathcal{E}$ [see Davis (1993)]. Moreover, if Hypothesis 2.3 holds, it is a bounded martingale. As usual, we have denoted by $(x_t)_{t \geq 0}$ the canonical process on $D(\mathbb{R}^+; E)$. As mentioned above, this statement is true for a much larger set of test functions.

The following result gives a uniqueness property for this martingale problem. It will enable us to characterize the asymptotic behavior of our stochastic regulatory networks.

THEOREM 2.5. *If Hypothesis 2.3 is satisfied, then the law of the PDP determined by F, λ, and Q is the unique solution of the martingale problem associated to the generator $\mathcal{A}$.*

The proof of the theorem is given in the Appendix.

Markov jump model for stochastic regulatory networks: *Known results.* We consider a set of chemical reactions R_r, $r \in \mathcal{R}$; $\mathcal{R}$ is supposed to be finite. These reactions involve species indexed by a set $S = 1, \ldots, M$, the number of molecules of the specie i is denoted by n_i and $X \in \mathbb{N}^M$ is the vector consisting of the n_i's. Each reaction R_r has a rate $\lambda_r(X)$ which depends on the state of the system, described by X, and corresponds to a change $X \to X + \gamma_r$, $\gamma_r \in \mathbb{Z}^M$.

Mathematically, this evolution can be described by the following Markov jump process. It is based on a sequence $(\tau_k)_{k \geq 1}$ of random waiting times with exponential distribution. Setting $T_0 = 0$, $T_i = \tau_1 + \cdots + \tau_i$, X is constant on $[T_{i-1}, T_i)$ and has a jump at T_i. The parameter of τ_i is given by $\sum_{r \in \mathcal{R}} \lambda_r(X(T_{i-1}))$:

$$\mathbb{P}(\tau_i > t) = \exp\left(-\sum_{r \in \mathcal{R}} \lambda_r(X(T_{i-1}))t\right).$$

At time T_i, a reaction $r \in \mathcal{R}$ is chosen with probability

$$\lambda_r(X(T_{i-1})) \Big/ \sum_{r \in \mathcal{R}} \lambda_r(X(T_{i-1}))$$

and the state changes according to $X \to X + \gamma_r$:

$$X(T_i) = X(T_{i-1}) + \gamma_r.$$

This Markov process has the following generator [see Ethier and Kurtz (1986)]:

$$Af(X) = \sum_{r \in \mathcal{R}} [f(X + \gamma_r) - f(X)]\lambda_r(X).$$

We do not need a precise description of the domain of A; the above definition holds, for instance, for functions in $C_b(\mathbb{R}^M)$.

In the applications we have in mind, the numbers of molecules have different scales. Some of the molecules are in small numbers and some are in large numbers.

Accordingly, we split the set of species into two sets C and D with cardinals M_C and M_D. This induces the decomposition $\mathbb{N}^M = \mathbb{N}^{M_C} \times \mathbb{N}^{M_D}$, $X = (X_C, X_D)$, $\gamma_r = (\gamma_r^C, \gamma_r^D)$. For $i \in D$, n_i is of order 1 while for $i \in C$, n_i is proportional to N where N is a large number. For $i \in C$, setting $\tilde{n}_i = n_i/N$, $\tilde{n}_i$ is of order 1. We define $x_C = \frac{1}{N}X_C$ and $x = (x_C, X_D)$. Then $x \in E = \mathbb{R}^{M_C} \times \mathbb{N}^{M_D}$.

We also decompose the set of reactions according to the species involved. We set $\mathcal{R} = \mathcal{R}_D \cup \mathcal{R}_C \cup \mathcal{R}_{DC}$. A reaction in $\mathcal{R}_D$ (resp., $\mathcal{R}_C$) produces or consumes only species in D (resp., C). Also, the rate of a reaction in $\mathcal{R}_D$ (resp., $\mathcal{R}_C$) depends only on X_D (resp., x_C). A reaction in $\mathcal{R}_{DC}$ has a rate depending on both x_C and X_D and produces or consumes, among others, species from C or D.

The rate of a reaction in $r \in \mathcal{R}_C$ is also large and of order N and we set $\tilde{\lambda}_r = \frac{\lambda_r}{N}$. In some applications, reactions in $\mathcal{R}_D$ or $\mathcal{R}_{DC}$ have a rate of order 1. Then introducing the new scaled variables, the generator has the form:

$$\tilde{\mathcal{A}}f(x_C, X_D) = \sum_{r \in \mathcal{R}_C} \left[f\left(x_C + \frac{1}{N}\gamma_r^C, X_D\right) - f(x_C, X_D) \right] N\tilde{\lambda}_r(x_C)$$

$$+ \sum_{r \in \mathcal{R}_{DC}} \left[f\left(x_C + \frac{1}{N}\gamma_r^C, X_D + \gamma_r^D\right) - f(x_C, X_D) \right] \lambda_r(x_C, X_D)$$

$$+ \sum_{r \in \mathcal{R}_D} [f(x_C, X_D + \gamma_r^D) - f(x_C, X_D)]\lambda_r(X_D).$$

Assuming that the scaled rates $\tilde{\lambda}_r$ are continuous with respect to x_C, it is not difficult to see that if $N \to \infty$, the two sets of species decouple. Indeed, reactions in $\mathcal{R}_{DC}$ do not happen sufficiently often and they do not change x_C in a sufficiently large manner. The limit would simply give a set of differential equations for the continuous variable x_C, which evolves without influence of X_D. The discrete variable would have its own dynamic made of jumps, with some rates depending on both x_C and X_D. These results have been shown by Kurtz (1971, 1978).

In the following sections, we consider more general systems where other types of reactions may happen and which yield different limiting systems.

3. Continuous piecewise deterministic process.

In this section, we assume that some of the reactions in a subset S_1 of $\mathcal{R}_{DC}$ are such that their rate is large and scales with N. We again set $\tilde{\lambda}_r = \frac{1}{N}\lambda_r$ for $r \in S_1$. We assume that these equations do not change X_D, in other words,

$$(4) \qquad\qquad\qquad \gamma_r^D = 0, \qquad r \in S_1.$$

However, the rate λ_r depends on X_D. Note that this is possible and even frequent in molecular biology, meaning that reactions of the type S_1 recover the reactant, like in the reaction $A \to A + B$, with A and B discrete and continuous species, respectively. The more complicated case $\gamma_r^D \neq 0$ is treated in Section 5.

With the decomposition $E = \mathbb{R}^{M_C} \times \mathbb{N}^{M_D}$ introduced above, the scaled generator has now the form

$$
\tilde{\mathcal{A}}_N f(x_C, X_D)
$$

$$
= \sum_{r \in \mathcal{R}_C} \left[f\left(x_C + \frac{1}{N}\gamma_r^C, X_D\right) - f(x_C, X_D) \right] N\tilde{\lambda}_r(x_C)
$$

$$
\text{(5)} \qquad + \sum_{r \in S_1} \left[f\left(x_C + \frac{1}{N}\gamma_r^C, X_D\right) - f(x_C, X_D) \right] N\tilde{\lambda}_r(x_C, X_D)
$$

$$
+ \sum_{r \in \mathcal{R}_{DC} \setminus S_1} \left[f\left(x_C + \frac{1}{N}\gamma_r^C, X_D + \gamma_r^D\right) - f(x_C, X_D) \right] \lambda_r(x_C, X_D)
$$

$$
+ \sum_{r \in \mathcal{R}_D} [f(x_C, X_D + \gamma_r^D) - f(x_C, X_D)] \lambda_r(X_D).
$$

For $f \in C_b^1(E)$, we may let $N \to \infty$ and obtain the limit generator

$$
\mathcal{A}_\infty f(x_c, X_D)
$$

$$
= \left(\sum_{r \in \mathcal{R}_C} \tilde{\lambda}_r(x_C)\gamma_r^C + \sum_{r \in S_1} \tilde{\lambda}_r(x_C, X_D)\gamma_r^C \right) \cdot \nabla_{x_C} f(x_C, X_D)
$$

$$
+ \sum_{r \in \mathcal{R}_{DC} \setminus S_1} [f(x_C, X_D + \gamma_r^D) - f(x_C, X_D)] \lambda_r(x_C, X_D)
$$

$$
+ \sum_{r \in \mathcal{R}_D} [f(x_C, X_D + \gamma_r^D) - f(x_C, X_D)] \lambda_r(X_D).
$$

This formal argument indicates that, as $N \to \infty$, the process converges to a continuous PDP [see (1)]. The state is described by a continuous variable x_C and a discrete variable X_D. The discrete variable is a jump process, with some rates depending on both x_C and X_D, and is piecewise constant. The continuous variable evolves according to differential equations depending on X_D. It is continuous but the vector field describing its evolution changes when X_D jumps.

This is rigorously justified by the following theorem.

THEOREM 3.1. *Let* $x^N = (x_C^N, X_D^N)$ *be a jump Markov process as above, starting at* $x^N(0) = (x_C^N(0), X_D^N(0))$. *Assume that the jump rates* $\tilde{\lambda}_r$, $r \in \mathcal{R}_C \cup S_1$ *and* λ_r, $r \in \mathcal{R}_{DC} \setminus S_1$ *are* C^1 *functions of* $x_C \in \mathbb{R}^{M_C}$. *We define* P_{x_0} *the law of the PDP starting at* $x_0 = (x_{C,0}, X_{D,0})$ *whose jump intensities are*

$$
\lambda(x) = \sum_{r \in \mathcal{R}_D \cup \mathcal{R}_{DC} \setminus S_1} \lambda_r(x);
$$

the transition measure is defined by

$$\int_E f(z)\,Q(dz;x)$$

$$= \frac{1}{\lambda(x)}\left(\sum_{r\in\mathcal{R}_{DC}\setminus S_1} f(x_C, X_D + \gamma_r^D)\lambda_r(x_C, X_D)\right.$$

$$\left. + \sum_{r\in\mathcal{R}_D} f(x_C, X_D + \gamma_r^D)\lambda_r(X_D)\right)$$

for $x = (x_C, X_D)$, and the vectors fields are given by

$$F_{X_D}(x_C) = \sum_{r\in\mathcal{R}_C} \gamma_r^C \tilde{\lambda}_r(x_C) + \sum_{r\in S_1} \gamma_r^C \tilde{\lambda}_r(x_C, X_D).$$

Assume that Hypothesis 2.1 is satisfied and $x^N(0)$ converges in distribution to x_0; then x^N converges in distribution to the PDP whose law is P_{x_0}.

PROOF. In the following, we work only with scaled variables and simplify the notation by omitting the tildes. In other words, we use λ_r to denote the rate of all reactions.

The proof is divided into three steps. We begin our proof by supposing that the jump rates and their derivatives with respect to x_C are bounded. Hypothesis 2.3 is then satisfied. We then prove Theorem 3.1 by a truncation argument.

Step 1: *Tightness for bounded reaction rates.* We first assume that all rates λ_r are bounded as well as their derivatives with respect to x_C.

Let x^N be a Markov jump process whose generator is given by $\tilde{\mathcal{A}}_N$.

Without loss of generality, we assume that the initial value of the process is deterministic: $x^N(0) = (x_C^N(0), X_D^N(0))$ and converges to $x_0 = (x_{C,0}, X_{D,0})$ in $\mathbb{R}^{M_C} \times \mathbb{N}^{M_D}$.

Let $(Y_r)_{r\in\mathcal{R}}$ be a sequence of independent standard Poisson processes. By Proposition 1.7, Part 4, and Theorem 4.1, Part 6, of Ethier and Kurtz (1986) we know that there exist stochastic processes $(\tilde{x}^N)_{N\in\mathbb{N}}$ in $D(\mathbb{R}^+; E)$ such that

$$\tilde{x}^N(t) = x^N(0) + \sum_{r\in\mathcal{R}} \gamma_r Y_r\left(\int_0^t \lambda_r(\tilde{x}^N(s))\,ds\right), \qquad t \geq 0.$$

Moreover, for each N, x^N and $\tilde{x}^N$ have the same distribution. Since we consider only the distributions of the processes, we only consider $\tilde{x}^N$ in the following and use the same notation for both processes.

Using the decomposition $x^N = (x_C^N, X_D^N)$, we have

$$x_C^N(t) = x_C^N(0) + \sum_{r \in \mathcal{R}_C} \frac{1}{N} \gamma_r^C Y_r \left(N \int_0^t \tilde{\lambda}_r(x_C^N(s)) \, ds \right)$$

$$+ \sum_{r \in S_1} \frac{1}{N} \gamma_r^C Y_r \left(N \int_0^t \tilde{\lambda}_r(x_C^N(s), X_D^N(s)) \, ds \right)$$

$$+ \sum_{r \in \mathcal{R}_{DC} \setminus S_1} \frac{1}{N} \gamma_r^C Y_r \left(\int_0^t \lambda_r(x_C^N(s), X_D^N(s)) \, ds \right)$$

and

$$X_D^N(t) = X_D^N(0) + \sum_{r \in \mathcal{R}_{DC} \setminus S_1} \gamma_r^D Y_r \left(\int_0^t \lambda_r(x_C^N(s), X_D^N(s)) \, ds \right)$$

$$+ \sum_{r \in \mathcal{R}_D} \gamma_r^D Y_r \left(\int_0^t \lambda_r(X_D^N(s)) \, ds \right).$$

We easily prove tightness in $D(\mathbb{R}^+; \mathbb{R}^{M_D})$ of the laws of $(X_D^N)_{N \in \mathbb{N}}$ by the same proof as for Proposition 3.1, Part 6, of Ethier and Kurtz (1986) and by using the fact that the law of Y_r is tight in $D(\mathbb{R}^+; \mathbb{N})$, for every $r \in \mathcal{R}$, according to Theorem 1.4 of Billingsley (1999).

To prove that the laws of $(x_C^N)_{N \in \mathbb{N}}$ are tight in $C(\mathbb{R}^+; \mathbb{R}^{M_C})$, we adapt the argument of Section 2, Part 11, in Ethier and Kurtz (1986).

Let $\tilde{Y}_r(u) = Y_r(u) - u$ be the standard Poisson process centered at its expectation; we have

$$x_C^N(t) = x_C^N(0) + \sum_{r \in \mathcal{R}_C} \frac{1}{N} \gamma_r^C \tilde{Y}_r \left(N \int_0^t \tilde{\lambda}_r(x_C^N(s)) \, ds \right)$$

$$+ \sum_{r \in S_1} \frac{1}{N} \gamma_r^C \tilde{Y}_r \left(N \int_0^t \tilde{\lambda}_r(x_C^N(s), X_D^N(s)) \, ds \right)$$

$$+ \int_0^t F(x_C^N(s), X_D^N(s)) \, ds$$

$$+ \sum_{r \in \mathcal{R}_{DC} \setminus S_1} \frac{1}{N} \gamma_r^C Y_r \left(\int_0^t \lambda_r(x_C^N(s), X_D^N(s)) \, ds \right).$$

Observe that

$$\sup_{u \in [0, A]} \frac{1}{N} \tilde{Y}_r(Nu) \to 0 \qquad \text{a.s.}$$

for any $A \geq 0$. Since λ_r are bounded, it follows that, for all $T > 0$,

$$\sup_{t \in [0,T]} \left| \sum_{r \in \mathcal{R}_C} \frac{1}{N} \gamma_r^C \tilde{Y}_r \left(N \int_0^t \tilde{\lambda}_r(x_C^N(s)) \, ds \right) \right.$$

$$\left. + \sum_{r \in S_1} \frac{1}{N} \gamma_r^C \tilde{Y}_r \left(N \int_0^t \tilde{\lambda}_r(x_C^N(s), X_D^N(s)) \, ds \right) \right|$$

$$\to 0 \qquad \text{a.s. when } N \to \infty.$$

Clearly,

$$\sup_{t \in [0,T]} \left| \sum_{r \in \mathcal{R}_{DC} \setminus S_1} \frac{1}{N} \gamma_r^C Y_r \left(\int_0^t \lambda_r(x_C^N(s), X_D^N(s)) \, ds \right) \right| \to 0 \qquad \text{a.s.}$$

It follows that there exists a random constant K_N going to zero such that, for $t, t_1, t_2 \in [0, T]$, and $\|F\|_\infty = \sup_{x \in \mathbb{R}^{M_C} \times \mathbb{N}^{M_D}} |F(x)|$,

$$|x_C^N(t)| \leq |x_C^N(0)| + \|F\|_\infty t + K_N \qquad \text{a.s.}$$

and

$$|x_C^N(t_1) - x_C^N(t_2)| \leq \|F\|_\infty |t_1 - t_2| + 2K_N \qquad \text{a.s.}$$

Tightness of $(x_C^N)_{N \in \mathbb{N}}$ in $C(\mathbb{R}^+; \mathbb{R}^{M_C})$ follows by classical criteria [see, e.g., Jacod and Shiryaev (1987), Chapter 6, Section 3b].

We conclude, from Jacod and Shiryaev (1987) (Chapter 6, Section 3b), that $\{x^N\}_N = \{(x_C^N, X_D^N)\}_N$ is tight in $D(\mathbb{R}^+; E)$.

Step 2: *Identification of limit points for bounded reaction rates.* Let $x = (x_t)_{t \geq 0}$ be the canonical process on $D(\mathbb{R}^+; E)$, and P_N the law of $(x_t^N)_{t \geq 0}$ on this space, for each $N \in \mathbb{N}$.

We know that for each $N \in \mathbb{N}$ and $\varphi \in \mathcal{E}$,

$$\varphi(x_t) - \varphi(x_0) - \int_0^t \tilde{\mathcal{A}}_N \varphi(x_s) \, ds$$

is a P_N-martingale. Equivalently, for each $n \in \mathbb{N}$, $t_1, \ldots, t_n \in [0, r]$, $t \geq r \geq 0$, $\psi \in (C_b(E))^n$ and $\varphi \in \mathcal{E}$,

$$(6) \quad \begin{aligned} \mathbb{E}_{P_N} &\left(\left(\varphi(x_t) - \varphi(x_0) - \int_0^t \tilde{\mathcal{A}}_N \varphi(x_s) \, ds \right) \psi(x_{t_1}, \ldots, x_{t_n}) \right) \\ &= \mathbb{E}_{P_N} \left(\left(\varphi(x_r) - \varphi(x_0) - \int_0^r \tilde{\mathcal{A}}_N \varphi(x_s) \, ds \right) \psi(x_{t_1}, \ldots, x_{t_n}) \right). \end{aligned}$$

Let $(P_{N_k})_k$ be a subsequence which converges weakly to a measure P on $D(\mathbb{R}^+; E)$. We know that x is P-almost surely continuous at every t except for a countable set D_P and that for $t_1, \ldots, t_n$ outside D_P, $P_{N_k} \pi_{t_1, \ldots, t_n}^{-1}$ converges weakly

to $P\pi_{t_1,\dots,t_n}^{-1}$ where $\pi_{t_1,\dots,t_n}$ is the projection that carries the point $x \in D(\mathbb{R}^+; E)$ to the point $(x_{t_1}, \dots, x_{t_n})$ of $\mathbb{R}^n$.

Therefore, it is easy, using the dominated convergence theorem and weak convergence properties, to let $k \to \infty$ in (6) and obtain for $t, t_1, \dots, t_n, r$ outside D_P:

(7)
$$
\mathbb{E}_P\left(\left(\varphi(x_t) - \varphi(x_0) - \int_0^t \mathcal{A}_\infty\varphi(x_s)\,ds\right)\psi(x_{t_1}, \dots, x_{t_n})\right)
$$
$$
= \mathbb{E}_P\left(\left(\varphi(x_r) - \varphi(x_0) - \int_0^r \mathcal{A}_\infty\varphi(x_s)\,ds\right)\psi(x_{t_1}, \dots, x_{t_n})\right).
$$

If $t \in D_P$, we choose a sequence (t^k) outside D_P such that $t^k \to t$ with $t^k > t$. Then $P\pi_{t^k}^{-1}$ converges weakly to $P\pi_t^{-1}$ since x is P-a.s. right-continuous in t and x_{t^k} converges almost surely to x_t. Then, we use (7) with t^k instead of t, let $k \to \infty$ and deduce that (7) also holds for $t \in D_P$. Similarly, we show that $t_1, \dots, t_n, r$ may be taken in D_P.

This shows that the measure P is a solution of the martingale problem associated to the generator $\mathcal{A}_\infty$ on the domain $\mathcal{E}$.

Hypothesis 2.3 enables us to apply Theorem 2.5. The martingale problem has then a unique solution. It follows that the limit P is equal to P_{x_0}, the law of the PDP, and that the whole sequence $(P_N)_N$ converges weakly to P_{x_0}.

Step 3: *Conclusion.* Now, we prove Theorem 3.1 with a truncation argument.
Let $\theta \in C^\infty(\mathbb{R}^+)$ such that
$$
\begin{cases} \theta(x) = 1, & x \in [0, 1], \\ \theta(x) = 0, & x \in [2, \infty), \end{cases}
$$
and, for $k \geq 1$ and $r \in \mathcal{R}$, define
$$
\theta_k(x) = \theta\left(\frac{|x|^2}{k^2}\right), \qquad x \in E,
$$
and
$$
\lambda_r^k(x) = \theta_k(x)\lambda_r(x).
$$

Then, the problem with λ_r^k instead of λ_r fulfills Hypothesis 2.3. We define $x_k^N = (x_{C,k}^N, X_{D,k}^N)$ the jump Markov process associated to the jump intensities λ_r^k, starting at $x^N(0)$. By the preceding result, we know that, for all $k \in \mathbb{N}$, $(x_k^N)_{N \in \mathbb{N}}$ converges weakly to the PDP x_k in $D(\mathbb{R}^+; E)$, whose characteristics are the jump intensities λ_r^k, with corresponding transition measure, and vector fields (obvious definitions as in Theorem 3.1).

Then, $((x_k^N)_{k \in \mathbb{N}})_{N \in \mathbb{N}}$ converges weakly to $(x_k)_{k \in \mathbb{N}}$ in $D(\mathbb{R}^+; E)^{\mathbb{N}}$.

By the Skorohod representation theorem [see Billingsley (1971), Theorem 3.3], up to a change of probability space, we may assume that for all $k \in \mathbb{N}$, $(x_k^N)_N$ converges a.s. to x_k in $D(\mathbb{R}^+; E)$.

Let $T > 0$ and the stopping times

$$\tau^k = \inf\{t \in [0, T], |x_k(t)| \geq k\}$$

with $\tau^k = T$ if $\{t \in [0, T], |x_k(t)| \geq k\} = \varnothing$.

Then, for $k, l \in \mathbb{N}$

$$x_k(t) = x_l(t), \qquad t \in [0, \tau^l \wedge \tau^k] \qquad \text{a.s.}$$

so that τ^k is a.s. nondecreasing.

Moreover, if x (resp., x^N) are the PDP associated to $\mathcal{A}_\infty$ (resp., the Markov jump process associated to $\tilde{\mathcal{A}}_N$), then

$$x_k(t) = x(t), \qquad t \in [0, \tau^k) \qquad \text{a.s.}$$

and if

$$\tau_N^k = \inf\{t \in [0, T], |x_k^N(t)| \geq k\}$$

with $\tau_N^k = T$ if $\{t \in [0, T], |x_k^N(t)| \geq k\} = \varnothing$, then

$$x_k^N(t) = x^N(t), \qquad t \in [0, \tau_N^k) \qquad \text{a.s.}$$

Let $\delta > 0$. Observing that if $\tau^{k-1} > T - \delta$ and $d_{T-\delta}(x_k^N, x_k) < \epsilon$, where $d_{T-\delta}$ is the distance on $D([0, T - \delta]; E)$, then, for enough small ϵ:

$$\sup_{t \in [0, T-\delta]} |x_k^N(t)| \leq k \qquad \text{a.s.,}$$

then a.s., $\tau_N^k \geq T - \delta$ and $x_k^N = x^N$ in $[0, T - \delta]$. Since $\tau^k \geq \tau^{k-1} > T - \delta$, we have also, a.s., $x_k = x$ in $[0, T - \delta]$ and

$$d_{T-\delta}(x^N, x) < \epsilon.$$

We deduce that

$$\forall \delta > 0 \qquad \mathbb{P}\big(d_{T-\delta}(x^N, x) \geq \epsilon\big) \leq \mathbb{P}(\tau^{k-1} \leq T - \delta) + \mathbb{P}\big(d_{T-\delta}(x_k^N, x_k) \geq \epsilon\big).$$

Finally, we have

$$\mathbb{P}(\tau^{k-1} \leq T - \delta) = \mathbb{P}\Big(\sup_{t \in [0, \tau^k]} |x_k(t)| \geq k - 1 \Big) \leq \mathbb{P}\Big(\sup_{t \in [0, T-\delta]} |x(t)| \geq k - 1 \Big).$$

By Hypothesis 2.1, the PDP x cannot explode in finite time. Thus for k large, this term is small. Then for large N, the second is small. We deduce that x^N converges in probability to x in the new probability space. Returning in the original probability space, we obtain that x^N converges in distribution to x. □

4. Piecewise deterministic process with jumps. In this section, we assume that some of the reactions in a subset S_2 of $\mathcal{R}_{DC}$ are such that γ_r^C is large and scales with N. For the other reactions, we assume that γ_r^C is still of order 1. We

set $\tilde{\gamma}_r^C = \frac{1}{N}\gamma_r^C$ for $r \in S_2$. The jump rates λ_r, $r \in S_2$ are still assumed to be of order 1. We define $S = S_1 \cup S_2$.

The scaled generator has now the form

$$\tilde{\mathcal{A}}_N f(x_C, X_D)$$

$$= \sum_{r \in \mathcal{R}_C} \left[f\left(x_C + \frac{1}{N}\gamma_r^C, X_D\right) - f(x_C, X_D) \right] N\tilde{\lambda}_r(x_C)$$

$$+ \sum_{r \in S_2} [f(x_C + \tilde{\gamma}_r^C, X_D + \gamma_r^D) - f(x_C, X_D)]\lambda_r(x_C, X_D)$$

(8)

$$+ \sum_{r \in S_1} \left[f\left(x_C + \frac{1}{N}\gamma_r^C, X_D\right) - f(x_C, X_D) \right] N\tilde{\lambda}_r(x_C, X_D)$$

$$+ \sum_{r \in \mathcal{R}_{DC}\setminus S} \left[f\left(x_C + \frac{1}{N}\gamma_r^C, X_D + \gamma_r^D\right) - f(x_C, X_D) \right]\lambda_r(x_C, X_D)$$

$$+ \sum_{r \in \mathcal{R}_D} [f(x_C, X_D + \gamma_r^D) - f(x_C, X_D)]\lambda_r(X_D).$$

For $f \in C_b^1(E)$, we may let $N \to \infty$ and obtain the limit generator

$$\tilde{\mathcal{A}}_\infty f(x_c, X_D) = \left(\sum_{r \in \mathcal{R}_C} \tilde{\lambda}_r(x_C)\gamma_r^C + \sum_{r \in S_1} \tilde{\lambda}_r(x_C, X_D)\gamma_r^C \right) \cdot \nabla_{x_C} f(x_C, X_D)$$

$$+ \sum_{r \in S_2} [f(x_C + \tilde{\gamma}_r^C, X_D + \gamma_r^D) - f(x_C, X_D)]\lambda_r(x_C, X_D)$$

$$+ \sum_{r \in \mathcal{R}_{DC}\setminus S} [f(x_C, X_D + \gamma_r^D) - f(x_C, X_D)]\lambda_r(x_C, X_D)$$

$$+ \sum_{r \in \mathcal{R}_D} [f(x_C, X_D + \gamma_r^D) - f(x_C, X_D)]\lambda_r(X_D).$$

This formal argument indicates that if $N \to \infty$, the process is a piecewise deterministic process with jumps in both x_C and X_D. In fact, the proof of this can be done easily thanks to the result of Section 3.

Indeed, let us introduce the following auxiliary system of reactions involving the variable $(x_C^{1,N}, x_C^{2,N}, X_D^N) \in \mathbb{R}^{M_C} \times \mathbb{R}^{M_C} \times \mathbb{N}^{M_D}$:

- If $r \in \mathcal{R}_C$, $(x_C^{1,N}, x_C^{2,N}, X_D^N) \to (x_C^{1,N} + \frac{1}{N}\gamma_C^r, x_C^{2,N}, X_D^N)$.
- If $r \in S_2$, $(x_C^{1,N}, x_C^{2,N}, X_D^N) \to (x_C^{1,N}, x_C^{2,N} + \tilde{\gamma}_C^r, X_D^N + \gamma_D^r)$.
- If $r \in \mathcal{R}_{DC} \setminus S_2$, $(x_C^{1,N}, x_C^{2,N}, X_D^N) \to (x_C^{1,N} + \frac{1}{N}\gamma_C^r, x_C^{2,N}, X_D^N + \gamma_D^r)$.
- If $r \in \mathcal{R}_D$, $(x_C^{1,N}, x_C^{2,N}, X_D^N) \to (x_C^{1,N}, x_C^{2,N}, X_D^N + \gamma_D^r)$.

The rates of these reactions are $\lambda_r(x_C^{1,N}, x_C^{2,N}, X_D^N) = \lambda_r(x_C^{1,N} + x_C^{2,N}, X_D^N)$.

If we choose the initial data $(x_C^N(0), 0, X_D^N(0))$, then this system and the original system are equivalent. Indeed, we recover the value of the original system by the relation

$$(x_C^N, X_D^N) = (x_C^{1,N} + x_C^{2,N}, X_D^N).$$

Conversely, notice that x_C^N is a sum of a pure jump part and of a continuous one so that given (x_C^N, X_D^N), $(x_C^{1,N}, x_C^{2,N}, X_D^N)$ can be recovered by isolating from x_C^N the small jumps to obtain $x_C^{1,N}$ and the jump of order 1 (from S_2) to obtain $x_C^{2,N}$.

The auxiliary system corresponds to the following generator:

$$\mathcal{A}_{\text{aux}} f(x_C^1, x_C^2, X^D)$$

$$= \sum_{r \in \mathcal{R}_C} \left[f\left(x_C^1 + \frac{1}{N}\gamma_r^C, x_C^2, X_D\right) - f(x_C^1, x_C^2, X_D) \right] N\tilde{\lambda}_r(x_C^1 + x_C^2)$$

$$+ \sum_{r \in S_2} [f(x_C^1, x_C^2 + \tilde{\gamma}_r^C, X_D + \gamma_r^D) - f(x_C^1, x_C^2, X_D)]\lambda_r(x_C^1 + x_C^2, X_D)$$

(9) $\qquad + \sum_{r \in S_1} \left[f\left(x_C^1 + \frac{1}{N}\gamma_r^C, x_C^2, X_D\right) - f(x_C^1, x_C^2, X_D) \right] N\tilde{\lambda}_r(x_C^1 + x_C^2, X_D)$

$$+ \sum_{r \in \mathcal{R}_{DC} \setminus S} \left[f\left(x_C^1 + \frac{1}{N}\gamma_r^C, x_C^2, X_D + \gamma_r^D\right) - f(x_C^1, x_C^2, X_D) \right]$$

$$\times \lambda_r(x_C^1 + x_C^2, X_D)$$

$$+ \sum_{r \in \mathcal{R}_D} [f(x_C^1, x_C^2, X_D + \gamma_r^D) - f(x_C^1, x_C^2, X_D)]\lambda_r(X_D).$$

This generator is of the form as in Section 3 but the discrete variable is now (x_C^2, X_D). Using the results proved in this section, we thus obtain that this auxiliary system of reaction converges in distribution to the piecewise continuous deterministic process given by the generator

$$\mathcal{A}_{\text{aux},\infty} f(x_C^1, x_C^2, X_D)$$

$$= \left(\sum_{r \in \mathcal{R}_C} \tilde{\lambda}_r(x_C^1 + x_C^2, X_D)\gamma_r^C + \sum_{r \in S_1} \tilde{\lambda}_r(x_C^1 + x_C^2, X_D)\gamma_r^C \right)$$

$$\times \nabla_{x_C^1} f(x_C^1, x_C^2, X_D)$$

$$+ \sum_{r \in S_2} [f(x_C^1, x_C^2 + \tilde{\gamma}_r^C, X_D + \gamma_r^D) - f(x_C^1, x_C^2, X_D)]\lambda_r(x_C^1 + x_C^2, X_D)$$

$$+ \sum_{r \in \mathcal{R}_{DC} \setminus S} [f(x_C^1, x_C^2, X_D + \gamma_r^D) - f(x_C^1, x_C^2, X_D)]\lambda_r(x_C^1 + x_C^2, X_D)$$

$$+ \sum_{r \in \mathcal{R}_D} [f(x_C^1, x_C^2, X_D + \gamma_r^D) - f(x_C^1, x_C^2, X_D)]\lambda_r(X_D).$$

Going back to the original variables, that is, setting $x_C = x_C^1 + x_C^2$, we deduce that the original system converges in distribution to the piecewise deterministic process with generator $\tilde{\mathcal{A}}_\infty$.

To obtain this convergence, we suppose that the jump intensities are C^1-differentiable with respect to the variable x_C^1 and that the limit PDP satisfies Hypothesis 2.1.

5. Averaging. In this section, we examine the case when (4) is not satisfied by all the discrete variables. In this case, the previous results are not valid. We introduce the decomposition $X_D = (X_D^1, X_D^2) \in \mathbb{N}^{M_{D,1}} \times \mathbb{N}^{M_{D,2}}$ and $\gamma_r^D = (\gamma_r^{D,1}, \gamma_r^{D,2})$ such that

$$\gamma_r^{D,1} = 0, \qquad r \in S_1.$$

This replaces (4). The continuous variable x_C follows the same characteristics as in Section 3.

In the set of the reactions S_1, the discrete variable X_D^2 has fast motion, its jump rates are of order N and its jumps are of order 1.

More precisely, for $r \in S_1$, the state changes according to

$$(x_C, X_D^1, X_D^2) \to \left(x_C + \frac{1}{N}\gamma_r^C, X_D^1, X_D^2 + \gamma_r^{D,2} \right)$$

with the rate $N\tilde{\lambda}_r(x_C, X_D^1, X_D^2)$.

We now have the following generator for the process:

$$\tilde{\mathcal{A}}_N f(x_C, X_D^1, X_D^2)$$

$$= \sum_{r \in \mathcal{R}_C} \left[f\left(x_C + \frac{1}{N}\gamma_r^C, X_D^1, X_D^2 \right) - f(x_C, X_D^1, X_D^2) \right] N\tilde{\lambda}_r(x_C)$$

$$+ \sum_{r \in S_1} \left[f\left(x_C + \frac{1}{N}\gamma_r^C, X_D^1, X_D^2 + \gamma_r^{D,2} \right) - f(x_C, X_D^1, X_D^2) \right]$$

$$\times N\tilde{\lambda}_r(x_C, X_D^1, X_D^2)$$

(10) $$+ \sum_{r \in \mathcal{R}_{DC} \setminus S_1} \left[f\left(x_C + \frac{1}{N}\gamma_r^C, X_D^1 + \gamma_r^{D,1}, X_D^2 + \gamma_r^{D,2} \right) \right.$$

$$\left. - f(x_C, X_D^1, X_D^2) \right]$$

$$\times \lambda_r(x_C, X_D^1, X_D^2)$$

$$+ \sum_{r \in \mathcal{R}_D} [f(x_C, X_D^1 + \gamma_r^{D,1}, X_D^2 + \gamma_r^{D,2}) - f(x_C, X_D^1, X_D^2)]$$

$$\times \lambda_r(X_D^1, X_D^2).$$

This new model combines slow and fast motions. This leads to double time scale evolution which can be further simplified. Contrary to the previous sections, we cannot take the formal limit of this generator. The idea is to average in fast discrete variables X_D^2, and focus on the slow variables (x_C, X_D^1), in order to obtain a much simpler averaged generator.

To this aim, we introduce the following generator depending on (x_C, X_D^1):

$$\mathcal{A}_{x_C, X_D^1} h(X_D^2) = \sum_{r \in S_1} [h(X_D^2 + \gamma_r^{D,2}) - h(X_D^2)] \tilde{\lambda}_r(x_C, X_D^1, X_D^2).$$

We assume that for all (x_C, X_D^1), the process associated to the generator $\mathcal{A}_{x_C, X_D^1}$ is ergodic with a unique invariant measure, denoted by ν_{x_C, X_D^1}.

Then we define the averaged jump rates:

$$\bar{\lambda}_r(x_C, X_D^1) = \int_{\mathbb{N}^{M_{D,2}}} \tilde{\lambda}_r(x_C, X_D^1, X_D^2) \nu_{x_C, X_D^1}(dX_D^2), \qquad r \in S_1,$$

$$(11) \quad \bar{\lambda}_r(x_C, X_D^1) = \int_{\mathbb{N}^{M_{D,2}}} \lambda_r(x_C, X_D^1, X_D^2) \nu_{x_C, X_D^1}(dX_D^2), \qquad r \in \mathcal{R}_{DC} \setminus S_1,$$

$$\bar{\lambda}_r(x_C, X_D^1) = \int_{\mathbb{N}^{M_{D,2}}} \lambda_r(X_D^1, X_D^2) \nu_{x_C, X_D^1}(dX_D^2), \qquad r \in \mathcal{R}_D.$$

We present now the main result of this section.

THEOREM 5.1. *Let $x^N = (x_C^N, X_{D,1}^N, X_{D,2}^N)$ be a jump Markov process with generator $\tilde{\mathcal{A}}_N$, starting at $x^N(0) = (x_C^N(0), X_{D,1}^N(0), X_{D,2}^N(0))$. Assume that*:

- *For all (x_C, X_D^1), the process associated to the generator $\mathcal{A}_{x_C, X_D^1}$ is ergodic with a unique invariant measure.*
- *There exist $K_1 > 0$ and $K_2 > 0$ such that, for every $s \in \mathbb{R}^+$, $(x_C, X_D^1, X_D^2) \in \mathbb{R}^{M_C} \times \mathbb{N}^{M_{D,1}} \times \mathbb{N}^{M_{D,2}}$, and for every bounded function g satisfying the centering condition*

$$(12) \qquad \int_{\mathbb{N}^{M_{D,2}}} g(X_D^2) \nu_{x_C, X_D^1}(dX_D^2) = 0,$$

we have

$$P_s^{x_C, X_D^1} g(X_D^2) < K_1 e^{-K_2 s} \|g\|_\infty,$$

where $(P_s^{x_C, X_D^1})_s$ is the semigroup associated to $\mathcal{A}_{x_C, X_D^1}$.

- *If g is a bounded function satisfying the centering condition (12), the Poisson equation*

$$\sum_{r \in S_1} [h(x_C, X_D^1, X_D^2 + \gamma_r^{D,2}) - h(x_C, X_D^1, X_D^2)] \tilde{\lambda}_r(x_C, X_D^1, X_D^2) = g(X_D^2)$$

has a solution given by

$$h(x_C, X_D^1, X_D^2) = -\int_0^\infty P_s^{x_C, X_D^1} g(X_D^2) \, ds$$

and this solution is Lipschitz with respect to x_C, uniformly in X_D^1.

- *The jump rates $\tilde{\lambda}_r$, $r \in \mathcal{R}_C \cap S_1$, and λ_r, $r \in \mathcal{R}_{DC} \setminus S_1$, and $\bar{\lambda}_r$, $r \in \mathcal{R}_{DC} \cup \mathcal{R}_D$, are C^1 with respect to x_C.*
- *For every bounded B set of $\mathbb{R}^{M_C} \times \mathbb{N}^{M_{D,1}}$ the functions $\tilde{\lambda}_r(\cdot, \cdot, X_D^2)$, $r \in \mathcal{R}_C \cap S_1$, and $\lambda_r(\cdot, \cdot, X_D^2)$, $r \in \mathcal{R}_{DC} \setminus S_1$, and $\bar{\lambda}_r(\cdot, \cdot, X_D^2)$, $r \in \mathcal{R}_{DC} \cup \mathcal{R}_D$, are bounded and Lipschitz with constants depending on B but not on X_D^2.*
- *λ, Q and F satisfy Hypothesis 2.1.*
- *$(x_C^N(0), X_D^{1,N}(0))$ converges in distribution to x_0 in $\mathbb{R}^{M_C} \times \mathbb{N}^{M_{D,1}}$.*

Define P_{x_0} the law of the PDP starting at $x_0 = (x_{C,0}, X_{D,0}^1)$ whose jump intensities are

$$\lambda(x_C, X_D^1) = \sum_{r \in (\mathcal{R}_{DC} \setminus S_1) \cup \mathcal{R}_D} \bar{\lambda}_r(x_C, X_D^1)$$

with transition measure:

$$\int_E f(z) Q(dz; x_C, X_D^1)$$

$$= \frac{1}{\lambda(x_C, X_D^1)} \sum_{r \in (\mathcal{R}_{DC} \setminus S_1) \cup \mathcal{R}_D} f(x_C, X_D^1 + \gamma_r^{D,1}) \bar{\lambda}_r(x_C, X_D^1)$$

and vectors fields:

$$F_{X_D^1}(x_C) = \sum_{r \in \mathcal{R}_C} \gamma_r^C \tilde{\lambda}_r(x_C) + \sum_{r \in S_1} \gamma_r^C \bar{\lambda}_r(x_C, X_D^1).$$

Then $(x_C^N, X_D^{1,N})$ converges in distribution in $D(\mathbb{R}^+; \mathbb{R}^{M_C} \times \mathbb{N}^{M_{D,1}})$ to the PDP whose law is P_{x_0}.

REMARK 5.2. The last two assumptions are similar to those made previously. The first three assumptions guarantee that it is possible to average with respect to the variable X_D^2. The first two state that the dynamic associated to the generator $\mathcal{A}_{x_C, X_D^1}$ is exponentially mixing, uniformly with respect to (x_C, X_D^1); the third one then follows if $\mathcal{A}_{x_C, X_D^1}$ depends smoothly on x_C.

PROOF OF THEOREM 5.1. The steps of the proof are similar to the proof of Theorem 3.1. We begin by supposing that all the jump rates are bounded as well

as their derivatives. This hypothesis will be removed at the end of the proof. We take $E = \mathbb{R}^{M_C} \times \mathbb{N}^{M_{D,1}}$.

Step 1: *Tightness.* The proof of the tightness of the laws of $(x_C^N, X_{D,1}^N)$ is similar to that in Section 3, and is left to the reader.

Step 2: *Identification of limit points.* Since we are interested in the evolution of (x_C, X_D^1), it is natural to consider, for the generator $\tilde{A}_N$, test functions $f(x_C, X_D^1)$ depending only on those variables. Unfortunately, some terms of order 1 depending on the variable X_D^2 remain. To overcome this problem, we introduce "perturbed test functions" [see Papanicolaou, Stroock and Varadhan (1976), Kushner (1994), Fouque et al. (2007)]:

$$f_N(x_C, X_D^1, X_D^2) = f(x_C, X_D^1) + \frac{1}{N} f^1(x_C, X_D^1, X_D^2)$$

with $f \in \mathcal{E}$ and f independent of X_D^2. We temporarily need that f is also C^2 with bounded second derivatives with respect to x_C. The function f^1 is chosen such that the formal limit of $\tilde{A}_N f_N(x_C, X_D^1, X_D^2)$ does not depend on X_D^2.

This can be done by taking f^1 such that

(13)
$$\sum_{r \in S_1} [f^1(x_C, X_D^1, X_D^2 + \gamma_r^{D,2}) - f^1(x_C, X_D^1, X_D^2)] \tilde{\lambda}_r(x_C, X_D^1, X_D^2)$$
$$= g_{x_C, X_D^1}(X_D^2)$$

with

$$g_{x_C, X_D^1}(X_D^2)$$
$$= \left(\sum_{r \in S_1} \gamma_r^C \bar{\lambda}_r(x_C, X_D^1) - \gamma_r^C \tilde{\lambda}_r(x_C, X_D^1, X_D^2) \right) \cdot \nabla_{x_C} f(x_C, X_D^1)$$
$$+ \sum_{r \in \mathcal{R}_{DC} \setminus S_1} [f(x_C, X_D^1 + \gamma_r^{D,1}) - f(x_C, X_D^1)]$$

(14)
$$\times (\bar{\lambda}_r(x_C, X_D^1) - \lambda_r(x_C, X_D^1, X_D^2))$$
$$+ \sum_{r \in \mathcal{R}_D} [f(x_C, X_D^1 + \gamma_r^{D,1}) - f(x_C, X_D^1)]$$
$$\times (\bar{\lambda}_r(x_C, X_D^1) - \lambda_r(X_D^1, X_D^2)).$$

Since the jump rates are assumed to be bounded, g_{x_C, X_D^1} is clearly bounded. Also, the function g_{x_C, X_D^1} is zero mean with respect to the invariant distribution, that is, g_{x_C, X_D^1} satisfy (12).

Then, the function f^1 satisfying (13) exists under our assumptions. Moreover, it is not difficult to check that f^1 is Lipschitz with respect to x_C, uniformly with respect to X_D^1.

Limiting infinitesimal generator. We now return to the analysis of the limiting problem. We denote by $P_{x,N}$ the law of $(x_C^N, X_{D,1}^N, X_{D,2}^N)$; then

$$f_N(x_C(t), X_D^1(t), X_D^2(t)) - f_N(x_C(0), X_D^1(0), X_D^2(0))$$

$$- \int_0^t \tilde{\mathcal{A}}_N f_N(x_C(s), X_D^1(s), X_D^2(s)) \, ds$$

is a $P_{x,N}$ martingale. Equivalently, for each $n \in \mathbb{N}$, $t_1, \ldots, t_n \in [0, r]$, $t \geq r \geq 0$, $\psi \in (C_b(E))^n$,

(15)
$$\mathbb{E}_{P_{x,N}}\left(\left(f_N(x_t) - f_N(x_0) - \int_0^t \tilde{\mathcal{A}}_N f_N(x_s) \, ds\right)\psi(x_{t_1}, \ldots, x_{t_n})\right)$$

$$= \mathbb{E}_{P_{x,N}}\left(\left(f_N(x_r) - f_N(x_0) - \int_0^r \tilde{\mathcal{A}}_N f_N(x_s) \, ds\right)\psi(x_{t_1}, \ldots, x_{t_n})\right).$$

Let us now consider a subsequence $(\tilde{P}_{x,N_k})_k$ of $(\tilde{P}_{x,N})_N$, the laws of $(x_C^N, X_{D,1}^N)$, which weakly converges to a measure $\tilde{P}_x$ on $D(\mathbb{R}^+; \mathbb{R}^{M_C} \times \mathbb{N}^{M_{D,1}})$.

It is clear that f_N converges to f, uniformly on E. Moreover, the continuity and boundedness properties of f^1 ensure that $\tilde{\mathcal{A}}_N f_N$ converges on E toward $\tilde{\mathcal{A}}_\infty f$, where

(16)
$$\tilde{\mathcal{A}}_\infty f(x_C, X_D^1)$$

$$= \left(\sum_{r \in \mathcal{R}_C} \gamma_r^C \tilde{\lambda}_r(x_C) + \sum_{r \in S_1} \gamma_r^C \bar{\lambda}_r(x_C, X_D^1)\right) \cdot \nabla_{x_C} f(x_C, X_D^1)$$

$$+ \sum_{r \in (\mathcal{R}_{DC} \setminus S_1) \cup \mathcal{R}_D} [f(x_C, X_D^1 + \gamma_r^{D,1}) - f(x_C, X_D^1)]\bar{\lambda}_r(x_C, X_D^1).$$

It is the generator of a PDP whose characteristics have been averaged in fast variable X_D^2.

Equation (15) gives, as $N_k \to \infty$:

(17)
$$\mathbb{E}_{\tilde{P}_x}\left(\left(f(x_t) - f(x_0) - \int_0^t \tilde{\mathcal{A}}_\infty f(x_s) \, ds\right)\psi(x_{t_1}, \ldots, x_{t_n})\right)$$

$$= \mathbb{E}_{\tilde{P}_x}\left(\left(f(x_r) - f(x_0) - \int_0^r \tilde{\mathcal{A}}_\infty f(x_s) \, ds\right)\psi(x_{t_1}, \ldots, x_{t_n})\right).$$

Recall that we have assumed that f is C^2 with respect to x_C. It is easy to prove that (17) holds for any $f \in \mathcal{E}$ by choosing a sequence (f_n) in $\mathcal{E}$ such that, for all n, f_n is C^2 with bounded second derivatives and f_n and ∇f_n are uniformly bounded and converge pointwise to f and ∇f.

The measure $\tilde{P}_x$ is then a solution of the martingale problem associated to the generator $\tilde{\mathcal{A}}_\infty$ on the domain of the functions in $\mathcal{E}$ independent of X_D^2.

By Theorem 2.5, the martingale problem has then a unique solution. We can then deduce that $(\tilde{P}_{x,N})_N$ converges weakly to P_{x_0}, the law of the PDP defined in Theorem 5.1.

Step 3: *Conclusion.* Since the jump rates are assumed to be bounded and Lipschitz on bounded sets of E uniformly in X_D^2, we can remove the more restrictive boundedness assumption on the jump rates as in Section 3. We then conclude as in Section 3. Theorem 5.1 is proved. $\square$

6. Discontinuous PDP and singular switching.

In this section, our system has two time scales. The switching between fast and slow dynamics is governed by the state of a discrete variable. For simplicity, we suppose now that there is no other discrete variable and denote by θ, taking values in $\{0, 1\}$, this unique discrete variable which governs the time scale of the continuous variables.

The number of molecules belonging to species C is again assumed to be of order N, and we continue to write $x_C = \frac{X_C}{N}$.

Let us split R_{DC} into two sets $R_{DC} = S_1 \cup S_0$. Reactions in S_1 have rates scaling with N. They change X_C ($\gamma_r^C \neq 0$) but do not change the value of θ. When $\theta = 0$, the rates of all reactions from $\mathcal{R}_C \cup S_1$ is of order N. When $\theta = 1$, some reactions from S_1 (inactive for $\theta = 0$) become active with much faster rates of order $\frac{N}{\epsilon}$, where ϵ is a new parameter, assumed to be small. On the other hand, for the same models, the reactions in S_0 do not change X_C ($\gamma_r^C = 0$, $r \in S_0$) but do change the value of θ and have rates that can depend on X_C (ensuring feedback control of X_C).

Then for $r \in S_1$, we introduce the rescaled reaction rates

$$\tilde{\lambda}_r(x_C, 0) = \frac{1}{N}\lambda_r(x_C, 0),$$

$$\tilde{\lambda}_r(x_C, 1) = \frac{\epsilon}{N}\lambda_r(x_C, 1).$$

Similarly for $r \in R_C$

$$\tilde{\lambda}_r(x_C) = \frac{1}{N}\lambda_r(x_C).$$

For models in which the discrete variables take only two states, $\theta \in \{0, 1\}$, S_0 contains two reactions. The rate of the reaction in S_0 changing θ from state 1 to state 0 is assumed to be of order $\frac{1}{\epsilon}$ and we also set

$$\tilde{\lambda}_\theta(x_C, 1) = \epsilon\lambda_\theta(x_C, 1).$$

The rate of the opposite reaction changing θ from state 0 to state 1 is assumed to be of order 1 and is written $\lambda_\theta(x_C, 0)$.

In these conditions, the generator of the process has the following form:

$$\tilde{\mathcal{A}}_{N,\epsilon} f(x_C, 0)$$

$$= \sum_{r \in S_1} \left[f\left(x_C + \frac{1}{N}\gamma_r^C, 0\right) - f(x_C, 0)\right] N\tilde{\lambda}_r(x_C, 0)$$

$$+ \sum_{r \in \mathcal{R}_C} \left[f\left(x_C + \frac{1}{N}\gamma_r^C, 0\right) - f(x_C, 0)\right] N\tilde{\lambda}_r(x_C)$$

$$+ [f(x_C, 1) - f(x_C, 0)]\lambda_\theta(x_C, 0)$$

and

$$\tilde{\mathcal{A}}_{N,\epsilon} f(x_C, 1)$$

$$= \frac{1}{\epsilon} \sum_{r \in S_1} \left[f\left(x_C + \frac{1}{N}\gamma_r^C, 1\right) - f(x_C, 1)\right] N\tilde{\lambda}_r(x_C, 1)$$

$$+ \sum_{r \in \mathcal{R}_C} \left[f\left(x_C + \frac{1}{N}\gamma_r^C, 1\right) - f(x_C, 1)\right] N\tilde{\lambda}_r(x_C)$$

$$+ \frac{1}{\epsilon}[f(x_C, 0) - f(x_C, 1)]\tilde{\lambda}_\theta(x_C, 1).$$

At the limit (for high N and small ϵ), we will show that the discrete process θ inducing kicks in the continuous variable x_C is almost surely equal to 0. We are interested in the limit distribution of the process x_C.

We introduce the vector fields

$$F(x_C) = \sum_{r \in S_1} \gamma_r^C \tilde{\lambda}_r(x_C, 0) + \sum_{r \in \mathcal{R}_C} \gamma_r^C \tilde{\lambda}_r(x_C),$$

$$F_1(x_C) = \sum_{r \in S_1} \gamma_r^C \tilde{\lambda}_r(x_C, 1),$$

and the flow $\phi_1(t, x_C)$ associated to F_1.

We now state the main result of this section.

THEOREM 6.1. *Let $x^{N,\epsilon} = (x_C^{N,\epsilon}, \theta^{N,\epsilon})$ be a jump Markov process with values in $\mathbb{R}^{M_C} \times \{0, 1\}$, with generator $\tilde{\mathcal{A}}_{N,\epsilon}$, starting at $x^{N,\epsilon}(0) = (x_C^{N,\epsilon}(0), \theta^{N,\epsilon}(0))$. Assume that the following assumptions hold:*

(i) *$x_C^{N,\epsilon}(0)$ converges in distribution to $x_C(0)$ in $\mathbb{R}^{M_C}$ as $N \to \infty$ and $\epsilon \to 0$.*
(ii) *$\theta^{N,\epsilon}(0)$ converges in distribution to 0 as $N \to \infty$ and $\epsilon \to 0$.*
(iii) *There exists $\alpha > 0$, such that*

$$\tilde{\lambda}_\theta(x_C, 1) \geq \alpha \qquad \forall x_C \in \mathbb{R}^{M_C}.$$

Assume also that either (iv) *or* (v) *holds:*

(iv) *The jump rates* $\tilde{\lambda}_r(\cdot)$, $r \in R_C$, *and* $\tilde{\lambda}_r(\cdot, 0)$, $\tilde{\lambda}_r(\cdot, 1)$, $r \in S_1$, *have at most linear growth and* $\lambda_\theta(x_C, 0)$ *is bounded*: *there exists* $M_1 \geq 0$ *such that, for* $x_C \in \mathbb{R}^{M_C}$,

$$\tilde{\lambda}_r(x_C) \leq M_1(|x_C| + 1), \qquad r \in R_C,$$
$$\tilde{\lambda}_r(x_C, 0) \leq M_1(|x_C| + 1), \qquad r \in S_1,$$
$$\tilde{\lambda}_r(x_C, 1) \leq M_1(|x_C| + 1), \qquad r \in S_1,$$

and

$$\lambda_\theta(x_C, 0) \leq M_1.$$

(v) *The jump rates* $\tilde{\lambda}_r(\cdot)$, $r \in R_C$, *and* $\tilde{\lambda}_r(\cdot, 0)$, $r \in S_1$, $\lambda_\theta(x_C, 0)$ *have at most linear growth, and* $\tilde{\lambda}_r(\cdot, 1)$, $r \in S_1$, *is bounded*: *there exists* $M_2 \geq 0$ *such that, for* $x_C \in \mathbb{R}^{M_C}$,

$$\tilde{\lambda}_r(x_C) \leq M_2(|x_C| + 1), \qquad r \in R_C,$$
$$\tilde{\lambda}_r(x_C, 0) \leq M_2(|x_C| + 1), \qquad r \in S_1,$$
$$\lambda_\theta(x_C, 0) \leq M_2(|x_C| + 1)$$

and

$$\tilde{\lambda}_r(x_C, 1) \leq M_2, \qquad r \in S_1.$$

Then $(x_C^{N,\epsilon})$ *converges in distribution in* $L^p([0, T]; \mathbb{R}^{M_C})$, *for any* $\infty > p \geq 1$, *to the PDP whose generator is given by*

$$\tilde{\mathcal{A}}_\infty \varphi(x_C) = F(x_C) \cdot \nabla_{x_C} \varphi(x_C)$$
$$+ \lambda_\theta(x_C, 0) \int_0^\infty (\varphi(\phi_1(t, x_C)) - \varphi(x_C)) \tilde{\lambda}_\theta(\phi_1(t, x_C), 1)$$
$$\times e^{-\int_0^t \tilde{\lambda}_\theta(\phi_1(s, x_C), 1) \, ds} \, dt$$

for every $\varphi \in C_b^1(\mathbb{R}^{M_C})$.

REMARK 6.2. The PDP associated to the generator $\tilde{\mathcal{A}}_\infty$ has discontinuities which are not present for the process $x_C^{N,\epsilon}$. They only appear at the limit $\epsilon \to 0$. Therefore, the convergence does not occur in $D(\mathbb{R}^+; \mathbb{R}^{M_C})$; indeed creation of discontinuities cannot occur in $D(\mathbb{R}^+; \mathbb{R}^{M_C})$. As shown below, it is possible to show tightness, then weak convergence of the process $x_C^{N,\epsilon}$ in $L^p([0, T]; \mathbb{R}^{M_C})$. Unfortunately, this topology does not imply a uniform bound on the processes and we cannot get rid easily of the boundedness assumptions on the reaction rates as was done in the preceding sections.

Therefore, we have to assume the extra assumption (iv) or (v). These are technical. They are used at the end of the proof to obtain a bound on (x_C^N) in $L^\infty([0, T]; \mathbb{R}^{M_C})$. The linear growth of the jump rate is verified for a large class of models in molecular biology.

The boundedness of $\lambda_\theta(x_C, 0)$ in (iv) is a restrictive assumption but it is satisfied in the case that $\lambda_\theta(x_C, 0)$ does not depend on x_C. This applies to systems with negative feedback such as self-repressed gene promoters, for which only the transition of θ from 1 to 0 is controlled by the protein repressor.

Also, the boundedness assumption of $\tilde{\lambda}_r(\cdot, 1)$ for $r \in S_1$ in (v) is restrictive. Such assumption is satisfied when $\tilde{\lambda}_r(\cdot, 1)$ do not depend on x_C. Lifting the boundedness of $\lambda_\theta(x_C, 0)$ in (v) is necessary for systems with positive feedback, such as self-induced gene promoters. In such cases, considering that $\tilde{\lambda}_r(\cdot, 1)$ do not depend on x_C stands for the rather usual situation when the concentration of the regulator influences the transitions between various states of the promoter, but not the production rates of these states.

Thus, in spite of the above restrictions, the results of this section cover a large class of models [e.g., all the models analyzed in Crudu, Debussche and Radulescu (2009)].

PROOF OF THEOREM 6.1. The proof is divided into four steps. Again, we start with the extra assumption that all the jump rates are uniformly bounded together with their derivatives with respect to x_C. This assumption will be relaxed in step 5.

Step 1: *Limit of* $\theta^{N,\epsilon}$. We begin with the test-function

$$f(x_C, 0) = 0, \qquad f(x_C, 1) = 1 \qquad \forall x_C \in \mathbb{R}^{M_C}.$$

Then, $\tilde{\mathcal{A}}_{N,\epsilon} f(x_C, \theta) = \lambda_\theta(x_C, 0)\mathbb{1}_{\theta=0} - \frac{1}{\epsilon}\tilde{\lambda}_\theta(x_C, 1)\mathbb{1}_{\theta=1}$ and

$$M(t) = \mathbb{1}_{\theta^{N,\epsilon}(t)=1} - \mathbb{1}_{\theta^{N,\epsilon}(0)=1}$$
$$- \int_0^t \left(\lambda_\theta(x_C^{N,\epsilon}(s), 0)\mathbb{1}_{\theta^{N,\epsilon}(s)=0} - \frac{1}{\epsilon}\tilde{\lambda}_\theta(x_C^{N,\epsilon}(s), 1)\mathbb{1}_{\theta^{N,\epsilon}(s)=1} \right) ds$$

is a martingale. We then deduce, for every $T > 0$ and $t \in [0, T]$,

(18) $$\frac{\alpha}{\epsilon}\mathbb{E}\left(\int_0^t \mathbb{1}_{\theta^{N,\epsilon}(s)=1} \, ds \right) \leq M_\lambda T + 1,$$

where M_λ is the supremum bound of $\lambda_\theta(\cdot, 0)$. Then, $\theta^{N,\epsilon} \to 0$ in $L^1(\Omega \times [0, T]; \{0, 1\})$ when $\epsilon \to 0$ and $N \to \infty$. In particular, $(\theta^{N,\epsilon})_{N,\epsilon}$ is tight in $L^1([0, T; \{0, 1\})$.

Step 2: *Tightness of* $x_C^{N,\epsilon}$. For all $t \in [0, T]$, we write

$$x_C^{N,\epsilon}(t) = x_C^{N,\epsilon}(0) + \sum_{r \in \mathcal{R}_C} \frac{\gamma_r^C}{N} Y_r \left(\int_0^t N\tilde{\lambda}_r(x_C^{N,\epsilon}(s)) \, ds \right)$$

$$+ \sum_{r \in S_1} \frac{\gamma_r^C}{N} Y_r \left(\int_0^t N\tilde{\lambda}_r(x_C^{N,\epsilon}(s), 0)\mathbb{1}_{\theta^{N,\epsilon}(s)=0} \right.$$

$$\left. + \frac{N}{\epsilon}\tilde{\lambda}_r(x_C^{N,\epsilon}(s), 1)\mathbb{1}_{\theta^{N,\epsilon}(s)=1} \, ds \right),$$

where $(Y_r)_{r \in R_C \cup S_1}$ are independent standard Poisson processes. Then

$$|x_C^{N,\epsilon}(t)| \leq |x_C^{N,\epsilon}(0)| + \sum_{r \in \mathcal{R}_C} \frac{|\gamma_r^C|}{N} Y_r(NMT)$$

(19)

$$+ \sum_{r \in S_1} \frac{|\gamma_r^C|}{N} Y_r\left(NM\left(T + \frac{1}{\epsilon} \int_0^T \mathbb{1}_{\theta^{\epsilon,N}(s)=1}\, ds\right)\right),$$

where M is the supremum bound of $\tilde{\lambda}_r, r \in R_C \cup S_1$. Since $\sup_{t \in [0,T]} \frac{1}{N} Y_r(Nt)$ is almost surely bounded, using (18) and (19), we have for all $K > 0$, for N large enough and ϵ small enough:

$$\mathbb{P}\left(\sup_{[0,T]} |x_C^{\epsilon,N}(t)| \geq K\right) \leq \varepsilon(K),$$

where $\varepsilon(K) \to 0$ if $K \to \infty$.

We introduce $\mathrm{BV}(0,T)$ the space of functions of bounded variations on $[0,T]$, with its norm defined for $f \in \mathrm{BV}(0,T)$ by

$$\|f\|_{\mathrm{BV}(0,T)} = \|f\|_{L^1([0,T])}$$

$$+ \sup\left\{\sum_i |f(t_{i+1}) - f(t_i)|, (t_i)_i \text{ finite subdivision of } [0,T]\right\}.$$

Since the processes Y_r are nondecreasing, we obtain for a subdivision of $[0,T]$, denoted by $0 = t_0 < t_1 < \cdots < t_n = T$,

$$\sum_{i=0}^{n-1} |x_C^{\epsilon,N}(t_{i+1}) - x_C^{\epsilon,N}(t_i)|$$

$$\leq \sum_{i=0}^{n-1} \sum_{r \in R_C} \frac{|\gamma_r^C|}{N}\left[Y_r\left(\int_0^{t_{i+1}} N\tilde{\lambda}_r(x_C^{\epsilon,N}(s))\, ds\right)\right.$$

$$\left. - Y_r\left(\int_0^{t_i} N\tilde{\lambda}_r(x_C^{\epsilon,N}(s), 0)\, ds\right)\right]$$

$$+ \sum_{i=0}^{n-1} \sum_{r \in S_1} \frac{|\gamma_r^C|}{N}\left[Y_r\left(\int_0^{t_{i+1}} N\tilde{\lambda}_r(x_C^{\epsilon,N}(s), 0)\mathbb{1}_{\theta^{\epsilon,N}(s)=0}\right.\right.$$

$$\left. + \frac{N}{\epsilon}\tilde{\lambda}_r(x_C^{\epsilon,N}(s), 1)\mathbb{1}_{\theta^{\epsilon,N}(s)=1}\, ds\right)$$

$$- Y_r\left(\int_0^{t_i} N\tilde{\lambda}_r(x_C^{\epsilon,N}(s), 0)\mathbb{1}_{\theta^{\epsilon,N}(s)=0}\right.$$

$$\left.\left. + \frac{N}{\epsilon}\tilde{\lambda}_r(x_C^{\epsilon,N}(s), 1)\mathbb{1}_{\theta^{\epsilon,N}(s)=1}\, ds\right)\right]$$

$$= \sum_{r \in R_C} \frac{|\gamma_r^C|}{N} Y_r \left(\int_0^T N \tilde{\lambda}_r(x_C^{\epsilon,N}(s), 0) \, ds \right)$$

$$+ \sum_{r \in S_1} \frac{|\gamma_r^C|}{N} Y_r \left(\int_0^T N \tilde{\lambda}_r(x_C^{\epsilon,N}(s), 0) \mathbb{1}_{\theta^{\epsilon,N}(s)=0} \right.$$

$$\left. + \frac{N}{\epsilon} \tilde{\lambda}_r(x_C^{\epsilon,N}(s), 1) \mathbb{1}_{\theta^{\epsilon,N}(s)=1} \, ds \right).$$

It is then easy to prove that for all $K > 0$, for N large enough and ϵ small enough:

$$\mathbb{P}\big(\|x_C^{N,\epsilon}\|_{\mathrm{BV}(0,T)} \geq K\big) \leq \varepsilon(K),$$

where $\varepsilon(K) \to 0$ if $K \to \infty$.

The set $\{f \in \mathrm{BV}(0, T), \text{ such that } \|f\|_{\mathrm{BV}(0,T)} \leq K\}$ is relatively compact in $L^1([0, T])$ [see, e.g., Giusti (1984)], and the set

$$\{f \in \mathrm{BV}(0, T), \text{ such that } \|f\|_\infty \leq K \text{ and } \|f\|_{\mathrm{BV}(0,T)} \leq K\}$$

is relatively compact in $L^p([0, T])$ for $1 < p < \infty$. We then conclude that the family of processes $(x_C^{N,\epsilon})_{N,\epsilon}$ is tight in $L^p([0, T])$, $1 \leq p < \infty$.

Step 3: *Identification of the limit distribution of $x_C^{N,\epsilon}$.* Since we are interested in the limit distribution of $x_C^{N,\epsilon}$, we introduce test-functions depending only on the continuous variable $x_C \in \mathbb{R}^{M_C}$. We then define

$$f(x_C, 0) = \varphi(x_C), \qquad x_C \in \mathbb{R}^{M_C},$$

where $\varphi \in C_b^1(\mathbb{R}^{M_C})$.

We want to define $f(\cdot, 1)$ such that

$$\sum_{r \in S_1} \left[f\left(x_C + \frac{1}{N}\gamma_r^C, 1\right) - f(x_C, 1) \right] N\tilde{\lambda}_r(x_C, 1) - \tilde{\lambda}_\theta(x_C, 1) f(x_C, 1)$$

$$(20) \qquad + \epsilon \sum_{r \in \mathcal{R}_C} \left[f\left(x_C + \frac{1}{N}\gamma_r^C, 1\right) - f(x_C, 1) \right] N\tilde{\lambda}_r(x_C)$$

$$= -\tilde{\lambda}_\theta(x_C, 1)\varphi(x_C).$$

Drawing inspiration from the preceding section, we introduce the process $y_{N,\epsilon}(\cdot, x)$ starting from x, and whose generator is

$$A_{N,\epsilon}\psi(y) = \sum_{r \in S_1} \left[\psi\left(y + \frac{1}{N}\gamma_r^C\right) - \psi(y) \right] N\tilde{\lambda}_r(y, 1)$$

$$+ \epsilon \sum_{r \in R_C} \left[\psi\left(y + \frac{1}{N}\gamma_r^C\right) - \psi(y) \right] N\tilde{\lambda}_r(y).$$

By a slight extension of a result of Kurtz (1971), we know that for all $x \in \mathbb{R}^{M_C}$, $T > 0$ and $\delta > 0$,

$$(21) \qquad \mathbb{P}\left(\sup_{s \leq T} |y_{N,\epsilon}(s, x) - \phi_1(s, x)| > \delta\right) \xrightarrow[N \to \infty, \epsilon \to 0]{} 0.$$

Remember that $\phi_1(\cdot, x)$ is the flow associated to the vector field $F_1(x) = \sum_{r \in R_{DC}} \gamma_r^C \tilde{\lambda}_r(x, 1)$ and starting at x.

We also introduce the semigroup $(P_t^{N,\epsilon})_{t \geq 0}$ defined on $\mathcal{B}_b(\mathbb{R}^{M_C})$ by

$$P_t^{N,\epsilon} \psi(x) := \mathbb{E}\left(\psi(y_{N,\epsilon}(t, x)) e^{-\int_0^t \tilde{\lambda}_\theta(y_{N,\epsilon}(s,x),1)\,ds}\right).$$

It is classical that $(P_t^{N,\epsilon})_{t \geq 0}$ satisfies the semigroup property and that

$$\frac{d}{dt} P_t^{N,\epsilon} \psi(x) = A_{N,\epsilon} P_t^{N,\epsilon} \psi(x) - \tilde{\lambda}_\theta(x, 1) P_t^{N,\epsilon} \psi(x),$$

$$x \in \mathbb{R}^{M_C}, \psi \in \mathcal{C}_b^1(\mathbb{R}^{M_C}), t \in \mathbb{R}^+.$$

Then we propose:

$$f(x_C, 1) := \int_0^\infty P_s^{N,\epsilon}(\tilde{\lambda}_\theta(\cdot, 1)\varphi)(x_C)\,ds.$$

Since $\tilde{\lambda}_\theta(\cdot, 1)$ is bounded below by $\alpha > 0$, $f(\cdot, 1)$ is well defined. Moreover $f(\cdot, 1)$ satisfies (20). Indeed,

$$A_{N,\epsilon} f(x_C, 1) - \tilde{\lambda}_\theta(x_C, 1) f(x_C, 1)$$

$$= \int_0^\infty A_{N,\epsilon} P_t^{N,\epsilon}(\tilde{\lambda}_\theta(\cdot, 1)\varphi)(x_C) - \tilde{\lambda}_\theta(x_C, 1) P_t^{N,\epsilon}(\tilde{\lambda}_\theta(\cdot, 1)\varphi)(x_C)\,dt$$

$$= \int_0^\infty \frac{d}{dt} P_t^{N,\epsilon}(\tilde{\lambda}_\theta(\cdot, 1)\varphi)(x_C)\,dt$$

$$= \lim_{t \to \infty} P_t^{N,\epsilon}(\tilde{\lambda}_\theta(\cdot, 1)\varphi)(x_C) - (\tilde{\lambda}_\theta(\cdot, 1)\varphi)(x_C)$$

$$= -(\tilde{\lambda}_\theta(\cdot, 1)\varphi)(x_C).$$

This test-function f satisfies:

$$\tilde{\mathcal{A}}_{N,\epsilon} f(x_C, 1) = 0$$

and

$$\tilde{\mathcal{A}}_{N,\epsilon} f(x_C, 0) = \sum_{r \in R_{DC}} \left[f\left(x_C + \frac{1}{N}\gamma_r^C, 0\right) - f(x_C, 0)\right] N\tilde{\lambda}_r(x_C, 0)$$

$$+ \left[\int_0^\infty P_t^{N,\epsilon}(\tilde{\lambda}_\theta(\cdot, 1)\varphi)(x_C)\,dt - \varphi(x_C)\right] \lambda_\theta(x_C, 0).$$

Since $f \in C_b(E)$, f is in the domain of $\tilde{\mathcal{A}}_{N,\epsilon}$, for all N, ϵ. Then, for all $0 \le t_1, \ldots, t_\ell \le t$ and $\psi \in C_b(\mathbb{R}^{M_C \times \ell})$

$$
(22) \quad \mathbb{E}\Bigg(\Bigg[f(x_C^{N,\epsilon}(t), \theta^{N,\epsilon}(t)) - f(x_C^{N,\epsilon}(0), \theta^{N,\epsilon}(0)) \\
- \int_0^t \tilde{\mathcal{A}}_{N,\epsilon} f(x_C^{N,\epsilon}(s), 0) \mathbb{1}_{\theta^{N,\epsilon}(s)=0}\, ds \Bigg] \psi(x_C^{N,\epsilon}(t_1), \ldots, x_C^{N,\epsilon}(t_\ell)) \Bigg) \\
= 0.
$$

We now are looking for a possible formulation for the limiting generator. For all $x \in \mathbb{R}^{M_C}$ and uniformly in $t \in [0, T]$, for all $T > 0$, we have

$$
P_t^{N,\epsilon}(\tilde{\lambda}_\theta(\cdot, 1)\varphi)(x) \xrightarrow[N \to \infty, \epsilon \to 0]{} (\tilde{\lambda}_\theta(\cdot, 1)\varphi)(\phi_1(t, x)) e^{-\int_0^t \tilde{\lambda}_\theta(\phi_1(s,x),1)\, ds}.
$$

Indeed,

$$
P_t^{N,\epsilon}(\tilde{\lambda}_\theta(\cdot, 1)\varphi)(x) - (\tilde{\lambda}_\theta(\cdot, 1)\varphi)(\phi_1(t, x)) e^{-\int_0^t \tilde{\lambda}_\theta(\phi_1(s,x),1)\, ds} \\
= \mathbb{E}\Big((\tilde{\lambda}_\theta(\cdot, 1)\varphi)(y_{N,\epsilon}(t, x)) e^{-\int_0^t \tilde{\lambda}_\theta(y_{N,\epsilon}(s,x),1)\, ds} \\
- (\tilde{\lambda}_\theta(\cdot, 1)\varphi)(\phi_1(t, x)) e^{-\int_0^t \tilde{\lambda}_\theta(y_{N,\epsilon}(s,x),1)\, ds}\Big) \\
+ \mathbb{E}\Big((\tilde{\lambda}_\theta(\cdot, 1)\varphi)(\phi_1(t, x)) e^{-\int_0^t \tilde{\lambda}_\theta(y_{N,\epsilon}(s,x),1)\, ds} \\
- (\tilde{\lambda}_\theta(\cdot, 1)\varphi)(\phi_1(t, x)) e^{-\int_0^t \tilde{\lambda}_\theta(\phi_1(s,x),1)\, ds}\Big) \\
= a + b
$$

and it can be shown that for all $\delta > 0$, and $t \le T$,

$$
|a| \le \delta L_{\tilde{\lambda}_\theta(\cdot,1)\varphi} + 2\|\tilde{\lambda}_\theta(\cdot, 1)\varphi\|_\infty \mathbb{P}\Big(\sup_{s \le T} |y_{N,\epsilon}(s, x) - \phi_1(s, x)| > \delta\Big)
$$

and

$$
|b| \le \|\tilde{\lambda}_\theta(\cdot, 1)\varphi\|_\infty \Big(T\delta L_{\tilde{\lambda}_\theta(\cdot,1)} + 2\mathbb{P}\Big(\sup_{s \le T} |y_{N,\epsilon}(s, x) - \phi_1(s, x)| > \delta\Big)\Big),
$$

where $L_{\tilde{\lambda}_\theta(\cdot,1)\varphi}$ and $L_{\tilde{\lambda}_\theta(\cdot,1)}$ are the Lipschitz constants of the functions $\tilde{\lambda}_\theta(\cdot, 1)\varphi$ and $\tilde{\lambda}_\theta(\cdot, 1)$.

By (21) and the dominated convergence theorem, we deduce that

$$
\tilde{\mathcal{A}}_{N,\epsilon} f(x_C, 0) \xrightarrow[N \to \infty, \epsilon \to 0]{} \tilde{\mathcal{A}}_\infty \varphi(x_C).
$$

The tightness in $L^p([0, T]) \times L^1([0, T]; \{0, 1\})$ of the family of processes $(x_C^{N,\epsilon}, \theta^{N,\epsilon})_{N,\epsilon}$, and the Skorohod representation theorem imply the existence of a subsequence which converges almost surely in $L^p([0, T]) \times L^1([0, T]; \{0, 1\})$

to $(x_C, 0) \in L^p([0, T]) \times L^1([0, T]; \{0, 1\})$. This almost sure convergence implies that a new subsequence can be extracted such that

$$(x_C^{N_k, \epsilon_k}(t), \theta^{N_k, \epsilon_k}) \xrightarrow[k \to \infty]{} (x_C(t), 0)$$

almost surely, and for almost all $t \in [0, T]$. To simplify notation, this subsequence is still denoted by $x_C^{N, \epsilon}$.

Taking $N \to \infty$ and $\epsilon \to 0$ in (22), we deduce by the dominated convergence theorem, the boundedness of f and ψ, and the hypothesis on $x_C^{N, \epsilon}(0)$ and $\theta^{N, \epsilon}(0)$:

$$\mathbb{E}\big([f(x_C^{N, \epsilon}(t), \theta^{N, \epsilon}(t)) - f(x_C^{N, \epsilon}(0), \theta^{N, \epsilon}(0))]\psi(x_C^{N, \epsilon}(t_1), \dots, x_C^{N, \epsilon}(t_\ell)))$$

$$\xrightarrow[N \to \infty, \epsilon \to 0]{} \mathbb{E}\big([\varphi(x_C(t)) - \varphi(x_C(0))]\psi(x_C(t_1), \dots, x_C(t_\ell)))$$

for almost all $t \in [0, T]$ and almost all $0 \le t_1, \dots, t_l \le t$. We now prove that

$$(23) \qquad \tilde{\mathcal{A}}_{N, \epsilon} f(x_C^{N, \epsilon}(s), 0)\mathbb{1}_{\theta^{N, \epsilon}(s) = 0} \xrightarrow[N \to \infty, \epsilon \to 0]{} \tilde{\mathcal{A}}_\infty \varphi(x_C(s))$$

almost surely, and for almost all $s \in [0, T]$.

First, for $s \in [0, T]$, $t \in \mathbb{R}^+$, we write

$$P_t^{N, \epsilon}(\tilde{\lambda}_\theta(\cdot, 1)\varphi)(x_C^{N, \epsilon}(s)) - (\tilde{\lambda}_\theta(\cdot, 1)\varphi)(\phi_1(t, x_C(s)))e^{-\int_0^t \tilde{\lambda}_\theta(\phi_1(u, x_C(s)), 1)\, du}$$

$$= \mathbb{E}\big[(\tilde{\lambda}_\theta(\cdot, 1)\varphi)(y_{N, \epsilon}(t, x_C^{N, \epsilon}(s)))e^{-\int_0^t \tilde{\lambda}_\theta(y_{N, \epsilon}(u, x_C^{N, \epsilon}(s)), 1)\, du}\big]$$

$$\quad - \mathbb{E}\big[(\tilde{\lambda}_\theta(\cdot, 1)\varphi)(\phi_1(t, x_C(s)))e^{-\int_0^t \tilde{\lambda}_\theta(y_{N, \epsilon}(u, x_C^{N, \epsilon}(s)), 1)\, du}\big]$$

$$\quad + \mathbb{E}\big[(\tilde{\lambda}_\theta(\cdot, 1)\varphi)(\phi_1(t, x_C(s)))e^{-\int_0^t \tilde{\lambda}_\theta(y_{N, \epsilon}(u, x_C^{N, \epsilon}(s)), 1)\, du}\big]$$

$$\quad - (\tilde{\lambda}_\theta(\cdot, 1)\varphi)(\phi_1(t, x_C(s)))e^{-\int_0^t \tilde{\lambda}_\theta(\phi_1(u, x_C(s)), 1)\, du}$$

$$= a_1 + b_1.$$

For all $\delta > 0$, $T_1 > 0$, and $0 \le t \le T_1$,

$$|a_1| \le \delta L_{\tilde{\lambda}_\theta(\cdot, 1)\varphi}$$

$$\quad + 2\|\tilde{\lambda}_\theta(\cdot, 1)\varphi\|_\infty \mathbb{P}\Big(\sup_{u \le T_1} |y_{N, \epsilon}(u, x_C^{N, \epsilon}(s)) - \phi_1(u, x_C(s))| > \delta\Big),$$

and if $t \ge T_1$,

$$|a_1| \le 2\|\tilde{\lambda}_\theta(\cdot, 1)\varphi\|_\infty e^{-\alpha T_1}.$$

Similarly, for $0 \le t \le T_1$,

$$|b_1| \le \|\tilde{\lambda}_\theta(\cdot, 1)\varphi\|_\infty$$

$$\quad \times \Big(T_1 \delta L_{\tilde{\lambda}_\theta(\cdot, 1)} + 2\mathbb{P}\Big(\sup_{u \le T_1} |y_{N, \epsilon}(s, x_C^{N, \epsilon}(s)) - \phi_1(s, x_C(s))| > \delta\Big)\Big),$$

and if $t \geq T_1$,

$$|b_1| \leq 2\|\tilde{\lambda}_\theta(\cdot, 1)\varphi\|_\infty e^{-\alpha T_1}.$$

We choose T_1 sufficiently large and then δ small enough and have almost surely, for all $s \in [0, T]$, $t \in \mathbb{R}^+$:

$$P_t^{N,\epsilon}(\tilde{\lambda}_\theta(\cdot, 1)\varphi)(x_C^{N,\epsilon}(s)) - (\tilde{\lambda}_\theta(\cdot, 1)\varphi)(\phi_1(t, x_C(s)))e^{-\int_0^t \tilde{\lambda}_\theta(\phi_1(u, x_C(s)), 1)\, du}$$

$$\xrightarrow[N\to\infty, \epsilon\to 0]{} 0.$$

By the dominated convergence theorem, we have almost surely, for almost all $s \in [0, T]$,

$$\int_0^\infty P_t^N(\tilde{\lambda}_\theta(\cdot, 1)\varphi)(x_C^{N,\epsilon}(s))\, dt$$

$$\xrightarrow[N\to\infty, \epsilon\to 0]{} \int_0^\infty (\tilde{\lambda}_\theta(\cdot, 1)\varphi)(\phi_1(t, x_C(s)))e^{-\int_0^t \tilde{\lambda}_\theta(\phi_1(u, x_C(s)), 1)\, du}\, dt.$$

Since $\varphi \in C_b^1(\mathbb{R}^{M_C})$, we obtain (23).

By the dominated convergence theorem, we conclude finally that

$$\mathbb{E}\left(\left[\varphi(x_C(t)) - \varphi(x_C(0)) - \int_0^t \tilde{\mathcal{A}}_\infty \varphi(x_C(s))\, ds\right]\psi(x_C(t_1), \ldots, x_C(t_\ell))\right) = 0$$

for almost all $t, t_1, \ldots, t_l$ such that $t \in [0, T]$ and $0 \leq t_1 \cdots t_l \leq t$.

Step 4: *Uniqueness of the solution of the martingale problem.* By Theorem 2.5, proved in a weaker sense in the Appendix, we conclude that the limit x_C, more exactly a version of x_C, is the PDP whose generator is given by $\tilde{\mathcal{A}}_\infty$.

Step 5: *Conclusion.* We now relax the assumptions on the boundedness of the jump rates and their derivatives. Without loss of generality, we assume that $x_C^{N,\epsilon}(0)$ are deterministic. We derive an estimate on the supremum of $|x_C^{N,\epsilon}(t)|$ on $[0, T]$ for $T > 0$.

We first assume that (iv) holds. We again write that

$$M(t) = \mathbb{1}_{\theta^{N,\epsilon}(t)=1} - \mathbb{1}_{\theta^{N,\epsilon}(0)=1} - \int_0^t \lambda_\theta(x_C^{N,\epsilon}(s), 0)\mathbb{1}_{\theta^{N,\epsilon}(s)=0}$$

$$- \frac{1}{\epsilon}\tilde{\lambda}_\theta(x_C^{N,\epsilon}(s), 1)\mathbb{1}_{\theta^{N,\epsilon}(s)=1}\, ds$$

is a martingale. We have as in step 1:

$$(24) \qquad \frac{\alpha}{\epsilon}\mathbb{E}\left(\int_0^T \mathbb{1}_{\theta^{N,\epsilon}(s)=1}\, ds\right) \leq 1 + M_1 T.$$

For all $t \in [0, T]$, we write using the same notations as above:

$$x_C^{N,\epsilon}(t) = x_C^{N,\epsilon}(0) + \sum_{r \in R_C} \frac{\gamma_r^C}{N} Y_r \left(\int_0^t N \tilde{\lambda}_r(x_C^{N,\epsilon}(s)) \, ds \right)$$

$$+ \sum_{r \in S_1} \frac{\gamma_r^C}{N} Y_r \left(\int_0^t N \tilde{\lambda}_r(x_C^{N,\epsilon}(s), 0) \mathbb{1}_{\theta^{N,\epsilon}(s)=0} \right.$$

$$\left. + \frac{N}{\epsilon} \tilde{\lambda}_r(x_C^{N,\epsilon}(s), 1) \mathbb{1}_{\theta^{N,\epsilon}(s)=1} \, ds \right),$$

where $(Y_r)_{r \in R_{DC}}$ are independent standard Poisson processes, and thanks to the linear growth assumptions on the jump rates

$$|x_C^{N,\epsilon}(t)| \leq |x_C^{N,\epsilon}(0)| + \sum_{r \in R_C} \frac{|\gamma_r^C|}{N} Y_r \left(N M_1 \int_0^t (|x_C^{N,\epsilon}(s)| + 1) \, ds \right)$$

$$+ \sum_{r \in S_1} \frac{|\gamma_r^C|}{N} Y_r \left(N M_1 \int_0^t (|x_C^{N,\epsilon}(s)| + 1) \, ds \right.$$

$$\left. + M_1 \frac{N}{\epsilon} \int_0^t (|x_C^{N,\epsilon}(s)| + 1) \mathbb{1}_{\theta^{N,\epsilon}(s)=1} \, ds \right).$$

By the law of large numbers $\frac{Y_r(t)}{t}$ are almost surely convergent to 1. They are therefore almost surely bounded on $[1, \infty)$. We deduce that there exists a random C_1 such that

$$|x_C^{N,\epsilon}(t)| \leq |x_C^{N,\epsilon}(0)|$$

$$+ C_1 \left(\int_0^t (|x_C^{N,\epsilon}(s)| + 1) \, ds + \frac{1}{\epsilon} \int_0^t (|x_C^{N,\epsilon}(s)| + 1) \mathbb{1}_{\theta^{N,\epsilon}(s)=1} \, ds \right).$$

Then by the Gronwall lemma:

$$\sup_{t \in [0,T]} |x_C^{N,\epsilon}(t)| \leq \exp \left(C_1 \int_0^T \left(1 + \frac{1}{\epsilon} \mathbb{1}_{\theta^{N,\epsilon}(s)=1} \right) ds \right) (|x_C^{N,\epsilon}(0)| + 1).$$

Since $|x_C^{N,\epsilon}(0)|$ is bounded with respect to N and ϵ, by Markov inequality and (24), we deduce easily

$$\mathbb{P} \left(\sup_{t \in [0,T]} |x_C^{N,\epsilon}(t)| \geq K \right) \to 0$$

as $K \to \infty$ uniformly in N and ϵ.

It is now easy to use a truncation argument similar to—and in fact easier than—the one used in step 3 of the proof of Theorem 3.1 to conclude.

Assume now that (v) holds. Inequality (24) is no more valid; instead we have

$$(25) \qquad \frac{\alpha}{\epsilon} \mathbb{E} \left(\int_0^T \mathbb{1}_{\theta^{N,\epsilon}(s)=1} \, ds \right) \leq 1 + M_2 \int_0^t \left(\mathbb{E}(|x_C^{N,\epsilon}(s)|) + 1 \right) ds.$$

We use the test function $f(x_C, 0) = f(x_C, 1) = |x_C|$ in the generator and deduce that

$$
\begin{aligned}
\tilde{M}(t) = {}& |x_C^{N,\epsilon}(t)| - |x_C^{N,\epsilon}(0)| \\
& - \int_0^t \sum_{r \in S_1} \left(\left| x_C^{N,\epsilon}(s) + \frac{\gamma_r^C}{N} \right| - |x_C^{N,\epsilon}(s)| \right) N \tilde{\lambda}_r(x_C^{N,\epsilon}, 0) \mathbb{1}_{\theta^{N,\epsilon}(s)=0} \, ds \\
& + \int_0^t \frac{1}{\epsilon} \sum_{r \in S_1} \left(\left| x_C^{N,\epsilon}(s) + \frac{\gamma_r^C}{N} \right| - |x_C^{N,\epsilon}(s)| \right) N \tilde{\lambda}_r(x_C^{N,\epsilon}, 1) \mathbb{1}_{\theta^{N,\epsilon}(s)=1} \, ds \\
& + \int_0^t \sum_{r \in R_C} \left(\left| x_C^{N,\epsilon}(s) + \frac{\gamma_r^C}{N} \right| - |x_C^{N,\epsilon}(s)| \right) N \tilde{\lambda}_r(x_C^{N,\epsilon}) \, ds
\end{aligned}
$$

is a martingale. By the triangle inequality and the assumptions on the jump rates, we get

$$
\begin{aligned}
\mathbb{E}(|x_C^{N,\epsilon}(t)|) \leq {}& \mathbb{E}(|x_C^{N,\epsilon}(0)|) + M_2 \int_0^t \sum_{r \in R_C \cup S_1} |\gamma_r^C| (\mathbb{E}(|x_C^{N,\epsilon}(s)|) + 1) \, ds \\
& + \frac{M_2}{\epsilon} \sum_{r \in S_1} |\gamma_r^C| \mathbb{E} \left(\int_0^t \mathbb{1}_{\theta^{N,\epsilon}(s)=1} \, ds \right).
\end{aligned}
$$

Then, using (25),

$$
\mathbb{E}(|x_C^{N,\epsilon}(t)|) \leq \mathbb{E}(|x_C^{N,\epsilon}(0)|) + M_3 + M_4 \int_0^t \mathbb{E}(|x_C^{N,\epsilon}(s)|) \, ds
$$

for $t \in [0, T]$ and with some constants M_3, M_4 which could be written explicitly in terms of γ_r^C, α, M_2 and T. By the Gronwall lemma, we obtain

$$
\mathbb{E}(|x_C^{N,\epsilon}(t)|) \leq e^{M_4 T} (\mathbb{E}(|x_C^{N,\epsilon}(0)|) + M_3).
$$

Inserting this in (25), we have obtained a similar bound as (24) was allowed to conclude with in the case when (iv) was the assumption. We thus can conclude similarly. $\quad\square$

APPENDIX: PROOF OF THEOREM 2.5

Let $x = (x_t)_{t \geq 0}$ denote the canonical process on $D(\mathbb{R}^+; E)$. Recall that we say that a probability measure P on $D(\mathbb{R}^+; E)$ is solution of the martingale problem associated to the generator $\mathcal{A}$ if, for all $f \in \mathcal{E}$,

$$
f(x_t) - f(x_0) - \int_0^t \mathcal{A} f(x_s) \, ds
$$

is a P-martingale.

Let $(P_t)_{t\geq 0}$, P_x and $\bar{\mathcal{A}}$ denote the semigroup, the probability measure and the infinitesimal generator associated to the PDP whose characteristics are F, λ, Q, and starting from $x \in E$.

We need the following:

PROPOSITION A.1. *Assume that Hypothesis 2.3 holds*; *then, for all* $t \in \mathbb{R}^+$, $f \in \mathcal{E}$, $P_t f$ *is also in* $\mathcal{E}$ *and*

$$(26) \qquad \|P_t f\|_\infty \leq \|f\|_\infty,$$

$$(27) \quad |P_t f(y_1, \nu) - P_t f(y_2, \nu)| \leq c\|f\|_{\mathcal{E}} e^{Kt} |y_1 - y_2|, \qquad y_1, y_2 \in \mathbb{R}^n, \nu \in \mathbb{N}^d,$$

for some constants c *and* K *depending only on the characteristics of the PDP.*

PROOF. Equation (26) is clear. Let us define, for $f \in \mathcal{E}$ and $\psi \in \mathcal{B}_b(\mathbb{R}^+ \times E)$:

$$G_f \psi(t, x) = \mathbb{E}_x\big(f(x_t)\mathbb{1}_{t < T_1} + \psi(t - T_1, x_{T_1})\mathbb{1}_{t \geq T_1}\big)$$

$$= f(\phi_\nu(t, y), \nu)H(t, x)$$

$$+ \int_0^t \int_E \psi(t - s, z)Q(dz; \phi_\nu(s, y), \nu)\lambda(\phi_\nu(s, y), \nu)H(s, x)\,ds$$

for $(t, x) \in \mathbb{R}^+ \times E$, with $x = (y, \nu)$. Then, according to Lemma 27.3 of Davis (1993),

$$G_f^n \psi(t, x) = \mathbb{E}_x\big(f(x_t)\mathbb{1}_{t < T_n} + \psi(t - T_n, x_{T_n})\mathbb{1}_{t \geq T_n}\big)$$

and

$$\lim_{n \to \infty} G_f^n \psi(t, x) = P_t f(x).$$

We then deduce by dominated convergence and the strong Markov property:

$$(28) \quad \begin{aligned} P_t f(x) &= f(\phi_\nu(t, y), \nu)H(t, x) \\ &+ \int_0^t \int_E P_{t-s} f(z)Q(dz; \phi_\nu(s, y), \nu)\lambda(\phi_\nu(s, y), \nu)H(s, x)\,ds. \end{aligned}$$

That is, $P_\cdot f : t \mapsto P_t f$ is a fixed point of G_f.

For $T > 0$, we introduce the Banach space $\mathcal{E}_T = L^\infty([0, T], \mathcal{E})$, with the norm

$$\|\psi\|_{\mathcal{E}_T} = \sup_{t \in [0, T]} e^{-\alpha t}\big(\|\psi(t, \cdot)\|_\infty + L_{\psi(t, \cdot)}\big),$$

where α will be fixed hereafter.

For $t \in [0, T]$, $(y, \nu) \in E$, we set $q_t(y, \nu) = f(\phi_\nu(t, y), \nu)H(t, y, \nu)$. Then $q : t \mapsto q_t$ is in $\mathcal{E}_T$. Indeed,

$$|H(t, y, \nu)| \leq 1, \qquad (y, \nu) \in E, t \in [0, T].$$

Also, from the Gronwall lemma, we have that ϕ_v is differentiable with respect to y and

$$|D_y\phi_v(t, y)| \leq e^{L_F t}, \qquad (y, v) \in E, t \in [0, T].$$

Then we have, for $(y, v) \in E, t \in [0, T]$,

$$|D_y H(t, y, v)| \leq \int_0^t |D_y\{\lambda(\phi_v(s, y), v)\}|\, ds \leq L_\lambda \int_0^t |D_y\phi_v(s, y)|\, ds$$

$$\leq \frac{L_\lambda}{L_F}(e^{L_F t} - 1).$$

Hence if $\alpha > L_F$,

$$\|q\|_{\mathcal{E}_T} \leq \left(L_f + \frac{L_\lambda}{L_F}\|f\|_\infty\right) + \|f\|_\infty.$$

If $\psi \in \mathcal{E}_T$, we then easily show that $G_f\psi \in \mathcal{E}_T$. Moreover, if $\psi_1, \psi_2 \in \mathcal{E}_T$, we have

$$(29) \quad \|G_f\psi_1(t, \cdot) - G_f\psi_2(t, \cdot)\|_\infty \leq M_\lambda \int_0^t \|\psi_1(t - s, \cdot) - \psi_2(t - s, \cdot)\|_\infty\, ds.$$

Also, it is easy to see that

$$\|D_y G_f\psi_1(t, \cdot) - D_y G_f\psi_2(t, \cdot)\|_\infty \leq \kappa \int_0^t e^{L_F s}\|\psi_1(t - s, \cdot) - \psi_2(t - s, \cdot)\|_{\mathcal{E}}\, ds,$$

where κ is a constant depending only on the characteristics of the PDP.

Then we easily deduce that

$$\|G_f\psi_1 - G_f\psi_2\|_{\mathcal{E}_T}$$

$$\leq \kappa_1 \sup_{t \in [0, T]} e^{-\alpha t}\left(\int_0^t e^{\alpha(t-s)}\, ds + \int_0^t e^{\alpha(t-s)}e^{L_F s}\, ds\right)\|\psi_1 - \psi_2\|_{\mathcal{E}_T}$$

$$\leq \kappa_1\left(\frac{1}{\alpha} + \frac{1}{\alpha - L_F}\right)\|\psi_1 - \psi_2\|_{\mathcal{E}_T},$$

where κ_1 is a constant depending only on the characteristics of the PDP. We now choose α sufficiently large and deduce from the Picard theorem that G_f has a unique fixed point in $\mathcal{E}_T$. This fixed point is the limit $G_f^n\psi$ for any $\psi \in \mathcal{E}_T$. Thus $P.f : t \mapsto P_t f$ is that fixed point, and is then in $\mathcal{E}_T$.

The Lipschitz constant $L_{P_t f}$ of $P_t f$ can easily be found with the preceding results and the Gronwall lemma. $\quad\square$

The infinitesimal generator $\bar{\mathcal{A}}$ of P_t is characterized by [cf., e.g., Ethier and Kurtz (1986), Part 1, Section 2]:

$$(30) \qquad (\mu - \bar{\mathcal{A}})^{-1} f = \int_0^\infty e^{-\mu t} P_t f\, dt, \qquad \mu > 0, f \in C_b(E).$$

Let $f \in \mathcal{E}$, $\mu > K$ and $\varphi = (\mu - \bar{\mathcal{A}})^{-1} f$; then $\varphi \in \mathcal{D}(\bar{\mathcal{A}})$ and

$$\frac{P_t \varphi(x) - \varphi(x)}{t} \xrightarrow[t \to 0]{} \bar{\mathcal{A}} \varphi(x), \qquad x \in E.$$

By Proposition A.1, it is easy to see that if Hypothesis 2.3 is satisfied, $(\mu - \bar{\mathcal{A}})^{-1}$ maps $\mathcal{E}$ into itself if $\mu > K$. In particular, $\varphi \in \mathcal{E}$ and for all $x \in E$:

$$(31) \qquad \varphi(x_t) - \varphi(x) - \int_0^t \mathcal{A}\varphi(x_s) \, ds$$

is a P_x-martingale.

It follows that, for all $x \in E$, and $t \in \mathbb{R}^+$,

$$\frac{P_t \varphi(x) - \varphi(x)}{t} = \frac{1}{t} \int_0^t \mathbb{E}_x (\mathcal{A}\varphi(x_s)) \, ds.$$

Since $\mathcal{A}\varphi(x_s)$ is P_x-a.s. bounded and continuous on $[0, T_1[$, we deduce that

$$\frac{P_t \varphi(x) - \varphi(x)}{t} \xrightarrow[t \to 0]{} \mathcal{A}\varphi(x), \qquad x \in E.$$

It follows that

$$\mathcal{A}\varphi(x) = \bar{\mathcal{A}}\varphi(x), \qquad x \in E.$$

REMARK A.2. In fact, we do not need that (31) defines a P_x-martingale. This can be weakened and replaced by the following:

For all $x \in E$, $\varphi \in \mathcal{E}$ and for $t \in \mathbb{T}$, where $\mathbb{T}$ is a dense subset of $\mathbb{R}^+$:

$$(32) \qquad P_t \varphi(x) - \varphi(x) = \int_0^t \mathbb{E}_x (\mathcal{A}\varphi(x_s)) \, ds.$$

In that case, we can define a sequence $(t_n)_n$ such that $t_n \in \mathbb{T}$ for all n and $t_n \to 0$ as $n \to \infty$. Taking $n \to \infty$ in the equality

$$\frac{P_{t_n} \varphi(x) - \varphi(x)}{t_n} = \frac{1}{t_n} \int_0^{t_n} \mathbb{E}_x (\mathcal{A}\varphi(x_s)) \, ds,$$

we obtain, since $\mathcal{A}\varphi(x_s)$ is P_x-a.s. bounded and continuous on $[0, T_1[$,

$$\bar{\mathcal{A}}\varphi(x) = \mathcal{A}\varphi(x), \qquad x \in E.$$

We now use a classical argument [see, e.g., Ethier and Kurtz (1986), Part 4, Section 4]. Let $\tilde{P}_x$ be another solution of the martingale problem. By the same reasoning as before, let $f \in \mathcal{E}$, $\mu > K$ and $\varphi = (\mu - \bar{\mathcal{A}})^{-1} f$; then

$$\varphi(x_t) - \varphi(x) - \int_0^t \mathcal{A}\varphi(x_s) \, ds$$

is a $\tilde{P}_x$-martingale.

Let us write

$$\tilde{\mathbb{E}}_x \left(\varphi(x_t) - \int_0^t \mathcal{A}\varphi(x_s)\, ds \right) = \varphi(x).$$

Multiply this identity by $\mu e^{-\mu t}$ and integrate on $[0, \infty)$ to obtain

$$\tilde{\mathbb{E}}_x \left(\int_0^\infty e^{-\mu t} (\mu \varphi(x_t) - \mathcal{A}\varphi(x_t))\, dt \right) = \varphi(x).$$

Since $\bar{\mathcal{A}}\varphi = \mathcal{A}\varphi$, we deduce, from (30),

$$\tilde{\mathbb{E}}_x \left(\int_0^\infty e^{-\mu t} f(x_t)\, dt \right) = \varphi(x) = \int_0^\infty e^{-\mu t} P_t f(x)\, dt, \qquad \mu > K.$$

By injectivity of the Laplace transform, this implies

$$\tilde{\mathbb{E}}_x(f(x_t)) = P_t f(x) \qquad \text{for almost all } t \geq 0, \text{ for all } x \in E.$$

We have proved that the law of a solution of the martingale problem at a fixed time t in a dense set of $\mathbb{R}^+$ is uniquely determined. This implies uniqueness for the martingale problem [see Billingsley (1999), Section 14].

REMARK A.3.   Again, the hypothesis that $\tilde{\mathbb{P}}_x$ is a solution of the martingale problem can be weakened and replaced by the following:

$$\tilde{\mathbb{E}}_x \left(\varphi(x_t) - \varphi(x) - \int_0^t \mathcal{A}\varphi(x_s)\, ds \right) = 0$$

for $x \in E$ and almost all $t \in \mathbb{R}^+$. The previous calculations are similar, since the Lebesgue integral is null over a negligible set.

REFERENCES

ALFONSI, A., CANCÈS, E., TURINICI, G., DI VENTURA, B. and HUISINGA, W. (2004). Exact simulation of hybrid stochastic and deterministic models for biochemical systems. Research Report RR-5435, INRIA.

ALFONSI, A., CANCÈS, E., TURINICI, G., DI VENTURA, B. and HUISINGA, W. (2005). Adaptive simulation of hybrid stochastic and deterministic models for biochemical systems. In *CEMRACS 2004—Mathematics and Applications to Biology and Medicine*. *ESAIM Proceedings* **14** 1–13 (electronic). EDP Sci., Les Ulis. MR2226797

BALL, K., KURTZ, T. G., POPOVIC, L. and REMPALA, G. (2006). Asymptotic analysis of multiscale approximations to reaction networks. *Ann. Appl. Probab.* **16** 1925–1961. MR2288709

BARKAI, N. and LEIBLER, S. (2000). Biological rhythms: Circadian clocks limited by noise. *Nature* **403** 267–268.

BILLINGSLEY, P. (1971). *Weak Convergence of Measures*: *Applications in Probability*. *Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics* **5**. SIAM, Philadelphia, PA. MR0310933

BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. Wiley, New York. MR1700749

BOXMA, O., KASPI, H., KELLA, O. and PERRY, D. (2005). On/off storage systems with state-dependent input, output, and switching rates. *Probab. Engrg. Inform. Sci.* **19** 1–14. MR2104547

BUJORIANU, M. and LYGEROS, J. (2004). General stochastic hybrid systems: Modelling and optimal control. In *Proc.* 43*th Conference in Decision and Control*. IEEE Press, New York.

CAI, L., FRIEDMAN, N. and XIE, X. S. (2006). Stochastic protein expression in individual cells at the single molecule level. *Nature* **440** 358–362.

CRUDU, A., DEBUSSCHE, A. and RADULESCU, O. (2009). Hybrid stochastic simplifications for multiscale gene networks. *BMC Syst. Biol.* **3** 89.

DAVIS, M. H. A. (1993). *Markov Models and Optimization. Monographs on Statistics and Applied Probability* **49**. Chapman & Hall, London. MR1283589

DELBRÜCK, M. (1940). Statistical fluctuations in autocatalytic reactions. *J. Chem. Phys.* **8** 120–124.

ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes. Characterization and Convergence*. Wiley, New York. MR0838085

FOUQUE, J.-P., GARNIER, J., PAPANICOLAOU, G. and SØLNA, K. (2007). *Wave Propagation and Time Reversal in Randomly Layered Media. Stochastic Modelling and Applied Probability* **56**. Springer, New York. MR2327824

GHOSH, M. K. and BAGCHI, A. (2005). Modeling stochastic hybrid systems. In *System Modeling and Optimization. IFIP* **166** 269–280. Kluwer Academic, Boston, MA. MR2155316

GILLESPIE, D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22** 403–434. MR0503370

GIUSTI, E. (1984). *Minimal Surfaces and Functions of Bounded Variation. Monographs in Mathematics* **80**. Birkhäuser, Basel. MR0775682

HASELTINE, E. L. and RAWLINGS, J. B. (2002). Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *J. Chem. Phys.* **117** 6959–6969.

JACOD, J. and SHIRYAEV, A. N. (1987). *Limit Theorems for Stochastic Processes. Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **288**. Springer, Berlin. MR0959133

KAERN, M., ELSTON, T. C., BLAKE, W. J. and COLLINS, J. J. (2005). Stochasticity in gene expression: From theories to phenotypes. *Nat. Rev. Genet.* **6** 451–464.

KAUFMANN, B. and VAN OUDENAARDEN, A. (2007). Stochastic gene expression: From single molecules to the proteome. *Current Opinion in Genetics and Development* **17** 107–112.

KAUFMANN, B. B., YANG, Q., METTETAL, J. T. and VAN OUDENAARDEN, A. (2007). Heritable stochastic switching revealed by single-cell genealogy. *Plos Biol.* **5** 1973–1980.

KEPLER, T. B. and ELSTON, T. C. (2001). Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. *Biophys. J.* **81** 3116–3136.

KIERZEK, A. M., ZAIM, J. and ZIELENKIEWICZ, P. (2001). The effect of transcription and translation initiation frequencies on the stochastic fluctuations in prokaryotic gene expression. *J. Biol. Chem.* **276** 8165–8172.

KRISHNA, S., BANERJEE, B., RAMAKRISHNAN, T. and SHIVASHANKAR, G. (2005). Stochastic simulations of the origins and implications of long-tailed distributions in gene expression. *PNAS* **102** 4771–4776.

KURTZ, T. G. (1971). Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *J. Appl. Probab.* **8** 344–356. MR0287609

KURTZ, T. G. (1978). Strong approximation theorems for density dependent Markov chains. *Stochastic Process. Appl.* **6** 223–240. MR0464414

KUSHNER, H. (1994). *Approximation and Weak Convergence Methods for Random Processes*. MIT Press, Cambridge.

PAPANICOLAOU, G. C., STROOCK, D. and VARADHAN, S. R. S. (1976). Martingale approach to some limit theorems. In *Statistical Mechanics and Dynamical Systems. Duke Turbulence Conf.*, *Duke Univ. Math. Series* **III** 1–120. Duke Univ. Press, Durham, NC.

PAULSSON, J. (2004). Summing up the noise in gene networks. *Nature* **427** 415–418.

RADULESCU, O., MULLER, A. and CRUDU, A. (2007). Théorèmes limites pour des processus de Markov à sauts. synthèse des resultats et applications en biologie moleculaire. *Tech. Sci. Inform.* **26** 443–469.

WARREN, P. B., TĂNASE-NICOLA, S. and TEN WOLDE, P. R. (2006). Exact results for noise power spectra in linear biochemical reaction networks. *J. Chem. Phys.* **125** 144904.

ZEISER, S., FRANZ, U., WITTICH, O. and LIEBSCHER, V. (2008). Simulation of genetic networks modelled by piecewise deterministic Markov processes. *Systems Biology, IET* **2** 113–135.

A. CRUDU
IRMAR—UMR 6625
UNIVERSITÉ DE RENNES 1
CAMPUS DE BEAULIEU
35042 RENNES
FRANCE

A. MULLER
IECN—UMR CNRS 7502
UNIVERSITÉ DE LORRAINE
54506 VANDOEUVRE-LÈS-NANCY
FRANCE

A. DEBUSSCHE
IRMAR—UMR 6625
ENS CACHAN BRETAGNE
CAMPUS DE KER LANN
35170 BRUZ
FRANCE

O. RADULESCU
DIMNP—UMR 5235 CNRS/UM1/UM2
UNIVERSITÉ DE MONTPELLIER 2
PLACE EUGÈNE BATAILLON
CP 107, 34095 MONTPELLIER
FRANCE