HAL open science

A study of interacting stochastic networks : large scale, long time behavior and fluctuations, with applications in communication networks and protein polymerization

Wen Sun

▶ To cite this version:

Wen Sun. A study of interacting stochastic networks : large scale, long time behavior and fluctuations, with applications in communication networks and protein polymerization. Dynamical Systems [math.DS]. Sorbonne Université, 2018. English. NNT : 2018SORUS075 . tel-02481792v2

HAL Id: tel-02481792

https://theses.hal.science/tel-02481792v2

Submitted on 17 Feb 2020

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.

Université Pierre et Marie Curie
École doctorale de Sciences Mathématiques de Paris Centre

Thèse de Doctorat

Spécialité : Mathématiques Appliquées

Présentée et soutenue publiquement le 11 juin 2018

par **Wen SUN**

sous la direction de
Philippe ROBERT

A study of interacting stochastic networks : large scale, long time behavior and fluctuations

with applications in communication networks and protein polymerization

Après avis de :

M. Amarjit Budhiraja & M. Pascal Moyal

devant le jury composé de :

M.	Francis Comets	Université Paris Diderot	Examinateur
M.	Carl Graham	École Polytechnique	Examinateur
Mme.	Irina Kourkova	Sorbonne Université	Examinateur
M.	Pascal Moyal	Université de Technologie de Compiègne	Rapporteur
M.	Philippe Robert	Inria Paris	Directeur de Thèse
M.	Pierre-André Zitt	Université de Marne-la-Vallée	Examinateur

INRIA RESEARCH CENTER OF PARIS
2 Rue Simone IFF,
75012 Paris

Université Pierre et Marie Curie
École doctorale de Sciences Mathématiques de Paris Centre (ED386)
4, Place Jussieu
75005 Paris

Abstract

This PhD document is devoted to the analyses of large stochastic networks used to study mathematical models in communication networks and in biology.

The first part consists of the analysis of three models used to evaluate the efficiency of duplication and placement algorithms in large distributed systems. These models are represented by large stochastic networks under different scaling regimes assumptions. In Chapter 2, the dynamic of the system can be described with the empirical measure associated to a multi-dimensional transient Markov process. We provide a detailed study of these processes on several fast time scales. Stochastic averaging principles with multiple time scales are investigated in particular. In Chapter 3 and Chapter 4, the interactions have unbounded jump sizes and occur within a limited random set of nodes. We develop a convenient mean field analysis in despite that the associated empirical measures do not have autonomous evolution equations. We also study the long time behavior of the corresponding limiting nonlinear jump diffusions.

The second part analyzes two models used to study the variability in the polymerization phenomena occurring in a biological context. In Chapter 5, we investigate the polymerization and fragmentation processes with an assumption of critical nucleus size. A scaling analysis of these stochastic models show that the sharp phase transition and, especially the large variance, observed in the experiments can be explained by these models. In Chapter 6, we provide a functional central limit theorem in the classical (infinite dimensional) stochastic Becker-Döring model.

Keywords :

Stochastic Averaging Principle ; Multi-dimensional Transient Process ; Interacting Particle System ; Propagation of Chaos ; Marked Jump Process ; Measure-valued Process ; Infinite-dimensional Fluctuations ; Coagulation and Fragmentation Processes ; Becker-Döring Model.

Résumé

Ce document de thèse est consacré aux analyses de grands réseaux stochastiques utilisés pour étudier des réseaux de communication et ainsi que certains phénomènes biologiques.

La première partie se compose de trois modèles pour évaluer l'efficacité des algorithmes de duplication et de placement dans les grands systèmes distribués. Ces modèles sont étudiés sous différents régimes d'échelle. Au chapitre 2, la dynamique du système est décrite l'aide de la mesure empirique associée à un processus de Markov transient multidimensionnel. Une étude détaillée de ces processus est effectuée sur plusieurs échelles de temps rapides. Des principes de moyenne stochastique avec plusieurs échelles de temps sont étudiées. Aux chapitres 3 et 4, les interactions considérées peuvent avoir des tailles de saut illimitées et se produire dans un ensemble aléatoire fini de nœuds. Le processus de la mesure empirique associé n'ayant pas d'équations d'évolution simples, nous développons une analyse de champ moyen spécifique pour étudier ces systèmes. Le comportement en temps long des processus de diffusions non linéaires correspondants est aussi analysé.

La deuxième partie présente deux modèles pour étudier la variabilité dans les modèles de polymérisation se produisant dans un contexte biologique. Dans le chapitre 5, nous étudions les processus de polymérisation et de fragmentation avec l'hypothèse d'un noyau critique pour la taille des polymères. Notre résultat principal montre que ces modèles semblent donner une explication raisonnable de la transition de phase courte du phénomène de polymérisation, et surtout de la grande variabilité de l'instant de transition, qui ont été observés dans de nombreuses expériences de biologie. Au chapitre 6, nous proposons un théorème de limite centrale fonctionnelle dans le modèle stochastique classique en dimension infinie de Becker-Döring.

Mots-clefs :

Principe de la Moyenne Stochastique ; Processus Transitoire Multidimensionnel ; Système de Particules en Interaction ; Propagation du Chaos ; Processus de Saut Marqué ; Processus à Valeur de Mesure ; Fluctuations Infinies ; Processus de Coagulation et de Fragmentation ; Modèle de Becker-Döring.

Acknowledgements

Most of all, I am eternally grateful to my supervisor, Philippe Robert, who evolved me from a senior student to a baby researcher. During my four PhD years, he had always been very supportive, enthusiasm and inspiring, leading me to enjoy the beauty and magic of probability theory. He had always answered my questions patiently, helped me to improve my technical writing and scientific presentation skills and encouraged me to overcome the difficulties in my research life. I was very lucky to have him as my supervisor. Words cannot express my gratitude.

I want to thank Amarjit Budhiraja and Pascal Moyal for their time reading my thesis and writing inspiring reviews. Their valuable comments were very helpful to the final version of this thesis. I also thank Francis Comets, Carl Graham, Irina Kourkova, Pierre-André Zitt for accepting to become the member of the jury. I acknowledge all the people who attend my defense.

I would like to thank Fondation Sciences Mathématiques de Paris that gave me the opportunity to study math in Paris, supported me from master to PhD. I acknowledge all the professors that I have had in jussieu and USTC. Due to them I discovered a wonderful world of mathematics and made my mind to pursue a PhD in probability.

I thank Kavita Ramanan for her hospitality during my visits to Brown University, for many fruitful discussions and for leading me to think what kind of research I want to do. It is my pleasure to work with my co-author Reza Aghajani and talk with Marie Doumic. They have helped me to broaden my horizons in research.

I would like to mention my academic siblings, Sarah Eugene, Guilherme Thompson, Renaud Dessalles; and my colleagues in the old Inria team RAP, Christine Fricker, Davit Martirosyan, Nicolas Broutin, Othmane Safsafi, for the cheerful conversations, cakes, coffee and tea in the afternoon, sometimes in company of some sheep, from pastoral Rocquencourt to the industrial 12th arrondissement.

My deepest thanks to my friends in Paris, Qilong Weng, Bingxiao Liu, Chaoyu Quan, Xianglong Duan, Yin Yi and Songyan Xie, especially for the reunions on our festivals that made a foreign city feels like home. I am also grateful to the people who have ever been nice to me, perhaps without their knowledge, that have helped me get over many depressed moments.

Finally, my parents have always been supportive and believed in their daughter. Their love makes me determined.

孙雯

01/06/2018

To my grandfather.

Contents

1 Introduction **1**
- 1.1 A General Overview . . . 1
- 1.2 Mathematical Background . . . 1
- 1.3 Main Contributions . . . 10
- 1.4 Summary of Part I . . . 14
- 1.5 Summary of Part II . . . 22
- 1.6 Notations . . . 27

Part I: Large stochastic networks with failures **29**

2 The Global Duplication Model **31**
- 2.1 Introduction . . . 32
- 2.2 The Stochastic Model . . . 38
- 2.3 First Order Asymptotic Behavior . . . 41
- 2.4 Evolution of Stable Network . . . 46
- 2.5 Second Order Asymptotics in the Stable Case . . . 56
- 2.6 A Local Equilibrium in the Overloaded Case . . . 61
- 2.A Appendix: Generalized Skorohod Problems . . . 66

3 The Local Duplication Model **69**
- 3.1 Introduction . . . 70
- 3.2 The Stochastic Model . . . 74
- 3.3 An Asymptotic Process . . . 77
- 3.4 Mean-Field Limit . . . 82
- 3.5 An Asymptotic Bound on the Decay of the Network . . . 89
- 3.6 The Case of Multiple Copies . . . 92

4 A study of allocation algorithms **101**
- 4.1 Motivations: allocation in large storage distributed system . . . 102
  - 4.1.1 Introduction . . . 102
  - 4.1.2 Placement policies . . . 104
  - 4.1.3 Simulations . . . 105
  - 4.1.4 Mathematical Models . . . 108
  - 4.1.5 Conclusion . . . 118
- 4.2 Large Urn Models with Local Mean-Field Interactions . . . 123

4.2.1 Introduction 123
4.2.2 The Stochastic Model 127
4.2.3 Evolution Equations 133
4.2.4 Mean-Field Convergence Results 138
4.2.5 Equilibrium of the McKean-Vlasov Process 148

Part II: Variations in polymerization models **153**

5 The asymptotic study of nucleation times **155**
5.1 Introduction 156
5.2 Stochastic Model 161
5.3 The First Instant of Nucleation 167
5.4 Fast Growth of Stable Polymers 176
5.5 Asymptotic Behavior of Lag Time 178
5.A Appendix: Biological background 181
5.B Appendix: Technical Details 182

6 Fluctuations in Becker-Döring model **187**
6.1 Introduction 187
6.2 The Stochastic Model 192
6.3 SDE in Hilbert space 195
6.4 Convergence of the Fluctuation Processes 201

References **209**

Chapter 1

Introduction

1.1 A General Overview

This PhD thesis consists of two independent parts. The first part is composed of three stochastic models in order to analyze the efficiency of duplication and placement algorithms in large communication networks. The results are based on four papers:

[154] Analysis of large unreliable stochastic networks. *Annals of Applied Probability* 26, 5 (2016). Joint work with M. Feuillet and P. Robert.

[1] A large scale analysis of unreliable stochastic networks. *Annals of Applied Probability* 28, 2 (2018). Joint work with M. Aghajani and P. Robert.

[156] Analysis of a stochastic model of replication in large distributed storage systems: A mean-field approach. *ACM-Sigmetrics 2017*. Joint work with V. Simon, S. Monnet, P. Robert and P. Sens.

[155] Analysis of Large Urn Models with Local Mean-Field Interactions. *Preprint*. Joint work with P. Robert.

The second part investigates stochastic models associated to biological polymerization processes, it corresponds to two papers

[139] On the asymptotic distribution of nucleation times of polymerization processes. *Preprint*. Joint work with P. Robert.

[153] A functional central limit theorem for the Becker-Döring model. *Journal of Statistical Physics* 171, 1 (2018).

In the following, we will describe briefly the mathematical context of this thesis, present our main contributions and give the outline of the two parts: describe the motivations, explain the models, present the mathematical results and compare our results with the real experiments.

1.2 Mathematical Background

From a mathematical point of view, this thesis presents the studies of five large stochastic networks that are used to study important phenomena occurring in communication networks or in a biological context. These phenomena of interest can be described through the asymptotic dynamics of the empirical measures associated with

these networks. In Chapter 3 and Chapter 4, mean-field approaches are developed in a quite different setting for the reason that our empirical measures do not have self-contained descriptions of evolutions. In Chapter 2 and Chapter 5, the interesting phenomena only happen on large time scales, *i.e.*, $t \mapsto N^k t$, for some integer $k \geq 1$. In this context, stochastic averaging principles have to be established in order to analyze the dynamics of empirical measure processes on large time scales. In Chapter 6, a functional central limit theorem around the re-scaled empirical measure is proved in the classical Becker-Döring model.

In this section, we will introduce the classical mathematical background of this thesis, the basic scaling methods: stochastic averaging principle and mean field theory.

1.2.1 Scaling Methods: Time Scales and Space Scales

Scaling methods have been widely used in Probability Theory. A basic example is the Brownian motion which can be roughly seen as random walks with very small size of steps (space scale) in very long time (time scale). The scaling techniques can provide useful approximations for large systems. For example, the estimations of first-order (functional law of large numbers), second order (functional central limits) and also the descriptions of long-time behaviors, *etc.* of a given system.

Here, we begin with the basic idea of the scaling methods. Let $(X^N(t))$ be a random process, given any sequence $(\phi_N(t))$ and (Ψ_N), one can define the re-scaled process

$$(\overline{X}^N(t)) := \left(\frac{X^N(\phi_N(t))}{\Psi_N} \right).$$

By careful choice of scaling parameters $\phi_N(t)$ and Ψ_N, one may expect that there exists a process (or function) $x(t)$, such that, in probability,

$$\lim_{N \to \infty} (\overline{X}^N(t)) = (x(t)).$$

We should remark that:

1. The choices of scalings are different case by case. Usually, it depends on the properties of the process $(X^N(t))$. For example, in Chapter 2, we consider the scaling limits when $\phi_N(t) = N^k t$ and $\Psi_N = N$ or $\sqrt{N}$.
2. The processes $(X^N(t))$ are not necessarily to be Markovian. In Chapter 3, we study a case when $(X^N(t))$ is a non-Markovian counting measure process.

In the following, we present some classical examples by using scaling limits.

Example: Donsker's Theorem

Let $(\xi_n)_{n \geq 1}$ be a sequence of i.i.d. random variables with mean 0 and variance 1. Let

$$(X(t)) = \left(\sum_{i=1}^{\lfloor t \rfloor} \xi_i \right).$$

The law of large numbers implies the re-scaled process $(X(Nt)/N)$ converges to 0 almost surely. For the case time scale $\phi_N(t) = Nt$, and space scale $\Psi_N = \sqrt{N}$, one has the Donsker's Theorem as follows.

Theorem 1.1. *For the convergence in distribution of continuous processes, one has*

$$\lim_{N\to\infty}\left(\frac{X(Nt)}{\sqrt{N}}\right) = (\mathcal{B}(t)),$$

where $(\mathcal{B}(t))$ *is a standard Brownian motion on* $\mathbb{R}$.

See the Section 2.8 in Billingsley [19] for example.

Example: M/M/1 queue

Let $(L^N(t))$ be the M/M/1 queue with arrival rate λ, departure rate μ and initial state $L^N(0) = \lfloor xN \rfloor$, where $x \in \mathbb{R}^+$. This process can also be seen as a reflected continuous time random walk in $\mathbb{N}$. It has the same distribution with the solution of the following stochastic differential equation,

$$L^N(t) = \lfloor xN \rfloor + \int_0^t \mathcal{N}_\lambda(\mathrm{d}s) - \int_0^t \mathbb{I}_{\{L^N(s-)>0\}}\mathcal{N}_\mu(\mathrm{d}s),$$

where $\mathcal{N}_\eta(\mathrm{d}s)$ denotes a Poisson point process on $\mathbb{R}$ with intensity η, for $\eta = \lambda, \mu$. The two Poisson point processes are assumed to be independent.

By using scaling methods, one will see that (in the next Theorem), on the long time interval $[0, Nt]$, for any $t<x/(\mu-\lambda)^+$, the $M/M/1$ queue behaves like a diffusion around a linear function:

$$L^N(Nt) \sim N\left(x + (\lambda-\mu)t\right)^+ + \sqrt{N}\left(\sqrt{\lambda+\mu}\mathcal{B}(t)\right).$$

It is actually a functional law of large numbers and a functional central limit theorem of the re-scaled process $(L^N(Nt)/N)$.

Theorem 1.2. *For the convergence of process associated with the uniform norm on compact sets, in probability, one has*

$$\lim_{N\to\infty}\left(\frac{L^N(Nt)}{N}\right) = \left(x + (\lambda-\mu)t\right)^+.$$

Moreover, with the convention that $0/0 = 0$, *on the time interval* $[0, x/(\mu-\lambda)^+)$, *one has*

$$\lim_{N\to\infty}\left(\frac{L^N(Nt) - N\left(x + (\lambda-\mu)t\right)^+}{\sqrt{N}}\right) = \left(\sqrt{\lambda+\mu}\mathcal{B}(t)\right).$$

We refer to Section 5.7 of Robert [138] for more scaling limits of queues.

Scaling limits in this thesis

In this thesis, various scaling problems have been investigated:

— in Chapter 2 and 5, we study the scaling limits with multiple time scales for two sequences of multi-dimensional transient Markov processes. They give quite precise descriptions of the long time behavior of these systems.

— In Chapter 3, 4 and 6, we study the scaling limits for counting measure processes, to obtain the first order and second order estimations.

1.2.2 Stochastic Averaging Principle

The averaging principle first appeared in the perturbation problems in celestial mechanics. Roughly speaking, it has provided an approximation of the dynamics of a coupled system with two kinds of motions: one evolving much faster than the other. The faster motion has effect on the slow motion, which can be seen as perturbations. The simplest example would be the Earth's rotation and revolution around the sun. The rotation is evolving faster than the revolution and has effects on the orbit. The main idea of the averaging principle is to use an averaged motion to replace the fast motion and make a prediction of the effects on the slow motion. See Chapter 5 in Arnold [12] for more details.

In the following, we start with two basic results of averaging principle in a deterministic dynamic system and in a stochastic dynamic system. The statements are taken from Chapter 7 of Freidlin and Wentzell [63]. For more references in this topic, we refer to Khasminskii [94], Papanicolaou, Stroock and Varadhan [124], Kurtz [98]. In the subsection 1.3.1, we will discuss our stochastic averaging problems associated with some transient multi-dimensional jump processes.

Averaging principle in an ODE system

Let $(X^N(t), Y^N(t))$ be the solution of the following ODEs in $\mathbb{R}^r \otimes \mathbb{R}^\ell$,

$$\begin{cases} \mathrm{d}X^N(t) = b(X^N(t), Y^N(t))\,\mathrm{d}t, & X^N(0) = x, \\ \mathrm{d}Y^N(t) = NB(X^N(t), Y^N(t))\,\mathrm{d}t, & Y^N(0) = y, \end{cases}$$

where $x \in \mathbb{R}^r$, $y \in \mathbb{R}^\ell$ and b, B are bounded, continuously differentiable functions from $\mathbb{R}^r \otimes \mathbb{R}^\ell$ to $\mathbb{R}^r$ and $\mathbb{R}^\ell$ respectively.

Due to the boundedness condition, one could roughly say that the velocity of the function $(X^N(t))$ is of the order of $O(1)$ and the velocity of the function $(Y^N(t))$ is of the order of N. For N sufficiently large, the function Y^N moves much faster than X^N. Hence, if at time t, $(X^N(t), Y^N(t)) = (x, y)$, on the small time interval $[t, t + \Delta t/N]$, the function $X^N(\cdot)$ is approximately x while

$$Y^N(t + \frac{\Delta t}{N}) \sim y + \int_0^{\Delta t} B(x, Y^N(t + \frac{s}{N}))\,\mathrm{d}s.$$

This gives us the idea of introduce an auxiliary function $Y^{xy}(t)$, which is the solution of the following ODE in $\mathbb{R}^\ell$,

$$\mathrm{d}Y^{xy}(t) = B(x, Y^{xy}(t))\,\mathrm{d}t, \qquad Y^{xy}(0) = y.$$

The averaging principle states that if the function $(Y^{xy}(t))$ satisfies the following condition, the solution of original ODE system can be approximated by the solution of an averaged system.

Assumptions 1.1. *We assume that there exist a function $\overline{b}$: $\mathbb{R}^r \to \mathbb{R}^r$, such that for any $x \in \mathbb{R}^r$ and $y \in \mathbb{R}^\ell$,*

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T b(x, Y^{xy}(s))\,\mathrm{d}s = \overline{b}(x).$$

For any $t > 0$, let (t_n) be a partition of $[0,t]$ and $\Delta_i = t_i - t_{i-1}$, if the norm of the partition is small enough, then under the Assumption 1.1,

$$\begin{aligned} X^N(t) &\sim x + \sum_{(t_n)} \int_{t_{i-1}}^{t_i} b(X^N(t_{i-1}), Y^N(s))\,\mathrm{d}s \\ &\sim x + \sum_{(t_n)} \frac{\Delta_i}{N} \sum_{k=1}^{N} b(X^N(t_{i-1}), Y^N(t_{i-1} + \frac{k}{N})) \\ &\sim x + \int_0^t \overline{b}(X^N(s))\,\mathrm{d}s. \end{aligned}$$

Theorem 1.3. *Let $\overline{X}(t)$ be the solution of the ordinary differential equation*

$$\mathrm{d}\overline{X}(t) = \overline{b}(\overline{X}(t))\,\mathrm{d}t, \qquad \overline{X}(0) = x.$$

If Assumption 1.1 holds, then for any $T>0$, $\delta>0$, $x\in\mathbb{R}^r$ and $y\in\mathbb{R}^\ell$, one has

$$\lim_{N\to\infty} \sup_{0\le t\le T} \left| X^N(t) - \overline{X}(t) \right| = 0.$$

Averaging principle in a SDE system

Let's consider a stochastic system driven by a standard Wiener process $(\mathcal{B}(t))$ in $\mathbb{R}^n$. For any $N\in\mathbb{N}$, let process $(X^N(t), Y^N(t))$ be the solution of the following SDE,

$$\begin{cases} \mathrm{d}X^N(t) = b(X^N(t), Y^N(t))\,\mathrm{d}t + \sigma(X^N(t))\,\mathrm{d}\mathcal{B}(t), & X^N(0) = x, \\ \mathrm{d}Y^N(t) = NB(X^N(t), Y^N(t))\,\mathrm{d}t + \sqrt{N}C(X^N(t), Y^N(t))\,\mathrm{d}\mathcal{B}(t), & Y^N(0) = y, \end{cases}$$

where $b(\cdot,\cdot)$ is a bounded and Lipschitz mapping from $\mathbb{R}^r\otimes\mathbb{R}^\ell$ to $\mathbb{R}^r$, $B(\cdot,\cdot)$ is a bounded and Lipschitz mapping from $\mathbb{R}^r\otimes\mathbb{R}^\ell$ to $\mathbb{R}^\ell$, and for any $x\in\mathbb{R}^r$ and $y\in\mathbb{R}^\ell$, $\sigma(x)$ is a $r\times n$ matrix and $C(x,y)$ is a $\ell\times n$ matrix. Moreover, all the entries of these matrices are assumed to be bounded and Lipschitz with respect to (x,y).

Analogously, we introduce an auxiliary process: for any $x\in\mathbb{R}^r$ and $y\in\mathbb{R}^\ell$, let $Y^{xy}(t)$ be the solution of the stochastic differential equation

$$\mathrm{d}Y^{xy}(t) = B(x, Y^{xy}(t))\,\mathrm{d}t + C(x, Y^{xy}(t))\,\mathrm{d}\mathcal{B}(t), \qquad Y^{xy}(0) = y.$$

The stochastic averaging principle states that if this auxiliary process satisfies the following condition, then there would be an averaged system whose solution could be a good approximation of the solution of the original system for N large.

Assumptions 1.2. *There exists a function $\overline{b}$ from $\mathbb{R}^r$ to $\mathbb{R}^r$, such that for any $t\ge 0$ and $x\in\mathbb{R}^r$, $y\in\mathbb{R}^\ell$, one has*

$$\lim_{T\to\infty} \mathbb{E} \left| \frac{1}{T} \int_0^T b(x, Y^{xy}(s))\,\mathrm{d}s - \overline{b}(x) \right| = 0.$$

Theorem 1.4. *Let $\overline{X}(t)$ be the solution of the stochastic differential equation*

$$\mathrm{d}\overline{X}(t) = \overline{b}(\overline{X}(t))\,\mathrm{d}t + \sigma(\overline{X}(t))\,\mathrm{d}\mathcal{B}(t), \qquad \overline{X}(0) = x.$$

If Assumption 1.2 holds, then for any $T>0$, $\delta>0$, $x\in\mathbb{R}^r$ and $y\in\mathbb{R}^\ell$, one has

$$\lim_{N\to\infty} \mathbb{P} \left(\sup_{0\le t\le T} \left| X^N(t) - \overline{X}(t) \right| > \delta \right) = 0.$$

We should remark that Assumption 1.2 is valid when $(Y^{xy}(t))$ is an ergodic process. For some specific cases, one can identify the limit function $\bar{b}$.

An Example: Large Loss network

Here we give an example of the application of stochastic averaging principle and scaling methods in large loss network (c.f. Kelly [93]). The results are based on the paper Hunt and Kurtz [77].

Definition 1.1. *(Loss Network) It is a network with J links and R kinds of calls, labeled by $j=1,\dots,J$ and $r=1,\dots,R$. The link j has C_j circuits. The arrival of the calls of type r can be described by a Poisson process with rate κ_r. Upon the arrival moment, the call will hold A_{jr} circuits from all $j=1,\dots,J$ links. After an exponential holding time with rate μ_r, it will release all these circuits simultaneously and leave the system. If some link doesn't have enough circuits at the arrival moment, the call will be rejected.*

The scaling methods (Kelly's scaling [93]) are used in order to analyze the behavior of the loss network when the arrival rates are large. Of course in this case, the capacity of the system should be large as well. Hence, one could introduce the scaling parameter N and it is assumed that the arrival rates κ_r^N and capacities C_j^N are in the order of N.

The state descriptor of such a system is $X^N(t)=(X_1^N(t),\dots,X_R^N(t))$ where $X_r^N(t)$ denotes the number of calls of type r at time t. In order to analyze the dynamic, one should keep track the free circuits in each link, i.e., for each link $j=1,\dots,J$, let

$$Y_j^N(t) = C_j^N - \sum_{r=1}^{R} A_{jr} X_r^N(t).$$

In the paper Hunt and Kurtz [77], it is proved that, in some cases, the fluid limit of the scaled process $(\overline{X}^N(t)):=(X^N(t)/N)$ can be obtained by a stochastic averaging principle for the coupled system consists of slow process $(\overline{X}^N(t))$ and fast process $(Y^N(t))$. The transition rate of the slow process $(\overline{X}^N(t))$ can be approximated by the equilibrium of the fast process when N is large. That gives the motivation to study the occupation measure of the fast process $(Y^N(t))$, i.e.

$$\nu^N((0,t)\times\cdot) := \int_0^t \mathbb{I}_{\{Y^N(u)\in\cdot\}}\,\mathrm{d}u.$$

By studying the associated martingale problem, one can have the convergence and then identify the limit of coupled system $(\overline{X}^N(t),\nu^N(t))$.

We remark that the convergence of occupation measure of fast process $(Y^N(t))$ is a analog of the Assumption 1.2 in the classical stochastic averging system. For more about averaging in martingale problems, see Kurtz [98].

In this thesis

In Chapter 2 and Chapter 5, we have investigated two stochastic averaging principles associated with multiple time scales in the analyses of the transition behaviors of multi-dimensional Markov processes. See Section 1.3.1 for a detailed introduction of our stochastic averaging problems.

1.2.3 Mean-Field Theory

Mean field theory has a long history in physics. It has been first derived by Van der Waals in the 1870s to understand the gas-liquid phase transition. In 1907, Pierre Curie and Pierre Weiss developed a mean field theory of ferromagnetism. The mathematically formulated study started from Kac [85] on stochastic model for the Vlasov equation of plasma and then by McKean [109]. After that, a wide variety of systems have been studied through mean field theory. For example, Sznitman [159] gave a full proof of propagation of chaos in Boltzmann equation; Vvedenskaya, Dobrushin and Karpelevich [165] introduced the mean field analysis in queuing theory and worked on the interacting jump diffusions and also the Ising model, the Kuramoto model, Ginzburg-Landau model, *etc*. In short, mean field models describe the dynamic of the system with a large number of particles where the force acting on a single particle can be approximated by an averaged interactions over all the other particles. A general survey of these methods is presented in the book Sznitman [158].

In the following, we will take a simple version of McKean-Vlasov model as example to explain briefly the classical mean field theory.

The McKean-Vlasov N-particle system

It is a N-body system with pair interactions. Namely, there are N particles, labeled by integers $1, \dots, N$. The evolution of state of the particle i is described by a stochastic process $X_i^N(t)$ taking values in the state space E. The interaction between particle i and particle j leads to a force acting on particle i, which is captured by some function $b(X_i^N, X_j^N)$. It is assumed that the total force on the particle i is approximated by the averaged force over all the particles in the system, i.e.

$$\frac{1}{N}\sum_{j=1}^{N} b(X_i^N, X_j^N).$$

We assume that the fluctuations on each particle are independent standard Brownian motions. Hence, the evolution equation of the system can be described by the following SDEs: for all integer i=1, ..., N,

$$\begin{cases} \mathrm{d}X_i^N(t) = \mathrm{d}\mathcal{B}_i(t) + \frac{1}{N}\sum_{j=1}^{N} b(X_i^N(t), X_j^N(t))\,\mathrm{d}t, \\ \mathrm{d}X_i^N(0) = x_i. \end{cases} \tag{1.1}$$

For any fixed N, the process $(X^N(t))$=$(X_1^N(t), \dots, X_N^N(t))$ lives in the high dimensional space $E^{\otimes N}$. It is difficult to analyze such a process when N is large. Mean field theory allows us to reduce the N-body problem to a one-body problem by introducing the process of empirical measure.

Empirical measure and One-body system

The empirical measure of the system is denoted by $(\Lambda^N(\cdot))$:

$$\Lambda^N(t) := \frac{1}{N}\sum_{i=1}^{N} \delta_{X_i^N(t)},$$

where $(\delta_\cdot)$ is the Dirac measure. Due to the randomness of the McKean-Vlasov system (1.1), $(\Lambda^N(\cdot))$ is a stochastic process taking values in the space of probabilities on E, noted by $M_1(E)$. Then by using Ito's lemma, for any reasonable test function ϕ on E, one could obtain the one-body problem

$$\langle \phi, \Lambda^N(t)\rangle = \langle \phi, \Lambda^N(0)\rangle + \int_0^t \langle L[\Lambda(s)]\phi, \Lambda^N(s)\rangle \, \mathrm{d}s + \mathcal{M}_\phi^N(t) \tag{1.2}$$

where $L[\cdot]$ is the nonlinear operator: for any $\sigma \in M_1(E)$,

$$L[\sigma]\phi(x) = \frac{1}{2}\Delta\phi(x) + \nabla\phi(x) \int_E b(x,y)\sigma(\mathrm{d}y)$$

and $(\mathcal{M}_\phi^N(t))$ is a continuous martingale with quadratic variation

$$\left\langle \mathcal{M}_\phi^N \right\rangle (t) = \frac{1}{N} \int_0^t \left\langle (\nabla\phi)^2, \Lambda^N(s) \right\rangle \mathrm{d}s.$$

One should notice that if the function $\nabla\phi$ is bounded, then by using Doob's inequality, the martingale $(\mathcal{M}_\phi^N(t))$ is vanishing in probability as N getting large.

Let $(\Lambda(\cdot))$ be a function taking values in $M_1(E)$, and for any reasonable test function ϕ, it satisfies an analog equation of (1.2) as follows:

$$\langle \phi, \Lambda(t)\rangle = \langle \phi, \Lambda(0)\rangle + \int_0^t \langle L[\Lambda(s)]\phi, \Lambda(s)\rangle \, \mathrm{d}s. \tag{1.3}$$

Thus, when N is essentially large, providing the convergence of the initial state, one may expect that the function $(\Lambda(t))$ (if it exists) is an approximation of the empirical measure $(\Lambda^N(\cdot))$.

The McKean-Vlasov process

In the probability point of view, the solution of (1.3) is the distribution of a nonlinear process taking values in E, so called McKean-Vlasov process.

Definition 1.2. *The McKean-Vlasov processes are the stochastic processes that can be described by the following form of SDEs,*

$$\mathrm{d}X(t) = \int_E a(X(t),y)\Lambda(t)(\mathrm{d}y)\, \mathrm{d}\mathcal{B}(t) + \int_E b(X(t),y)\Lambda(t)(\mathrm{d}y)\, \mathrm{d}t, \tag{1.4}$$

where $(\Lambda(\cdot))$ is the distribution of the process $(X(\cdot))$. These processes are also called non-linear diffusions.

When $a \equiv 1$, the distribution of the McKean-Vlasov process (1.4) with initial distribution $\Lambda(0)$ satisfies the equation (1.3). For the existence and uniqueness for the process (1.4) when function b is bounded and Lipschitz and $a \equiv 1$, we refer to Theorem 1.1 in Sznitman [158].

Weak compactness of the sequence of processes of random measures

Clearly, for any finite time $T>0$, the sequence of processes $(\Lambda^N(t), t\leq T)$ and $(\Lambda(t), t\leq T)$ live in the space $C_T:=C([0,T]; M_1(E))$, which is the continuous functional space on time interval $[0,T]$ and taking values in $M_1(E)$. If E is a Polish space, then so is the space C_T. See Billingsley [19] for details. In order to show the convergence

$$\lim_{N\to\infty}(\Lambda^N(t), t\leq T) = (\Lambda(t), t\leq T),$$

one need to provide tightness arguments of the sequence of processes of random measures $(\Lambda^N(t))$ in C_T first. We refer to the Theorem 3.7.1 of Dawson [42] for the sufficient and necessary condition for the weak convergence of the measure-valued processes.

Propagation of chaos

In the particle system (1.1), if all the diffusion $(X_i^N(t))$ starts from an *i.i.d.* distribution, then for any i, j, it is easy to see that

$$(X_i^N(t), t\leq T) \overset{law}{=} (X_j^N(t), t\leq T).$$

However, on the time interval $[0,T]$, the processes $(X_i^N(t))$ and $(X_j^N(t))$ are not independent for any fixed N. It is natural to ask, in the large system, how to describe the dependency among the particles? Or more precisely, could the particles be asymptotically *i.i.d.* after time 0? The concept of "Propagation of chaos" has been proposed by Kac [85] to derive the homogeneous Boltzmann equation from a random walk on the energy sphere. In the McKean-Vlasov particle system (1.1), it means that under certain conditions, for any finite subset of particles, they are asymptotically (i.e. $N\to\infty$) *i.i.d.* distributed on any finite time interval.

For a finite time T, the joint distribution of $(X_i^N(t), t\leq T)_{i=1}^N$ is denoted by $\mathcal{P}^N$, which is a probability on $C([0,T],E)^{\otimes N}$. Thus, we first introduce the notion of the property "chaotic" for the sequence of probabilities.

Definition 1.3. *(u-chaotic)*
Let (u_N) be a sequence of exchangeable probabilities on $S^{\otimes N}$, where S is a separable metric space. We say that u_N is u-chaotic, for a $u\in M_1(S)$, if for any $k\in\mathbb{N}^*$, $\phi_1,\ldots,\phi_k\in C_b(S)$,

$$\lim_{N\to\infty}\left\langle u_N, \phi_1\otimes\cdots\otimes\phi_k\underbrace{\otimes\mathbb{1}\cdots\otimes\mathbb{1}}_{N-k}\right\rangle = \prod_{i=1}^k\langle u,\phi_i\rangle\,.$$

By using the notion of "chaotic", the "propagation of chaos" means that if the sequence of initial states $(X_i^N(0))_{i=1}^N$ are *i.i.d.* random variables with distribution $\Lambda_0\in M_1(E)$, then $\mathcal{P}^N$ is Λ-chaotic, where Λ is the solution of (1.3) on the time $[0,T]$ with initial condition Λ_0.

To prove this, one need the following theorem, which depicts the relation between $\mathcal{P}^N\in M_1(C([0,T],E)^{\otimes N})$, the joint distribution of all particles, and $\Lambda^N\in M_1(C([0,T],E))$, the empirical distribution.

Theorem 1.5. *(Proposition 2.2 in Sznitman [158])*
Assume that $(X_i^N(0))$ *are* i.i.d. *random variables with distribution* Λ_0*, then,* $(\mathcal{P}^N)$ *is* Λ*-chaotic is equivalent to the sequence of empirical measures* (Λ^N) *converges in law to* Λ*, where* Λ *is the solution of* (1.3) *with initial condition* Λ_0*.*

In conclusion, in the McKean-Vlasov system (1.1), if all particles starts independently from the same distribution, then the empirical distribution of the system converges to the distribution of a non-linear process. It implies that asymptotically, any finite subset of test particles behave like *i.i.d.* processes.

Central limit theorem

The mean field limit can be interpreted as a law of large numbers in probability, it is natural to study the fluctuations along this limit. The first result on the central limit theorem for the Vlasov limit seems to have been given by Braun and Hepp [24] in 1977. They investigated the fluctuations around the trajectory of a test particle. The CLT for the general McKean-Vlasov model when the initial measures of the system are products of N i.i.d. measures is proved by Sznitman [160] in 1984. He proved that, the process

$$Y_f^N := \sqrt{N}\left(\left\langle f, \Lambda^N(t)\right\rangle - \langle f, \Lambda(t)\rangle\right)$$

converges in the finite distribution sense towards a centered Gaussian field Y_f, for any reasonable test function f.

After these, many people have worked in the central limit problems around the hydrodynamic limits. For instance, the classical work of Shiga and Tanaka [146]; Zhu [173], Chang and Yau [33] proved the CLT in the Ginzburg-Landau model; Graham [71] gave the CLT in a large queuing system with power of choices policy.

In this thesis

In Chapter 3 and 4, two symmetrically interacting particles systems have been studied. Unlike the classical mean field models, the interactions among particles in our cases can not be interpreted explicitly by the empirical measures, which are the main difficulties in our studies. In Chapter 6, we present a study of the fluctuation problem around the re-scaled empirical measure in the classical stochastic Becker-Döring model. For an exhaustive introduction of our work related to mean field theory, see Section 1.3.2.

1.3 Main Contributions

1.3.1 Stochastic Averaging Principle and Multiple Time Scales

In Chapter 2 and Chapter 5, we have investigated two large interacting particle systems with an absorbing state. We could apply the classic mean field method to translate the problem with large number of particles into an one-body problem for the empirical measure. However, in our cases, the scaling limits of the empirical measures $(X^N(t)/N)$ do not capture the absorption phenomenon. It indicates that the absorption could only happen after "a very long time". Thus, in order to have an estimation of

the instant of absorption, instead of studying the fluid limit of the normal time scale of the empirical measures $(X^N(t)/N)$, we study the asymptotic behavior of the process on larger time scales: *i.e.*, for a $k\geq 1$, the scaling limit of the process in the form

$$\left(\overline{X}^N(t)\right) := \left(\frac{X^N(N^k t)}{N}\right).$$

The tightness of the sequence of processes $(\overline{X}^N(t))$ is difficult to obtain mainly due to the large fluctuations occurring when $k>1$. Naturally, by using Doob's inequality, the martingale part of process $(X^N(N^k t))$ should be of the order $N^{k/2}$ as N gets large which is not obviously vanishing when scaled by N. Recall that the coordinates of $(X^N(t))$ are counting measures and then upper bounded by the total number of the particles which is of the order of N.

To control the fluctuations, we make a careful analysis of the transition of the process $(X^N(N^k t))$ and find that the coordinates processes $(X_i^N(t))$ evolves on multiple different scales in our systems. For this reason, there are stochastic averaging phenomena arising in the dynamic of the empirical measure processes. We should remark that the multiple scales of evolving speeds make the stochastic averaging principles much more difficult to study than the dynamic with only "fast" and "slow" processes.

In the following, we will describe briefly the main problems in Chapter 2 and Chapter 5 and give the sketch of the proofs. Roughly speaking, the empirical measure processes in these two systems share some common properties:

1. they have absorbing states;
2. they live in multi-dimensional spaces;
3. asymptotically, the process moves much faster away from than towards the absorbing state.

To be more specific on the speed of absorbing, the speed of moving away and the consequences on the multiple time scales, we will take the processes in Chapter 2 for instance.

The Transient Process *v.s.* The normal time Fluid Limit

In Chapter 2, a large system with $F_N = \lfloor \beta N \rfloor$ particles has been investigated. In this system, each particle has $d+1$ types of states, labeled by $0, 1, \ldots, d$. For a particle at state k, where $d\geq k\geq 1$, it will jump to the state $k-1$ at rate $k\mu$. Once it reaches state 0, it is dead. There is a strong force that favors the upwards jumps: it is assumed that at rate λN, one of the particle at the lowest state (≥ 1) will have a upward jump.

The empirical measure of this system can seen as a $d+1$-dimensional transient process $(X^N(t)/F_N)$, where $(X^N(t)) = (X_0^N(t), X_1^N(t), \ldots, X_d^N(t))$ is a process taking values in $\mathbb{N}^{d+1}$, satisfies the conservation law

$$\sum_{i=0}^{d} X_i^N(t) \equiv F_N,$$

where $F_N = \lfloor \beta N \rfloor$. For all $i=0,\ldots,d$, $X_i^N(t)$ represents the number of particles of type i at time t. Clearly, this process has an absorbing state $(F_N, 0, \ldots, 0)$.

In Section 2.3, we have proved the tightness and convergence of the empirical measure $(X^N(t)/F_N)$ and obtained the limit through a generalized Skorohod problem. [1] Moreover, for the case $2\mu\beta < \lambda$, one has

$$\lim_{N\to\infty}\left(\frac{X_0^N(t)}{F_N}\right) \equiv (0).$$

This means that the fluid limit of the normal time scaled empirical measures is not adequate to describe the absorption.

The Fast Process *v.s.* The Slow Process

In this model, the transition matrix $(q^N(\cdot,\cdot))$ is given by follows

$$\begin{cases} q^N(x, x - e_p + e_{p-1}) = \mu p x_p & \forall\ p = 1, \dots, d, \\ q^N(x, x + e_p - e_{p-1}) = \lambda N \mathbb{I}_{\{x_1 = \cdots = x_{p-2} = 0, x_{p-1} > 0\}} & \forall\ p = 2, \dots, d. \end{cases}$$

where $(e_p, 0 \le p \le d)$ is the standard orthonormal basis of $\mathbb{N}^{d+1}$. Intuitively, one can observe from this matrix, for the case $\lambda N > \mu p F_N$, if N is large enough, the upward rates are much larger than the downward rates. It means that the particles are more likely staying above the critical level p. In Proposition 2.7, we have proved this observation by using an auxiliary process

$$Z_{p-1}^N(t) := (p-1)X_1^N(t) + (p-2)X_2^N(t) + \cdots + X_{p-1}^N(t)$$

and showing that it is upper bounded by an ergodic process with transition rates in the order of N. That gives us the idea to decouple the process by

$$\begin{cases} Y^N(t) := (X_1^N(t), \dots, X_{p-1}^N(t)), & \text{the fast part,} \\ \overline{S}^N(t) := \frac{1}{N}(X_p^N(t), \dots, X_d^N(t)), & \text{the slow part.} \end{cases}$$

If we follow the classical procedures described in Section 1.2.2, we need to prove the convergence and then identify the limit of the occupation measures,

$$\nu_p^N((0,t) \times \cdot) := \int_0^t \mathbb{I}_{\left\{(X_1^N(s), \dots, X_{p-1}^N(s)) \in \cdot\right\}}\, \mathrm{d}s,$$

which are random measures in multi-dimensional space $\mathbb{R}^+ \times \mathbb{N}^{p-1}$. The local times in the transition and the multi-dimensionality make the stochastic averaging problem complicated.

The Multiple Time Scale and Scaling limit

In order to investigate the fast process, we have analyzed the marginal distributions of these occupation measures. In Proposition 2.3 and Proposition 2.7, we find the critical

1. For more about Skorohod problems, see the appendix in Chapter 2.

time scale $t :\to N^{p-1}t$ and prove the local times of these marginal processes on $[0, N^{p-1}t]$ are of different scale: *i.e.* for any $1 \leq k \leq p-1$ and any $\gamma > 0$

$$\int_0^{N^{p-1}t} \mathbb{I}_{\{X_k^N(s)>0\}}\, \mathrm{d}s \leq \int_0^{N^{p-1}t} X_k^N(s)\, \mathrm{d}s \sim o\left(N^{k+\gamma}\right).$$

It means that the process $(X_{p-1}^N(\cdot))$ is the dominating component in the fast process. More precisely, we can give a sequence of identical relations for the scaled processes

$$\frac{\lambda}{\mu}\left(\frac{1}{N^k}\int_0^{N^{p-1}t} X_k^N(s)\, \mathrm{d}s\right) \sim \left(\frac{1}{N^{k+1}}\int_0^{N^{p-1}t} X_{k+1}^N(s)\, \mathrm{d}s\right), \qquad \forall\, k = 1, \ldots, p-2,$$

in Proposition 2.4 and Proposition 2.7. It indicates that one should study the scaled marginal occupation measure

$$\omega_{p-1}^N((0,t) \times \cdot) := \frac{1}{N^{p-1}} \int_0^{N^{p-1}t} \mathbb{I}_{\left\{X_{p-1}^N(s)>0\right\}}\, \mathrm{d}s.$$

Then by studying the stochastic averaging principle in the coupled system:

$$(\overline{S}^N(N^{p-1}), \omega_{p-1}^N),$$

the slow dynamic and the occupation measure based on the dominating component of the fast process, we give the limit in Theorem 2.2 and Theorem 2.4,

$$\lim_{N\to\infty}\left(\frac{X^N(N^{p-1}t)}{N}\right) = (\Phi(t))$$

where $(\Phi(t))$ is a non-trivial function. It means that under condition $\mu p\beta < \lambda < \mu(p+1)\beta$, the absorbing moment is in the order of N^{p-1}. The explicit expression of the estimation of the absorbing moment is stated in Corollary 2.2. In addition, we also provide the central limit results along this convergence (c.f. Theorem 2.3).

1.3.2 Mean-Field limits

This scaling regime has been used only in the communication network framework when the number of nodes of the network goes to infinity. In a classical mean-field context, the evolution equations of the empirical distribution of the states of the particles can be expressed through a set of stochastic differential equations which can be amenable to an asymptotic analysis. Our two studies in this domain, Chapter 3 and 4 respectively, have differed from this context in two ways.

1. In Chapter 3, we have studied a system where, in a Markovian context, the natural state space of a particle is a space of dimension N, and therefore not really suitable for a classical mean-field analysis. The approach we have used has consisted in taking a reduced (non-Markovian) representation of the state of a node so that the dimension of the state space is fixed. In this setting, we have proved through various estimates that a mean-field result holds. In the limit, this non-Markovian representation is converging to a Markovian McKean-Vlasov process which can be analyzed.

2. In Chapter 4, a Markovian system with local interactions has been investigated. For these systems, a node does not interact with all the other nodes in the system as in a classical mean-field context, but only with a subset nodes, called "neighborhood" whose size h_N is converging to infinity as N gets large. Intuitively, the associated asymptotic McKean-Vlasov process should not depend on h_N provided that h_N grows to infinity sufficiently fast. It turns out that this result is not easy to prove. One of the main difficulties is that the evolution equation of the empirical distribution of all nodes highly depends on the structure of the "neighborhoods" and therefore cannot be expressed in a set of autonomous equations. Another difficulty is that the interactions (1) are unbounded and follow distributions depending on the current state; (2) occur among random sets of nodes.

   The (quite technical) method we have used here is to (1) for a given node, introduce the local empirical measures associated to the nodes in neighborhood of this node and study the global empirical measure of these local empirical measures. An autonomous evolution equation holds for the corresponding process; (2) prove the convergence for this process with values in a measure space on another measure space. With this method, a mean-field result has been proved for all growth rates of (h_N), the size of neighborhood, that tends to infinity.

1.3.3 Functional Central Limit Theorem

Outside these two main scaling regimes described above, in Chapter 6, we have proved a functional central limit theorem for the Becker-Döring model, which is a classical model in coagulation-fragmentation theory. In this model, the state space of a Markovian description is an infinite dimensional space. The main difficulty here is of controlling the fluctuations of the first coordinate since it has interactions with all the other coordinates. To handle this problem, we have proposed a convenient Hilbert space such that: (1) the potential limiting SDE is well-defined; (2) tightness and weak convergence of the fluctuations can be proved in this space.

We are now going to describe more precisely the stochastic models analyzed in this document. Section 1.4 presents the stochastic models of the communication networks context and Section 1.5 the analogous section for the polymerization processes.

1.4 Summary of Part I: Large stochastic networks with failures

In the first part, we propose three stochastic models to analyze the long time performance of algorithms in large unreliable stochastic networks. Here, we will begin with the introduction of the framework of large network with failures, duplication algorithms and allocation algorithms. Then we will explain the stochastic models to discribe the dynamic of these algorithms. We will state our main results and outline the plan of Chapter 2, 3, 4 briefly in advance.

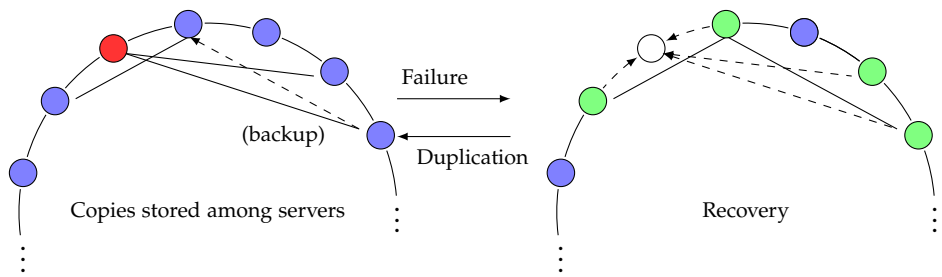
A large unreliable network is a system in charge of managing a large number of files in a network of servers in the Internet. Files have copies stored on the servers. Each server may have a failure and, in this case, all copies on it are lost. If all copies of a given file are lost then the file is lost for good. Despite the fact that failures of servers rarely

occur, in a system with a large number of servers, having failures is not a negligible phenomenon. For example, in a data center with 200,000 servers, in average five servers fail every day. See the extensive study Pinheiro et al. [130] in this domain at Google.

In order to prevent file losses, it is necessary to have copies of files on different servers. Consequently, each server should devote a part of its processing capacity to duplicate files on other servers. Duplication mechanism has thus a cost on the processing capacities of the server to achieve other tasks. One hand, maintaining too many copies of the same file is a waste of processing capacity, bandwidth and storage of servers, on the other hand having too few of them may lead to the loss of all copies of a file if the servers having its copies fail before any recovery procedure can happen.

Another requirement of these systems is load balancing. If a heavily loaded server, i.e. having many copies, fails then it will trigger a lot of activity (processing and network) because recovery mechanisms have to be launched for all copies lost on it. If, because of congestion, the procedure is too long, it may end-up in losing files if other servers fail in the meantime. Therefore, a good allocation policy should avoid unbalanced loads on servers and, at the same time, keep as many files as the capacity of the system permits.

See the following figure for example. The circles represent servers in the network. There is an edge between different servers if there exists at least one file that has copies on these two servers. Servers can make copies on other servers. When the red server failed, copies on it are lost. For any lost copy, the servers that have a copy of the same file (green servers) may help to recover the data on the failed server. If there is no another copy available, the copy could not be retrieved and the file would be lost for good.



In general, duplication algorithms is defined by the rate at which a copy is done and by the allocation algorithm that determines the location of the new copy. Concerning the duplication rate, we will study two situations.

Global duplication The capacity of all servers is assumed to be available to copy any file in the system on a given server. The policy is to duplicate files with the least number of copies in the system. For example, in a system of N servers, each server devotes capacity λ for duplication, then the system has capacity λN to duplicate one file at any moment. It is the optimistic duplication policy for the whole system since the most vulnerable files are taken care of.

Local duplication The duplication capacities are assumed to be local in the sense that each server is making copies of files that it has stored.

Allocation Algorithms When a server generates a new copy, a server is chosen to receive this copy. Computer scientists have done experiments and have shown that allocation policies may lead to different distributions of loads on servers. The unbalance loads on servers may speed up the losses of files and is thought to be a waste of capacity. Heavily loaded servers may lose most of their files before making copies. At the opposite, the bandwidth and storage of an almost empty server are wasted. We introduce three allocation policies: Random, Least loaded and Power of Choices. These policies can be applied among all servers, or in a subset of servers, called "neighborhood".

— **Random Policy** Under Random Policy, a server generates a new copy on a server chosen with uniform probability on all the other nodes.

— **Least Loaded Policy** The Least Loaded Policy is the most effective way that can prevent heavily loaded servers. In this case, the new copy will be sent to the server with least number of copies in the system. This policy manages to keep the loads on servers almost constant. However, it also requires the knowledge of all the loads of all nodes of the network which is expensive in terms of communication.

— **Power of choices** For this policy when a new copy is generated, the server will randomly choose m servers and send the new copy to the least loaded one. This policy is a compromise between communication costs (m messages) and the choice of a node with a minimal load. It can lower the load by just checking some finite subset of servers instead of all of them. In the context of static allocation schemes of balls into bins in computer science, this policy has been first studied by Mitzenmacher and others, see [112] for a survey.

— **Allocation in neighborhood** Each server is connected with a subset of servers, called "neighborhood". When a new copy is generated from a given server, it will be placed in a server chosen from this neighborhood, following some Choice Policy.

The problem of reliability of such large systems has been studied mainly in theoretical computer science. For example, Rhea et al. [137] and Rowstron and Druschel [144] studied duplication policies by simulations for the design the large scale distributed system. Simple mathematical models using birth and death processes have been proposed by Chun et al. [36], Picconi et al. [128] and Ramabhadran and Pasquale [134]. In Feuillet and Robert [61], a stochastic model of the case when each file has at most two copies has been analyzed in a simplified setting.

The efficiency of a duplication mechanism can be expressed as a trade off between the number of files a system can maintain and the lifetime of these files. If there are many files with few copies, then very likely consecutive failures of servers will lead quite quickly to losses of files. On the other hand, if there are few files but with many copies, their durability will be preserved with high probability at the expense of a reduced capacity of the system.

1.4.1 The Global Duplication Model

The performance of the global duplication is investigated in Chapter 2. We briefly describe the associated stochastic model. This is a closed system with N servers and F_N files, there is no external file arriving in this system.

1. The total capacity of N servers is λN. For this model, a new copy of a file occurs at rate λN and only for a file with the lowest number of copies in the system.
2. Each copy has a lifetime which exponentially distributed with parameter μ.
3. The maximum number of copies for a file is d.
4. Initially, all files have d copies in the system.
5. Scaling assumption for $\beta>0$,
$$\lim_{N\to\infty} \frac{F_N}{N} = \beta.$$

Recall that a file that has no copy is lost for good. For this model, we do not need to keep track of the locations of copies in particular the impact of allocation policy is meaningless in this context. Note that this model provides, in some way, an upper bound on the efficiency of these mechanisms since the most vulnerable files (the files with the least number of copies) are duplicated in priority.

The evolution of the system can be expressed by a $d+1$-dimensional Markov jump process $(X_i^N(t), i = 0, \ldots, d)$. For $i = 0, 1, \ldots, d$, $X_i^N(t)$ denotes the number of files with i copies at time t. See Figure 1.1 for the possible jumps of the coordinates of this process. It is assumed that the initial state is optimal, all files have the maximum number of copies, i.e. $X_d^N(0) = F_N$ and $X_i^N(0) = 0$ for $0 \leq i < d$. It is transient with a unique absorbing state $(F_N, 0, \ldots, 0)$ for which all files are eventually lost.

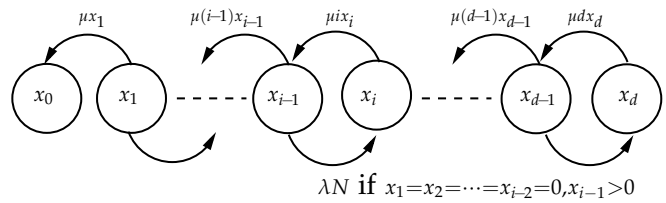


Figure 1.1 – Jump Rates between the coordinates of the Markov Process $(X^N(t))$

We are interested in the decay rate of the system, the duration of time after which a fraction of files is lost. For any $\delta \in (0,1)$, let $T_N(\delta)$ be the first time that a fraction δ of the $F_N \sim \beta N$ initial files are lost:

$$T_N(\delta) = \inf\left\{t \geq 0 : X_0^N(t)/N \geq \delta\beta\right\}.$$

We study the transient Markovian process $(X^N(t))$ and obtain the following results.

— Under the condition $\lambda > d\mu\beta$, we have shown that the system is stable in the following way, for the convergence in distribution of processes,

$$\lim_{N\to+\infty} \left(\frac{X_d^N(t)}{N}\right) = (\beta),$$

at the first order none of the files have been lost and all of them have the maximal number of copies d.

— Under the condition $\lambda < d\mu\beta$, i.e. $p\mu\beta < \lambda < (p+1)\mu\beta$ for some $2 \leq p \leq d-1$, then we have proved that

$$\lim_{N\to+\infty}\left(\frac{X_p^N(t)}{N}, \frac{X_{p+1}^N(t)}{N}, t>0\right) = \left((p+1)\beta - \frac{\lambda}{\mu}, \frac{\lambda}{\mu} - p\beta, t>0\right),$$

all files have immediately either p or $p+1$ copies.

Hence if $\lambda > d\mu\beta$ holds the decay of the network does not happen on the "normal" time scale. The main result of this work shows that the variable $T_N(\delta)$ is of the order of N^{d-1}. It leads us to study the asymptotic behavior of a scaled process on the time scale $t \mapsto N^{d-1}t$ and prove the following convergence result for the number of lost files,

$$\lim_{N\to+\infty}\left(\frac{X_0^N(N^{d-1}t)}{N}\right) = (\Phi(t)), \tag{1.5}$$

where $(\Phi(t))$ is non-trivial deterministic function such that $\Phi(0) = 0$ and

$$\lim_{t\to+\infty}\Phi(t) = \beta,$$

all files are eventually lost on this time scale. This is done by using the associated stochastic differential equations (SDEs) satisfied by $X^N(t))$ and a series of technical arguments which I describe briefly.

— Relation (1.5) gives a convergence to a deterministic process, in particular the stochastic fluctuations vanish at infinity. Because of the very rapid time scale, this phenomenon is not easy to establish, a direct use of the SDEs gives that the martingale terms are of the order of N^{d-1}/N^2 and, consequently, does not vanish when $d \geq 3$. The problem is taken care of in several steps.

  A first argument consists in establishing a stochastic uniform upper bounded for the processes $(X_i^N(t), i = 1, \ldots, d-1)$, it is obtained by a coupling with an ergodic Markov processes. Then, by using a series of quadratic equations, we prove a key (highly) technical result that shows a kind of balance of flows between the components of the state variable, for all $1 \leq i \leq d-2$,

$$\lim_{N\to\infty}\left(\frac{1}{N^{i+1}}\int_0^{N^{d-1}t}\left[(i+1)\mu X_{i+1}^N(u) - \lambda N X_i^N(u)\right]\mathrm{d}u\right) = 0. \tag{1.6}$$

  Finally these results give an asymptotic relation between the scaled processes $(X_0^N(N^{d-1}t)/N)$ and $(X_{d-1}^N(N^{d-1}t))$.

— To establish (1.5) one has then to remark that $(X_{d-1}^N(N^{d-1}t))$ is a "fast" process interacting with a "slow" process, $(X_0^N(N^{d-1}t)/N)$. The evolution between the fast process and the slow process generates a stochastic average problem which we solve. By combining all these arguments, Convergence (1.5) is established.

— A central limit theorem associated to (1.5), due to the stochastic averaging context it is not straightforward. A refined version of the convergence (1.6) is derived to get the corresponding result.

1.4.2 The Local Duplication Model

In Chapter 3, we have studied the long time performance of the local duplication algorithm in a closed system with N servers, F_N files. The maximum number of copies that a file can have is denoted by d.

1. The duplication processes are assumed to be local. Each server generates a new copy of one of its files after any exponential distributed amount of time with parameter λ. The copy is done on another server following the Random Choice Policy.
2. On each server, the duplication capacity is used only on the files with the least number of copies.
3. Each server fails after any independent exponentially distributed amount of time with rate μ. A failed server is immediately replaced by a new, empty, server. Upon this failure time, it loses all the copies on it and is replaced by a new server immediately.
4. Scaling assumption,
$$\lim_{N\to\infty} \frac{F_N}{N} = \beta.$$

In this model, the locations of different copies of a given file play an important role. However, keeping track of the locations of all copies may lead to large state space which is very complex to study. For example, for $d = 2$, let $X^N_{i,j}(t)$ be the number of files that have copies on server i and server j. The process $(X^N_{i,j}(t))_{i=1,\dots,N;i\le j\le N}$ is a $N(N+1)/2$ dimensional Markov process. In general case, the dimension of the state space is in the order of N^d and a huge number of equations are required.

To avoid the large dimension space, we introduce a reduced representation of the system for the case $d = 2$.

1. Let $R^N_{i,1}(t)$ be the number of files that only have 1 copy on server i,
$$R^N_{i,1}(t) = X^N_{i,0}(t).$$
2. Let $R^N_{i,2}(t)$ be the number of files that have 1 copy on server i and have 1 copy on another server at time t,
$$R^N_{i,2}(t) = \sum_{j\neq i} X^N_{i,j}(t).$$
3. Instead of studying the $N(N+1)/2$ dimensional Markov process $(X^N_{i,j}(t))$, we focus on the $2N-$dimensional non-Markovian process
$$R^N(t) := (R^N_{i,1}(t), R^N_{i,2}(t))^N_{i=1}.$$

The loss of the Markov property is the price for the dimension reduction. However, through the process $R^N(t)$, we can obtain an estimate of the decay rate of files in the model. A possible approach to study $R^N(t)$ is of using mean-field convergence results by considering the evolution of the empirical distribution

$$\Lambda^N_t(\cdot) := \frac{1}{N}\sum_{i=1}^{N} \delta_{(R^N_{i,1}(t),R^N_{i,2}(t))}(\cdot).$$

The main results are the following. We prove that this non-Markovian description of the network is asymptotically a nonlinear Markov jump process, by the following steps. First, we show the existence and uniqueness of a possible asymptotic nonlinear Markov jump process. In this case, a fixed point problem in the probability space of the Skorohod space is studied. A technical result is then proved : on any finite time interval $[0, T]$, for a given server, with high probability, the load on it is bounded and all other servers have almost surely one file in common with it, i.e. there exists a constant $C(T)$, such that for any given i,

$$\mathbb{P}\left(\sup_{j\neq i, 0\leq t\leq T} X_{i,j}^N(t) \geq 2\right) \leq \frac{C(T)}{N}.$$

This technical result makes it possible to handle the evolution equations related to the reduced process $(R^N(t))$ without using the transitions of the complete vector $(X_{i,j}^N(t))$. Finally, it is shown that the sequence of empirical distributions of reduced process $(R^N(t))$ is converging to the distribution of the possible asymptotic nonlinear Markov jump process. With these results, we can obtain a lower bound for the asymptotic exponential decay rate of the total number of files.

For $d > 2$, we introduce a related simplified model based on the reduced processes and apply a mean-field method. In this case, we show that asymptotically, the evolution of total number of files is upper bounded by a function $K_0 e^{-\mu\kappa_d^+(\lambda/\mu)t}$, where $K_0, \kappa_d^+(\lambda/\mu)$ are positive constants and $\kappa_d^+(\cdot) < 1$. The value of $-\kappa_d^+(\lambda/\mu)$ is the largest eigenvalue of the matrix associated with the limiting nonlinear Markov process. Although there is no explicit expression of $\kappa_d^+(\lambda/\mu)$, We obtain a good estimate for it. Moreover, we prove that $\kappa_d^+(\lambda/\mu)$ is also the lower bound on the exponential decay of the number of files in the non simplified system.

1.4.3 Allocation algorithms

In Chapter 4, we investigate the long time performances of several allocation policies in neighborhoods in a large stochastic network with failures. As before it is a closed system with N servers, labeled by $1, \ldots, N$. Initially, there are F_N files and the scaling condition

$$\lim_{N\to\infty} \frac{F_N}{N} = \beta$$

is assumed for some $\beta>0$. The life time of each server is exponentially distributed with parameter 1, a failed server is immediately replaced by an empty new server at the same time. For $i = 1, \ldots, N$, the "neighborhood" of server i is given by

$$\mathcal{H}_i^N := \left\{ j \middle| \text{server } i \text{ and server } j \text{ are connected} \right\},$$

the cardinality of $\mathcal{H}_\cdot^N$ is denoted by h_N, it is assumed the (h_N) is converging to infinity. To evaluate the allocation algorithms, the following assumption is done, once a server has a failure, the files that were on it are immediately assigned to the servers in its neighborhood according to the allocation policy; the failed server is replaced by a new empty server immediately, inheriting the same neighborhood.

A measure-valued Markovian description Intuitively, the simplest description of the system is the Markov processes $L^N(t) = (L_1^N(t), \ldots, L_N^N(t))$, where $L_i^N(t)$ is the load on server i at time t. By the assumption, at rate 1, a server i will have a failure and the files on server i are re-placed on the servers of $\mathcal{H}_i^N$. Therefore, the transition matrix $(q^N(\cdot,\cdot))$ at state l can be given by

$$q^N(l, l - l_i e_i + z_i^j e_j) = p^N(z = (z_i^j), l, \mathcal{H}_i^N),$$

where $p^N(\cdot)$ is a probability function depends on the neighborhood $\mathcal{H}_i^N$. Since the the interactions in neighborhoods $(\mathcal{H}_i^N, i = 1, \ldots, N)$ cannot be expressed only in terms of the empirical measure of the process $L^N(\cdot)$, the classical mean-field approach cannot be used in this case. After careful estimations on the probability function $p^N(z, l, \mathcal{H}_i^N)$, we show that this expression can be approximated by a simple function of z, l, and of the local empirical measure

$$\Lambda_i^N(t) := \frac{1}{h_N} \sum_{j \in \mathcal{H}_i^N} \mathbb{I}_{\left\{L_j^N(t)\right\}}.$$

This fact suggests the study of the measure-valued interacting system

$$\Lambda^N(t) = (\Lambda_1^N(t), \ldots, \Lambda_N^N(t))$$

and the evolution equation of the global empirical measure of $\Lambda^N(t)$, i.e.,

$$\mathcal{P}^N(t) := \frac{1}{N} \sum_{i=1}^{N} \delta_{\Lambda_i^N(t)}.$$

We may then expect that the limit of $(\mathcal{P}^N)$ is a Dirac measure at the distribution of a McKean-Vlasov process $(L(t))$. Moreover, note that the empirical measure of $L^N(t)$ is actually the arithmetic average of processes $(\Lambda_i^N(t))$, therefore, this mean field convergence of $(\Lambda^N(t))$ may give the desired mean-field convergence of the system $(L^N(t))$ whose limit is the process $(L(t))$.

The main results are the following.

1. For all h_N, which tend to infinity along with N, under certain conditions, the local empirical measure processes converge in distribution to the distribution of a non-linear jump process.
2. The mean field limit and propagation of chaos hold for the whole system $(L_i^N(t), 1 \leq i \leq N)$.
3. When h_N is linear with N, we have the convergence of the stationary distributions of the N servers system to the stationary distribution of the limit system.
4. For some allocation algorithms, the stationary distributions of the limit non-linear jump processes have nice properties in a heavy load regime.

We present some details of the proofs of the results for the power of d Choices. For Random policy, the approach is similar. One of the most important quantity in an assignment policy is the probability that a given server being chosen. We give quite precise estimates of this quantity for N large. At any failure moment, the increments of the number of files at the servers follows a multinomial distribution. We show that

with a high probability, these increments have the value 0 or 1. This property allows us to approach the increments by Poisson random measures. We then give the definition and prove the existence and uniqueness for the possible limiting process $(L(t))$, the McKean-Vlasov process,

$$\mathrm{d}L(t) = \int_0^\infty \mathbb{I}_{\left\{0 \le r \le \beta \frac{\Lambda_{t-}[L(t-),\infty)^d - \Lambda_{t-}[L(t-)+1,\infty)^d}{\Lambda_{t-}\{L(t-)\}}\right\}} \mathcal{N}(\mathrm{d}r, \mathrm{d}t) - L(t-)\mathcal{N}(\mathrm{d}t),$$

where $\Lambda_t(\cdot) = \mathbb{P}(L(t) \in \cdot)$. We show a mean field convergence holds for $(\Lambda^N(t))$ (and therefore of $(L^N(t))$). We prove that the processes $L^N(t)$ and the limit McKean-Vlasov process $L(t)$ are ergodic and that the empirical distribution of the stationary distribution of $L^N(\cdot)$ converges to the stationary distribution of the limit McKean-Vlasov process when h_N is linear with N. The main striking result is the following convergence result for the stationary distribution of the McKean-Vlasov process. If X_β be the random variable on $\mathbb{N}$ whose distribution is the stationary distribution of the limit McKean-Vlasov process, then, for the convergence in distribution,

$$\lim_{\beta \to \infty} \frac{X_\beta}{\beta} \overset{\text{law}}{=} \frac{d}{d-1}(1 - U^{d-1}),$$

where U is a uniform random variable on $[0, 1]$. This result implies in particular that, the load of an arbitrary node of the Mc-Kean Vlasov model has an asymptotic finite support of the order of $\beta d/(d-1)$ when β is large.

1.5 Summary of Part II: Variations in coagulation and fragmentation models

In the second part of this thesis, we investigate two stochastic models to understand the variability observed in experimental data of biological polymerization processes.

These polymerization phenomena arise in a large variety of domains, like in aerosol science, atmospheric physics, polymer science and also biology. The purpose of coagulation and fragmentation models is the description of the reactions involving growth and degradation of polymers. To be accurate, a complete model for such reactions should involve the size, position and velocity of each particle, which is too complicated for analysis. As a simplification, we consider models where a monomer is a molecule of size 1 and a polymer is a macro-molecule, composed by monomers. Furthermore the dynamic of the system depends only on numbers of polymers of a given size. The growth mechanism can be seen as polymers and monomers assembling to form a larger polymer, for example, n polymers/monomers of size x_k, k=1, . . . , n may form a a polymer of size x=x_1+x_2+ · · · +x_n,

$$(x_1) + (x_2) + \cdots + (x_n) \longrightarrow (x)\,.$$

The fragmentation mechanism occurs when an unstable polymer breaks into smaller pieces, for example, a polymer of size x may break into polymers/monomers of sizes x_k, k=1, . . . , n,

$$(x) \longrightarrow (x_1) + (x_2) + \cdots + (x_n),$$

if $x=x_1+x_2+\cdots+x_n$. In a deterministic setting, these reactions can be described by a set of differential equations. The literature of these dynamical systems is very rich, mainly concerning the well-posedness, long-time behavior, metastability. See [15] for example. These ODEs can be seen as the thermodynamic limit of the time evolution of polymers when the initial number of monomers is large, yet the thermal fluctuations are not represented by these deterministic dynamical systems. These aspects are important, for instance, it is known from experiments that the lag time, i.e. the first time a fraction of monomers is polymerized, exhibits a large variance. See Szavits-Nossan et al. [157] and Xue et al. [168] for example. Related phenomena also occur in different contexts, such as the lag time of nucleation of ice in supercooled small water droplets, see Heneghan et al. [74]. See Aldous [5] for a general survey on the coalescence models.

In the following, we first introduce two classical models, the Smoluchowski Model and the Becker-Döring Model, for both a deterministic and a stochastic descriptions. We will bring up some useful properties and literatures of these two models. Then we present our contribution concerning the fluctuations in coagulation and fragmentation models.

The Smoluchowski Model

Only binary interactions are considered in the classical Smoluchowski model (1917) [163]. This model focus on the case when coagulation occurs between only two particles and the unstable polymers can only break into two particles. The quantity $K(i,j)$ denotes the chemical rate at which a polymer of size i reacts with a polymer of size j and the quantity $F(i,j)$ is the chemical rate at which a polymer of size $(i+j)$ breaks into two polymers of size i and size j. The chemical reactions thus can be represented as

$$(i)+(j) \underset{F(i,j)}{\overset{K(i,j)}{\rightleftharpoons}} (i+j).$$

The kernel $K(\cdot,\cdot)$ and $F(\cdot,\cdot)$ are (clearly) symmetrical, i.e., $K(i,j)=K(j,i)$ and $F(i,j)=F(j,i)$. The deterministic description of this system is an infinite set of non-linear ordinary differential equations given by

$$\frac{\mathrm{d}c_j}{\mathrm{d}t}(t) = \frac{1}{2}\sum_{k=1}^{j-1}\left(K(j-k,k)c_{j-k}(t)c_k(t) - F(j-k,k)c_j(t)\right) - \sum_{k=1}^{\infty}\left(K(j,k)c_j(t)c_k(t) - F(j,k)c_{j+k}(t)\right),$$

where $c(t) = (c_j(t), j \in \mathbb{N}^+)$ is the expected number of cluster of size j per volume at time t.

If initially there are only N monomers, N is in particular the total mass of the system, a corresponding stochastic model is an N-dimensional Markov process $(X^N(t)) = (X_k^N(t))$ where $X_k^N(t)$ is the number of polymer of size k at time t. The associated transition matrix $q_s^N(\cdot,\cdot)$ is given by

$$\begin{cases} q_s^N(x, x-e_i-e_j+e_{i+j}) = K(i,j)x_ix_j/N, \\ q_s^N(x, x+e_i+e_j-e_{i+j}) = F(i,j)x_{i+j}, \end{cases}$$

where $(e_k, k \in \mathbb{N}^+)$ is the standard orthonormal basis of $\mathbb{N}^{\mathbb{N}^+}$.

It has been proved by Jeon (1998) [83] that, that under condition

(a) $\lim_{i+j\to\infty} K(i,j)/(ij) = 0$,

(b) there exists $G(i+j)$ such that $F(i,j) \leq G(i+j)$ and $\lim_{i+j\to\infty} G(i+j) = 0$,

the scaled process $(X^N(t)/N)$ converges to the solution of the Smoluchowski equation (thermodynamic limit) in the L^2-norm. The gelation phenomenon corresponding to an asymptotic loss of mass after finite time is also investigated in this reference.

The Becker-Döring Model

The Becker-Döring model (1935) [16] can be seen as a simplified version of the Smoluchowski model. It describes the system where only additions (coagulation) or removals (fragmentation) of one monomer from a particle are possible, i.e.,

$$(1) + (k) \underset{b_{k+1}}{\overset{a_k}{\rightleftharpoons}} (k+1),$$

where $a_k{=}K(k,1)$ and $b_{k+1} = F(k,1)$ are the chemical rates.

The ODEs associated to this model is a (classical) infinite system of ordinary differential equations of $c(t){=}(c_k(t), k \in \mathbb{N}^+)$, given by

$$\begin{cases} \dfrac{\mathrm{d}c_1}{\mathrm{d}t}(t) = -2J_1(c(t)) - \displaystyle\sum_{k\geq 2} J_k(c(t)), \\ \dfrac{\mathrm{d}c_k}{\mathrm{d}t}(t) = J_{k-1}(c(t)) - J_k(c(t)),\ k{>}1, \end{cases} \tag{BD}$$

with $J_k(c){=}a_k c_1 c_k - b_{k+1} c_{k+1}$ if $c{=}(c_k){\in}\mathbb{R}_+^{\mathbb{N}^+}$. For $k{\geq}1$, $c_k(t)$ represents the concentration of clusters of size k at time t. The conditions on existence/uniqueness of solutions for the Becker-Döring equations (BD) have been extensively investigated. See Ball et al. [15], Niethammer [118] and Penrose [125, 126]. Moreover, under certain assumption the equilibrium point $\widetilde{c}$ of the Becker-Döring equation (BD) has an explicit form.

The stochastic description of this model can be represented by a N-dimensional Markov jump process $(X^N(t)){=}(X_k^N(t), 1{\leq}k{\leq}N)$ where, as before, $X_k^N(t)$ is the total number of polymers of size k at time t and N is the initial mass of monomers. The transition matrix $Q_{\mathrm{BD}}^N = (q_{\mathrm{BD}}^N(\cdot,\cdot))$ is given by

$$\begin{cases} q_{\mathrm{BD}}^N(x, x - e_1 - e_k + e_{k+1}) = a_k x_1 x_k / N, \\ q_{\mathrm{BD}}^N(x, x + e_1 + e_k - e_{k+1}) = b_{k+1} x_{k+1}, \end{cases}$$

where $(e_k, k{\in}\mathbb{N}^+)$ is the standard orthonormal basis of $\mathbb{N}^{\mathbb{N}^+}$. As a special case, the first order convergence for this model is given by Jeon's result. In Chapter 6, we study the second-order of this thermodynamic limit by providing a functional central limit theorem.

1.5.1 The large variability of the lag time for nucleation

In Chapter 5, we investigate the nucleation phenomenon in a model where the coagulation mechanism occurs only through the successive additions of monomers

while the fragmentation mechanism has a generalized form, i.e.,

$$\begin{cases} (1)+(k) & \xrightarrow{\kappa^k_{\text{on}}} (k+1), \\ (k) & \xrightarrow{\kappa^{k,a}_{\text{off}}} (a_1)+(a_2)+\cdots+(a_p), \qquad p\geq 2,\ a_1+\cdots+a_p=k, \end{cases}$$

where κ^k_{on} and $\kappa^{k,a}_{\text{off}}$ are the corresponding reaction rates. Let κ^k_{off} be the degradation rate of a polymer of size k, i.e. the sum of all $\kappa^{k,a}_{\text{off}}$ over all possible fragmentation vectors a.

The key assumption for nucleation is that there exists a critical size n_c, called *nucleus size* such that, for a polymer of size less than n_c, the aggregation on an additional monomer is thermodynamically unfavorable while its degradation into smaller pieces is more likely due the random fluctuations of the environment. When the size is larger than n_c, the polymer is more stable even if it can still be broken at a relatively smaller rate. This mechanism with a nucleus assumption is also called nucleation process in the literature in physics, see Kashchiev [89]. This assumption will be translated in a mathematical model as follows,

$$\begin{cases} \kappa^k_{\text{off}} \gg \kappa^k_{\text{on}} & \text{for k} < \text{n}_\text{c}, \\ \kappa^k_{\text{off}} \sim \kappa^k_{\text{on}} & \text{for k} \geq \text{n}_\text{c}. \end{cases} \tag{NU}$$

Assuming that there are initially only monomers, the main quantity of interest is the duration of time after which a significant amount of stable polymers are generated in this system. This quantity is referred to as the *lag time* of the nucleation. Experiments show that there is a sharp phase transition, the polymerized mass starts and stays at 0 until it reaches quickly its final value after the lag time. Moreover, the measured lag time exhibits a significant variability in experiments. Our stochastic analysis gives a qualitative and quantitative explanation of these experimental observations.

We represent the evolution of this system by an infinite dimensional Markov jump process $(U^N(t)) = (U^N_k(t), k \in \mathbb{N}^+)$ where $U^N_k(t)$ denotes the number of polymers of size k at time t and N is the number of monomers at time 0, i.e., $U^N(0) = (N, 0, \dots)$. The transition matrix $q^N(\cdot,\cdot)$ is given by

$$\begin{cases} q^N(u, u - e_1 - e_k + e_{k+1}) & = \kappa^k_{\text{on}} u_1 u_k / N, \\ q^N(u, u - e_k + \sum_p e_{a_p}) & = \kappa^{k,a}_{\text{off}} u_k, \end{cases}$$

where $(e_k, k\in\mathbb{N}^+)$ is the standard orthonormal basis of $\mathbb{N}^{\mathbb{N}^+}$. The lag time can be expressed as, for any $\delta \in (0,1)$,

$$L^N_\delta = \inf\{t \geq 0 : \sum_{k\geq n_c} U^N_k(t) \geq \delta N\}.$$

We denote $\lambda_k = \kappa^k_{\text{on}}$ and the assumption on the nucleation is translated in the scaling representation

$$\kappa^k_{\text{off}} = \begin{cases} N\mu_k & \text{if } k<n_c, \\ \mu_k & \text{if } k\geq n_c. \end{cases}$$

Our main result is the following convergence in distribution

$$\lim_{N\to+\infty} \frac{L^N_\delta}{N^{n_c-3}} = E_\rho,$$

where E_ρ is an exponential random variable with parameter

$$\rho = \lambda_1 \prod_{k=2}^{n_c-1} \frac{\lambda_k}{\mu_k}.$$

In particular, because of the exponential distribution, the standard deviation of the lag time is of the same order as its expected value. Note that the limit does not depend on δ which shows the quick transition for the polymerized mass. To the best of our knowledge this is the first rigorous result of such a phenomenon. The proof relies on coupling techniques, careful stochastic calculus with Poisson processes, and the use of several results on branching processes.

1.5.2 Fluctuations in the Becker-Döring model

In Chapter 6, we prove a functional central limit theorem for the Becker-Döring model. As we state above, the Becker-Döring equation (BD) is the thermodynamic limit of the Markov process with transition matrix Q^N_{BD} when the total mass tends to infinity. We are interested in the second order approximation of this limit, i.e., the asymptotic behavior of the $\mathbb{R}^N$-valued process

$$W^N(t) = \left(\frac{1}{\sqrt{N}} \left(X^N(t) - Nc(t) \right) \right).$$

We prove that, under appropriate conditions, this fluctuation process $(W^N(t))$ converges for the Skorohod topology to a infinite dimensional process $W(t)$, which is the strong solution of the following SDE

$$\mathrm{d}W(t) = \tau\Big(\nabla s(c(t)) \cdot W(t) \Big)\, \mathrm{d}t + \tau\Big(\mathrm{Diag}\left(\sqrt{s(c(t))} \right) \cdot \mathrm{d}\beta(t) \Big), \qquad \text{(BD-fluc)}$$

where

1. s is a mapping from $\mathbb{R}_+^{\mathbb{N}^+} \to \mathbb{R}_+^{\mathbb{N}^+}$: for any $k \in \mathbb{N}^+$ and $c \in \mathbb{R}_+^{\mathbb{N}^+}$

$$s_{2k-1}(c) = a_k c_1 c_k \quad \text{and} \quad s_{2k}(c) = b_{k+1} c_{k+1};$$

2. and τ is a linear mapping from $\mathbb{R}_+^{\mathbb{N}^+} \to \mathbb{R}_+^{\mathbb{N}^+}$: for any $z \in \mathbb{R}_+^{\mathbb{N}^+}$ and $k \geq 2$,

$$\begin{cases} \tau_1(z) = -\sum_{i\geq 1}(1 + \mathbb{I}_{\{i=1\}}) z_{2i-1} + \sum_{i\geq 2}(1 + \mathbb{I}_{\{i=2\}}) z_{2i-2}, \\ \tau_k(z) = z_{2k-3} - z_{2k-2} - z_{2k-1} + z_{2k}. \end{cases}$$

3. $\nabla s(c)$ is the Jacobian matrix

$$\nabla s(c) = \left(\frac{\partial s_i}{\partial c_j}(c) \right), c \in \mathbb{R}_+^{\mathbb{N}^+}$$

4. $\beta(t) = (\beta_k(t), k \in \mathbb{N}^+)$ is a sequence of independent standard Brownian motions in $\mathbb{R}$;

5. $\mathrm{Diag}(v)$ is a diagonal matrix whose diagonal entries are the coordinates of the vector $v \in \mathbb{R}^{\mathbb{N}^+}$.

With these notations, the (BD) equations can be expressed in a more compact form as

$$\frac{\mathrm{d}c}{\mathrm{d}t}(t) = \tau\Big(s\big(c(t)\big)\Big).$$

It should be noted that the drift part of the SDE (BD-fluc) is actually the gradient of the Becker-Döring equation. At the thermal equilibrium, it can be seen as the fluctuation-dissipation relation in physics.

We first investigate the existence and uniqueness of the corresponding infinite dimensional stochastic differential equation (BD-fluc). For this a weighted L^2 space $L_2(w)$, see Section 6.2), has to be introduced. We also prove that, at equilibrium point of the first order equations (BD), the solution of this SDE is a Gaussian process. The proofs consists in proving Lipschitz conditions for several functionals and that the martingale part of the evolution equation is a well defined stochastic integral in the space $L_2(w)$.

To prove the tightness of the sequence of process $(W^N(t))$ one of the main difficulties is the fact that the first coordinate $(W_1^N(t))$ is interacting with all the other coordinates (because monomers are required to increase the size of polymers). For this reason, there are several infinite series of stochastic fluctuations in the evolution equation of $(W^N(t))$ which have to be controlled. This aspect had implications in the choice of the Hilbert space $L_2(w)$ to formulate the SDE (BD-fluc). Another difficulty is that the fluctuations of $(W^N(t))$ are driven by a set of independent Poisson processes whose intensities are state dependent. For this reason, the tightness argument highly relies on the boundedness properties of both Becker-Döring equation and processes. With the help of quite technical estimations of scaled processes and martingales in the Hilbert space $L_2(w)$, we are able to obtain the desired tightness and convergence results.

1.6 Notations

Here we introduce some notations, used repeatedly in the rest of the thesis.

$\mathbb{R}$ (*resp.* $\mathbb{R}_+$)	the set of (*resp.* non-negative) real numbers,
$\mathbb{Z}$, $\mathbb{N}$, $\mathbb{N}_+$	the set of integers, non-negative integers, positive integers,
i.i.d.	independent and identically distributed,
a.s.	almost surely,
$M_1(\cdot)$	the space of the probability on a space,
$\|\cdot\|_{\mathrm{TV}}$	total variation norm,
$a \wedge b$ (*resp.* $a \vee b$)	$\min(a,b)$ (*resp.* $\max(a,b)$),
$\mathrm{Card}(\cdot)$	the cardinality of a set,
$\mathbb{I}_{\{\cdot\}}$	Indicator function.

With a slight abuse of notation, to make the document more readable, when a subsequence of some sequence (X^N) is converging, we do not use an index (X^{N_p}) to indicate the converging subsequence but keep the notation (X^N). We use the

term "convergence in distribution" even in the cases where the convergence proved is stronger, when the convergence in probability holds for example.

Part I

Large Stochastic Networks with failures

Chapter 2

Large Unreliable Stochastic Network: Global Duplication

Contents

2.1 Introduction **32**
2.1.1 Large Distributed Systems 32
2.1.2 Mathematical Models 33
2.1.3 Related Mathematical Models 35
2.1.4 Presentation of the Results 36
2.2 The Stochastic Model **38**
2.3 First Order Asymptotic Behavior **41**
2.4 Evolution of Stable Network **46**
2.4.1 Some Technical Results 47
2.4.2 Convergence of Occupation Measures 52
2.4.3 The Decay of the Network occurs on the Time Scale $t \mapsto N^{d-1}t$. 54
2.5 Second Order Asymptotics in the Stable Case **56**
2.6 A Local Equilibrium in the Overloaded Case **61**
2.A Appendix: Generalized Skorohod Problems **66**

Abstract In this chapter a stochastic model of a large distributed system where users' files are duplicated on unreliable data servers is investigated. Due to a server breakdown, a copy of a file can be lost, it can be retrieved if another copy of the same file is stored on other servers. In the case where no other copy of a given file is present in the network, it is definitively lost. In order to have multiple copies of a given file, it is assumed that each server can devote a fraction of its processing capacity to duplicate files on other servers to enhance the durability of the system.

A simplified stochastic model of this network is analyzed. It is assumed that a copy of a given file is lost at some fixed rate and that the initial state is optimal: each file has the maximum number d of copies located on the servers of the network. The capacity of duplication policy is used by the files with the lowest number of copies. Due to random losses, the state of the network is transient and all files will be eventually lost. As a consequence, a transient d-dimensional Markov process $(X(t))$ with a unique absorbing

state describes the evolution this network. By taking a scaling parameter N related to the number of nodes of the network, a scaling analysis of this process is developed. The asymptotic behavior of $(X(t))$ is analyzed on time scales of the type $t \mapsto N^p t$ for $0 \leq p \leq d-1$. The chapter derives asymptotic results on the decay of the network: Under a stability assumption, the main results state that the critical time scale for the decay of the system is given by $t \mapsto N^{d-1}t$. In particular the duration of time after which a fixed fraction of files are lost is of the order of N^{d-1}. When the stability condition is not satisfied, i.e. when it is initially overloaded, it is shown that the state of the network converges to an interesting local equilibrium which is investigated. As a consequence it sheds some light on the role of the key parameters λ, the duplication rate and d, the maximal number of copies, in the design of these systems. The techniques used involve careful stochastic calculus for Poisson processes, technical estimates and the proof of a stochastic averaging principle.

2.1 Introduction

2.1.1 Large Distributed Systems

In this chapter the problem of reliability of large distributed system is analyzed via mathematical models. A typical framework is a cloud computing environment where users' files are duplicated on several data servers. When a server breaks down, all copies of files stored on this server are lost but they can be retrieved if copies of the same files are stored on other servers. In the case where no other copy of a given file is present in the network, it is definitively lost. Failures of disks occur naturally in this context, these events are quite rare but, given the large number of nodes of these large systems, this is not a negligible phenomenon at all at network scale. For example, in a data center with 200 000 servers, in average five disks fail every day. See the extensive study Pinheiro et al. [130] in this domain at Google. A natural consequence of these failures is the potential loss of some files if several servers holding copies of these files fail during a small time interval. For this reason this is a critical issue for companies deploying these large data centers.

Duplication Policies. In order to maintain copies on distant servers, a fraction λ of the bandwidth of each server has to be devoted to the duplication mechanism of its files to other servers. If, for a short period of time, several of the servers break down, it may happen that files will be lost for good just because all the available copies were on these servers and because a recovery procedure was not completed before the last copy disappeared. A second parameter of importance is d the maximal number of copies of a given file in different servers. The general problem can then be presented as follows: On the one hand, d should be sufficiently large, so that any file has a copy available on at least one server at any time. On the other hand, the maximum number of copies for a given should not be too large, otherwise the necessary fraction of the server capacity for maintaining the number of copies would be very large and could impact other functions of the server.

2.1.2 Mathematical Models

The natural critical parameters of such a distributed system with N servers are the failure rate μ of servers, the bandwidth λ allocated to duplication and the total number of files F_N. To design such a system, it is therefore desirable to have a duplication policy which maximizes the average number of files $\beta = F_N/N$ per server and the first instant $T_N(\delta)$ when a fraction $\delta \in (0,1)$ of files is lost. The main goal of this chapter is to give some insight on the role of these parameters through a simplified stochastic model.

A lot of work has been done in computer science concerning the implementation of duplication algorithms. These systems are known as distributed hash tables (DHT). They play an important role in the development of some large scale distributed systems, see Rhea et al. [137] and Rowstron and Druschel [144] for example.

Curiously, except extensive simulations, little has been done to evaluate the performances of these algorithms. Simplified models using birth and death processes have been used. See Chun et al. [36], Picconi et al. [128] and Ramabhadran and Pasquale [134]. In Feuillet and Robert [61], a mathematical model of the case of d=2 copies has been investigated. In [61], the main stochastic process of interest lives in dimension 1 which simplifies somewhat the analysis. As it will be seen, in our case, one has to investigate the more challenging problem of estimating some transient characteristics of a $d-1$-dimensional Markov process.

To the best of our knowledge, there has not been any mathematical study investigating the dependence of the decay of the network, represented by the variable $T_N(\delta)$, with respect to the maximal number of copies d and β the average number of files per server. As it will be seen, even with a simplified model of this chapter, the problem is already quite challenging. One has to derive estimates of transient characteristics of a transient d-dimensional Markov process on $\mathbb{N}^d$ with a reflection mechanism on the boundary of the state space.

A Possible Mathematical Model. Without simplifying assumptions, a mathematical model could use a state descriptor $(Y_j(t), 1 \leq j \leq F_N)$, where $Y_j(t)$ is the subset of $\{1, \ldots, N\}$ of servers having a copy of file j at time t. Note that the cardinality of $Y_i(t)$ is at most d and that file i is lost if $Y_i(t) = \emptyset$. The transitions can be described as follows.

1. Loss: If, for $1 \leq i \leq N$, node i breaks down in state (Y_j) then the value of Y_j does not change if $i \notin Y_j$ and, otherwise, $Y_j \mapsto Y_j \backslash \{i\}$.
2. Duplication: if $1 \leq i_1 \neq i_2 \leq N$ and $1 \leq j \leq F_N$ are such that $|Y_j| < d$, $i_1 \in Y_j$ and $i_2 \notin Y_j$, if the duplication policy at node i_1 does a copy of j at i_2, then $Y_j \mapsto Y_j \cup \{i_2\}$ and the other coordinates are not affected by this change. Depending on the duplication policy at node i_1, the choice of the node i_2 and of file j to copy may depend in a complicated way of the current state (Y_j).

As it can be seen the state space is quite complicated and, moreover, its dimension is growing with N which is a difficulty to investigate the asymptotics for N large. It does not seem to lead to a tractable mathematical model to study for example the first instant when a fraction $\delta \in (0,1)$ of files are lost,

$$\inf \left\{ t \geq 0 : \sum_{1}^{F_N} \mathbb{I}_{\{Y_j(t)=\emptyset\}} \geq \delta F_N \right\}.$$

Simplifying Assumptions. We present the mathematical model to be studied. The model has been chosen so that the role of the parameter d on the decay of the network can be investigated. To keep mathematics tractable, simplifications for some of the other aspects of these systems have been done. We review the main features of our model and the assumptions we have done.

1. *Capacity for duplication.*
   If there are N servers and each of them has an available bandwidth λ to duplicate files, then the maximal capacity for duplication is λN. One will assume that the duplication capacity can be used globally, i.e. the rate at which copies are created is λN.
2. *Duplication Policy.*
   Moreover, the duplication capacity is used on the files with the *lowest number* of copies. The duplication capacity is in fact used at best, on the files that, potentially, are the most likely to be lost.
3. *Failures.*
   Any copy of a given file is lost at rate μ. With this assumption, failures are more frequent but only a copy is lost at each event. In a more realistic setting, when a server breaks down, copies of several different files are lost at the same time.
4. *Topological Aspects.*
   In practice, in DHT, servers are located on a logical ring and, in order to limit the communication overhead, the location of copies of a file owned by a given server i are done at random on a fixed subset A_i of nodes, the leaf set of i. In our model, we assume that A_i is the whole set of servers.
5. *Statistical Assumptions.*
   For mathematical convenience, the random variables used for the duration between two breakdowns of a server or of a duplication of a file are assumed to be exponentially distributed.

For this simplified model, the use of the total capacity of duplication is optimal, see items (1) and (2) below. Our results gives therefore an upper bound on the efficiency of duplication mechanisms in a general context.

The Corresponding Markovian Model. With our assumptions, the state space can be embedded in a fixed state space of dimension $d+1$. If, for $0 \leq i \leq d$ and $t \geq 0$, $X_i^N(t)$ is the number of files with i copies, then the vector $X^N(t) = (X_0^N(t), X_1^N(t), \ldots, X_d^N(t))$ is a Markov process on $\mathbb{N}^{d+1}$.

Transitions. The model starts initially with F_N files, each of them having a maximal number of copies d, i.e. $X^N(0) = (0, 0, \ldots, 0, F_N)$. If $X^N(t)$ is in state $x=(x_i) \in \mathbb{N}^{d+1}$ and, for $0 \leq i \leq d$, e_i is the ith unit vector, there are two types of transitions for the Markov process. See Figure 2.1.

1. Loss: for $1 \leq i \leq d$, $x \to x + e_{i-1} - e_i$.
   A copy of a file with i copies is lost at rate $ix_i\mu$.
2. Duplication: for $1 \leq i < d$, $x \to x - e_i + e_{i+1}$, $1 \leq i < d$.
   It occurs at rate λN under the condition $x_1 = x_2 = \cdots = x_{i-1} = 0$, which means that there are no files with between 1 and i copies.

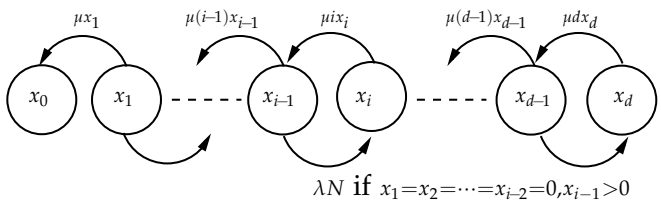


Figure 2.1 – Jump Rates for transfers of one unit between the coordinates of the Markov Process $(X^N(t))$ in state $(x_0, x_1, \ldots, x_d)$

Clearly enough, this system is transient, due to the random losses, all files are eventually lost, the state $\emptyset = (F_N, 0, \ldots, 0)$ is an absorbing state. The aim of this chapter is to describe the decay of the network, i.e. how the number $X_0^N(t)$ of lost files is increasing with respect to time.

For fixed F_N and N, this problem is related to the analysis of the transient behavior of a multi-dimensional Markov process. In our case, because of reflection on boundaries of $\mathbb{N}^{d+1}$ due to the duplication mechanism, the distribution of the evolution of the Markov process $(X_k(t))$ is not easy to study. For this reason, a scaling approach is used, with N converging to infinity and F_N being kept proportional to N.

It will be assumed that the average number of files per server F_N/N converges to some $\beta > 0$. For $\delta > 0$, the decay of the system can be represented by the random variable

$$T_N(\delta) = \inf\left\{t \geq 0 : \frac{X_0^N(t)}{N} \geq \delta\beta\right\}$$

the time it take to have a fraction δ of the files lost.

2.1.3 Related Mathematical Models

Ehrenfest Urn Models. The Markov process $(X(t))$ can be seen as a particle system with $d+1$ boxes and any particle in box $1 \leq i \leq d$ moves to box $i-1$ at rate μ. Box with index 0 is a cemetery for particles. A "pushing" process moves the particle the further on the left (box 0 excluded) to the next box on its right at a high rate λN. The model can be seen as a variation of the classical Ehrenfest urn model, see Karlin and McGregor [86] and Diaconis et al. [48] for example.

Polymerization Processes in Biology. It turns out that this model has some similarities with stochastic processes representing polymerization processes of some biological models. The simplest model starts with a set of monomers (some proteins) that can aggregate to form polymers. Due to random fluctuations within the cell, a polymer of size i and a monomer can produce a polymer of size $i+1$ at some fixed rate. In this context, as long as the size i of the polymer is below some constant i_0, the polymer is not stable, it will lose monomers very quickly, at a high rate, it breaks into a polymer of size $i-1$ and a monomer. When the size is greater or equal to i_0 (nucleation phase) the polymer is much more stable, it is assumed that it remains in this state. Again due

to the random fluctuations, all particles will end up in polymers of sizer greater than i_0. These "large" polymers correspond to our box 0 for the duplication process and the monomers are the equivalent of files with d copies. The *lag time* is the first instant when a positive fraction (half say) of the monomers have been consumed into stable polymers. Note that it is analogous to our $T_N(1/2)$. In this framework, the fluctuations of the lag time have important consequences on biological processes. See Prigent et al. [132], Xue et al. [168] and Szavits-Nossan et al. [157] for example.

2.1.4 Presentation of the Results

The model starts with $F_N \sim \beta N$ files, all of them with the maximum number of copies d. The loss rate of a copy is μ and the duplication rate is λN but only for the files with the minimum number of copies. For $0 \leq i \leq d$, $X_i^N(t)$ denotes the number of files with i copies.

It is first shown in Theorem 2.1 that, as N gets large, for the convergence of stochastic process

$$\lim_{N\to+\infty}\left(\frac{X_k^N(t)}{N}, 0 \leq k \leq d\right) = (x_k(t), 0 \leq k \leq d). \tag{2.1}$$

The limit $(x_k(t), 0 \leq k \leq d)$ can be expressed as the solution of a deterministic generalized Skorohod problem. See Section 2.3 for the definition.

Stable Case: $\lambda > d\mu\beta$.
With the resolution of the generalized Skorohod problem, Proposition 2.1 of Section 2.3, one proves that if $\lambda > d\mu\beta$ then the network is stable in the sense that the limiting process $(x_k(t), 0 \leq k \leq d)$ is constant and equal to $(0, \ldots, 0, \beta)$. In other words, on the normal time scale, the fraction of lost files is zero and, at the fluid level, all files have the maximal number of copies d.

The key results of the stable case are Theorem 2.2 of Section 2.4 and Theorem 2.3 of Section 2.5 . These quite technical and delicate results rely on stochastic calculus and various technical estimates related to the flows between coordinates of the process $(X_k^N(t))$. They are proved in several propositions of Section 2.4, the important Proposition 2.4 in particular. A stochastic averaging result completes the proof of these convergence results.

Theorem 2.2 shows that the network is in fact beginning to lose files only on the time scale $t \mapsto N^{d-1}t$, i.e. that the convergence in distribution

$$\lim_{N\to+\infty}\left(\frac{X_0^N(N^{d-1}t)}{N}\right) = (\Phi(t)) \tag{2.2}$$

where $\Phi(t)$ is the unique solution $y \in [0, \beta]$ of the equation

$$\left(1 - \frac{y}{\beta}\right)^{\rho/d} e^y = \exp\left(-\lambda \frac{(d-1)!}{\rho^{d-1}} t\right),$$

where $\rho = \lambda/\mu$. On this time scale, the fluid state of the network evolves from $(0, \ldots, 0, \beta)$ to the absorbing state: $\Phi(0) = 0$ and $\Phi(t)$ converges to β as t goes to infinity.

The second order fluctuations are described by the convergence in distribution

$$\lim_{N\to+\infty}\left(\frac{X_0^N(N^{d-1}t)-N\Phi(t)}{\sqrt{N}}\right)=(W(t)),$$

where $\Phi(t)$ is the solution of Equation (2.31) and the process $(W(t))$ is the solution of a stochastic differential equation, Relation (2.34).

Overloaded Case: $p\mu\beta<\lambda<(p+1)\mu\beta$ for some $2\le p\le d-1$.
In this case, the limiting process $(x_k(t), 0\le k\le d)$ of Relation (2.1) is not trivial, i.e. different from its initial state $(0,0,\dots,0,\beta)$. Its explicit expression is given in Proposition 2.1. Moreover, it is shown that

$$\lim_{t\to+\infty}(x_p(t),x_{p+1}(t))=\left((p+1)\beta-\frac{\lambda}{\mu},\frac{\lambda}{\mu}-p\beta\right).$$

This can be interpreted as follows: In the limit, on the normal time scale, at the fluid level all files have either p or $p+1$ copies. The network exhibits therefore an interesting property of local equilibrium.

If one starts from this local equilibrium, it is shown that the system begins to lose files only the time scale $t\mapsto N^{p-1}t$. A result analogous to Relation (2.2) is proved by Theorem 2.4, for the convergence in distribution,

$$\lim_{N\to+\infty}\left(\frac{X_0^N(N^{p-1}t)}{N},\frac{X_p^N(N^{p-1}t)}{N},\frac{X_{p+1}^N(N^{p-1}t)}{N}\right)=\big(\Phi_0(t),\Phi_p(t),\Phi_{p+1}(t)\big)$$

holds, where $(\Phi_0(t),\Phi_p(t),\Phi_{p+1}(t))$ is deterministic, with the property that

$$\lim_{t\to+\infty}\big(\Phi_0(t),\Phi_p(t),\Phi_{p+1}(t)\big)=\left(\beta-\frac{\rho}{p+1},0,\frac{\rho}{p+1}\right),$$

i.e. asymptotically all files are either lost or have $p+1$ copies.

These results give the main phenomena concerning the evolution of a stable network towards the absorbing state. It should be noted that we do not consider the special cases when the parameters satisfy the relation $\lambda{=}d\mu\beta$ for the following reason. When $\lambda{<}d\mu\beta$, the analysis involves a stochastic averaging principle with an underlying ergodic Markov process. See Section 2.4.2 below. With equality $\lambda{=}d\mu\beta$, the corresponding Markov process is in fact null recurrent and proving a stochastic averaging principle in this context turns out to be more delicate. There are few examples in this domain to the best of our knowledge. See Khasminskii and Krylov [95] in the case of diffusions. The same remark applies to similar identities, like $\lambda=p\mu\beta$ for $1\le p\le d$.

Choice of Parameters. As a consequence, the parameters β and d should be chosen so that $\lambda/(\beta\mu)>2$ and $d=\lfloor\lambda/(\beta\mu)\rfloor$ to maximize the time of decay of the network and at the same time to preserve the stability of the network. For $\delta\in(0,1)$ the variable $T_N(\delta)$, the first instant when a fraction δ of files is lost, is then of the order of N^{d-1}.

Outline of the Chapter. Section 2.2 introduces the main notations and the stochastic evolution equations of the network. Section 2.3 shows that the Markov process can be expressed as the solution of a generalized Skorohod problem, presented in

Appendix 2.A. A convergence result on the evolution of the network on the normal time scale is established and an explicit expression for the limiting process is provided. Section 2.4 investigates the decay of the network on the time scale $t \mapsto N^{d-1}t$ in the stable case. A central limit theorem on this time scale is established in Section 2.5. The overloaded case is analyzed in Section 2.6, the asymptotic evolution of the local equilibrium is studied on several time scales.

2.2 The Stochastic Model

In this section we introduce the notations used throughout this chapter as well as the statistical assumptions. The stochastic differential equations describing the evolution of the network are introduced. It is shown that, via a change of coordinates, the state descriptor of the process can be expressed as the solution of a generalized Skorohod problem. See Section 2.A. The convergence results at the normal time scale $t \mapsto t$ proved in the next section use this key property.

A given file has a maximum of d copies and each of them vanishes after an independent exponential time with rate rate μ. A file with 0 copy is lost for good. The recovery policy works as follows. The total capacity λN of the network is allocated to the files with the minimum number of copies. Consequently, if at a given time all non-lost files present have at least $k \geq 1$ copies and there are x_k files with k copies, then each of these is duplicated after an independent exponential time with rate $\lambda N/x_k$. Initially it is assumed that there are F_N files and that the network starts from the optimal state where each file has d copies.

For $0 \leq k \leq d$, $X_k^N(t)$ denotes the number of files with k copies at time t. The quantity $X_0^N(t)$ is the number of lost files at time t, the function $t \mapsto X_0^N(t)$ is in particular non-decreasing.

The conservation relation $X_0^N(t)+X_1^N(t)+\cdots+X_d^N(t)=F_N$ gives that the stochastic process $(X_0^N(t), X_1^N(t), \ldots, X_{d-1}^N(t))$ on $\mathbb{N}^d$ has the Markov property. Its Q-matrix $Q^N = (q^N(\cdot,\cdot))$ is given by

$$\begin{cases} q^N(x, x - e_k + e_{k-1}) &= \mu k x_k, \qquad 1 \leq k \leq d-1, \\ q^N(x, x + e_{d-1}) &= \mu d \left(F_N - x_0 - x_1 - \cdots - x_{d-1}\right), \\ q^N(x, x + e_k - e_{k-1}) &= \lambda N \mathbb{I}_{\{x_{k-1}>0, x_i=0, 1\leq i<k-1\}}, \qquad 2 \leq k \leq d-1, \\ q^N(x, x - e_{d-1}) &= \lambda N \mathbb{I}_{\{x_{d-1}>0, x_i=0, 1\leq i<d-1\}}, \end{cases} \tag{2.3}$$

where e_k is the kth unit vector of $\mathbb{N}^d$. The first two relations come from the independence of losses of various copies of files, note that $F_N - x_0 - x_1 - \cdots - x_{d-1}$ is the number of files with d copies. The last two equations are a consequence of the fact that the capacity is devoted to the smallest index $k \geq 1$ such that $x_k \neq 0$. The coordinate $X_d^N(t)$ is of course given by

$$X_d^N(t) = F_N - X_0^N(t) - X_1^N(t) - \cdots - X_{d-1}^N(t),$$

$X_d^N(t)$ is the number of files with the maximal number d of copies at time t. The initial condition is such that $X_k^N(0) = 0$ for $0 \leq k \leq d-1$ and $X_d^N(0) = F_N \in \mathbb{N}$.

Scaling Condition

It is assumed that there exist some $\beta > 0$ and $\gamma \geq 0$ such that

$$\lim_{N\to\infty} \frac{F_N - N\beta}{\sqrt{N}} = \gamma. \tag{2.4}$$

Equations of Evolution

To analyze the asymptotic behavior of the process $(X^N(t))$, it is convenient to introduce the processes $(S^N(t))=((S_k^N(t), 1\leq k\leq d-1))$ and $(R^N(t))=((R_k^N(t), 1\leq k\leq d-1))$. For $1 \leq k \leq d-1$ and $t \geq 0$, $S_k^N(t)$ is the number of files with no more than k copies at time t and $R_k^N(t)$ is the local time at 0 of the process $(S_k^N(t))$,

$$S_k^N(t) = \sum_{i=1}^{k} X_i^N(t) \text{ and } R_k^N(t) = \int_0^t \mathbb{I}_{\{S_k^N(u)=0\}}\, \mathrm{d}u.$$

For any function $h \in \mathcal{D}(\mathbb{R}_+, \mathbb{R}_+)$, i.e. h is continuous on the right and has left limits on $\mathbb{R}_+$, one denotes by $\mathcal{N}_h$ denotes a point process on $\mathbb{R}_+$ defined as follows

$$\mathcal{N}_h([0,t]) = \int_0^t \mathcal{P}([0, h(u-)] \times \mathrm{d}u) \tag{2.5}$$

where $h(u-)$ is the left limit of h at u and $\mathcal{P}$ is a Poisson process in $\mathbb{R}_+^2$ whose intensity is the Lebesgue measure on $\mathbb{R}_+^2$. In particular if h is deterministic, then $\mathcal{N}_h$ is a Poisson process with intensity $(h(t-))$. When several such processes $\mathcal{N}_h$ are used as below in the evolution equations, then the corresponding Poisson processes $\mathcal{P}$ used are assumed to be independent. The equations of evolution can then be written as,

$$\begin{cases} \mathrm{d}S_{d-1}^N(t) = \mathcal{N}_{d\mu(F_N - S_{d-1}^N - X_0^N)}(\mathrm{d}t) - \mathcal{N}_{\mu S_1^N}(\mathrm{d}t) \\ \qquad\qquad\qquad\qquad\qquad\qquad -\mathbb{I}_{\{S_{d-2}^N(t-)=0, S_{d-1}^N(t-)>0\}}\mathcal{N}_{\lambda N}(\mathrm{d}t), \\ \\ \mathrm{d}S_k^N(t) = \mathcal{N}_{(k+1)\mu X_{k+1}^N}(\mathrm{d}t) - \mathcal{N}_{\mu S_1^N}(\mathrm{d}t) - \mathbb{I}_{\{S_{k-1}^N(t-)=0, S_k^N(t-)>0\}}\mathcal{N}_{\lambda N}(\mathrm{d}t), \end{cases}$$

for $1 \leq k \leq d-2$, with the convention that $(S_0^N(t)) \equiv 0$ is the null process and also that $(R_0(t)) = (t)$. By integrating and compensating these equations, one gets that

$$S_k^N(t) = Z_k^N(t) - \lambda N(R_{k-1}^N(t) - R_k^N(t)), \quad 1 \leq k \leq d-1, \tag{2.6}$$

and the first coordinates $(X_0^N(t))$ satisfies the relation

$$X_0^N(t) = \mu \int_0^t S_1^N(u)\, \mathrm{d}u + U_0^N(t), \tag{2.7}$$

with

$$Z_{d-1}^N(t) = d\mu \int_0^t \left[F_N - S_{d-1}^N(u) - X_0^N(u)\right] \mathrm{d}u - \mu \int_0^t S_1^N(u)\, \mathrm{d}u + U_{d-1}^N(t)$$
$$Z_k^N(t) = (k+1)\mu \int_0^t (S_{k+1}^N(u) - S_k^N(u))\, \mathrm{d}u - \mu \int_0^t S_1^N(u)\, \mathrm{d}u + U_k^N(t),$$

for $1 \leq k \leq d-2$, where the $(U^N(t)) = (U_k^N(t), 1 \leq k \leq d-1)$ are the martingales associated to the jumps of these processes, for $1 \leq k \leq d-2$,

$$U_k^N(t) = \int_0^t \left[\mathcal{N}_{\mu(k+1)X_{k+1}^N}(\mathrm{d}u) - \mu(k+1)X_{k+1}^N(u)\ \mathrm{d}u\right]$$
$$-\int_0^t \left[\mathcal{N}_{\mu X_1^N}(\mathrm{d}u) - \mu X_1^N(u)\,\mathrm{d}u\right] - \int_0^t \mathbb{I}_{\left\{S_{k-1}^N(u-)=0, S_k^N(u-)>0\right\}}\left[\mathcal{N}_{\lambda N}(\mathrm{d}u) - \lambda N\,\mathrm{d}u\right],$$

and its increasing process is given by

$$\left\langle U_k^N \right\rangle(t) = \mu(k+1)\int_0^t X_{k+1}^N(u)\ \mathrm{d}u + \mu\int_0^t X_1^N(u)\ \mathrm{d}u$$
$$+\lambda N \int_0^t \mathbb{I}_{\left\{S_{k-1}^N(u)=0, S_k^N(u)>0\right\}}\ \mathrm{d}u. \quad (2.8)$$

The martingales $(U_0^N(t))$ and $(U_{d-1}^N(t))$ have similar expressions,

$$U_{d-1}^N(t) = \int_0^t \left[\mathcal{N}_{d\mu(F_N - S_{d-1}^N - X_0^N)}(\mathrm{d}u) - d\mu(F_N - S_{d-1}^N(u) - X_0^N(u))\,\mathrm{d}u\right]$$
$$-\int_0^t \left[\mathcal{N}_{\mu X_1^N}(\mathrm{d}u) - \mu X_1^N(u)\,\mathrm{d}u\right] - \int_0^t \mathbb{I}_{\left\{S_{d-2}^N(u-)=0, S_{d-1}^N(u-)>0\right\}}\left[\mathcal{N}_{\lambda N}(\mathrm{d}u) - \lambda N\,\mathrm{d}u\right],$$

with

$$\left\langle U_{d-1}^N \right\rangle(t) = d\mu\int_0^t \left(F_N - S_{d-1}^N(u) - X_0^N(u)\right)\mathrm{d}u$$
$$+\int_0^t \mu X_1^N(u)\,\mathrm{d}u + \lambda N\int_0^t \mathbb{I}_{\left\{S_{d-2}^N(u-)=0, S_{d-1}^N(u-)>0\right\}}\,\mathrm{d}u,$$

and

$$U_0^N(t) = \int_0^t \left[\mathcal{N}_{\mu S_1^N}(\mathrm{d}u) - \mu S_1^N(u)\right]\ \mathrm{d}u, \text{ with } \left\langle U_0^N \right\rangle(t) = \mu\int_0^t S_1^N(u)\ \mathrm{d}u.$$

A Generalized Skorohod Problem Representation

For $h = (h_i)$ an element of $\mathcal{D}(\mathbb{R}_+, \mathbb{R}^{d-1})$, $\eta > 0$ and $F \in \mathbb{N}$, denote

$$\begin{cases} G_1(h,F,\eta)(t) = \mu\int_0^t (2h_2(u) - 3h_1(u))\ \mathrm{d}u - \eta t, \\ G_k(h,F,\eta)(t) = \mu\int_0^t ((k+1)h_{k+1} - (k+1)h_k(u) - h_1(u))\,\mathrm{d}u,\ 1<k<d-1, \\ G_{d-1}(h,F,\eta)(t) = d\mu\int_0^t \left[F - h_{d-1}(u) - \mu\int_0^u h_1(v)\,\mathrm{d}v\right]\ \mathrm{d}u - \mu\int_0^t h_1(u)\,\mathrm{d}u, \end{cases} \quad (2.9)$$

and $G(h,F,\eta) = (G_k(h,F,\eta), 1 \leq k \leq d-1)$. Equations (2.6) and (2.7) give the relations

$$S^N(t) = G\left(S^N, F_N, \lambda N\right)(t) + U^N(t)$$
$$-\,d\mu\int_0^t U_0^N(u)\ \mathrm{d}u \cdot e_{d-1} + \lambda N(I-P)R^N(t), \quad (2.10)$$

where P is the matrix $P = (P_{ij}, 1 \leq i, j \leq d-1)$ whose non zero coefficients are the $P_{i,i-1} = 1$ for $2 \leq i \leq d-1$.

In other words, for a fixed N, the couple $(S^N, \lambda N R^N)$ is the solution of the generalized Skorohod problem associated to the matrix P and the functional

$$\overline{G} : h \to G(h, F_N, \lambda N) + U^N - d\mu \int_0^{\cdot} U_0^N(u)\, \mathrm{d}u \cdot e_{d-1}. \tag{2.11}$$

See the appendix for the definition and a result of existence and uniqueness.

2.3 First Order Asymptotic Behavior

In this section, the asymptotic behavior of the sequence of processes $(X_k^N(t)/N)$ at the "normal" time scale is investigated. As a consequence, it is shown that if $\lambda > \beta d\mu$ then the network is stable at the fluid level, i.e. the fraction of lost files is 0 at any time. Otherwise, a positive fraction of files is lost, an explicit expression for this quantity is provided.

More precisely the convergence of the sequence of stochastic processes

$$\left(\frac{X_k^N(t)}{N}, 0 \leq k \leq d \right),$$

is investigated. One first shows that this sequence is tight and the limit is identified as the solution of a deterministic generalized Skorohod problem. An explicit computation of this limit concludes the section.

Tightness

Due to Assumption (2.4), there exists some constant C_0 such that the relation $F_N \leq C_0 N$ holds for all N. Since $0 \leq X_k^N(t) \leq F_N$ for any $0 \leq k \leq d-1$ and $t \geq 0$, Relation (2.8) gives the existence of a constant C_1 such that

$$\mathbb{E}\left(U_k^N(t)^2\right) = \mathbb{E}\left(\left\langle U_k^N \right\rangle(t)\right) \leq C_1 N t, \quad \forall 1 \leq k < d-1 \text{ and } t \geq 0. \tag{2.12}$$

with Doob's Inequality one gets that, for $1 \leq k \leq d-1$ and $\varepsilon > 0$,

$$\mathbb{P}\left(\sup_{0 \leq s \leq t} \frac{U_k^N(s)}{N} \geq \varepsilon \right) \leq \frac{1}{(\varepsilon N)^2} \mathbb{E}(U_k^N(t)^2) \leq \frac{C_1 t}{\varepsilon^2 N},$$

which shows that, for $0 \leq k \leq d-1$, the martingale $(U_k^N(t)/N)$ converges in distribution to 0 uniformly on compact sets.

For $T > 0$, $\delta > 0$ and for Z a function in the space $\mathcal{D}(\mathbb{R}_+, \mathbb{R})$ of càdlàg functions, i.e. right continuous functions on $\mathbb{R}_+$ with left limits at every point, define $w_Z(\delta)$ as the modulus of continuity of the process $(Z(t))$ on the interval $[0, T]$,

$$w_Z(\delta) = \sup_{0 \leq s \leq t \leq T, |t-s| \leq \delta} |Z(t) - Z(s)|. \tag{2.13}$$

By using again the relation $X_k^N(t) \leq C_0 N$ for all $N \in \mathbb{N}$, $1 \leq k \leq d-1$ and $t \geq 0$, the above equations and the convergence of the martingales to 0 give that, for any $\varepsilon > 0$ and $\eta > 0$, there exists $\delta > 0$ such that

$$\mathbb{P}(w_{S_k^N/N}(\delta) \geq \eta) \leq \varepsilon, \quad \mathbb{P}(w_{X_0^N/N}(\delta) \geq \eta) \leq \varepsilon, \quad \forall N \text{ and } 1 \leq k \leq d-1.$$

This implies that the sequence of stochastic processes

$$\left(\frac{X_0^N(t)}{N}, \frac{S^N(t)}{N}\right) = \left(\frac{X_0^N(t)}{N}, \frac{S_k^N(t)}{N}, 1 \leq k \leq d-1\right)$$

is tight and that any of its limiting points is almost surely a continuous processes. See Billingsley [19] for example.

Convergence

Let $(x_0(t), (s_k(t), 1 \leq k \leq d-1))$ denote a limiting point of the sequence

$$(X_0^{N_p}(t)/N_p, S^{N_p}(t)/N_p)$$

associated to some non-decreasing subsequence (N_p). By choosing an appropriate probability space, it can be assumed that the convergence holds almost surely. By Equation (2.7), one gets that

$$x_0(t) = \mu \int_0^t s_1(u)\, \mathrm{d}u,$$

From Definition (2.11) of the functional $\overline{G}$ and by convergence of the sequence of processes $(X_0^{N_p}(t)/N_p, S^{N_p}(t)/N_p)$ and of the martingale (U^{N_p}/N_p) to 0, one gets that the convergence

$$\lim_{p \to +\infty} \frac{1}{N_p} \overline{G}\left(S^{N_p}, F_{N_p}, \lambda N_p\right) = G\left(S, \beta, \lambda\right)$$

holds uniformly on compact sets, where G is defined by Relation (2.9). As it has been seen in the previous section, Equation (2.10), the couple $(S^N/N, R^N/N)$ is the solution of a classical Skorohod problem associated to the matrix P introduced in Equation (2.10) and the free process $(\overline{G}(S^{N_p}, F_{N_p}, \lambda N_p)/N_p$. By continuity of the solutions of a classical Skorohod problem, see Proposition 5.11 of Robert [138] for example, one concludes that $(S^N/N, R^N)$ converges to the solution (S, R) of the Skorohod problem associated to P and $h \mapsto G(h, \beta, \lambda)$. Hence (S, R) is the unique solution of the *generalized* Skorohod problem for the matrix P and the functional $h \to G(h, \beta, \lambda)$. The convergence of the sequence $(S^N/N, X_0^N/N)$ has been therefore established.

Theorem 2.1. *If $S(t) = (s_k(t), 1 \leq k \leq d-1)$ is the unique solution of the generalized Skorohod problem associated to the matrix $P = (\mathbb{I}_{\{(i,j)=(i,i-1)\}}, 1 \leq i, j \leq d-1)$ and the functional $h \mapsto G(h, \beta, \lambda)$ defined by Equation (2.9), then the sequence of processes*

$$\left(\frac{X_k^N(t)}{N}, 0 \leq k \leq d\right)$$

converges in distribution uniformly on compact sets to $(x_k(t))$ *defined by*

$$\begin{aligned} x_0(t) &= \mu \int_0^t s_1(u)\, \mathrm{d}u, \quad x_1(t) = s_1(t), \\ x_k(t) &= s_k(t) - s_{k-1}(t),\ 2 \le k \le d-1, \\ x_d(t) &= \beta - s_{d-1}(t) - x_0(t). \end{aligned}$$

If the limiting processes is uniquely determined as the solution of a Skorohod problem, it is not always easy to have an explicit representation of the solution of a Skorohod problem. The classical example of Jackson networks, see Chen and Mandelbaum [35], shows that this is not always easy to have an explicit expression for the solutions of these problems in dimension greater than 2. The linear topology of the network simplifies this question as the following proposition shows.

Proposition 2.1 (Characterization of fluid limits)**.**

1. $2\mu\beta < \lambda < d\mu\beta$.
   Let $p = \lfloor \rho/\beta \rfloor$ *with* $\rho = \lambda/\mu$. *the fluid limits* $(s(t)) = (s_1(t), \dots, s_d(t))$ *of Theorem 2.1 are defined as follows. There exist a sequence* (t_k),

$$0 = t_d < t_{d-1} < \cdots < t_{p+1} < t_p = \infty,$$

   such that, for all $l = d-1, \dots, p$ *and for* $t_{l+1} \le t \le t_l$,

$$\begin{cases} s_k(t) = 0, & 1 \le k \le l-1, \\ s_l(t) = (l+1)\beta - \rho + \xi_{l,1} e^{-\mu t} + \sum_{i=l+2}^{d} \xi_{l,i} e^{-\mu i t}, \\ s_k(t) = \beta \left(1 - \sum_{i=k+1}^{d} \alpha_{k,i} e^{-i\mu t}\right), & l+1 \le k \le d, \end{cases}$$

   where $\alpha_{d,d} = 1$ *and, for* $j > l+1$

$$\alpha_{l,j} = \frac{l+1}{l+1-j} \alpha_{l+1,j}, \quad \alpha_{l,l+1} = e^{(l+1)\mu t_l} \left(1 - \frac{\rho}{l\beta} - \sum_{k=l+2}^{d} \alpha_{l,k} e^{-k\mu t_l}\right),$$

$$\xi_{l,j} = \frac{\beta(l+1)}{j-1} \alpha_{l+1,j}, \quad \xi_{l,1} = -\left((l+1)\beta - \rho + \sum_{j=l+2}^{d} \xi_{l,j} e^{-\mu(j-1)t_{l+1}}\right),$$

   with $\alpha_{l,l} = 0$ *and* t_l *is the unique solution of* $s_l(t) = \lambda/(l\mu)$.

2. $\lambda > d\mu\beta$.
   For all $t \ge 0$, $(x_1(t), \dots, x_d(t)) = (0, \dots, 0, \beta)$.

Proof. The vector $(s_k(t))$ is solution of the following equation:

$$\begin{aligned} s_k(t) &= \mu(k+1) \int_0^t (s_{k+1}(u) - s_k(u))\, \mathrm{d}u - \mu \int_0^t s_1(u)\, \mathrm{d}u - \lambda(r_{k-1}(t) - r_k(t)), \\ s_d(t) &= \beta - \int_0^t \mu s_1(u)\, \mathrm{d}u, \end{aligned}$$

where the $(r_k(t))$ are the reflection processes such that

$$\int_0^t s_k(u)\,\mathrm{d}r_k(u) = 0.$$

By uniqueness of the solution of a generalized Skorohod problem given by Proposition 2.8 of the Appendix, it is enough to exhibit a solution to the above equations.

We assume the conditions of the case (1) of the proposition. We will prove in fact that there exists $t_d = 0 < t_{d-1} < t_{d-2} < \cdots < t_p < t_{p-1} = +\infty$ such that, for all $p \le l \le d-1$ and $t_{l+1} < t < t_l$, the s_k and the t_k have the following equations:

$$\begin{cases} s_k(t) = 0, & r_k(t) = t, \qquad 1 \le k \le l-2, \\ s_{l-1}(t) = 0, & \dot{r}_{l-1}(t) = 1 - l\mu/\lambda s_l(t), \\ \dot{s}_l(t) = \mu(l+1)s_{l+1}(t) - \mu s_l(t) - \lambda, & \dot{r}_l(t), = 0 \\ \dot{s}_k(t) = \mu(k+1)(s_{k+1}(t) - s_k(t)), & \dot{r}_k(t) = 0, \qquad l+1 \le k \le d-1, \\ s_d(t) = \beta, & r_d(t) = 0. \end{cases} \tag{2.14}$$

The t_k are defined such that $s_k(t_k) = \lambda/(\mu k)$.

We start with the case $d-1$. It is easy to check that $((s_k), (r_k))$ defined by the following equations is the solution of the generalized Skorohod problem,

$$\begin{cases} s_k(t) = 0, \quad r_k(t) = t, \qquad \text{for } 1 \le k \le d-3, \\ s_{d-2} = 0, \qquad r_{d-2}(t) = t - \dfrac{(d-1)\mu}{\lambda}\displaystyle\int_0^t s_{d-1}(u)\,\mathrm{d}u, \\ s_{d-1}(t) = (d\beta - \lambda/\mu)\,(1 - e^{-\mu t}), \quad r_{d-1}(t) = 0. \end{cases}$$

This is valid for all $0 \le t < t_{d-1}$ with

$$t_{d-1} = \frac{1}{\mu}\log\left(\frac{d\beta - \rho}{d\beta - d\rho/(d-1)}\right).$$

Now, we proceed by using a recursion. Assume that there exists $l > p$ such that the system of equation (2.14) is verified until t_l. Moreover, we assume that, for all $k \ge l$ and $t_k \le t \le t_{k-1}$,

$$s_k(t) = \beta\left(1 - \sum_{i=k+1}^{d} \alpha_{k,i} e^{-\mu i t}\right);$$

and

$$s_{k-1}(t) = k\beta - \rho + \xi_{k-1,1}e^{-\mu t} + \sum_{i=k+1}^{d} \xi_{k-1,i}e^{-\mu i t};$$

and t_{k-1} is the only solution of

$$s_{k-1}(t_{k-1}) = \frac{\lambda}{(k-1)\mu}.$$

We define

$$\alpha_{l,i} = \frac{l+1}{l+1-i}\alpha_{l+1,i}, \text{ for } d \ge i > l+1,$$

$$\alpha_{l,l+1} = e^{(l+1)\mu t_l}\left(1 - \frac{\rho}{l\beta} - \sum_{k=l+2}^{d} \alpha_{l,k} e^{-k\mu t_l}\right),$$

$$\xi_{l-1,i} = \frac{\beta l}{i-1}\alpha_{l,i}, \text{ for } d \geq i > l+1,$$

and

$$\xi_{l-1,1} = -\left(l\beta - \rho + \sum_{i=l+1}^{d} \xi_{l-1,i} e^{-\mu(i-1)t_l}\right).$$

It is easy to check that s_l is then solution of the equation

$$\dot{s}_l = \mu(l+1)(s_{l+1}(t) - s_l(t))$$

when $t \geq t_l$ and s_{l-1} is the solution of the equation

$$\dot{s}_{l-1}(t) = \mu l s_l(t) - \mu s_{l-1}(t) - \lambda,$$

when $t_{l-1} \geq t \geq t_l$, t_{l-1} is the solution of the equation

$$s_{l-1}(t_{l-1}) = \frac{\lambda}{(l-1)\mu}.$$

The recursion is proved and therefore the assertion of case (1) of the proposition.

Concerning the case (2), one has only to check that the couple

$$\begin{cases} (x_1(t), \ldots, x_d(t)) \stackrel{\text{def.}}{=} (0, \ldots, 0, \beta) \\ (r_1(t), \ldots, r_{d-2}(t), r_{d-1}(t), r_d(t)) \stackrel{\text{def.}}{=} (t, \ldots, t, (1 - d\mu\beta/\lambda)\, t, 0) \end{cases}$$

is indeed the solution of the generalized Skorohod problem. □

The following corollary shows that in the overloaded cases, asymptotically, there is an equilibrium where most of files will have either p or $p+1$ copies for some convenient p. This situation is investigated in Section 2.6.

Corollary 2.1 (Stable Fluid State of the Overloaded System)**.** *In the case (1) of Proposition 2.1, then*

$$\lim_{t \to +\infty} (x_p(t), x_{p+1}(t)) = ((p+1)\beta - \rho, \rho - p\beta),$$

and $x_k(t) \to 0$ as $t \to +\infty$ for all $1 \leq k \leq d$, $k \notin \{p, p+1\}$.

Example

To illustrate the results of Proposition 2.1, one considers the case $d = 4$ and with the condition $2\beta\mu < \lambda < 3\beta\mu$. One easily gets that

$$t_3 = \frac{1}{\mu}\log\left(\frac{3}{4} \cdot \frac{4\beta - \rho}{3\beta - \rho}\right)$$

and

$$s_3(t) = \begin{cases} (4\beta - \rho)(1 - e^{-\mu t}) & \text{if } t \leq t_3, \\ \beta - \dfrac{27}{256}\dfrac{(4\beta - \rho)^4}{(3\beta - \rho)^3} e^{-4\mu t} & \text{if } t \geq t_3, \end{cases}$$

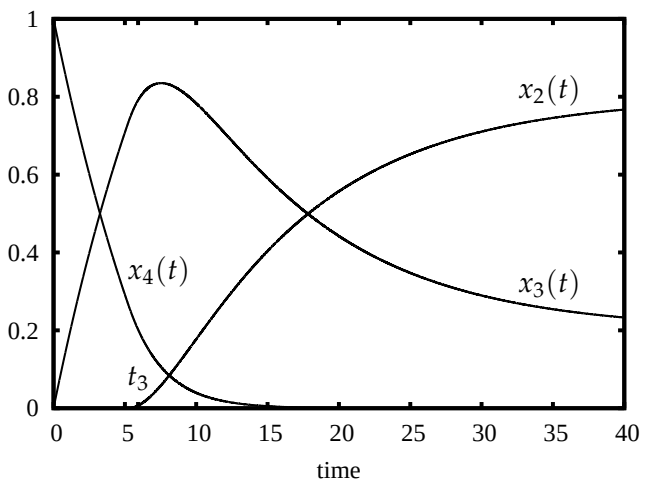


Figure 2.2 – Fluid Limits of an Overloaded Network, $2\beta\mu<\lambda<3\beta\mu$ with $d = 4$, $\mu = 0.1$, $\lambda = 0.22$, $\beta = 1$. In this case $t_3 = 5.23$.

and for $t < t_3$, $s_2(t) = 0$ and if $t > t_3$,

$$s_2(t) = (3\beta - \rho) - (4\beta - \rho)e^{-\mu t} - \frac{27}{256}\frac{(\rho - 4\beta)^4}{(\rho - 3\beta)^3}e^{-4\mu t}.$$

Finally $s_1(t) = x_0(t) = 0$, for all $t \geq 0$. Figure 2.2 presents a case with $d = 4$ and where, asymptotically, a local equilibrium holds: files have either 2 or 3 copies as t goes to infinity.

2.4 Evolution of Stable Network

In this section, the asymptotic properties of the sequence of processes

$$(X^N(t)) = (X_0^N(t), X_1^N(t), \ldots, X_{d-1}^N(t))$$

are investigated under the condition $\rho = \lambda/\mu > d\beta$ and with the initial state $X^N(0) = (0, \ldots, 0, F_N)$. Section 2.3 has shown that, in this case, the system is stable at the first order, i.e. that the fraction of lost files is 0. This does not change the fact that the system is transient with one absorbing state $(F_N, 0, \ldots, 0)$. The purpose of this section is of showing that the decay of this networks occurs on the time scale $t \mapsto N^{d-1}t$.

The section is organized as follows, preliminary results, Lemma 2.2 and Proposition 2.1 partially based on couplings show that the coordinates in the middle, i.e. with index between 1 and $d - 1$, cannot be very large on any time scale. In a second step, Proposition 2.3 and Proposition 2.4 show that the flows between the coordinates of the Markov process are "small". Proposition 2.4 is the crucial technical result of this section. Finally, the asymptotic study of a random measure on $\mathbb{N} \times \mathbb{R}_+$ gives the last element to

establish the main result, Theorem 2.2, on the evolution of the network on the time scale $t \mapsto N^{d-1}t$.

Stochastic Differential Equations
The SDE satisfied by the process $(X_k^N(t))$ are recalled. As before, if $1 \leq k \leq d$, $S_k^N(t) = X_1^N(t) + \cdots + X_k^N(t)$ and the convention that $S_0^N \equiv 0$ and $S_{-1}^N \equiv -1$, then

$$X_0^N(t) = \mu \int_0^t X_1^N(u)\ \mathrm{d}u + M_0^N(t), \tag{2.15}$$

$$X_k^N(t) = \mu(k+1) \int_0^t X_{k+1}^N(u)\ \mathrm{d}u - \mu k \int_0^t X_k^N(u)\ \mathrm{d}u \tag{2.16}$$
$$+ \lambda N \int_0^t \mathbb{I}_{\{S_{k-2}^N(u)=0, X_{k-1}^N(u)>0\}}\ \mathrm{d}u - \lambda N \int_0^t \mathbb{I}_{\{S_{k-1}^N(u)=0, X_k^N(u)>0\}}\ \mathrm{d}u$$
$$+ M_k^N(t), \quad \text{for } 1 \leq k \leq d-1,$$

where, for $0 \leq k \leq d-1$, $(M_k^N(t))$ is a square integrable martingale whose previsible increasing process is given by

$$\left\langle M_0^N \right\rangle (t) = \mu \int_0^t X_1^N(u)\ \mathrm{d}u, \tag{2.17}$$

$$\left\langle M_k^N \right\rangle (t) = \mu(k+1) \int_0^t X_{k+1}^N(u)\ \mathrm{d}u + \mu k \int_0^t X_k^N(u)\ \mathrm{d}u \tag{2.18}$$
$$+ \lambda N \int_0^t \mathbb{I}_{\{S_{k-2}^N(u)=0, X_{k-1}^N(u)>0\}}\ \mathrm{d}u + \lambda N \int_0^t \mathbb{I}_{\{S_{k-1}^N(u)=0, X_k^N(u)>0\}}\ \mathrm{d}u$$

2.4.1 Some Technical Results

We start with two preliminary results on a coupling of the network.

Lemma 2.1. *If $(L(t))$ is the process of the number of customer of an $M/M/1$ queue with arrival rate α and service rate $\gamma > \alpha$ and with initial condition $L(0) = x_0 \in \mathbb{N}$ then, for the convergence in distribution of continuous processes,*

$$\lim_{N \to +\infty} \left(\frac{1}{N} \int_0^{Nt} L(u)\ \mathrm{d}u \right) = \left(\frac{\alpha}{\gamma - \alpha} t \right).$$

Proof. The proof is standard, see the end of Proof of Proposition 9.14 page 272 of Robert [138] for example. □

The next proposition presents an important property of the network. Roughly speaking it states that the F_N files have either 0 or d copies on the time scale $t \mapsto N^{d-1}t$. Coordinates with index between 1 and $d-1$ of the vector $(X^N(t))$ remain small.

Proposition 2.2. *Let the condition $d\beta\mu < \lambda$ hold and the initial state is $X^N(0) = (0, \ldots, 0, F_N)$.*
(I) (Coupling). *One can find a probability space so that the relation*

$$(d-1)X_1^N(t) + (d-2)X_2^N(t) + \cdots + X_{d-1}^N(t) \leq L_0(Nt), \qquad \forall t > 0,$$

holds, where the vector $(X_k^N(t), 1 \leq k \leq d-1)$ has the same distribution as the state of our network and $(L_0(t))$ is the process of the number of customers of an $M/M/1$ queue with arrival rate $d\mu\beta_0$ and service rate λ and with the initial condition $L_0(0) = 0$ for some β_0 satisfying $d\mu\beta_0 < \lambda$.

(II) (Convergence). *For all $i = 1, 2, \ldots, d-1$ and $\alpha > 0$ then, for the convergence in distribution of continuous processes, the relation*

$$\lim_{N\to\infty}\left(\frac{X_i^N(N^{d-1}t)}{N^\alpha}\right) = 0 \tag{2.19}$$

holds.

Proof. The existence of N_0 and β_0 such that $d\mu\beta_0 < \lambda$ and $F_N \leq \beta_0 N$ for $N \geq N_0$ is clear. Define

$$Z^N(t) = (d-1)X_1^N(t) + (d-2)X_2^N(t) + \cdots + X_{d-1}^N(t),$$

then the possible jumps of $(Z^N(t))$ are either 1, -1 or $-(d-1)$. If $X^N(t) = (x_k)$, jumps of size 1 occur at rate $\mu[2x_2 + \cdots + (d-1)x_{d-1} + dx_d] \leq \mu d F_N \leq \mu d\beta_0 N$. Similarly jumps of size -1 occurs at rate λN provided that $Z^N(t) \neq 0$. A simple coupling gives therefore that $(Z_N(t))$ is upper bounded by an $M/M/1$ queue with service rate λN and arrival rate $\mu d\beta_0 N$. The first part of the proposition is proved.

By ergodicity of the $M/M/1$ process $(L_0(t))$, one has, for the convergence in distribution on compact sets,

$$\lim_{N\to+\infty}\left(\frac{L_0(N^\kappa t)}{N^\alpha}\right) = 0 \tag{2.20}$$

for all $\kappa > 0$ and $\alpha > 0$. Indeed, if

$$T_K = \inf\{s \geq 0 : L_0(s) \geq K\},$$

then, if $\delta = d\mu\beta_0/\lambda$, the random variable $\delta^K T_K$ is converging in distribution to an exponential random variable as K goes to infinity. See Proposition 5.11 page 119 of Robert [138] for example.

For $T > 0$ and $\varepsilon > 0$, one has

$$\mathbb{P}\left(\sup_{0\leq s\leq T}\frac{L_0(N^\kappa t)}{N^\alpha} \geq \varepsilon\right) = \mathbb{P}\left(T_{\lceil \varepsilon N^\alpha\rceil} \leq N^\kappa T\right)$$

and since $\delta < 1$, this last term is converging to 0 as N goes to infinity. Convergence (2.20) has therefore been proved. One concludes that the sequence of processes $(Z^N(N^{d-1}t)/N^\alpha)$ converges in distribution to 0. The proposition is proved. □

Proposition 2.3. *Under the condition $d\beta\mu < \lambda$ and if $X^N(0) = (0, \ldots, 0, F_N)$, then, for $1 \leq k \leq d-1$ and any $\gamma > 0$. one has*

$$\lim_{N\to\infty}\left(\frac{1}{N^{k+\gamma}}\int_0^{N^{d-1}t} X_k^N(u)\,\mathrm{d}u\right) = 0. \tag{2.21}$$

for the convergence in distribution of continuous processes, and, for any $t \geq 0$,

$$\lim_{N\to\infty}\frac{1}{N^{k+\gamma}}\int_0^{N^{d-1}t}\mathbb{E}\left(X_k^N(u)\right)\mathrm{d}u = 0. \tag{2.22}$$

Proof. One proceeds by induction on $1 \leq k \leq d-1$. Let $k=1$, if $t \geq 0$, Equation (2.15) gives the relation

$$\mu \frac{1}{N^{1+\gamma}} \int_0^{N^{d-1}t} X_1^N(u)\ \mathrm{d}u = \frac{X_0^N(N^{d-1}t)}{N^{1+\gamma}} - \frac{M_0^N(N^{d-1}t)}{N^{1+\gamma}}, \tag{2.23}$$

by Doob's Inequality and Equation (2.17), for $\varepsilon > 0$,

$$\mathbb{P}\left(\sup_{0\leq u\leq t} \frac{|M_0^N(N^{d-1}u)|}{N^{1+\gamma}} \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2}\mu\frac{1}{N^{2+2\gamma}}\mathbb{E}\left(\int_0^{N^{d-1}t} X_1^N(N^{d-1}u)\ \mathrm{d}u\right) = \frac{1}{\varepsilon^2}\mathbb{E}\left(\frac{X_0^N(N^{d-1}t)}{N^{2+2\gamma}}\right)$$

by using again Equation (2.23). The variable X_0^N being upper bounded by F_N, relation (2.23) shows that Convergence (2.22) holds in this case. Additionally one gets that the martingale term of Equation (2.23) vanishes at infinity. The convergence (2.21) is therefore proved. Induction assumption is thus true for $k=1$.

Assume by that induction assumption holds up to index $k<d-1$. Equation (2.16) gives

$$\begin{aligned}\frac{\mu(k+1)}{N^{k+1+\gamma}} \int_0^{N^{d-1}t} X_{k+1}^N(u)\ \mathrm{d}u &= \frac{X_k^N(N^{d-1}t)}{N^{k+1+\gamma}} + \frac{\mu k}{N^{k+1+\gamma}} \int_0^{N^{d-1}t} X_k^N(u)\ \mathrm{d}u \\ &- \frac{\lambda}{N^{k+\gamma}} \int_0^{N^{d-1}t} \mathbb{I}_{\{S_{k-2}^N(u)=0, X_{k-1}^N(u)>0\}}\ \mathrm{d}u \\ &+ \frac{\lambda}{N^{k+\gamma}} \int_0^{N^{d-1}t} \mathbb{I}_{\{S_{k-1}^N(u)=0, X_k^N(u)>0\}}\ \mathrm{d}u - \frac{M_k^N(N^{d-1}t)}{N^{k+1+\gamma}}. \end{aligned} \tag{2.24}$$

Note that, for $i = k-1, k$,

$$\frac{\lambda}{N^{k+\gamma}} \int_0^{N^{d-1}t} \mathbb{I}_{\{S_{i-1}^N(u)=0, X_i^N(u)>0\}}\ \mathrm{d}u \leq \frac{\lambda}{N^{k+\gamma}} \int_0^{N^{d-1}t} X_i^N(u)\ \mathrm{d}u.$$

By integrating Equation (2.24) and using the induction assumption, one obtains that Convergence (2.22) holds for $k+1$. Back to Equation (2.24), by induction again, the first four terms of the right hand side of Equation (2.24) converges to 0 and the martingale term vanishes since the expected value of its previsible increasing process is converging to 0 by Relation (2.18) and Convergence (2.22) which has been established. □

Proposition 2.4. *Under the condition $d\beta\mu < \lambda$ and if $X^N(0) = (0, \ldots, 0, F_N)$, then the relations, for $1 \leq k \leq d-2$,*

$$\lim_{N\to\infty} \left(\frac{1}{N^{k+1/2}} \int_0^{N^{d-1}t} \left[(k+1)\mu X_{k+1}^N(u) - \lambda N X_k^N(u)\right]\ \mathrm{d}u\right) = 0 \tag{2.25}$$

holds for the convergence in distribution of continuous processes.

Proof. One proves Convergence (2.25) for $1 \leq k \leq d-2$. With the evolution equation (2.16) and the same notation (2.5) as in Section 2.2 for the Poisson processes, for any function $f : \mathbb{N} \to \mathbb{R}_+$, one has

$$\begin{aligned} f\left(X_k^N(t)\right) =& f\left(X_k^N(0)\right) + \int_0^t \left[f\left(X_k^N(u-)+1\right) - f\left(X_k^N(u-)\right)\right] \mathcal{N}_{\mu(k+1)X_{k+1}^N}(\mathrm{d}u) \\ & \int_0^t \left[f\left(X_k^N(u-)-1\right) - f\left(X_k^N(u-)\right)\right] \mathcal{N}_{\mu k X_k^N}(\mathrm{d}u) \\ & + \int_0^t \left[f\left(X_k^N(u-)+1\right) - f\left(X_k^N(u-)\right)\right] \mathbb{I}_{\left\{S_{k-2}^N(u-)=0, X_{k-1}^N(u-)>0\right\}} \mathcal{N}_{\lambda N}(\mathrm{d}u) \\ & + \int_0^t \left[f\left(X_k^N(u-)-1\right) - f\left(X_k^N(u-)\right)\right] \mathbb{I}_{\left\{S_{k-1}^N(u-)=0, X_k^N(u-)>0\right\}} \mathcal{N}_{\lambda N}(\mathrm{d}u). \end{aligned}$$

By taking $f(x)=x^2$ and by compensating the Poisson processes, one gets the relation

$$\begin{aligned} X_k^N(N^{d-1}t)^2 = \mu(k+1) & \int_0^{N^{d-1}t} (2X_k^N(u)+1) X_{k+1}^N(u)\ \mathrm{d}u \\ & + \mu k \int_0^{N^{d-1}t} (-2X_k^N(u)+1) X_k^N(u)\ \mathrm{d}u \\ & + \lambda N \int_0^{N^{d-1}t} (2X_k^N(u)+1) \mathbb{I}_{\left\{S_{k-2}^N(u)=0, X_{k-1}^N(u)>0\right\}}\ \mathrm{d}u \\ & + \lambda N \int_0^{N^{d-1}t} (-2X_k^N(u)+1) \mathbb{I}_{\left\{S_{k-1}^N(u)=0, X_k^N(u)>0\right\}}\ \mathrm{d}u + M_{k,2}^N(t). \end{aligned} \tag{2.26}$$

The process $(M_{k,2}^N(t))$ is a martingale with a previsible increasing process given by

$$\begin{aligned} \left\langle M_{k,2}^N \right\rangle (t) = \mu(k+1) & \int_0^{N^{d-1}t} (2X_k^N(u)+1)^2 X_{k+1}^N(u)\ \mathrm{d}u \\ & + \mu k \int_0^{N^{d-1}t} (-2X_{k+1}^N(u)+1)^2 X_k^N(u)\ \mathrm{d}u \\ & + \lambda N \int_0^{N^{d-1}t} (2X_k^N(u)+1)^2 \mathbb{I}_{\left\{S_{k-2}^N(u)=0, X_{k-1}^N(u)>0\right\}}\ \mathrm{d}u \\ & + \lambda N \int_0^{N^{d-1}t} (-2X_k^N(u)+1)^2 \mathbb{I}_{\left\{S_{k-1}^N(u)=0, X_k^N(u)>0\right\}}\ \mathrm{d}u. \end{aligned} \tag{2.27}$$

By adding up Equations (2.16) and (2.26), after some straightforward calculations, one gets

$$\begin{aligned} X_k^N(N^{d-1}t) + X_k^N(N^{d-1}t)^2 = 2 & \int_0^{N^{d-1}t} \left[\mu(k+1) X_{k+1}^N(u) - \lambda N X_k^N(u)\right]\ \mathrm{d}u \\ & + 2\mu(k+1) \int_0^{N^{d-1}t} X_k^N(u) X_{k+1}^N(u)\ \mathrm{d}u - 2\mu k \int_0^{N^{d-1}t} X_k^N(u)^2\ \mathrm{d}u \\ & + 2\lambda N \int_0^{N^{d-1}t} (X_k^N(u)+1) \mathbb{I}_{\left\{S_{k-2}^N(u)=0, X_{k-1}^N(u)>0\right\}}\ \mathrm{d}u \\ & + 2\lambda N \int_0^{N^{d-1}t} X_k^N(u) \mathbb{I}_{\left\{S_{k-1}^N(u)>0, X_k^N(u)>0\right\}}\ \mathrm{d}u + M_k^N(t) + M_{k,2}^N(t). \end{aligned} \tag{2.28}$$

It will be shown that, when this relation is scaled by the factor $N^{k+1/2}$, except the first integral in the right hand side, all terms of this identity vanish as N gets large. The proposition will be then proved.

For the terms of the left hand side this is clear. For $1 \leq k \leq d-2$ and $j \in \{k, k+1\}$, the relation

$$\frac{1}{N^{k+1/2}} \int_0^{N^{d-1}t} X_k^N(u) X_j^N(u) \, \mathrm{d}u \leq \sup_{0 \leq u \leq t} \left(\frac{X_j^N(N^{d-1}u)}{N^{1/4}} \right) \frac{1}{N^{k+1/4}} \int_0^{N^{d-1}t} X_k^N(u) \, \mathrm{d}u,$$

and Propositions 2.2 and 2.3 show that the second term of the right hand side of Equation (2.28) scaled by $N^{k+1/2}$ vanishes for the convergence of processes when N gets large. By using the inequalities

$$\int_0^{N^{d-1}t} X_k^N(u) \mathbb{I}_{\{S_{k-2}^N(u)=0, X_{k-1}^N(u)>0\}} \, \mathrm{d}u \leq \int_0^{N^{d-1}t} X_k^N(u) X_{k-1}^N(u) \, \mathrm{d}u,$$

and

$$\int_0^{N^{d-1}t} X_k^N(u) \mathbb{I}_{\{S_{k-1}^N(u)>0, X_k^N(u)>0\}} \, \mathrm{d}u \leq \sum_{i=1}^{k-1} \int_0^{N^{d-1}t} X_i^N(u) X_k^N(u) \, \mathrm{d}u,$$

the same property can be established in a similar way for the third, fourth and fifth terms.

By using Equations (2.18) and (2.27) and similar methods one gets that for any $t \geq 0$, for the convergence in distribution,

$$\lim_{N \to +\infty} \left(\frac{\langle M_k^N \rangle (N^{d-1}t)}{N^{2k+1}} \right) = 0, \quad \lim_{N \to +\infty} \left(\frac{\left\langle M_{k,2}^N \right\rangle (N^{d-1}t)}{N^{2k+1}} \right) = 0.$$

From Relation (3.31) of Lemma I.3.30 of Jacod and Shiryaev [81], these convergences of previsible increasing processes show that the martingale terms of Relation (2.28) scaled by $N^{k+1/2}$ vanish for the convergence of processes when N gets large. The proposition is proved. □

Proposition 2.5. *Under the condition $d\beta\mu < \lambda$ and if $X^N(0) = (0, \ldots, 0, F_N)$ then, for the convergence in distribution of continuous processes, the relations*

$$\lim_{N \to +\infty} \left(\sqrt{N} \left(\frac{\mu}{N} \int_0^{N^{d-1}t} X_1^N(u) \, \mathrm{d}u - \lambda \frac{(d-1)!}{\rho^{d-1}} \int_0^t X_{d-1}^N(N^{d-1}u) \, \mathrm{d}u \right) \right) = 0 \tag{2.29}$$

and

$$\lim_{N \to +\infty} \left(\frac{X_0^N(N^{d-1}t)}{N} - \lambda \frac{(d-1)!}{\rho^{d-1}} \int_0^t X_{d-1}^N(N^{d-1}u) \, \mathrm{d}u \right) = 0 \tag{2.30}$$

hold.

Proof. By Relation (2.25), one gets that, for the convergence in distribution of continuous processes,

$$\lim_{N\to+\infty}\left(\sqrt{N}\left(\int_0^{N^{d-1}t}\left[\mu(k+1)\frac{X_{k+1}^N(u)}{N^{k+1}}-\lambda\frac{X_k^N(u)}{N^k}\right]\,\mathrm{d}u\right)\right)=0,$$

holds for $1\le k\le d-2$, and therefore that

$$\lim_{N\to+\infty}\left(\sqrt{N}\left(\int_0^{N^{d-1}t}\left[\frac{(k+1)!}{\rho^{k+1}}\frac{X_{k+1}^N(u)}{N^{k+1}}-\frac{k!}{\rho^k}\frac{X_k^N(u)}{N^k}\right]\,\mathrm{d}u\right)\right)=0.$$

By summing up these relations, one finally gets that

$$\lim_{N\to+\infty}\left(\sqrt{N}\left(\frac{(d-1)!}{\rho^{d-1}}\frac{1}{N^{d-1}}\int_0^{N^{d-1}t}X_{d-1}^N(u)\,\mathrm{d}u-\frac{\mu}{\lambda}\int_0^{N^{d-1}t}\frac{X_1^N(u)}{N}\,\mathrm{d}u\right)\right)=0.$$

Relation (2.29) is proved.

SDE (2.15) for $(X_0^N(t))$ gives the relation

$$\frac{X_0^N(N^{d-1}t)}{N}=\frac{M_0^N(N^{d-1}t)}{N}+\frac{\mu}{N}\int_0^{N^{d-1}t}X_1^N(u)\,\mathrm{d}u,$$

where $(M_0^N(N^{d-1}t)/N)$ is a martingale whose previsible increasing process is given by

$$\left(\left\langle\frac{M_0^N}{N}\right\rangle(N^{d-1}t)\right)=\left(\mu\frac{1}{N^2}\int_0^{N^{d-1}t}X_1^N(u)\,\mathrm{d}u\right),$$

it is converging in distribution to 0 by Proposition 2.3, one concludes that the martingale is also converging to 0. The proposition is thus proved. □

We now turn to the proof of an averaging principle. It relies on the martingale characterization of Markov processes as used in Papanicolau et al. [124] in a Brownian setting, see also Kurtz [98].

2.4.2 Convergence of Occupation Measures

For $x\in\mathbb{N}$ and $N\ge 1$, the random measure Λ_x^N on $\mathbb{R}_+$ is defined as, for a measurable function $g:\mathbb{R}_+\to\mathbb{R}_+$,

$$\left\langle\Lambda_x^N,g\right\rangle=\int_{\mathbb{R}_+}g(t)\mathbb{I}_{\{X_{d-1}^N(N^{d-1}t)=x\}}\,\mathrm{d}t.$$

Clearly Λ_x^N is the random Radon measure associated with the local time of $(X_{d-1}^N(N^{d-1}t))$ at x. For a given x, the sequence (Λ_x^N) of random Radon measures on $\mathbb{R}_+$ is tight. See Dawson [42, Lemma 3.28, page 44] for example. Note that the null measure can be a possible limit of this sequence. By using a diagonal argument, one can fix (N_k) such that, for any $x\in\mathbb{N}$, $(\Lambda_x^{N_k})$ is a converging subsequence whose limit is ν_x.

Since, for $N\ge 1$, Λ_x^N is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}_+$, the same property holds for a possible limiting measure ν_x. Let $(x,t)\to\pi_t(x)$

denote its (random) density. It should be remarked that, one can choose a version of $\pi_t(x)$ such that the map $(\omega, x, t) \to \pi_t(x)(\omega)$ on the product of the probability space and $\mathbb{N} \times \mathbb{R}_+$ is measurable by taking $\pi_t(x)$ as a limit of measurable maps,

$$\pi_t(x) = \limsup_{s \to 0} \frac{1}{s} \nu_x([t, t+s]).$$

See Chapter 8 of Rudin [145] for example. See also Lemma 1.4 of Kurtz [98]. One denotes by π_t the measure on $\mathbb{N}$ defined by the sequence $(\pi_t(x), x \in \mathbb{N})$.

Proposition 2.6. *For any function $f : \mathbb{N} \to \mathbb{R}_+$ such that the sequence $(f(x)/x)$ is bounded then, with the subsequence (N_k) defined above, for the convergence in distribution of continuous processes,*

$$\lim_{k \to +\infty} \left(\frac{1}{N_k^{d-1}} \int_0^{N_k^{d-1} t} f\left(X_{d-1}^{N_k}(u)\right) \, \mathrm{d}u \right) = \left(\int_0^t \langle \pi_u, f \rangle \, \mathrm{d}u \right),$$

In particular, almost surely, for all $t \geq 0$,

$$\sum_{x \geq 0} \int_0^t \pi_u(x) \, \mathrm{d}u = \int_0^t \pi_u(\mathbb{N}) \, \mathrm{d}u = t.$$

Proof. Denote $K = \sup\{f(x)/x : x \geq 1\}$ and

$$\Psi_f^N(t) = \frac{1}{N^{d-1}} \int_0^{N^{d-1} t} f\left(X_{d-1}^N(u)\right) \, \mathrm{d}u,$$

the stochastic domination results of Proposition 2.2 gives that, for any $0 \leq s \leq t$,

$$\Psi_f^N(t) - \Psi_f^N(s) \leq K \frac{1}{N^d} \int_{N^d s}^{N^d t} L_0(u) \, \mathrm{d}u,$$

where $(L_0(t))$ is the process of the number of customers of an $M/M/1$ queue with arrival rate $d\mu\beta_0$ and service rate λ for some convenient β_0 such that $d\mu\beta_0 < \lambda$ and with the initial condition $L_0(0) = 0$. The convergence result of Lemma 2.1 implies then that the sequence of processes $(\Psi_f^N(t))$ is tight by the criteria of the modulus of continuity.

For $C \geq 1$ and $t > 0$

$$\begin{aligned} \frac{1}{N^{d-1}} \int_0^{N^{d-1} t} f(X_{d-1}^N(u)) \mathbb{I}_{\{X_{d-1}^N(u) \geq C\}} \, \mathrm{d}u &\leq \frac{K}{N^{d-1}} \int_0^{N^{d-1} t} X_{d-1}^N(u) \mathbb{I}_{\{X_{d-1}^N(u) \geq C\}} \, \mathrm{d}u \\ &\leq \frac{K}{N^d} \int_0^{N^d t} L_0(u) \mathbb{I}_{\{L_0(u) \geq C\}} \, \mathrm{d}u. \end{aligned}$$

The last term is converging in distribution to $Kt\mathbb{E}(\bar{L}_0)\mathbb{I}_{\{\bar{L}_0 \geq C\}})$, where $\bar{L}_0$ is a random variable with geometric distribution with parameter $d\mu\beta_0/\lambda$, the invariant distribution of the process $(L_0(t))$. In particular for $T > 0$, if C is sufficiently large, this term can be made arbitrarily small uniformly for $t \leq T$.

By using the fact that, for $x \in \mathbb{N}$,

$$\frac{1}{N^{d-1}} \int_0^{N^{d-1} t} f(X_{d-1}^N(u)) \mathbb{I}_{\{X_{d-1}^N(u) = x\}} \, \mathrm{d}u = f(x) \left\langle \Lambda_x^N, \mathbb{I}_{\{[0,t]\}} \right\rangle,$$

one gets the desired convergence in distribution.

□

2.4.3 The Decay of the Network occurs on the Time Scale $t \mapsto N^{d-1}t$

We have all the necessary technical results to prove the main result concerning the behavior of the system on the time scale $t \mapsto N^{d-1}t$.

Theorem 2.2 (Rate of Decay of the Network). *Under the condition $d\beta\mu < \lambda$ and if $X^N(0) = (0, \ldots, 0, F_N)$, then the sequence of processes $(X_0^N(N^{d-1}t)/N)$ converges in distribution to $(\Phi(t))$ where, for $t \geq 0$, $\Phi(t)$ is the unique solution $y \in [0, \beta]$ of the equation*

$$\left(1 - \frac{y}{\beta}\right)^{\rho/d} e^y = \exp\left(-\lambda \frac{(d-1)!}{\rho^{d-1}} t\right). \tag{2.31}$$

Proof. Let f be a function on $\mathbb{N}$ with finite support then, the SDE (2.16) associated to the evolution equations give

$$\begin{aligned}
&\frac{f(X_{d-1}^N(N^{d-1}t)) - f(0) - M_f^N(N^{d-1}t)}{N^d} = \\
&\int_0^t \Delta^+(f)(X_{d-1}^N(N^{d-1}u)) \left(d\mu \frac{X_d^N(N^{d-1}u)}{N} + \lambda \mathbb{I}_{\{S_{d-3}^N = 0, X_{d-2}^N(N^{d-1}u) > 0\}} \right) \mathrm{d}u \\
&+ \int_0^t \Delta^-(f)(X_{d-1}^N(N^{d-1}u)) \left((d-1)\mu \frac{X_{d-1}^N(N^{d-1}u)}{N} + \lambda \mathbb{I}_{\{S_{d-2}^N = 0, X_{d-1}^N(N^{d-1}u) > 0\}} \right) \mathrm{d}u,
\end{aligned}$$

where $\Delta^+(f)(x) = f(x+1) - f(x)$ and $\Delta^-(f)(x) = f(x-1) - f(x)$. The convergence of the various components of this identity are now examined.

Clearly enough, f being bounded, the process $([f(X_{d-1}^N(N^{d-1}t)) - f(0)]/N^d)$ is converging in distribution to 0 as N gets large. By calculating the previsible increasing process of the martingale $(M_f^N(N^{d-1}t)/N^d)$, it is not difficult to show that this process vanishes at infinity.

Note that

$$\begin{aligned}
&\left| \int_0^t \Delta^+(f)(X_{d-1}^N(N^{d-1}u)) \mathbb{I}_{\{S_{d-3}^N = 0, X_{d-2}^N(N^{d-1}u) > 0\}} \, \mathrm{d}u \right| \\
&\qquad \leq 2\|f\|_\infty \int_0^t X_{d-2}^N(N^{d-1}u) \, \mathrm{d}u = 2\|f\|_\infty \frac{1}{N^{d-1}} \int_0^{N^{d-1}t} X_{d-2}^N(u) \, \mathrm{d}u,
\end{aligned}$$

the process associated to the last term is converging in distribution to 0 by Proposition 2.3. Similarly,

$$\left| \int_0^t \Delta^-(f)(X_{d-1}^N(N^{d-1}u)) \mathbb{I}_{\{S_{d-2}^N > 0, X_{d-1}^N(N^{d-1}u) > 0\}} \, \mathrm{d}u \right| \leq 2\|f\|_\infty \sum_{k=1}^{d-2} \int_0^t X_k(N^{d-1}u) \, \mathrm{d}u$$

and the last term is also converging to 0 in distribution. In the same way

$$\left| \int_0^t \Delta^-(f)(X_{d-1}^N(N^{d-1}u)) \frac{X_{d-1}^N(N^{d-1}u)}{N} \, \mathrm{d}u \right| \leq 2\|f\|_\infty \int_0^t \frac{X_{d-1}^N(N^{d-1}u)}{N} \, \mathrm{d}u$$

which converges to 0 by the last assertion of Proposition 2.2.

To summarize, we have proved that the following convergence in distribution

$$\lim_{N\to+\infty}\left(\int_0^t \Delta^+(f)(X_{d-1}^N(N^{d-1}u))d\mu\frac{X_d^N(N^{d-1}u)}{N}\,\mathrm{d}u\right.$$
$$\left.+\int_0^t \Delta^-(f)(X_{d-1}^N(N^{d-1}u))\lambda\mathbb{I}_{\{X_{d-1}^N(N^{d-1}u)>0\}}\,\mathrm{d}u\right)=0. \quad (2.32)$$

By using again Propositions 2.5 and 2.6, one gets that the convergence of the sequence of processes is converging to a continuous process $(\Phi(t))$ such that

$$(\Phi(t)) \stackrel{\text{def.}}{=} \lim_{N\to+\infty}\left(\frac{X_0^{N_k}(N_k^{d-1}t)}{N_k}\right)=\left(\lambda\frac{(d-1)!}{\rho^{d-1}}\int_0^t\langle\pi_u,I\rangle\,\mathrm{d}u\right),$$

where $I(x)=x$ for $x\geq 0$. By the Skorohod representation theorem, one can take a convenient probability space such that, for all $x\in\mathbb{N}$, this convergence also holds almost surely as well as the convergence of the processes $(\left\langle\Lambda_x^{N_k},\mathbb{I}_{\{[0,t]\}}\right\rangle)$. The identity

$$\frac{X_d^N(N^{d-1}t)}{N}=\frac{F_N}{N}-\frac{X_0(N^{d-1}t)}{N}-\sum_{k=1}^{d-1}\frac{X_k^N(N^{d-1}t)}{N},$$

and Equation (2.32) give that the relation

$$\int_0^t\sum_{x\in\mathbb{N}}\pi_u(x)\left[d\mu(\beta-\Phi(u))\Delta^+(f)(x)+\lambda\mathbb{I}_{\{x>0\}}\Delta^-(f)(x)\right]\mathrm{d}u=0$$

holds almost surely for all $t\geq 0$ and all functions $f=f_k$, $k\geq 0$ with $f_k(x)=\mathbb{I}_{\{k\}}(x)$ for $x\in\mathbb{N}$. One concludes from this relation and Proposition 2.6, for $u\in\mathbb{R}_+$ outside a set $\mathcal{S}$ negligible for the Lebesgue measure, one has for all $k\geq 0$

$$\sum_{x\in\mathbb{N}}\pi_u(x)\left[d\mu(\beta-\Phi(u))\Delta^+(f_k)(x)+\lambda\mathbb{I}_{\{x>0\}}\Delta^-(f_k)(x)\right]=0$$

and $\pi_u(\mathbb{N})=1$. Hence, if $u\in\mathcal{S}$, $(\pi_u(x))$ is a geometric distribution, the invariant distribution of an $M/M/1$ queue with arrival rate $d\mu(\beta-\Phi(u))$ and service rate λ. The definition of $\Phi(t)$ gives therefore the fixed point equation, for all $t\geq 0$,

$$\Phi(t)=\lambda\frac{(d-1)!}{\rho^{d-1}}\int_0^t\frac{d\mu\,(\beta-\Phi(u))}{\lambda-d\mu\,(\beta-\Phi(u))}\,\mathrm{d}u, \quad (2.33)$$

one gets the relation

$$\left(1-\frac{\Phi(t)}{\beta}\right)^{\rho/d}e^{\Phi(t)}=\exp\left(-\lambda\frac{(d-1)!}{\rho^{d-1}}t\right).$$

The theorem is proved. □

One concludes this section with the asymptotic of the first instant when the network has lost a fraction $\delta\in(0,1)$ of its file. It generalizes Corollary 1 of Feuillet and Robert [61]. This is a direct consequence of the above theorem.

Corollary 2.2. *If, for $\delta \in (0,1)$,*

$$T_N(\delta) = \inf\left\{t \geq 0 : \frac{X_0^N(t)}{N} \geq \delta\beta\right\}$$

then, under the condition $\lambda > d\beta\mu$, the relation

$$\lim_{N\to+\infty} \frac{T_N(\delta)}{N^{d-1}} = \frac{\rho^{d-1}}{\lambda(d-1)!}\left(-\frac{\rho}{d}\log(1-\delta) - \beta\delta\right)$$

holds for the convergence in distribution.

2.5 Second Order Asymptotics in the Stable Case

This section is devoted to the study of the second order fluctuations associated to the law of large numbers proved in Theorem 2.2. As it will be seen the proof relies on careful stochastic calculus, technical estimates and Proposition 2.4 proved in Section 2.4.

Notations

— If $(Y_N(t))$ and $(Z_N(t))$ are sequences of stochastic process, with a slight abuse of notation, we will write $Z_N(t) = Y_N(t) + \mathcal{O}_d(1)$ when the sequence $(Z_N(t) - Y_N(t))$ converges in distribution to 0 when N goes to infinity.

Lemma 2.2. *Let*

$$Y_{d-1}^N(t) \overset{def.}{=} \frac{1}{\sqrt{N}}\int_0^t \left[(X_{d-1}^N(N^{d-1}u)+1)d\mu X_d^N(N^{d-1}u) - \lambda N X_{d-1}^N(N^{d-1}u)\right] \mathrm{d}u$$

then $(Y_{d-1}^N(t))$ converges in distribution to 0 as N goes to infinity.

Proof. By using the SDE satisfied by the process $(X_{d-1}^N(t))$, as in the proof in Proposition 2.4, one gets

$$\begin{aligned}
\frac{X_{d-1}^N(N^{d-1}t)}{N^{d-1/2}} + \frac{X_{d-1}^N(N^{d-1}t)^2}{N^{d-1/2}} &= \frac{M_{d-1}^N(t)}{N^{d-1/2}} + \frac{M_{d-1,2}^N(t)}{N^{d-1/2}} \\
&+ \frac{2}{N^{d-1/2}}\int_0^{N^{d-1}t} \Big[(X_{d-1}^N(u)+1)\lambda N \mathbb{I}_{\{S_{d-3}^N(u)=0, X_{d-2}^N(u)>0\}} \\
&\qquad - X_{d-1}^N(u)((d-1)X_{d-1}^N(u) - \lambda N \mathbb{I}_{\{S_{d-2}^N(u)>0\}})\Big] \mathrm{d}u \\
&+ \frac{2}{N^{d-1/2}}\int_0^{N^{d-1}t} [(X_{d-1}^N(u)+1)d\mu X_d^N(u) - \lambda N X_{d-1}^N(u)] \mathrm{d}u
\end{aligned}$$

where $(M_{d-1}^N(t))$ and $(M_{d-1,2}^N(t))$ are the associated local martingales. The processes of left hand side of this relation vanishes as N gets large due to Proposition 2.2. With similar arguments as in the proof of Proposition 2.4, one obtains that the martingale terms and the first integral of the right hand side vanish too. This is again a consequence of Propositions 2.2 and 2.3.

Therefore, the last term

$$\left(\frac{1}{N^{d-1/2}}\int_0^{N^{d-1}t} \left[(X_{d-1}^N(u)+1)d\mu X_d^N(u) - \lambda N X_{d-1}^N(u)\right] \mathrm{d}u\right)$$

converges to 0 in distribution when N gets large. The lemma is proved. □

Theorem 2.3 (Central Limit Theorem). *If $d\beta\mu<\lambda$ and if Condition (2.4) holds and the initial state is $X^N(0) = (0, \dots, 0, F_N)$, then the following convergence in distribution holds*

$$\lim_{N\to+\infty} \left(\frac{X_0^N(N^{d-1}t) - N\Phi(t)}{\sqrt{N}} \right) = (W(t)),$$

where $\Phi(t)$ is the solution of Equation (2.31) and the process $(W(t))$ is the solution of the stochastic differential equation

$$\mathrm{d}W(t) = \sqrt{\Phi'(t)}\,\mathrm{d}B(t) - \frac{\lambda^2\mu d!}{\rho^{d-1}} \frac{W(t)-\gamma}{(\lambda - d\mu(\beta - \Phi(t)))^2}\,\mathrm{d}t, \tag{2.34}$$

with $W_0(0) = 0$, where $(B(t))$ is a standard Brownian motion and $\Phi(t)$ is the unique solution of Equation (2.31).

Proof. We denote by

$$W_0^N(t) = \frac{X_0^N(N^{d-1}t) - N\Phi(t)}{\sqrt{N}} \text{ and } W_d^N(t) = \frac{X_d^N(N^{d-1}t) - N(\beta - \Phi(t))}{\sqrt{N}}.$$

The strategy of the proof consists in starting from the convergence proved in the above lemma to write an integral equation for the process $(W_0^N(t))$, this is Equation (2.37) below. Technical results of Section 2.4 are used repeatedly in the proof of this identity. The last part of the proof consists in proving the tightness and identifying the possible limits of this sequence.

The total sum of the coordinates of $(X_k^N(t))$ being F_N, scaling Condition (2.4) and Relation (2.19) of Proposition 2.2 give the identity

$$\begin{aligned} W_d^N(t) = \frac{X_d^N(N^{d-1}u) - N(\beta - \Phi(t))}{\sqrt{N}} &= \frac{F_N - N\beta}{\sqrt{N}} - \sum_{k=1}^{d-1} \frac{X_k^N(N^{d-1}u)}{\sqrt{N}} \\ &- \frac{X_0^N(N^{d-1}u) - N\Phi(t)}{\sqrt{N}} = -W_0^N(t) + \gamma + \mathcal{O}_d(1). \end{aligned} \tag{2.35}$$

The SDE (2.15) gives the relation

$$X_0^N(N^{d-1}t) = \mu \int_0^t X_1^N(N^{d-1}u) N^{d-1}\,\mathrm{d}u + M_0^N(N^{d-1}t).$$

The previsible increasing process of the martingale $(M_0^N(N^{d-1}t)/\sqrt{N})$ is given by

$$\left(\left\langle \frac{M_0^N}{\sqrt{N}} \right\rangle (N^{d-1}t) \right) = \left(\mu \frac{1}{N} \int_0^{N^{d-1}t} X_1^N(u)\,\mathrm{d}u \right),$$

and it is converging in distribution to $(\Phi(t))$, see the proof of Proposition 2.5. Consequently, by using Theorem 1.4 page 339 of Ethier and Kurtz [57] for example, for the convergence in distribution of processes, one has

$$\lim_{N\to+\infty} \left(\frac{M_0^N}{\sqrt{N}} \right) = \left(\int_0^t \sqrt{\Phi'(u)}\,\mathrm{d}B(u) \right) \overset{\text{dist.}}{=} (B(\Phi(t))),$$

where $(B(t))$ is a standard Brownian motion on $\mathbb{R}$.

Let

$$H_N(t) = \int_0^t X_{d-1}^N(N^{d-1}u)\ \mathrm{d}u,$$

Relation (2.29) of Proposition 2.5 shows that

$$\frac{\lambda(d-1)!}{\rho^{d-1}}\sqrt{N}H_N(t) = \sqrt{N}\mu\int_0^{N^{d-1}t}\frac{X_1^N(u)}{N}\ \mathrm{d}u + \mathcal{O}_d(1)$$

holds and SDE (2.15) gives

$$\mu\int_0^{N^{d-1}t} X_1^N(u)\ \mathrm{d}u = X_0^N(N^{d-1}t) - M_0^N(N^{d-1}t)$$
$$= N\Phi(t) + \sqrt{N}W_0^N(t) - M_0^N(N^{d-1}t).$$

One obtains therefore the following expansion for $(\sqrt{N}H_N(t))$,

$$\sqrt{N}\left(\frac{\lambda(d-1)!}{\rho^{d-1}}H_N(t) - \Phi(t)\right) = W_0^N(t) - \frac{M_0^N(N^{d-1}t)}{\sqrt{N}} + \mathcal{O}_d(1). \tag{2.36}$$

Lemma 2.2 gives the relation

$$\frac{1}{\sqrt{N}}\int_0^t\left[(X_{d-1}^N(N^{d-1}u)+1)d\mu X_d^N(N^{d-1}u) - \lambda N X_{d-1}^N(N^{d-1}u)\right]\mathrm{d}u = \mathcal{O}_d(1),$$

which can be rewritten as

$$d\mu\sqrt{N}\int_0^t X_{d-1}^N(N^{d-1}u)\left(\frac{X_d^N(N^{d-1}u)}{N} - (\beta - \Phi(u)))\right)\mathrm{d}u$$
$$+\sqrt{N}\int_0^t X_{d-1}^N(N^{d-1}u)\left(d\mu(\beta-\Phi(u)) - \lambda\right)\mathrm{d}u$$
$$+ d\mu\sqrt{N}\int_0^t(\beta-\Phi(u))\,\mathrm{d}u + d\mu\frac{1}{\sqrt{N}}\int_0^t(X_d^N(N^{d-1}u) - N(\beta-\Phi(u))\,\mathrm{d}u = \mathcal{O}_d(1).$$

If one plugs the integration by part

$$\int_0^t X_{d-1}^N(N^{d-1}u)\left[\lambda - d\mu(\beta-\Phi(u))\right]\mathrm{d}u$$
$$= H_N(t)\left[\lambda - d\mu(\beta-\Phi(t))\right] - d\mu\int_0^t H_N(u)\Phi'(u)\,\mathrm{d}u,$$

into this identity, this gives the relation

$$d\mu\int_0^t X_{d-1}^N(N^{d-1}u)W_d^N(u)\,\mathrm{d}u$$
$$-\sqrt{N}H_N(t)\left[\lambda - d\mu(\beta-\Phi(t))\right] + d\mu\int_0^t\sqrt{N}H_N(u)\Phi'(u)\,\mathrm{d}u$$
$$+ d\mu\sqrt{N}\int_0^t(\beta-\Phi(u))\,\mathrm{d}u + d\mu\int_0^t W_d^N(u)\,\mathrm{d}u = \mathcal{O}_d(1).$$

The expansion (2.36) for $(\sqrt{N}H_N(t))$ yields

$$\begin{aligned} d\mu \int_0^t X_{d-1}^N(N^{d-1}u)W_d^N(u)\,\mathrm{d}u + \sqrt{N}\Delta_N(t) \\ - \frac{\rho^{d-1}}{\lambda(d-1)!}\left(W_0^N(t) - \frac{M_0^N(N^{d-1}t)}{\sqrt{N}}\right)[\lambda - d\mu(\beta - \Phi(t))] \\ + d\mu\frac{\rho^{d-1}}{\lambda(d-1)!}\int_0^t \left(W_0^N(u) - \frac{M_0^N(N^{d-1}u)}{\sqrt{N}}\right)\Phi'(u)\,\mathrm{d}u \\ + d\mu\int_0^t W_d^N(u)\,\mathrm{d}u = \mathcal{O}_d(1) \end{aligned}$$

with

$$\begin{aligned} \Delta_N(t) &= \frac{\rho^{d-1}}{\lambda(d-1)!}\left(-\Phi(t)\,[\lambda - d\mu(\beta - \Phi(t))] + d\mu\int_0^t \Phi(u)\Phi'(u)\,\mathrm{d}u\right) \\ &\quad + d\mu\int_0^t (\beta - \Phi(u))\,\mathrm{d}u \\ &= -\frac{\rho^{d-1}}{\lambda(d-1)!}\int_0^t [\lambda - d\mu(\beta - \Phi(u))]\,\Phi'(u)\;\mathrm{d}u + d\mu\int_0^t (\beta - \Phi(u))\,\mathrm{d}u \\ &= 0, \end{aligned}$$

by Relation (2.33). By using Equation (2.35), one gets finally the relation

$$\begin{aligned} - d\mu \int_0^t X_{d-1}^N(N^{d-1}u)W_0^N(u)\,\mathrm{d}u + d\mu\gamma\int_0^t X_{d-1}^N(N^{d-1}u)\;\mathrm{d}u \\ - \frac{\rho^{d-1}}{\lambda(d-1)!}\left(W_0^N(t) - \frac{M_0^N(N^{d-1}t)}{\sqrt{N}}\right)[\lambda - d\mu(\beta - \Phi(t))] \\ + d\mu\frac{\rho^{d-1}}{\lambda(d-1)!}\int_0^t \left(W_0^N(u) - \frac{M_0^N(N^{d-1}u)}{\sqrt{N}}\right)\Phi'(u)\,\mathrm{d}u \\ - d\mu\int_0^t W_0^N(u)\,\mathrm{d}u + d\mu\gamma t = \mathcal{O}_d(1). \end{aligned} \tag{2.37}$$

Starting from the above equation, one can now complete the proof of the theorem in four steps.

1. Local boundedness.
   By using the convergence in distribution of $(M_0^N(N^{d-1}u)/\sqrt{N})$ and Gronwall's Inequality, one gets that, for $\varepsilon > 0$ and $T > 0$, there exists some $K > 0$ and N_0 such that if $N \geq N_0$, then
   $$\mathbb{P}\left(\sup_{0\leq s\leq T}|W_0^N(s)| \geq K\right) \leq \varepsilon. \tag{2.38}$$

2. Tightness.
   One first note that the two sequences of processes
   $$\left(\int_0^t X_{d-1}^N(N^{d-1}u)\,\mathrm{d}u\right) \text{ and } \left(\int_0^t X_{d-1}^N(N^{d-1}u)W_0^N(u)\,\mathrm{d}u\right)$$

satisfy the criterion of the modulus of continuity: for the first sequence this is a consequence of Proposition 2.5 and Theorem 2.2. Relation (2.38) and the fact that, for $0 \leq s \leq t \leq T$,

$$\left| \int_s^t X_{d-1}^N(N^{d-1}u) W_0^N(u)\, \mathrm{d}u \right| \leq \left(\sup_{0 \leq u \leq T} |W_0^N(u)| \right) \int_s^t X_{d-1}^N(N^{d-1}u)\, \mathrm{d}u,$$

give this property for the second sequence. As it has been seen this is also the case for $(M_0^N(N^{d-1}u)/\sqrt{N})$. Relation (2.37) can thus be rewritten as

$$W_0^N(t) + \int_0^t W_0^N(u) F(u)\, \mathrm{d}u = H^N(t), \tag{2.39}$$

where $(F(t))$ is a deterministic continuous function and $(H^N(t))$ is a sequence of processes which satisfies the criterion of the modulus of continuity. As before, See relation (2.13), denote w_Z as the modulus of continuity of the process $(Z(t))$ on $[0, T]$, Relation (2.39) gives the inequality,

$$w_{W_0^N} \leq w_{H^N} + \delta \|F\|_\infty \sup_{0 \leq s \leq T} |W_0^N(s)|$$

with $\|F\|_\infty = \sup(|F(s)|, 0 \leq s \leq T)$. One deduces the tightness of $(W_0^N(t))$ by the criterion of the modulus of continuity. In particular any limiting point is a continuous process.

3. Convergence of the first term of Equation (2.37).
Let $(W(t))$ be a limit of some subsequence $(W_0^{N_k}(t))$. By Skorohod's representation theorem, on can assume that the convergence

$$\lim_{k \to +\infty} \left(\int_0^t X_{d-1}^{N_k}(N_k^{d-1}u)\, \mathrm{d}u, W_0^{N_k}(t) \right) = \left(\frac{\rho^{d-1}}{\lambda(d-1)!} \Phi(t), W(t) \right)$$

holds almost surely for the uniform norm on compact sets of $\mathbb{R}_+$. If f is a $C^1(\mathbb{R}_+)$ function, by integration par parts, one has the convergence

$$\lim_{k \to +\infty} \left(\int_0^t X_{d-1}^{N_k}(N_k^{d-1}u) f(u)\, \mathrm{d}u \right) = \left(\frac{\rho^{d-1}}{\lambda(d-1)!} \int_0^t \Phi'(u) f(u)\, \mathrm{d}u \right),$$

which can be extended to any arbitrary continuous function f by a regularization procedure. Since

$$\lim_{k \to +\infty} \left(\int_0^t X_{d-1}^{N_k}(N_k^{d-1}u)(W_0^{N_k}(u) - W(u))\, \mathrm{d}u \right) = 0,$$

one finally gets the convergence

$$\lim_{k \to +\infty} \left(\int_0^t X_{d-1}^{N_k}(N_k^{d-1}u) W_0^{N_k}(u)\, \mathrm{d}u \right) = \left(\frac{\rho^{d-1}}{\lambda(d-1)!} \int_0^t \Phi'(u) W(u)\, \mathrm{d}u \right),$$

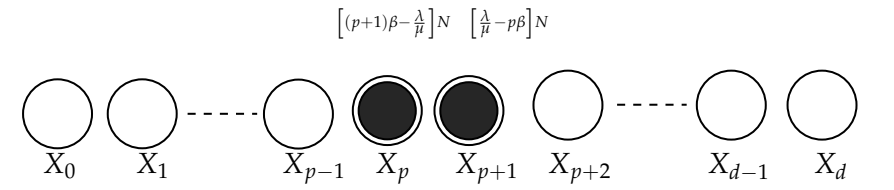


Figure 2.3 – Stable Asymptotic Fluid State in an Overloaded Network with $p\beta\mu < \lambda < (p+1)\beta\mu$ for some $1 < p < d$

4. Identification of the limit.
   A possible limit $(W(t))$ satisfies therefore the integral equation

$$
\begin{aligned}
-d\mu \int_0^t \left(\frac{\rho^{d-1}}{\lambda(d-1)!}\Phi'(u)+1\right) W(u)\,\mathrm{d}u + d\gamma\frac{\rho^{d-2}}{(d-1)!}\Phi(t) \\
- \frac{\rho^{d-1}}{\lambda(d-1)!}\left(W(t)-B(\Phi(t))\right)\left[\lambda - d\mu(\beta-\Phi(t))\right] \\
+ d\mu\frac{\rho^{d-1}}{\lambda(d-1)!}\int_0^t \left(W(u)-B(\Phi(u))\right)\Phi'(u)\,\mathrm{d}u + d\mu\gamma t = 0,
\end{aligned}
$$

   and, with Relation (2.33), it can be rewritten as

$$
\begin{aligned}
-\lambda d\mu \int_0^t \frac{W(u)}{\lambda-(\beta-\Phi(u))}\,\mathrm{d}u + d\gamma\frac{\rho^{d-2}}{(d-1)!}\Phi(t) + d\mu\gamma t \\
- \frac{\rho^{d-1}}{\lambda(d-1)!}\int_0^t \left[\lambda - d\mu(\beta-\Phi(u))\right]\left(\mathrm{d}W(u) - \sqrt{\Phi'(u)}\,\mathrm{d}B(u)\right) = 0.
\end{aligned}
$$

   The theorem is proved.

□

2.6 A Local Equilibrium in the Overloaded Case

We have seen in Corollary 2.1 that if for some $2 \le p < d$, one has $p\beta \le \rho < (p+1)\beta$ and if the initial state is $X_N(0) = (0,\dots,0,F_N)$ then one has the convergence in distribution

$$\lim_{N\to+\infty}\frac{1}{N}(X_p^N(t), X_{p+1}^N(t)) = (x_p(t), x_{p+1}(t))$$

and

$$\lim_{t\to+\infty}(x_p(t), x_{p+1}(t)) = ((p+1)\beta-\rho, \rho-p\beta).$$

The system started with $\sim\beta N$ files with d copies and it ends up, on the normal time scale, in a state where there are still βN files but with either p or $p+1$ copies.

In this section we start from this "equilibrium", Proposition 2.7 shows that this fluid state does not change on the time scale $t \mapsto N^{p-2}t$. Theorem 2.4 proves that, on the time

scale $t \mapsto N^{p-1}t$, a positive fraction of files are lost. It is also shown that the number of files with p copies decreases to end up in a state where, for the fluid state, there are only files with $p+1$ copies.

One starts with an elementary result concerning the $M/M/\infty$ queue.

Lemma 2.3. *If $(L_N(t))$ is the Markov process associated to an $M/M/\infty$ queue with arrival rate λN and service rate μ, and initial condition such that*

$$\lim_{N\to+\infty} \frac{L_N(0)}{N} = \frac{\lambda}{\mu},$$

then, for any $\ell \in \mathbb{N}$, the convergence in distribution

$$\lim_{N\to+\infty} \left(\frac{L_N(N^\ell t)}{N}\right) = \frac{\lambda}{\mu}$$

holds.

Proof. For $\varepsilon > 0$, by bounding the rate of jumps -1 of the process, a coupling can be constructed such that

$$L_N(t) \leq (\rho+\varepsilon)N + \bar{L}_N(Nt)$$

holds for all $t \geq 0$, where $(\bar{L}_N(t))$ is an $M/M/1$ queue with input rate λ and service rate $\lambda + \mu\varepsilon$, with initial condition $\bar{L}_N(0) = 0$. If $\tau_N = \inf\{t \geq 0 : \bar{L}_N(t) \geq \varepsilon N\}$ then, Proposition 5.11 page 119 of Robert [138], gives that for any $\ell \geq 1$ and $x > 0$,

$$\lim_{N\to+\infty} \mathbb{P}(\tau_N \leq N^\ell x) = 0.$$

This proves that, for any $T > 0$,

$$\lim_{N\to+\infty} \mathbb{P}\left(\sup_{0\leq t\leq T} \frac{L_N(N^\ell t)}{N} \leq \rho + 2\varepsilon\right) = 1.$$

With a similar argument for a lower bound one gets finally the convergence in distribution, for any $\ell \geq 0$,

$$\lim_{N\to+\infty} \frac{L_N(N^\ell t)}{N} = \rho.$$

The lemma is proved. □

One shows in the next proposition that the fluid state of the network does not change on the time scale $t \mapsto N^{p-2}t$.

Proposition 2.7 (Stability of Local Equilibrium on the time scale $t \mapsto N^{p-2}t$)**.** *If for some $2 \leq p < d$, one has $p\beta < \rho < (p+1)\beta$, and the initial state $X_N(0)$ is such that $X_i^N(0) = 0$ for $1 \leq i \leq d$, $i \notin \{p, p+1\}$ and*

$$\lim_{N\to+\infty} \left(\frac{X_p^N(0)}{N}, \frac{X_{p+1}^N(0)}{N}\right) = ((p+1)\beta - \rho, \rho - p\beta)$$

then for any $q \leq p-2$, for the convergence in distribution,

$$\lim_{N\to+\infty} \left(\frac{X_p^N(N^q t)}{N}, \frac{X_{p+1}^N(N^q t)}{N}\right) = ((p+1)\beta - \rho, \rho - p\beta)$$

Proof. Clearly it is enough to show the proposition for $q = p - 2$. Let

$$Z^N(t) = \sum_{k=1}^{p-1} (p-k) X_k^N(t),$$

then, if $Z^N(t) = z$, there is a jump of size $+1$ for Z^N at rate

$$\mu \sum_{k=2}^{p} k X_k^N(t) \leq p\mu\beta N,$$

and of size -1 at rate λN if $z > 0$. In the same way as in in the proof of Proposition 2.2, one can construct a coupling, for which

$$Z^N(t) \leq L_0(Nt),$$

where $(L_0(t))$ is a stable $M/M/1$ queue with input rate $p\mu\beta$ and output rate λ. In particular, the convergence in distribution

$$\lim_{N\to+\infty} \left(\frac{X_i^N(N^{p-2}t)}{N} \right) = 0, \quad 1 \leq i \leq p-1, \tag{2.40}$$

holds.

Because of Relation $p\mu\beta < \lambda$, one can extend the results of Propositions 2.3 and 2.5 to get that, for $1 \leq k \leq p-2$,

$$\lim_{N\to\infty} \left(\int_0^{N^{p-2}t} \left[\frac{(k+1)!}{\rho^{k+1}} \frac{X_{k+1}^N(u)}{N^{k+1}} - \frac{k!}{\rho^k} \frac{X_k^N(u)}{N^k} \right] \mathrm{d}u \right) = 0$$

holds for the convergence in distribution. By summing up all these relations for $1 \leq k \leq p-2$, one gets

$$\lim_{N\to+\infty} \left(\frac{(p-1)!}{\rho^{p-1}} \frac{1}{N^{p-1}} \int_0^{N^{p-2}t} X_{p-1}^N(u)\, \mathrm{d}u - \frac{\mu}{\lambda} \int_0^{N^{p-2}t} \frac{X_1^N(u)}{N}\, \mathrm{d}u \right) = 0.$$

Relation (2.40) gives the convergence in distribution

$$\lim_{N\to+\infty} \left(\int_0^{N^{p-2}t} \frac{X_1^N(u)}{N}\, \mathrm{d}u \right) = 0,$$

consequently

$$\lim_{N\to+\infty} \left(\frac{X_0^N(N^{p-2}t)}{N} \right) = 0,$$

by using the SDE associated to $(X_0^N(t))$ as in the proof of Proposition 2.5.

One concludes that

$$\lim_{N\to+\infty} \left(\frac{1}{N} \sum_{k=p}^{d} X_k^N(N^{p-2}t) \right) = \beta \tag{2.41}$$

Let

$$Y^N(t) = \sum_{k=1}^{d} kX_k^N(t),$$

then, if $Y^N(t) = y$, there is a jump of size -1 for Y^N at rate μy, and of size $+1$ at rate λN if $X_1^N(t) + \cdots + (d-1)X_{d-1}^N(t) > 0$. Hence, in the same way as in the proof of Proposition 2.2, a coupling can be constructed such that the process $(Y^N(t))$ is dominated by the process $(L_N(t))$ of the number of customers in an $M/M/\infty$ queue with arrival rate λN and service rate λ, and with initial condition such that

$$\lim_{N\to+\infty} \frac{L_N(0)}{N} = p((p+1)\beta - \rho) + (p+1)(\rho - p\beta) = \rho.$$

By using the relation

$$\sum_{k=1}^{p-1} kX_k^N(t) + pX_p^N(t) + (p+1)\left(\sum_{k=p}^{d} X_k^N(t) - X_p^N(t)\right) \leq Y^N(t) \leq L_N(t),$$

Equations (2.40) (2.41) and the above lemma, one gets that, for any $\varepsilon > 0$ and $T > 0$,

$$\lim_{N\to+\infty} \mathbb{P}\left(\inf_{0\leq t\leq T} \frac{X_p^N(N^{p-2}t)}{N} \geq (p+1)\beta - \rho - \varepsilon\right) = 1.$$

Relation $\lambda < (p+1)\beta\mu$, gives that $(X_p^N(N^{p-2}t))$ is strictly positive on any finite interval with high probability. Consequently,

$$\lim_{N\to+\infty} \mathbb{P}\left(\inf_{0\leq t\leq T} X_1^N(N^{p-2}t) + \cdots + (d-1)X_{d-1}^N(N^{p-2}t) > 1\right) = 1$$

this implies that the two processes $(Y^N(N^{p-2}t))$ and $(L_N(N^{p-2}t))$ are identical with probability close to 1 when N is large. Secondly, since the duplication capacity cannot be used at any node with index greater than $p+1$, for any $p+2 \leq k \leq d$, for the convergence in distribution, the relation

$$\lim_{N\to+\infty} \left(\frac{X_k^N(N^{p-2}t)}{N}\right) = 0$$

holds. One deduces therefore the convergence in distribution

$$\lim_{N\to+\infty} \left(\frac{pX_p^N(N^{p-2}t) + (p+1)X_{p+1}^N(N^{p-2}t)}{N}\right) = \rho$$
$$\lim_{N\to+\infty} \left(\frac{X_p^N(N^{p-2}t) + X_{p+1}^N(N^{p-2}t)}{N}\right) = \beta.$$

The proposition is proved. □

We can now state the main result of this section.

Theorem 2.4 (Evolution of the Local Equilibrium). *If for some $2 \leq p < d$, one has $p\beta < \rho < (p+1)\beta$, and the initial state $X_N(0)$ is such that $X_i^N(0) = 0$ for $1 \leq i \leq d$, $i \notin \{p, p+1\}$ and*

$$\lim_{N\to+\infty} \left(\frac{X_p^N(0)}{N}, \frac{X_{p+1}^N(0)}{N} \right) = ((p+1)\beta - \rho, \rho - p\beta)$$

then, for the convergence in distribution,

$$\lim_{N\to+\infty} \left(\frac{X_0^N(N^{p-1}t)}{N}, \frac{X_p^N(N^{p-1}t)}{N}, \frac{X_{p+1}^N(N^{p-1}t)}{N} \right) = \big(\Phi_0(t), \Phi_p(t), \Phi_{p+1}(t)\big)$$

where, for $t \geq 0$,

$$\Phi_p(t) = (p+1)(\beta - \Phi_0(t)) - \rho \text{ and } \Phi_{p+1}(t) = \rho - p(\beta - \Phi_0(t))$$

and $\Phi_0(t)$ is the unique solution y of the fixed point equation

$$\left(1 - \frac{y}{\beta - \rho/(p+1)}\right)^{\rho/(p(p+1))} e^y = \exp\left(-\lambda \frac{(p-1)!}{\rho^{p-1}} t\right). \tag{2.42}$$

In particular,

$$\lim_{t\to+\infty} \big(\Phi_0(t), \Phi_p(t), \Phi_{p+1}(t)\big) = \left(\beta - \frac{\rho}{p+1}, 0, \frac{\rho}{p+1}\right). \tag{2.43}$$

Remark. Relation (2.43) shows that a fraction $\beta - \rho/(p+1)$ of the files is lost asymptotically on the time scale $t \mapsto N^{p-1}t$. The corresponding asymptotic state consists then of files which are either lost and, at the first order in N, $\rho/(p+1) \cdot N$ files with $p+1$ copies. This suggests that β is changed to $\beta' = \rho/(p+1)$ and p replaced by $p' = p+1$. Unfortunately, this is the case of equality $\beta' = p'\rho$ which is not covered by our theorem. This suggests nevertheless the following evolution on the time scale $t \mapsto N^q t$, $p-1 \leq q \leq d-2$, for t going to infinity, there remain $\rho/(q+2)N$ files alive with $q+2$ copies. Some of the files are therefore lost and the number of copies of the remaining files is increasing, until the maximum number of copies is reached which is the framework of Section 2.4.

Proof. The proofs use the same arguments as in the proof of Theorem 2.2 and of the above proposition. We give a quick overview of it. By using again the results of Propositions 2.3 and 2.5 and Relation (2.25), one gets that, for $1 \leq k \leq p-2$,

$$\lim_{N\to\infty} \left(\int_0^{N^{p-1}t} \left[\frac{(k+1)!}{\rho^{k+1}} \frac{X_{k+1}^N(u)}{N^{k+1}} - \frac{k!}{\rho^k} \frac{X_k^N(u)}{N^k} \right] \mathrm{d}u \right) = 0$$

holds for the convergence in distribution. By summing up all these relations for $1 \leq k \leq p-2$, one gets

$$\lim_{N\to+\infty} \left(\frac{(p-1)!}{\rho^{p-1}} \frac{1}{N^{p-1}} \int_0^{N^{p-1}t} X_{p-1}^N(u)\, \mathrm{d}u - \frac{\mu}{\lambda} \int_0^{N^{p-1}t} \frac{X_1^N(u)}{N}\, \mathrm{d}u \right) = 0.$$

From there one gets that

$$\lim_{N\to+\infty}\left(\frac{X_0^N(u)}{N}-\frac{(p-1)!}{\rho^{p-1}}\frac{\lambda}{N^{p-1}}\int_0^{N^{p-1}t}X_{p-1}^N(u)\,\mathrm{d}u\right)=0.$$

As in the proof of Proposition 2.6, one can define a similar $(\Psi_f^N(t))$ and prove the same stochastic averaging property associated to the coordinate $(X_p^N(t))$. The rest of the proof is then similar to the proof of the last proposition with β replaced by $\beta-\phi(t)$ where $(\phi(t))$ is the limit of some converging subsequence of $(X_0^N(N^{p-1}t)/N)$. The convergence follows from the uniqueness of the fixed point equation satisfied by $(\phi(t))$. □

2.A Appendix: Generalized Skorohod Problems

For the sake of self-containedness, this section presents quickly the more or less classical material necessary to state and prove the convergence results used in this chapter. The general theme concerns the rigorous definition of a solution of a stochastic differential equation constrained to stay in some domain and also the proof of the existence and uniqueness and regularity properties of such a solution. See Skorohod [149], Anderson and Orey [9], Chaleyat-Maurel and El Karoui [55] and, in a multi-dimensional context, Harrison and Reiman [73] and Taylor and Williams [161] and, in a more general context, Ramanan [135]. See Appendix D of Robert [138] for a brief account.

We first recall the classical definition of Skorohod problem in dimension K. If $z=(z_k)\in\mathbb{R}^K$, one denotes $\|z\|=|z_1|+|z_2|+\cdots+|z_k|$. If $(Z(t))=(Z_k(t))$ is some function of the set $\mathcal{D}(\mathbb{R}_+,\mathbb{R}^K)$ of càdlàg functions defined on $\mathbb{R}_+$ and P is a $K\times K$ non-negative matrix, the couple of functions $[(X(t)),(R(t))]=[((X_k(t))),((R_k(t)))]$ is said to be a solution of the Skorohod problem associated to $(Z(t))$ and P whenever

1. $X(t)=Z(t)+(I-P)\cdot R(t)$, for all $t\geq 0$,
2. $X_k(t)\geq 0$, for all $t\geq 0$ and $1\leq k\leq d$,
3. For $1\leq k\leq K$, $t\to R_k(t)$ is non-decreasing, $R_k(0)=0$ and

$$\int_{\mathbb{R}_+}X_k(t)\,\mathrm{d}R_k(t)=0.$$

In the important case of dimension 1, Conditions (1) and (3) are

1. $X(t)=Z(t)+R(t)$, for all $t\geq 0$,
3. $t\to R(t)$ is non-decreasing, $R(0)=0$ and

$$\int_{\mathbb{R}_+}X(t)\,\mathrm{d}R(t)=0.$$

The generalization used in this chapter corresponds to the case when $(Z(t))$ depends on $(X(t))$.

Definition 2.1 (Generalized Skorohod Problem)**.**
If $G:\mathcal{D}(\mathbb{R}_+,\mathbb{R}^K)\to\mathcal{D}(\mathbb{R}_+,\mathbb{R}^K)$ is a Borelian function and P a non-negative $K\times K$ matrix,

$((X(t)),(R(t)))$ is a solution of the generalized Skorohod Problem (GSP) associated to G and P if $((X(t)),(R(t)))$ is the solution of the Skorohod Problem associated to $G(X)$ and P, in particular, for all $t \geq 0$,

$$X(t) = G(X)(t) + (I - P) \cdot R(t),$$

and

$$\int_{\mathbb{R}_+} X_k(t)\, \mathrm{d}R_k(t) = 0, \quad 1 \leq k \leq K.$$

The classical Skorohod problem described above corresponds to the case when the functional G is constant and equal to $(Z(t))$. In dimension one, if one takes

$$G(x)(t) = \int_0^t \sigma(x(u))\ \mathrm{d}B(u) + \int_0^t m(x(u))\ \mathrm{d}u,$$

where $(B(t))$ is a standard Brownian motion and σ and m are Lipschitz functions on $\mathbb{R}$. The first coordinate $(X(t))$ of a possible solution to the corresponding GSP can be described as the solution of the SDE

$$\mathrm{d}X(t) = \sigma(X(t))\ \mathrm{d}B(t) + m(X(t))\ \mathrm{d}t$$

reflected at 0.

Proposition 2.8. *If $G : \mathcal{D}(\mathbb{R}_+, \mathbb{R}^K) \to \mathcal{D}(\mathbb{R}_+, \mathbb{R}^K)$ is such that, for any $T > 0$, there exists a constant C_T such that, for all $(x(t)) \in \mathcal{D}(\mathbb{R}_+, \mathbb{R}^K)$ and $0 \leq t \leq T$,*

$$\sup_{0 \leq s \leq t} \|G(x)(s) - G(y)(s)\| \leq C_T \int_0^t \|x(u) - y(u)\|\ \mathrm{d}u \tag{2.44}$$

and if the matrix P is nilpotent, then there exists a unique solution to the generalized Skorohod problem associated to the functional G and the matrix P.

Proof. Define the sequence $(X_N(t))$ by induction $(X^0(t), R^0(t)) = 0$ and, for $N \geq 1$, (X^{N+1}, R^{N+1}) is the solution of the Skorohod problem (SP) associated to $G(X^N)$, in particular,

$$X^{N+1}(t) = G\left(X^N\right)(t) + (I - P) \cdot R^{N+1}(t) \text{ and } \int_{\mathbb{R}_+} X^{N+1}(u)\, \mathrm{d}R^{N+1}(u) = 0.$$

The existence of such a solution is a consequence of a result of Harrison and Reiman [73]. Fix $T > 0$. The Lipschitz property of the solutions of a classical Skorohod problem, see Proposition D.4 of Robert [138], gives the existence of some constant K_T such that, for all $N \geq 1$ and $0 \leq t \leq T$,

$$\left\|X^{N+1} - X^N\right\|_{\infty,t} \leq K_T \left\|G\left(X^N\right) - G\left(X^{N-1}\right)\right\|_{\infty,t},$$

where $\|h\|_{\infty,T} = \sup\{\|h(s)\| : 0 \leq s \leq T\}$. From Relation (2.44), this implies that

$$\left\|X^{N+1} - X^N\right\|_{\infty,t} \leq \alpha \int_0^t \left\|X^N - X^{N-1}\right\|_{\infty,u}\ \mathrm{d}u,$$

with $\alpha = K_T C_T$. The iteration of the last relation yields the inequality

$$\left\| X^{N+1} - X^N \right\|_{\infty,t} \leq \frac{(\alpha t)^N}{N!} \int_0^t \left\| X^1 \right\|_{\infty,u} \, \mathrm{d}u, \quad 0 \leq t \leq T.$$

One concludes that the sequence $(X^N(t))$ is converging uniformly on compact sets and consequently the same is true for the sequence $(R^N(t))$. Let $(X(t))$ and $(R(t))$ be the limit of these sequences. By continuity of the SP, the couple $((X(t)), (R(t)))$ is the solution of the SP associated to $G(X)$, and hence a solution of the GSP associated to G.

Uniqueness. If $(Y(t))$ is another solution of the GSF associated to G. In the same way as before, one gets by induction, for $0 \leq t \leq T$,

$$\|X - Y\|_{\infty,t} \leq \frac{(\alpha t)^N}{N!} \int_0^t \|X - Y\|_{\infty,u} \, \mathrm{d}u,$$

and by letting N go to infinity, one concludes that $X = Y$. The proposition is proved. $\square$

Chapter 3

Large Unreliable Stochastic Network: Local Duplication

Contents

3.1 Introduction **70**
3.1.1 Models with Independent Losses of Copies and Global Duplication Capacity 70
3.1.2 Stochastic Models with Local Duplication Features 71
3.1.3 Rate of Convergence to Equilibrium 73
3.2 The Stochastic Model **74**
3.3 An Asymptotic Process **77**
3.4 Mean-Field Limit **82**
3.4.1 Uniform Bound for $(\mathbf{R}_i^N(t))$ 82
3.4.2 Evolution Equations for the Empirical Distribution 84
3.4.3 The negligible terms 85
3.4.4 The Martingale 86
3.5 An Asymptotic Bound on the Decay of the Network **89**
3.6 The Case of Multiple Copies **92**
3.6.1 The Additional Complexity of the Model 92
3.6.2 Introduction of a Dominating Process 93

Abstract The problem of reliability of a large distributed system is analyzed via a new mathematical model. A typical framework is a system where a set of files are duplicated on several data servers. When one of these servers breaks down, all copies of files stored on it are lost. In this way, repeated failures may lead to losses of files. The efficiency of such a network is directly related to the performances of the mechanism used to duplicate files on servers. In this chapter we study the evolution of the network using a natural duplication policy giving priority to the files with the least number of copies.

We investigate the asymptotic behavior of the network when the number N of servers is large. The analysis is complicated by the large dimension of the state space of the empirical distribution of the state of the network. A stochastic model of the evolution of the network which has values in state space whose dimension does not depend on

N is introduced. Despite this description does not have the Markov property, it turns out that it is converging in distribution, when the number of nodes goes to infinity, to a nonlinear Markov process. The rate of decay of the network, which is the key characteristic of interest of these systems, can be expressed in terms of this asymptotic process. The corresponding mean-field convergence results are established. A lower bound on the exponential decay, with respect to time, of the fraction of the number of initial files with at least one copy is obtained.

3.1 Introduction

The problem of reliability of a large distributed system is analyzed in the present chapter via a new mathematical model. A typical framework is a system where files are duplicated on several data servers. When a server breaks down, all copies of files stored on this server are lost but they can be retrieved if copies of the same files are stored on other servers. In the case when no other copy of a given file is present in the network, it is definitively lost. Failures of disks occur naturally in this context, these events are quite rare but, given the large number of nodes of these large systems, this is not a negligible phenomenon at all at network scale. See the measurements at Google in Pinheiro et al. [130] for example.

In order to maintain copies on distant servers, a fraction of the bandwidth of each server has to be devoted to the duplication mechanism of its files to other servers. If, for a short period of time, several of the servers break down, it may happen that files will be lost for good just because all available copies were on these servers and that the recovery procedure was not completed before the last copy disappeared. The natural critical parameters of such a distributed system with N servers are the failure rate μ of servers, the bandwidth λ allocated to duplication of a given server, and the total number of initial files F_N. The quantity λ represents the amount of capacity that a server allocates to make duplication to enhance the durability of the network. If there are initially too many files in the system, the duplication capacity at each node may not be able to cope with the losses due to successive failures of servers and, therefore, a significant fraction of files will be lost very quickly. An efficient storage system should be able to maximize both the average number of files $\beta = F_N/N$ per server and the durability, i.e. the first instant $T^N(\delta)$ when a fraction $\delta \in (0,1)$ of files which are definitely lost.

3.1.1 Models with Independent Losses of Copies and Global Duplication Capacity

A large body of work in computer science has been devoted to the design and the implementation of duplication algorithms. These systems are known as distributed hash tables (DHT). They play an important role in the development of some large scale distributed systems and are quite popular right now like in cloud computing, see Rhea et al. [137] and Rowstron and Druschel [144] for example. Except extensive simulations, little has been done to evaluate the performances of these algorithms.

Several approaches have been used to investigate the corresponding mathematical models. Simplified models using birth and death processes have been used often, see Chun et al. [36], Picconi et al. [128] and Ramabhadran and Pasquale [134]. In Feuillet

and Robert [61] and Sun et al. [154], the authors studied how the durability $T(\delta)$ scales with the number of servers N and the maximum number of copies d of each file, under simplifying assumptions on file losses and the duplication mechanism. Firstly, in their model, each file copy is assumed to be lost at a certain fixed rate, independent of other copies. Secondly, they assumed that the duplication capacity can be used globally. This means that when each of N servers has an available bandwidth λ to duplicate files, the total capacity for duplication, λN, can be used to create a copy of any file from any server in the system to another. With these assumptions, the mathematical representation of the network is significantly simplified because it is not necessary to know the locations of copies of files to derive the dynamics of the system. In particular, in [154], a Markovian model with a fixed state space of dimension $d+1$ has been investigated: if for $0 \leq i \leq d$ and $t \geq 0$, $X_i^N(t)$ is the number of files with i copies, then the vector $X^N(t) = (X_0^N(t), X_1^N(t), \ldots, X_d^N(t))$ is a Markov process on $\mathbb{N}^{d+1}$ under the hypothesis that the global capacity λN is devoted to a file with the least number of copies. They have shown that the durability $T^N(\delta)$ is of the order N^{d-1} for a large N under certain conditions. Limit theorems were established for the rescaled process $(X^N(N^{d-1}t))$ using various technical estimations. In Sun et al. [156] the impact of placement policies, i.e. policies determining the location of the node to make a copy of a given file, is investigated.

3.1.2 Stochastic Models with Local Duplication Features

In this chapter, we consider a more realistic stochastic model for these systems, dropping the two main simplifying assumptions of previous works on copy losses and duplication capacity.

i. *Simultaneous losses due to server failure.* Each server can fail with a constant rate μ, and independent of other servers. When a server fails, all copies on that server are lost simultaneously, and therefore, the copy losses are not independent anymore. This dependency and bursty losses of file copies has a crucial effect on system performance.

ii. *Local duplication capacity.* The duplication capacity is assumed to be local, that is, each server has a capacity λ to duplicate the copies of files present on that server. In particular, this capacity cannot be used to copy files of other servers, as it is case for models with a global duplication capacity.

iii. *Duplication Policy: Priority to files with smallest number of copies.* The capacity of a server is allocated to duplicating one of its own files which has the smallest number of copies alive in the network. It is copied, uniformly at random, onto one of the servers which does not have such a copy.

Without a duplication mechanism, it is not difficult to see that the probability that a given file with d initial copies has still at least one copy at time t is $\mathcal{O}(d\exp(-\mu t))$ for a large t, when μ is the failure rate of servers. If, initially, there are $\lfloor \beta N \rfloor$ files, all with $d \geq 2$ copies scattered randomly in the network, the average fraction of files with at least a copy at time t is thus $\mathcal{O}(\beta d \exp(-\mu t))$. The central question is how much a duplication mechanism can improve these (poor) performances.

One cannot expect, intuitively, that the average lifetime of a file will grow significantly with N as in the case of a global duplication capacity, see Sun et al. [154] where the decay occurs only on the "fast" time scale $t \mapsto N^{d-1}t$. In contrast, as it will

be seen, the decay of our system occurs in fact on the "normal" time scale $t \mapsto t$. The main aim of this chapter is of investigating the exponential decay rate of the fraction of the number of files alive at time t with bounds of the form

$$Ce^{-\mu\kappa t}.$$

Of course duplication is of interest only if $\kappa<1$ and in fact is as close to 0 as possible. The goal of this chapter is to investigate the decay of the system described above via a mean-field approach. This is the key problem of these systems in our view.

With these assumptions, our mathematical model turns out to have stark differences compared to previous stochastic models. For a system of fixed size N, the exact dynamics of the system under above duplication mechanism is quite intricate, and hence, obtaining mathematical quantitative results to estimate the coefficient κ is quite challenging. A natural approach is of studying the performance of the system when the number of servers N goes to infinity.

To illustrate in a simpler setting the difficulties of these models, we first consider the case where there are at most two copies of each file stored on the system (d=2). In this case, a Markovian representation of the state of the system can be given by

$$(X^N(t)) = (X^N_{i,j}(t), i, j = 0, 1, ..., N), \tag{3.1}$$

where for $1 \leq i \neq j \leq N$, $X^N_{ij}(t)$ is the number of files which have copies on server i and j at time t, and $X^N_{i0}(t)=X^N_{0i}(t)$ is the number of files having only one copy located on server i. The state space of the state of a given node is therefore of dimension of the order of N which does not seem to be not amenable to analysis since the dimension of the basic state space is growing with N.

To overcome this difficulty, we introduce a reduced state representation in which each node i is described by only two variables: the number of files whose unique copy is on server i and the number of files with two copies and one of the copies is on i. The empirical distribution associated with such a representation has values in a state space of probability distributions on $\mathbb{N}^2$, two integer valued variables give the state of the node. Instead of a state of dimension N. This dimension reduction comes at a price, the loss of the Markov property. We prove nevertheless that this non-Markovian description of the network is converging in distribution, as N goes to infinity, to *a nonlinear Markov process*, $(\overline{R}(t)=(\overline{R}_1(t), \overline{R}_2(t)) \in \mathbb{N}^2$ satisfying the following Fokker-Planck Equations

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left(f(\overline{R}_1(t), \overline{R}_2(t))\right) &= \lambda\mathbb{E}\left(\left(f(\overline{R}(t)+e_2-e_1) - f(\overline{R}(t))\right)\mathbb{I}_{\{\overline{R}_1(t)>0\}}\right) \\ &\quad + \lambda\mathbb{P}(\overline{R}_1(t) > 0)\mathbb{E}\left(f(\overline{R}(t)+e_2) - f(\overline{R}(t))\right) \\ &\quad + \mu\mathbb{E}\left(f(0,0) - f(\overline{R}(t))\right) + \mu\mathbb{E}\left(\left(f(\overline{R}(t)+e_1-e_2) - f(\overline{R}(t))\right)\overline{R}_2(t)\right),\end{aligned} \tag{3.2}$$

with $e_1 = (1,0)$ and $e_2 = (0,1)$, and f is a function with finite support on $\mathbb{N}^2$. In this setting the asymptotic fraction of the number of files alive at time t is given by $\mathbb{E}(\overline{R}_1(t)) + \mathbb{E}(\overline{R}_2(t))/2$.

The asymptotic process $(\overline{R}(t) = (\overline{R}_1(t), \overline{R}_2(t))$ is a jump process with a type of jump, $x \mapsto x+e_2$ having time-dependent and distribution-dependent rate given by $\lambda\mathbb{P}(\overline{R}_1(t) > 0)$ which is the nonlinear term of this evolution equation.

3.1.3 Rate of Convergence to Equilibrium

It will be shown that the non-homogeneous Markov process defined by Relation (3.2) is converging to the unique distribution $\pi=\delta_{(0,0)}$, the Dirac measure at $(0,0)$, corresponding to a system with all files lost. The decay of the network is then closely related to the convergence rate to equilibrium of this Markov process.

In our case, as we have seen before, the problem is of finding a constant $\kappa>0$ for which the asymptotic fraction of the number of files alive at time t has an exponential decay with parameter $\mu\kappa$, i.e.

$$\mathbb{E}\left(\overline{R}_1(t)\right) + \frac{1}{2}\mathbb{E}\left(\overline{R}_2(t)\right) \leq \left(\overline{R}_1(0) + \frac{1}{2}\overline{R}_2(0)\right) e^{-\mu\kappa t}, \quad \forall t \geq 0. \tag{3.3}$$

The convergence rate can be defined in terms of the *Wasserstein distance* between the distribution P_t of the distribution at time t and the equilibrium distribution π,

$$\mathcal{W}_1(P_t, \pi) \overset{\text{def.}}{=} \inf\left\{\mathbb{E}(\mathrm{d}(X,Y)) : X \overset{\text{dist.}}{=} P_t, Y \overset{\text{dist.}}{=} \pi\right\},$$

where $\mathrm{d}(\cdot,\cdot)$ is some distance on the state space. One has to find the best possible constant α such that the relation

$$\mathcal{W}_1(P_t, \pi) \leq \mathcal{W}_1(P_0, \pi)e^{-\alpha t} \tag{3.4}$$

holds for all $t\geq 0$.

For *homogeneous*, i.e. "standard", Markov processes, this is already a difficult problem. For finite state spaces, tight estimates are known for some random walks, see Aldous and Diaconis [4] for example. When the state space is countable, results are more scarce. Lyapunov functions techniques to prove the existence of finite exponential moments of hitting times of finite sets can give a lower bound on the exponential decay α. This is, in general, a quite rough estimate for α, furthermore it does not give an estimate of the form (3.4). See Section 6.5 of Nummelin [120], see also Chapter 15 of Meyn and Tweedie [111].

In the continuous case, i.e. with Brownian motions instead of Poisson processes, some functional inequalities have been successfully used to obtain Relations of the form (3.4), see Markowich and Villani [108] and Desvillettes and Villani [46] for surveys on this topic. This approach for the case of the discrete state space turns out to be more difficult to use. Some generalizations have been proposed by Caputo et al. [30], Joulin [84] and Ollivier [121] for some jump processes. They can be used with some success, see Alfonsi et al. [6] and Thai [162] for example. For classical birth and death processes on $\mathbb{N}$, the assumptions for these results lead to some quite specific (and strong) conditions on the birth and death rates in order to have a positive exponential decay α.

For *non-linear* Markov processes, which is our case, the situation is, of course, much more complicated. Recall that, in this context, there may be several invariant distributions, so that convergence to equilibrium is more delicate notion. Note that this is not our case however. Ideas using the functional inequalities mentioned before have been also used but for specific stochastic models. See Carrillo et al. [31] and Cattiaux et al. [32] for a class of diffusions and Thai [162] for a class of birth and death processes. They do not seem of any help for the class of models we consider. To the best of

our knowledge, the results of optimality concerning the exponential decay α are quite scarce. Only lower bounds are provided generally. For the non-linear Markov processes associated to the mean-field results in this chapter, our approach will mainly use some monotonicity properties to derive lower bounds on the exponential decay.

In a first step, the present chapter develops a mathematical framework to get a convenient asymptotic description of our network (Equations (3.2)), and, secondly, to obtain explicit lower bounds on its exponential decay. This program is completed in the case $d=2$. In particular it is shown that Equation (3.3) holds some a constant $\kappa=\kappa_2^+$. Note that, however, as it can be expected in such a complicated context, we are not able to show that the constant κ_2^+ is optimal. As it will be seen, the case $d>2$ is more awkward in terms of an asymptotic picture but results on the exponential decay of the network can be nevertheless obtained by studying a non-linear Markov process dominating, in some sense, the original Markov process.

Outline of the Chapter

Section 3.2 introduces the main evolution equations of the state of the network. Section 3.3 investigates the existence and uniqueness properties of a nonlinear Markov process, the main result is Theorem 3.1. This process turns out to be the limit of a reduced description of the network. Section 3.4 shows the mean-field convergence of the reduced description of the network to this asymptotic process, this is Theorem 3.2. Section 3.5 studies the asymptotic behavior of the nonlinear Markov process. A lower bound on the exponential decay, with respect to time, of the fraction of the number of initial files still alive at time t is derived. These results are obtained in the case when the maximal number of copies for a given file is 2. Section 3.6 investigates the case of a general d. By using a simpler stochastic model, for which a mean-field limit result also holds, a multi-dimensional analogue of the set of equations (3.2) is introduced and analyzed. It gives a lower bound on the exponential decay of the number of files present in the network. It is expressed as the maximal eigenvalue of a $d\times d$ matrix. The proofs of the main results rely essentially on several ingredients: careful stochastic calculus with marked Poisson processes, several technical estimates, Lemmas 3.1 and 3.2, and mean-field techniques.

3.2 The Stochastic Model

In this section, we describe the dynamics of our system. Recall that the system has N servers, and until Section 3.6, it is assumed that each file has at most two copies in the system. Recall the Markovian representation $(X^N(t))$ defined in (3.1), that is,

$$X^N(t) = (X^N_{i,j}(t), 1 \leq i, j \leq N),$$

where, for $1\leq i\leq N$, $X^N_{i,0}(t)$ is the number of files whose unique copy is located in server i at time t, and $X^N_{i,j}(t)$ is the number of files with a copy on server i and on server j, $1\leq j\leq N$, $j\neq i$. Note the symmetry $X^N_{i,j}(t)=X^N_{j,i}(t)$, and by convention, $X^N_{i,i}(\cdot)\equiv 0$. We assume that all files have initially two copies and are randomly scattered on the network, as described below.

Assumptions 3.1. *(Initial State) For $1 \leq i \leq N$, there are A_i files on server i and each file $\ell=1,\ldots,A_i$ has another copy on server $V_{i,\ell}$, where*

— *$A_i, i=1,\ldots,N$, are i.i.d. square integrable random variables on $\mathbb{N}$,*

— *For each i, $V^N_{i,\ell}, \ell \geq 1$, are i.i.d. random variables with uniform distribution over $\{1,\ldots,N\}\backslash\{i\}$.*

Hence, $X^N_{i,0}(0)=0$ and

$$X^N_{i,j}(0) = \sum_{\ell=1}^{A_i} \mathbb{I}_{\{V^N_{i,\ell}=j\}} + \sum_{\ell=1}^{A_j} \mathbb{I}_{\{V^N_{j,\ell}=i\}}.$$

The total number of initial files is therefore $F_N \stackrel{\text{def.}}{=} A_1+A_2+\cdots+A_N$, and the initial average load of the system is

$$\beta \stackrel{\text{def.}}{=} \lim_{N\to+\infty} \frac{F_N}{N} = \mathbb{E}\left(A_1\right).$$

The initial mean number of copies of files per server is therefore 2β.

The initial state described in Assumption 3.1 have two main properties. First, it is exchangeable, in the sense that the distribution of $X^N(0)$ is invariant under permutations of server indices, and second, the two copies of each file are uniformly distributed over all servers. Alternatively, one can also assume that the total number of files F_N is a fixed number, without changing the results of this chapter.

Transitions of the State Representation

The transitions of the Markov process $(X^N(t))$ is governed by server failures and file duplications, as described below. Throughout this chapter, $f(t-)$ denotes the left limit of a function f at time t.

— *Server failure.* Each server i breaks down after an exponential time with parameter μ. At each breakdown, all copies on server i are lost, and the server restarts immediately but empty. It is in fact replaced by a new one. If a breakdown happens at time t,

$$\begin{cases} X^N_{i,j}(t) = 0, \quad \text{for all } j = 1,..,N, \\ X^N_{i,0}(t) = 0, \\ X^N_{j,0}(t) = X^N_{j,0}(t-)+X^N_{i,j}(t-), \quad \text{for all } j \neq i. \end{cases}$$

— *Duplication.* If there are files on server i with only one copy (i.e. $X^N_{i,0}(t)>0$), one of such files is copied at a rate λ onto another server $j \in \{1,\ldots,N\}\backslash\{i\}$ chosen uniformly at random. If the duplication is completed at time t,

$$\begin{cases} X^N_{i,0}(t) = X^N_{i,0}(t-)-1, \\ X^N_{i,j}(t) = X^N_{i,j}(t-)+1. \end{cases}$$

Note that $(X^N(t))$ is a transient Markov process converging to the state with all coordinates being 0 (all copies are lost).

Stochastic Evolution Equations

We can describe the dynamics of $(X^N(t))$ using stochastic integrals with respect to Poisson processes. Throughout the chapter, we use the following notations for Poisson processes.

— $\mathcal{N}_\xi$ is a Poisson point process on $\mathbb{R}_+$ with parameter $\xi > 0$ and $(\mathcal{N}_{\xi,i})$ is an i.i.d. sequence of such processes.

— $\overline{\mathcal{N}}_\xi$ is a Poisson point process on $\mathbb{R}_+^2$ with intensity $\xi\,\mathrm{d}t\,\mathrm{d}h$ and $(\overline{\mathcal{N}}_{\xi,i})$ is an i.i.d. sequence of such processes.

— For $1 \le j \le N$, the random variable $\mathcal{N}_{\lambda,j}^{U,N} = (t_n^j, U_n^j)$ is a marked Poisson process, (t_n^j) is a Poisson process on $\mathbb{R}_+$ with rate λ and (U_n^j) is an i.i.d. random variables with uniform distribution over $\{1,\dots,N\}\backslash\{j\}$. In particular, for $1 \le i \le N$, $(\mathcal{N}_{\lambda,j}^{U,N}(\cdot,\{i\}))$ is an i.i.d. sequence of Poisson processes with rate $\lambda/(N-1)$.

With a slight abuse of notation, we denote $\mathcal{N}_{\lambda,j} \stackrel{\text{def.}}{=} \mathcal{N}_{\lambda,j}^{U,N}(\cdot,\mathbb{N})$, which is a Poisson process with rate λ. See Kingman [96] and [25] for an introduction on ordinary and marked Poisson processes. All Poisson processes used are assumed to be independent.

For every $j=1,\dots,N$, failure times of server j are given by the epoch times of a Poisson process $\mathcal{N}_{\mu,j}$. A marked Poisson process $\mathcal{N}_{\lambda,j}^{U,N}$ captures duplications of files from server j as follows: for $n \ge 1$, at the nth event time t_n^j of $\mathcal{N}_{\lambda,j}^{U,N}$, if $X_{j,0}^N(t_n^j-) > 0$, a file on server j is copied onto the server whose index is given by the mark U_n^j.

The process $(X^N(t))$ can then be characterized as the solution of the following system of stochastic differential equations (SDEs): for $1 \le i, j \le N$, $j \ne i$ and $t \ge 0$,

$$\begin{aligned}\mathrm{d}X_{i,j}^N(t) = -X_{i,j}^N(t-)\left[\mathcal{N}_{\mu,i}(\mathrm{d}t) + \mathcal{N}_{\mu,j}(\mathrm{d}t)\right] \\ + \mathbb{I}_{\left\{X_{j,0}^N(t-)>0\right\}}\mathcal{N}_{\lambda,j}^{U,N}(\mathrm{d}t,\{i\}) + \mathbb{I}_{\left\{X_{i,0}^N(t-)>0\right\}}\mathcal{N}_{\lambda,i}^{U,N}(\mathrm{d}t,\{j\}),\end{aligned} \tag{3.5}$$

and

$$\mathrm{d}X_{i,0}^N(t) = -X_{i,0}^N(t-)\mathcal{N}_{\mu,i}(\mathrm{d}t) - \mathbb{I}_{\left\{X_{i,0}^N(t-)>0\right\}}\mathcal{N}_{\lambda,i}(\mathrm{d}t) + \sum_{j=1}^N X_{i,j}^N(t-)\mathcal{N}_{\mu,j}(\mathrm{d}t). \tag{3.6}$$

Classical results on Poisson processes show that the process

$$\left(\int_0^t \mathbb{I}_{\left\{X_{i,0}^N(s-)>0\right\}}\mathcal{N}_{\lambda,i}^{U,N}(\mathrm{d}s,\{j\}) - \frac{\lambda}{N-1}\int_0^t \mathbb{I}_{\left\{X_{i,0}^N(s)>0\right\}}\,\mathrm{d}s\right)$$

is a martingale whose previsible increasing process is

$$\left(\frac{\lambda}{N-1}\int_0^t \mathbb{I}_{\left\{X_{i,0}^N(s)>0\right\}}\,\mathrm{d}s\right).$$

See, for example, Section 4 of Chapter IV of Rogers and Williams [142]. Therefore, for $1 \le i \ne j \le N$,

$$\begin{aligned}X_{i,j}^N(t) = X_{i,j}^N(0) - 2\mu\int_0^t X_{i,j}^N(s)\,\mathrm{d}s \\ + \frac{\lambda}{N-1}\int_0^t \left(\mathbb{I}_{\left\{X_{i,0}^N(s)>0\right\}} + \mathbb{I}_{\left\{X_{j,0}^N(s)>0\right\}}\right)\mathrm{d}s + \mathcal{M}_{i,j}^N(t),\end{aligned} \tag{3.7}$$

and

$$X_{i,0}^N(t) = X_{i,0}^N(0) - \mu \int_0^t X_{i,0}^N(s)\ \mathrm{d}s - \lambda \int_0^t \mathbb{I}_{\{X_{i,0}^N(s)>0\}}\ \mathrm{d}s + \mu \sum_{j=1}^N \int_0^t X_{i,j}^N(s)\ \mathrm{d}s + \mathcal{M}_{i,0}^N(t), \quad (3.8)$$

where $(\mathcal{M}_{i,0}^N(t))$, $1 \leq i \leq N$, and $(\mathcal{M}_{i,j}^N(t))$, $1 \leq i < j \leq N$, are local martingales with the respective previsible increasing processes:

$$\left\langle \mathcal{M}_{i,j}^N \right\rangle (t) = 2\mu \int_0^t \left(X_{i,j}^N(s)\right)^2 \mathrm{d}s + \frac{\lambda}{N-1} \int_0^t \mathbb{I}_{\{X_{i,0}^N(s)>0\}}\ \mathrm{d}s + \frac{\lambda}{N-1} \int_0^t \mathbb{I}_{\{X_{j,0}^N(s)>0\}}\ \mathrm{d}s,$$

and

$$\left\langle \mathcal{M}_{i,0}^N \right\rangle (t) = \mu \int_0^t \left(X_{i,0}^N(s)\right)^2 \mathrm{d}s + \frac{\lambda}{N-1} \int_0^t \mathbb{I}_{\{X_{i,0}^N(s)>0\}}\ \mathrm{d}s + \mu \int_0^t \sum_{j=1}^N \left(X_{i,j}^N(s)\right)^2 \mathrm{d}s.$$

3.3 An Asymptotic Process

As mentioned in the introduction, for $1 \leq i \leq N$, the state of each server i at time t can alternatively be described by the pair

$$R_i^N(t) = (R_{i,1}^N(t), R_{i,2}^N(t)), \quad (3.9)$$

where $R_{i,1}^N(t)$ (resp. $R_{i,2}^N(t)$) is the number of files with one copy (resp. two copies) at node i. This reduced representation can be obtained from the full Markovian representation $(X^N(t))$ via

$$R_{i,1}^N(t) = X_{i,0}^N(t) \quad \text{and} \quad R_{i,2}^N(t) = \sum_{j=1}^N X_{i,j}^N(t).$$

Therefore, the evolution equations of $(R_i^N(t))$ can be deduced from the SDEs (3.5) and (3.6):

$$\mathrm{d}R_{i,1}^N(t) = -R_{i,1}^N(t-)\mathcal{N}_{\mu,i}(\mathrm{d}t) - \mathbb{I}_{\{R_{i,1}^N(t-)>0\}}\mathcal{N}_{\lambda,i}(\mathrm{d}t) + \sum_{j\neq i} X_{i,j}^N(t-)\mathcal{N}_{\mu,j}(\mathrm{d}t) \quad (3.10)$$

$$\mathrm{d}R_{i,2}^N(t) = -R_{i,2}^N(t-)\mathcal{N}_{\mu,i}(\mathrm{d}t) + \mathbb{I}_{\{R_{i,1}^N(t-)>0\}}\mathcal{N}_{\lambda,i}(\mathrm{d}t) - \sum_{j\neq i} X_{i,j}^N(t-)\mathcal{N}_{\mu,j}(\mathrm{d}t) + \sum_{j\neq i} \mathbb{I}_{\{R_{j,1}^N(t-)>0\}}\mathcal{N}_{\lambda,j}^{U,N}(\mathrm{d}t, \{i\}). \quad (3.11)$$

The process $(R^N(t)) = (R_i^N(t), 1 \leq i \leq N)$ lives on a state space of dimension $2N$ instead of N^2. The process $(R^N(t))$ still captures the information on the decay of the system since, for example, the total number of files which are still available in the network at time t can be expressed as

$$\sum_{i=1}^N R_{i,1}^N(t) + \frac{1}{2} R_{i,2}^N(t).$$

This dimension reduction comes at the price of the loss of the Markov property. The evolution equations of $(R_i^N(t))$ are not autonomous, they depend on the process $(X^N(t))$, and, consequently, the process $(R^N(t))$ does not have the Markov property. However, as it will be seen, the limit in distribution of $(R_{i,1}^N(t), R_{i,2}^N(t))$ turns out to be a nonlinear Markov process, or a so-called McKean-Vlasov process. See e.g. Sznitman [158]. In this section we characterize this limiting process, while the proof of convergence as N goes to infinity is given in the next section.

An Intuitive Introduction of the Asymptotic Process

The purpose of this section is only of motivating the asymptotic process; rigorous arguments to establish the convergence results are given later. Fix some $1 \leq i \leq N$ and assume for the moment that $(R_{i,1}^N(t), R_{i,2}^N(t))$ is converging in distribution to a process $(\overline{R}_1(t), \overline{R}_2(t))$. Define the positive random measure

$$\mathcal{P}_i^N([0,t]) \stackrel{\text{def.}}{=} \int_0^t \sum_{j \neq i} X_{i,j}^N(s-) \mathcal{N}_{\mu,j}(\mathrm{d}s).$$

It will be shown later in Lemma 3.1 that for a fixed $1 \leq i \leq N$, with high probability when N is large, all the variables $(X_{i,j}^N(t), 1 \leq j \leq N)$ are either 0 or 1 on a fixed time interval. In particular, $\mathcal{P}_i^N$ is asymptotically a counting process, i.e. an increasing process with jumps of size 1, with compensator given by

$$\mu \int_0^t \sum_{j \neq i} X_{i,j}^N(s) \, \mathrm{d}s = \mu \int_0^t R_{i,2}^N(s) \, \mathrm{d}s.$$

See Jacod [80] or Kasahara and Watanabe [88] for example. The convergence in distribution of the process $(R_{i,2}^N(s))$ to $(\overline{R}_2(s))$ and standard results on convergence of point processes give that $\mathcal{P}_i^N$ converges to $\mathcal{P}^\infty$, an inhomogeneous Poisson process with intensity $(\overline{R}_2(t))$ which can be represented as

$$\mathcal{P}^\infty(\mathrm{d}t) = \int_{\mathbb{R}_+} \mathbb{I}_{\{0 \leq h \leq \overline{R}_2(t-)\}} \overline{\mathcal{N}}_\mu(\mathrm{d}t, \mathrm{d}h).$$

See e.g. Kasahara and Watanabe [88] and Brown [26]. Recall that $\overline{\mathcal{N}}_\mu$ is a Poisson process on $\mathbb{R}_+^2$ with intensity $\mu \, \mathrm{d}t \, \mathrm{d}h$ (see Section 3.2). By formally taking the limit on both sides of Equation (3.10) as N gets large, this yields that the process $(\overline{R}_1(t), \overline{R}_2(t))$ satisfies the relation

$$\mathrm{d}\overline{R}_1(t) = -\overline{R}_1(t-) \mathcal{N}_\mu(\mathrm{d}t) - \mathbb{I}_{\{\overline{R}_1(t-)>0\}} \mathcal{N}_\lambda(\mathrm{d}t) + \int_{\mathbb{R}_+} \mathbb{I}_{\{0 \leq h \leq \overline{R}_2(t-)\}} \overline{\mathcal{N}}_\mu(\mathrm{d}t, \mathrm{d}h). \quad (3.12)$$

A similar work can be done with Equation (3.11). Consider the counting measure

$$\mathcal{Q}_i^N([0,t]) \stackrel{\text{def.}}{=} \int_0^t \sum_{j \neq i} \mathbb{I}_{\{R_{j,1}^N(s-)>0\}} \mathcal{N}_{\lambda,j}^{U,N}(\mathrm{d}s, \{i\}),$$

which has the compensator, see Jacod [80],

$$\lambda \int_0^t \frac{1}{N-1} \sum_{j \neq i} \mathbb{I}_{\{R_{j,1}^N(s)>0\}} \, \mathrm{d}s.$$

Again formally, it follows from the asymptotic independence of different servers and the law of large numbers limit for the processes $(R^N_{j,1}(t))$ that

$$\lim_{N\to+\infty}\left(\frac{1}{N-1}\sum_{j\neq i}\mathbb{I}_{\left\{R^N_{j,1}(t)>0\right\}}\right)=\left(\mathbb{P}\left(\overline{R}_1(t)>0\right)\right),$$

and therefore, $\mathcal{Q}^N_i$ converges in distribution to an inhomogeneous Poisson process with intensity $(\mathbb{P}(\overline{R}_1(s)>0))$. Therefore, taking limit from both sides of Equation (3.11) as N gets large, one obtains that the process $(\overline{R}_2(t))$ satisfies

$$\begin{aligned}\mathrm{d}\overline{R}_2(t)=&-\overline{R}_2(t-)\mathcal{N}_\mu(\mathrm{d}t)+\mathbb{I}_{\left\{\overline{R}_1(t-)>0\right\}}\mathcal{N}_\lambda(\mathrm{d}t)\\&-\int_{\mathbb{R}_+}\mathbb{I}_{\left\{0\le h\le\overline{R}_2(t-)\right\}}\overline{\mathcal{N}}_\mu(\mathrm{d}t,\mathrm{d}h)+\int_{\mathbb{R}_+}\mathbb{I}_{\left\{0\le h\le\mathbb{P}(\overline{R}_1(t)>0)\right\}}\overline{\mathcal{N}}_\lambda(\mathrm{d}t,\mathrm{d}h).\end{aligned}\tag{3.13}$$

The first result establishes the existence and uniqueness of a stochastic process satisfying the SDEs (3.12) and (3.13). For $T>0$, let $\mathcal{D}_T\overset{\text{def.}}{=}\mathcal{D}([0,T],\mathbb{N}^2)$ be the set of càdlàg functions from $[0,T]$ to $\mathbb{N}^2$ and $d_T(\cdot,\cdot)$ denotes the distance associated with the Skorohod topology on $\mathcal{D}_T$ and under which $\mathcal{D}_T$ is complete; see Page 125, Section 12 Chapter 3 of Billingsley [19].

Theorem 3.1. *(McKean-Vlasov Process) For every $(x,y)\in\mathbb{N}^2$, the equations*

$$\begin{cases}\overline{R}_1(t)=x-\displaystyle\int_0^t\overline{R}_1(s-)\mathcal{N}_\mu(\mathrm{d}s)-\int_0^t\mathbb{I}_{\left\{\overline{R}_1(s-)>0\right\}}\mathcal{N}_\lambda(\mathrm{d}s)\\\qquad\qquad+\displaystyle\iint_{[0,t]\times\mathbb{R}_+}\mathbb{I}_{\left\{0\le h\le\overline{R}_2(s-)\right\}}\overline{\mathcal{N}}_\mu(\mathrm{d}s,\mathrm{d}h),\\\overline{R}_2(t)=y-\displaystyle\int_0^t\overline{R}_2(s-)\mathcal{N}_\mu(\mathrm{d}s)-\iint_{[0,t]\times\mathbb{R}_+}\mathbb{I}_{\left\{0\le h\le\overline{R}_2(s-)\right\}}\overline{\mathcal{N}}_\mu(\mathrm{d}s,\mathrm{d}h)\\+\displaystyle\int_0^t\mathbb{I}_{\left\{\overline{R}_1(s-)>0\right\}}\mathcal{N}_\lambda(\mathrm{d}s)+\iint_{[0,t]\times\mathbb{R}_+}\mathbb{I}_{\left\{0\le h\le\mathbb{P}(\overline{R}_1(s)>0)\right\}}\overline{\mathcal{N}}_\lambda(\mathrm{d}s,\mathrm{d}h).\end{cases}\tag{3.14}$$

have a unique solution $(\overline{R}_1(t),\overline{R}_2(t))$ in $\mathcal{D}_T$.

The set of probability distributions on $\mathcal{D}_T$ is denoted as $\mathcal{P}(\mathcal{D}_T)$. Theorem 3.1 states that there exists a unique $\pi\overset{\text{dist.}}{=}(\overline{R}_1(t),\overline{R}_2(t))$ in $\mathcal{P}(\mathcal{D}_T)$ which satisfies Equation (3.14). See Rogers and Williams [142] for definitions of existence and uniqueness of a solution. Note that the solution to Equation (3.14) solves the Fokker-Planck Equation (3.2) of the introduction.

Proof. Define the uniform norm $\|\cdot\|_{\infty,T}$ on $\mathcal{D}_T$, if $f=(f_1,f_2)\in\mathcal{D}_T$,

$$\|f\|_{\infty,T}=\sup\{\|f(t)\|:0\le t\le T\}=\sup\{|f_1(t)|+|f_2(t)|:0\le t\le T\}.$$

One can introduce the *Wasserstein metrics* on $\mathcal{P}(\mathcal{D}_T)$ as follows, for $\pi_1,\pi_2\in\mathcal{P}(\mathcal{D}_T)$

$$W_T(\pi_1,\pi_2)=\inf_{\pi\in\mathcal{C}_T(\pi_1,\pi_2)}\int_{\omega=(\omega_1,\omega_2)\in\mathcal{D}_T^2}\left[d_T(\omega_1,\omega_2)\wedge1\right]\,\mathrm{d}\pi(\omega),\tag{3.15}$$

$$\rho_T(\pi_1,\pi_2)=\inf_{\pi\in\mathcal{C}_T(\pi_1,\pi_2)}\int_{\omega=(\omega_1,\omega_2)\in\mathcal{D}_T^2}\left[\|\omega_1-\omega_2\|_{\infty,T}\wedge1\right]\,\mathrm{d}\pi(\omega),\tag{3.16}$$

where $a \wedge b = \min\{a, b\}$ for a, $b \in \mathbb{R}$ and $\mathcal{C}_T(\pi_1, \pi_2)$ is the subset of couplings of π_1 and π_2, i.e. the subset of $\mathcal{P}(\mathcal{D}_T \times \mathcal{D}_T)$ whose first (resp. second) marginal is π_1 (resp. π_2). Since $(\mathcal{D}_T, d_T)$ is separable and complete, the space $(\mathcal{P}(\mathcal{D}_T), W_T)$ is complete, which gives the topology of convergence in distribution on $\mathcal{P}(\mathcal{D}_T)$. Clearly, for any $\pi_1, \pi_2 \in \mathcal{P}(\mathcal{D}_T)$, one has the relation $W_T(\pi_1, \pi_2) \leq \rho_T(\pi_1, \pi_2)$.

Let $\Psi : (\mathcal{P}(\mathcal{D}_T), W_T) \to (\mathcal{P}(\mathcal{D}_T), W_T)$ be the mapping that takes π to the distribution $\Psi(\pi)$ of R_π, where $(R_\pi(t)) = (R_{\pi,1}(t), R_{\pi,2}(t))$ is the unique solution to the SDEs

$$R_{\pi,1}(t) = x - \int_0^t R_{\pi,1}(s-)\mathcal{N}_\mu(\mathrm{d}s) - \int_0^t \mathbb{I}_{\{R_{\pi,1}(s-)>0\}}\mathcal{N}_\lambda(\mathrm{d}s) + \iint_{[0,t]\times\mathbb{R}_+} \mathbb{I}_{\{0\leq h\leq R_{\pi,2}(s-)\}}\overline{\mathcal{N}}_\mu(\mathrm{d}s, \mathrm{d}h),$$

$$R_{\pi,2}(t) = y - \int_0^t R_{\pi,2}(s-)\mathcal{N}_\mu(\mathrm{d}s) - \iint_{[0,t]\times\mathbb{R}_+} \mathbb{I}_{\{0\leq h\leq R_{\pi,2}(s-)\}}\overline{\mathcal{N}}_\mu(\mathrm{d}s, \mathrm{d}h) + \int_0^t \mathbb{I}_{\{R_{\pi,1}(s-)>0\}}\mathcal{N}_\lambda(\mathrm{d}s) + \iint_{[0,t]\times\mathbb{R}_+} \mathbb{I}_{\{0\leq h\leq \pi(r_1(s)>0)\}}\overline{\mathcal{N}}_\lambda(\mathrm{d}s, \mathrm{d}h),$$

with initial condition $(R_{\pi,1}(0), R_{\pi,2}(0)) = (x, y)$. Note that

$$\pi(r_1(t) > 0) = \int_{\omega=(r_1,r_2)\in\mathcal{D}_T} \mathbb{I}_{\{r_1(t)>0\}}\, \mathrm{d}\pi(\omega).$$

The existence and uniqueness of a solution to Equations (3.14) is equivalent to the existence and uniqueness of a fixed point $\pi = \Psi(\pi)$.

For any $\pi_a, \pi_b \in \mathcal{P}(\mathcal{D}_T)$ then, let R_{π_a} and R_{π_b} both be solutions to the equations of the display above driven by same Poisson processes. Therefore, the distribution of the pair $(R_{\pi_a}(t), R_{\pi_b}(t))$ is a coupling of $\Psi(\pi_a)$ and $\Psi(\pi_b)$, and hence,

$$\rho_t(\Psi(\pi_a), \Psi(\pi_b)) \leq \mathbb{E}\left(\|R_{\pi_a} - R_{\pi_b}\|_{\infty,t}\right). \tag{3.17}$$

For $t \leq T$, using the definition of R_{π_a} and R_{π_b},

$$\begin{aligned}
\|R_{\pi_a} - R_{\pi_b}\|_{\infty,t} &= \sup_{s\leq t}\left(|R_{\pi_a,1}(s) - R_{\pi_b,1}(s)| + |R_{\pi_a,2}(s) - R_{\pi_b,2}(s)|\right) \\
&\leq \int_0^t \left(|R_{\pi_a,1}(s-) - R_{\pi_b,1}(s-)| + |R_{\pi_a,2}(s-) - R_{\pi_b,2}(s-)|\right)\mathcal{N}_\mu(\mathrm{d}s) \\
&\quad + 2\int_0^t \left|\mathbb{I}_{\{R_{\pi_a,1}(s-)>0\}} - \mathbb{I}_{\{R_{\pi_b,1}(s-)>0\}}\right|\mathcal{N}_\lambda(\mathrm{d}s) \\
&\quad + 2\int_0^t\int_0^\infty \mathbb{I}_{\{R_{\pi_a,2}(s-)\wedge R_{\pi_b,2}(s-)\leq h\leq R_{\pi_a,2}(s-)\vee R_{\pi_b,2}(s-)\}}\overline{\mathcal{N}}_\mu(\mathrm{d}s, \mathrm{d}h) \\
&\quad + \int_0^t\int_0^\infty \mathbb{I}_{\{\pi_a(r_1(s)>0)\wedge\pi_b(r_1(s)>0)\leq h\leq \pi_a(r_1(s)>0)\vee\pi_b(r_1(s)>0)\}}\overline{\mathcal{N}}_\lambda(\mathrm{d}s, \mathrm{d}h).
\end{aligned} \tag{3.18}$$

We bound the expected value of each of the terms of the right-hand side above. First, for $\ell = 1, 2$,

$$\begin{aligned}
\mathbb{E}\left(\int_0^t \left|R_{\pi_a,\ell}(s-) - R_{\pi_b,\ell}(s-)\right|\mathcal{N}_\mu(\mathrm{d}s)\right) &= \mu\mathbb{E}\left(\int_0^t \left|R_{\pi_a,\ell}(s) - R_{\pi_b,\ell}(s)\right|\,\mathrm{d}s\right) \\
&\leq \mu\int_0^t \mathbb{E}\left(\|R_{\pi_a} - R_{\pi_b}\|_{\infty,s}\right)\,\mathrm{d}s.
\end{aligned} \tag{a}$$

For the second term on the right-hand side of (3.18), since $R_{\pi_a,1}(s)$ and $R_{\pi_b,1}(s)$ are integer valued,

$$\left|\mathbb{I}_{\{R_{\pi_a,1}(s)>0\}} - \mathbb{I}_{\{R_{\pi_b,1}(s)>0\}}\right| \leq |R_{\pi_a,1}(s) - R_{\pi_b,1}(s)|,$$

and hence, using (a), we have the bound

$$\mathbb{E}\left(\int_0^t \left|\mathbb{I}_{\{R_{\pi_a,1}(s-)>0\}} - \mathbb{I}_{\{R_{\pi_b,1}(s-)>0\}}\right| \mathcal{N}_\lambda(\mathrm{d}s)\right) \quad \leq \quad \mu \int_0^t \mathbb{E}\left(\|R_{\pi_a} - R_{\pi_b}\|_{\infty,s}\right) \mathrm{d}s. \quad \text{(b)}$$

Similarly, for the third term on the right-hand side of (3.18), we have

$$\begin{aligned}\mathbb{E}\left(\int_0^t \int_0^\infty \mathbb{I}_{\{R_{\pi_a,2}(s-)\wedge R_{\pi_b,2}(s-)\leq h\leq R_{\pi_a,2}(s-)\vee R_{\pi_b,2}(s-)\}} \overline{\mathcal{N}}_\mu(\mathrm{d}s,\mathrm{d}h)\right)\\ = \mu \int_0^t \mathbb{E}\left(|R_{\pi_a,2}(s) - R_{\pi_b,2}(s)|\right) \mathrm{d}s \leq \mu \int_0^t \mathbb{E}\left(\|R_{\pi_a} - R_{\pi_b}\|_{\infty,s}\right) \mathrm{d}s. \quad \text{(c)}\end{aligned}$$

Finally, for the last term on the right-hand side of (3.18),

$$\begin{aligned}\mathbb{E}\left(\int_0^t \int_0^\infty \mathbb{I}_{\{\pi_a(r_1(s)>0)\wedge\pi_b(r_1(s)>0)\leq h\leq \pi_a(r_1(s)>0)\vee\pi_b(r_1(s)>0)\}} \overline{\mathcal{N}}_\lambda(\mathrm{d}s,\mathrm{d}h)\right)\\ = \lambda \int_0^t |\pi_a(r_1(s)>0) - \pi_b(r_1(s)>0)| \, \mathrm{d}s. \quad \text{(d)}\end{aligned}$$

Note that for every coupling $\pi \in \mathcal{C}_T(\pi_a, \pi_b)$ of π_a and π_b,

$$\begin{aligned}\int_0^t |\pi_a(r_1(s)>0) - \pi_b(r_1(s)>0) \, \mathrm{d}s \\ &= \int_0^t \left|\pi\left((r^a, r^b) : r_1^a(s)>0\right) - \pi\left((r^a, r^b) : r_1^b(s)>0\right)\right| \mathrm{d}s \\ &\leq \int_0^t \int_{\omega=(r^a,r^b)\in\mathcal{D}_T^2} |\mathbb{I}_{\{r_1^a(s)>0\}} - \mathbb{I}_{\{r_1^b(s)>0\}}| \pi(\mathrm{d}\omega) \, \mathrm{d}s \\ &\leq \int_0^t \int_{\omega=(r^a,r^b)\in\mathcal{D}_T^2} |r_1^a(s) - r_1^b(s)| \wedge 1 \, \pi(\mathrm{d}\omega) \, \mathrm{d}s.\end{aligned}$$

By taking the infimum among all the couplings of π_a and π_b, we have

$$\int_0^t |\pi_a(r_1(s)>0) - \pi_b(r_1(s)>0)| \, \mathrm{d}s \leq \int_0^t \rho_s(\pi_a, \pi_b) \, \mathrm{d}s. \quad \text{(e)}$$

Now, by combining the estimates (a), (b), (c), (d), (e), we conclude

$$\mathbb{E}\left((\|R_{\pi_a} - R_{\pi_b}\|_{\infty,t}\right) \quad \leq \quad (2\lambda + 3\mu) \int_0^t \mathbb{E}\left(\|R_{\pi_a} - R_{\pi_b}\|_{\infty,s}\right) \mathrm{d}s + \lambda \int_0^t \rho_s(\pi_a, \pi_b) \, \mathrm{d}s,$$

Grönwall's inequality then gives

$$\mathbb{E}\left((\|R_{\pi_a} - R_{\pi_b}\|_{\infty,t}\right) \leq C_T \int_0^t \rho_s(\pi_a, \pi_b) \, \mathrm{d}s, \qquad \forall t \in [0, T],$$

with $C_T = \lambda \exp(2\lambda + 3\mu)T$. Hence using (3.17), we have

$$\rho_t(\Psi(\pi_a), \Psi(\pi_b)) \leq C_T \int_0^t \rho_s(\pi_a, \pi_b)\, \mathrm{d}s, \qquad \forall t \in [0, T]. \tag{3.19}$$

Uniqueness of the fixed point for the equation $\Psi(\pi) = \pi$ follows immediately from (3.19). Also, a typical iterative argument proves the existence: pick any $\pi_0 \in \mathcal{P}(\mathcal{D}_T)$, and define the sequence (π_n) inductively by $\pi_{n+1} = \Psi(\pi_n)$. It follows from Relation (3.19) that

$$W_T(\pi_{n+1}, \pi_n) \leq \rho_T(\pi_{n+1}, \pi_n) \leq \frac{(TC_T)^n}{n!} \int_0^T \rho_s(\pi_1, \pi_0)\, \mathrm{d}s.$$

The metric space $(\mathcal{P}(\mathcal{D}_T), W_T)$ is complete, and therefore the sequence (π_n) converges. Since Ψ is continuous with respect to the Skorohod topology, its limit is necessarily a fixed point of Ψ. This completes the proof. □

3.4 Mean-Field Limit

The empirical distribution $\Lambda^N(t)$ of $(R_i^N(t), 1 \leq i \leq N)$ is defined by, for f a function on $\mathbb{N}^2$,

$$\Lambda^N(t)(f) = \frac{1}{N} \sum_{i=1}^N f(R_i^N(t)) = \frac{1}{N} \sum_{i=1}^N f\left((R_{i,1}^N(t), R_{i,2}^N(t))\right).$$

As it has already been remarked, at the beginning of Section 3.3, the process $(\Lambda^N(t))$ does not have the Markov property. The goal of this section is to prove that the stochastic process $(\Lambda^N(t))$ is converging in distribution as N goes to infinity, that is, for any function f with finite support, the sequence of stochastic processes $(\Lambda^N(t)(f))$ converges in distribution. See Billingsley [19] and Dawson [42].

The main result of this section is the following theorem.

Theorem 3.2. *(Mean-Field Convergence Theorem) Suppose the process $(X^N(t))$ is initialized according to Assumption 3.1. The sequence of empirical distribution process $(\Lambda^N(t))$ converges in distribution to a process $(\Lambda(t)) \in \mathcal{D}(\mathbb{R}_+, \mathcal{P}(\mathbb{N}^2))$ which is defined as follows: for f with finite support on $\mathbb{N}^2$,*

$$\Lambda(t)(f) \stackrel{def.}{=} \mathbb{E}\left(f\left(\overline{R}_1(t), \overline{R}_2(t)\right)\right),$$

where $(\overline{R}_1(t), \overline{R}_2(t))$ is the unique solution of Equations (3.14).

Moreover, for any $p \geq 1$, the sequence of finite marginals $(R_{i,1}^N(t), R_{i,2}^N(t), 1 \leq i \leq p)$ converges in distribution to $((\overline{R}_{i,1}(t), \overline{R}_{i,2}(t)), 1 \leq i \leq p)$, where $(\overline{R}_{i,1}(t), \overline{R}_{i,2}(t))$ are i.i.d. processes with the same distribution as $(\overline{R}_1(t), \overline{R}_2(t))$.

The last statement is the "propagation of chaos" property.

3.4.1 Uniform Bound for $(\mathbf{R_i^N(t)})$

We start with a technical result which will be used to establish mean-field convergence. It states that, uniformly on a compact time interval, the number of files with a copy at a given server i is stochastically bounded and that, with a high probability, all other servers have at most one file in common with server i. This is a key ingredient

to prove that the non-Markovian process $(R^N_{i,1}(t), (R^N_{i,2}(t))$ is converging in distribution to the nonlinear Markov process described in Theorem 3.1.

Lemma 3.1. *If the initial state of the process $(X^N(t))$ is given by Assumption 3.1, for $1 \leq i \leq N$ and $T > 0$ then, for $i \in \mathbb{N}$,*

$$\sup_{N \geq 1} \mathbb{E}\left(\sup_{0 \leq t \leq T}\left(R^N_{i,1}(t) + R^N_{i,2}(t)\right)^2\right) < +\infty \tag{3.20}$$

and if

$$\mathcal{E}_i{}^N(T) \stackrel{def.}{=} \left\{\sup_{0 \leq t \leq T, 1 \leq j \leq N} X^N_{i,j}(t) \geq 2\right\},$$

then there exists a constant $C(T)$ independent of i such that $\mathbb{P}\left(\mathcal{E}_i{}^N(T)\right) \leq C(T)/N$.

Proof. For $i = 1, ..., N$, the total number of files $D^N_{i,0}$ initially on server i satisfies

$$D^N_{i,0} \stackrel{\text{def.}}{=} R^N_{i,1}(0) + R^N_{i,2}(0) = X^N_{i,0}(0) + \sum_{j \neq i}^{N} X^N_{i,j}(0) = A_i + \sum_{j \neq i}^{N} \sum_{\ell=1}^{A_j} \mathbb{I}_{\left\{V^N_{j,\ell} = i\right\}}, \tag{3.21}$$

and hence, $\mathbb{E}\left(D^N_{i,0}\right) = 2\mathbb{E}\left(A_1\right)$ and $\text{var}(D^N_{i,0}) = 2\text{var}(A_1) + \mathbb{E}\left(A_1\right) N/(N-1)$. Also, the total number of files $D^N_{i,1}(t)$ copied on server i from all other servers during the interval $[0, t]$ verifies

$$D^N_{i,1}(t) \stackrel{\text{def.}}{=} \sum_{j \neq i}^{N} \int_0^t \mathbb{I}_{\left\{R^N_{j,0}(s) > 0\right\}} \mathcal{N}^{U,N}_{\lambda,j}(\mathrm{d}s, \{i\}) \leq \sum_{j \neq i}^{N} \mathcal{N}^{U,N}_{\lambda,j}(t, \{i\}). \tag{3.22}$$

Therefore, for every $t \leq T$, $\mathbb{E}\left(D^N_{i,1}(t)\right) \leq \lambda T$ and $\mathbb{E}\left(D^N_{i,1}(t)^2\right) \leq 2\lambda T$. The bound (3.20) then follows from the inequality

$$\sup_{0 \leq t \leq T}\left(R^N_{i,1}(t) + R^N_{i,2}(t)\right) \leq D^N_{i,0}(T) + D^N_{i,1}(T)$$

For the next part, note that on $\mathcal{E}_i{}^N(T)$, there exists $1 \leq j \leq N$ such that either server i or j makes two copies on the other one or both i and j make one copy on the other during the time interval $[0, T]$. Recall again that server i initially copies A_i files on other servers, and that the total number of files copied from server i onto server j during $(0, T]$ is upper bounded by $\mathcal{N}^{U,N}_{\lambda,i}(T, \{j\})$. Define the sequence $(Z^N_{i,\ell}, 1 \leq i \leq N, \ell \geq 1)$ as follows: $Z^N_{i,\ell} = V^N_{i,\ell}$ when $1 \leq \ell \leq A_i$, and $Z^N_{i,\ell} = U^i_{\ell - A_i}$ when $\ell > A_i$. For the first A_i indices ℓ, $Z^N_{i,\ell}$s are therefore the indices of servers which received an initial copy of a file of server i, while the subsequent $Z^N_{i,\ell}$s are the server indices on which (potential) duplications from server i can take place. $(Z^N_{i,\ell})$ is therefore a sequence of i.i.d. random variables uniformly

distributed on $\{1,\ldots,N\}\backslash\{i\}$. Therefore, $\mathbb{P}\left(\mathcal{E}_i^N(T)\right)\leq\mathbb{P}\left(\mathcal{B}_i^N\right)$, where

$$\mathcal{B}_i^N \stackrel{\text{def.}}{=} \bigcup_{j=1,j\neq i}^{N}\Bigg(\bigcup_{\substack{1\leq\ell\leq A_i+L_i(T)\\1\leq\ell'\leq A_j+L_j(T)}}\{Z_{i,\ell}^N=j, Z_{j,\ell'}^N=i\}$$
$$\bigcup_{1\leq\ell\neq\ell'\leq A_i+L_i(T)}\{Z_{i,\ell}^N=j, Z_{i,\ell'}^N=j\}\bigcup_{1\leq\ell\neq\ell'\leq A_j+L_j(T)}\left\{V_{j,\ell}^N=i, V_{j,\ell'}^N=i\right\}\Bigg),$$

with $L_i(T)=\mathcal{N}_{\lambda,i}([0,T])+A_i$. Since the probability of each of the elementary events of the right hand side of this relation is $1/(N-1)^2$, $Z_{k,\ell}$s are independent of $L_{k'}(T)$ for all k,k', and $\mathbb{E}\left(L_i(T)\right)=\lambda T+\mathbb{E}\left(A_1\right)$. It is then easy to conclude. □

3.4.2 Evolution Equations for the Empirical Distribution

Denote $e_1=(1,0)$ and $e_2=(0,1)$, and define the operators

$$\Delta^{\pm}(f)(x)=f(x+e_1-e_2)-f(x),\quad \Delta^{\mp}(f)(x)=f(x-e_1+e_2)-f(x),$$
$$\Delta_2^{+}(f)(x)=f(x+e_2)-f(x),$$

for $x\in\mathbb{N}^2$ and $f:\mathbb{N}^2\to\mathbb{R}_+$. For every function $f:\mathbb{N}^2\to\mathbb{R}_+$ with finite support, it follows from Equations (3.10) and (3.11) and using martingale decomposition for the Poisson processes, we have

$$\begin{aligned}\mathrm{d}f(R_i^N(t)) &= \mathrm{d}\mathcal{M}_{f,i}^N(t)+\Delta^{\mp}(f)(R_i^N(t))\mathbb{I}_{\left\{R_{i,1}^N(t)>0\right\}}\lambda\,\mathrm{d}t\\ &+\Delta_2^{+}(f)(R_i^N(t))\frac{\lambda}{N-1}\sum_{j\neq i}\mathbb{I}_{\left\{R_{j,1}^N(t)>0\right\}}\,\mathrm{d}t+\left[f(0,0)-f(R_i^N(t))\right]\mu\,\mathrm{d}t\\ &+\sum_{j\neq i}\left[f(R_i^N(t)+X_{i,j}^N(t)(e_1-e_2))-f(R_i^N(t))\right]\mu\,\mathrm{d}t,\end{aligned}\tag{3.23}$$

where $\mathcal{M}_{f,i}^N$ is a martingale. The jth term of the last sum on the right-hand side above corresponds to the event when server j breaks down and therefore the copies of $X_{i,j}^N(t)$ files at node j are lost, and the remaining copies are only located at node i. Using the notation of Lemma 3.1 then, outside the event $\mathcal{E}_i^N(T)$, $X_{i,j}^N(t)$ is either 0 or 1, and hence, $t\in[0,T]$,

$$\sum_{j\neq i}\left[f(R_i^N(t)+X_{i,j}^N(t)(e_1-e_2))-f(R_i^N(t))\right]=R_{2,i}^N(t)\Delta^{\pm}(f)(R_i^N(t)).$$

By summing up both sides of Relation (3.23) over i and denoting $\mathbb{N}^*=\mathbb{N}\backslash\{0\}$, we have

$$\begin{aligned}\Lambda^N(t)(f)=\Lambda^N(0)(f)+\mathcal{M}_f^N(t)+\lambda\int_0^t\int_{\mathbb{N}^2}\Delta^{\mp}(f)(x,y)\mathbb{I}_{\{x>0\}}\Lambda^N(s)(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}s\\ +\frac{\lambda N}{N-1}\int_0^t\Lambda^N(s)(\mathbb{N}^*\times\mathbb{N})\int_{\mathbb{N}^2}\Delta_2^+(f)(x,y)\Lambda^N(s)(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}s-H_1^N(t)\\ +\mu\int_0^t\int_{\mathbb{N}^2}(f(0,0)-f(x,y))\Lambda^N(s)(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}s\\ +\mu\int_0^t\int_{\mathbb{N}^2}y\Delta^{\pm}(f)(x,y)\,\Lambda^N(s)(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}s+H_2^N(t),\end{aligned}\tag{3.24}$$

where

$$\begin{aligned}\mathcal{M}_f^N(t)&=\frac{1}{N}\left(\mathcal{M}_{f,1}^N(t)+\mathcal{M}_{f,2}^N(t)+\cdots+\mathcal{M}_{f,N}^N(t)\right),\\ H_1^N(t)&=\frac{\lambda}{N-1}\int_0^t\int_{\mathbb{N}^2}\Delta_2^+(f)(x,y)\mathbb{I}_{\{x>0\}}\Lambda^N(s)(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}s,\\ H_2^N(t)&=\mu\frac{1}{N}\sum_{i=1}^N\int_0^t h_{2,i}^N(s)\,\mathrm{d}s,\end{aligned}$$

with

$$h_{2,i}^N(t)=\sum_{j\neq i}\Big(f(R_i^N(t)+X_{ij}^N(t)(e_1-e_2))-f(R_i^N(t))\Big)-\int_{\mathbb{N}^2}y\Delta^{\pm}(f)(x,y)\,\Lambda^N(t)(\mathrm{d}x,\mathrm{d}y).$$

Now, we investigate the asymptotic properties of the terms of the right hand side of Equation (3.24).

3.4.3 The negligible terms

We first prove that the two processes $(H_1^N(t))$ and $(H_2^N(t))$ converge to zero in distribution as N goes to infinity. For the former, the result follows immediately from the simple bound

$$\|H_1^N\|_{\infty,T}\leq\frac{2\lambda T}{N}\|f\|_\infty.$$

For $(H_2^N(t))$, first note that, for $0\leq t\leq T$ and $1\leq i\leq N$, $h_{2,i}^N(t)$ is non-zero only on the event $\mathcal{E}_i^N(T)$, and hence,

$$\begin{aligned}\left\|\frac{1}{N}\sum_{i=1}^N h_{2,i}^N(s)\right\|_{\infty,T}\leq\sup_{0\leq s\leq T}\frac{1}{N}\sum_{i=1}^N\Bigg|\sum_{j\neq i}\Big[f(R_i^N(s)+X_{ij}^N(s)(e_1-e_2))-f(R_i^N(s))\Big]\\ -R_{2,i}^N(s)\Delta^{\pm}(f)(R_i^N(s))\Bigg|\leq 4\mu\|f\|_\infty\frac{1}{N}\sum_{i=1}^N|R_{2,i}^N(s)|_{\infty,T}\mathbb{I}_{\{\mathcal{E}_i^N(T)\}}.\end{aligned}$$

By an application of Cauchy-Schwartz inequality and using Lemma 3.1, there exists a constant $C_1(T)$ such that

$$\frac{1}{N}\sum_{i=1}^N\mathbb{E}\left(|R_{2,i}^N|_{\infty,T}\mathbb{I}_{\{\mathcal{E}_i^N(T)\}}\right)\leq\frac{1}{N}\sum_{i=1}^N\sqrt{\mathbb{E}\left(|R_{2,i}^N|_{\infty,T}^2\right)}\sqrt{\mathbb{P}\left(\mathcal{E}_i^N(T)\right)}\leq\frac{C_1(T)}{\sqrt{N}}.$$

Consequently,

$$\lim_{N\to+\infty} \mathbb{E}\left(\sup_{0\le t\le T}\left|\frac{1}{N}\sum_{i=1}^{N}\int_0^t h_{2,i}^N(s)\ \mathrm{d}s\right|\right) = 0,$$

which implies that the process $(H_2^N(t))$ is also vanishing in distribution.

3.4.4 The Martingale

Careful calculations show that the previsible increasing process of $(\mathcal{M}_f^N(t))$ is given by

$$\left(\left\langle \mathcal{M}_f^N \right\rangle (t)\right) = \left(\frac{\lambda}{N^2}G_1^N(t) + \frac{\mu}{N^2}G_2^N(t)\right),$$

with

$$G_1^N(t) = \sum_{i=1}^{N}\int_0^t \left(\Delta^{\mp}(f)(R_i^N(s)) + \frac{N}{N-1}\Lambda^N(s)(\Delta_2^+(f)) - \frac{N}{N-1}\Delta_2^+(f)(R_i^N(t))\right)^2 \mathbb{I}_{\left\{R_{i,1}^N(s-)>0\right\}}\ \mathrm{d}s,$$

and

$$G_2^N(t) = \sum_{i=1}^{N}\int_0^t \left(f(0,0) - f(R_i^N(s)) + \sum_{j\neq i}[f(R_j^N(s)+X_{i,j}^N(s)(e_1-e_2)) - f(R_j^N(s))]\mathbb{I}_{\left\{X_{ij}^N(s)>0\right\}}\right)^2 \mathrm{d}s.$$

From the simple bounds $\|G_1^N\|_{\infty,T} \le 16\cdot NT\|f\|_\infty^2$ and

$$\|G_2^N\|_{\infty,T} \le 8NT\|f\|_\infty^2\left(1+\|R_{i,2}^N\|_{\infty,T}^2\right),$$

and, by using Relation (3.20) of Lemma 3.1, we get

$$\lim_{N\to+\infty}\mathbb{E}\left(\left\langle \mathcal{M}_f^N\right\rangle (T)\right) = 0.$$

Therefore, by Doob's inequality, the martingale $(\mathcal{M}_f^N(t))$ converges to zero in distribution as N goes to infinity.

Proposition 3.1. *(Tightness of the Empirical Distribution Process) The sequence $(\Lambda^N(t))$ is tight with respect to the convergence in distribution in $\mathcal{D}(\mathbb{R}_+, \mathcal{P}(\mathbb{N}^2))$. Any limiting point $(\Lambda(t))$ is a continuous process which satisfies*

$$\begin{aligned}\Lambda(t)(f) = \Lambda(0)(f) &+ \lambda\int_0^t\int_{\mathbb{N}^2}\Delta^{\mp}(f)(x,y)\mathbb{I}_{\{x>0\}}\Lambda(s)(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}s \\ &+ \lambda\int_0^t \Lambda(s)(\mathbb{N}^*\times\mathbb{N})\int_{\mathbb{N}^2}\Delta_2^+(f)(x,y)\Lambda(s)(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}s \\ &+ \mu\int_0^t\int_{\mathbb{N}^2}(f(0,0)-f(x,y))\Lambda(s)(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}s \\ &+ \mu\int_0^t\int_{\mathbb{N}^2} y\Delta^{\pm}(f)(x,y)\,\Lambda(s)(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}s \qquad (3.25)\end{aligned}$$

for every function f with finite support on $\mathbb{N}^2$.

Note that the Fokker-Planck Equation (3.2) of the introduction is the functional form of the stochastic Equation (3.25).

Proof. Theorem 3.7.1 of Dawson [42] states that it is enough to prove that, for any function f on $\mathbb{N}^2$ with finite support, the sequence of processes $(\Lambda_N(\cdot)(f))$ is tight with respect to the topology of the uniform norm on compact sets. Using the criterion of the modulus of continuity (see e.g. Theorem 7.2, page 81 of Billingsley [19]), we need to show that for every $\varepsilon>0$ and $\eta>0$, there exists a $\delta_0>0$ such that if $\delta<\delta_0$ then,

$$\mathbb{P}\left(\sup_{\substack{0\le s\le t\le T\\ |t-s|\le\delta}} |\Lambda_N(t)(f)-\Lambda_N(s)(f)|\ge\eta\right)\le\varepsilon \tag{3.26}$$

holds for all $N\in\mathbb{N}$. Fix $0\le s,t\le T$ with $|t-s|\le\delta$, and remember the equality (3.24) for the process $(\Lambda^N(t)(f))$. We have already shown that the processes $(H_1^N(t))$, $(H_2^N(t))$, and $(\mathcal{M}_f^N(t))$ vanish as N goes to infinity. For the remaining terms on the right-hand side of (3.24), note that there exists a finite constant C_0 such that

$$\left|\int_s^t\int_{\mathbb{N}^2}\Delta^{\mp}(f)(x,y)\mathbb{I}_{\{x>0\}}\Lambda^N(u)(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}u\right|\le C_0\delta\|f\|_\infty,$$

$$\left|\int_s^t\Lambda^N(u)(\mathbb{N}^*\times\mathbb{N})\int_{\mathbb{N}^2}\Delta_2^{+}(f)(x,y)\Lambda^N(u)(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}u\right|\le C_0\delta\|f\|_\infty,$$

and

$$\left|\int_s^t\int_{\mathbb{N}^2}(f(0,0)-f(x,y))\Lambda^N(u)(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}u\right|\le C_0\delta\|f\|_\infty.$$

Also, by Relation (3.20) of Lemma 3.1 shows that there exists $C_1<\infty$ independent of N such that

$$\mathbb{E}\left(\left|\int_s^t\int_{\mathbb{N}^2}y\Delta^{\pm}(f)(x,y)\,\Lambda^N(u)(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}u\right|\right)$$
$$\le 2\|f\|_\infty\delta\frac{1}{N}\sum_{i=1}^N\mathbb{E}\left(\sup_{0\le u\le T}R_{i,2}^N(u)\right)\le C_1\delta\|f\|_\infty.$$

If follows from the Chebishev's inequality that the sequence $(\Lambda^N(\cdot)(f))$ satisfies Relation (3.26), and hence it is tight.

Moreover, if Λ is a limiting point, from Relation (3.24) and the fact that the processes H_1^N, H_2^N, and $\mathcal{M}_f^N$ vanish as N gets large, one obtains that Relation (3.25) holds, Finally, it is straightforward to show that all the terms on the right-hand side of (3.25) are continuous in t. □

We now show that Equation (3.25) that characterized the limits of $(\Lambda^N(t))$ has a unique solution.

Lemma 3.2. *Let $(\Lambda(t))$ be a solution to equation (3.25) with an initial condition $\Lambda(0)$, a probability on $\mathbb{N}^2$ with bounded support. Then, for any $T>0$, there exists a constant C_T such that for all $K\geq 2\log(2)$,*

$$\sup_{0\leq t\leq T}\int_0^t\int_{\mathbb{N}^2} y\mathbb{I}_{\{y\geq K\}}\Lambda(s)(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}s\leq C_T e^{-K/2}. \tag{3.27}$$

Proof. For all $t\leq T$, since $y\leq \exp(y/2)$ if $y\geq 2\log(2)$, then for $K\geq 2\log(2)$,

$$\int_{\mathbb{N}^2} y\mathbb{I}_{\{y\geq K\}}\Lambda(t)(\mathrm{d}x,\mathrm{d}y)\leq e^{-K/2}\int_{\mathbb{N}^2} e^y\Lambda(t)(\mathrm{d}x,\mathrm{d}y). \tag{3.28}$$

For every $K_1\geq 0$, using equation (3.25) for Λ with f replaced by

$$\tilde{f}(x,y)=e^{x+y}\mathbb{I}_{\{x+y\leq K_1\}},$$

and since $\Delta^{\mp}(\tilde{f})=\Delta^{\pm}(\tilde{f})=0$, we have

$$\begin{aligned}\int_{\mathbb{N}^2} e^{x+y}\mathbb{I}_{\{x+y\leq K_1\}}\Lambda(t)(\mathrm{d}x,\mathrm{d}y)\leq&\int_{\mathbb{N}^2} e^{x+y}\Lambda(0)(\mathrm{d}x,\mathrm{d}y)\\ &+\lambda(e-1)\int_0^t\int_{\mathbb{N}^2} e^{x+y}\mathbb{I}_{\{x+y\leq K_1\}}\Lambda(s)(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}s\\ &+\mu\int_0^t\left(1-\int_{\mathbb{N}^2} e^{x+y}\mathbb{I}_{\{x+y\leq K_1\}}\Lambda(s)(\mathrm{d}x,\mathrm{d}y)\right)\mathrm{d}s.\end{aligned}$$

By an application of Grönwall's inequality, there exists a constant c_T independent of K_1 such that

$$\sup_{0\leq t\leq T}\int_{\mathbb{N}^2} e^y\mathbb{I}_{\{x+y\leq K_1\}}\Lambda(t)(\mathrm{d}x,\mathrm{d}y)\leq c_T.$$

The bound (3.27) can be obtained by letting K_1 go to infinity in the above inequality, and substituting it in Relation (3.28). □

Proposition 3.2 (Uniqueness)**.** *For every Λ_0 a probability on $\mathbb{N}^2$ with finite support, Equation (3.25) has at most one solution $(\Lambda(t))$ in $\mathcal{D}(\mathbb{R}^+,\mathcal{P}(\mathbb{N}^2))$, with initial condition Λ_0.*

Proof. Let $(\Lambda^1(t))$ and $(\Lambda^2(t))\in\mathcal{D}(\mathbb{R}_+,\mathcal{P}(\mathbb{N}^2))$ be solutions of (3.25) with initial condition Λ_0. Let f be a bounded function on $\mathbb{N}^2$ and $t\geq 0$, we have

$$\begin{aligned}\Lambda^1(t)(f)-\Lambda^2(t)(f)=&\lambda\int_0^t\int_{\mathbb{N}^2}\Delta^{\mp}(f)(x,y)\mathbb{I}_{\{x>0\}}\left(\Lambda^1(s)-\Lambda^2(s)\right)(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}s\\ &+\lambda\int_0^t\Lambda^1(s)(\mathbb{N}^*\times\mathbb{N})\int_{\mathbb{N}^2}\Delta_2^+(f)(x,y)\left(\Lambda^1(s)-\Lambda^2(s)\right)(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}s\\ &+\lambda\int_0^t\left(\Lambda^1(s)-\Lambda^2(s)\right)(\mathbb{N}^*\times\mathbb{N})\int_{\mathbb{N}^2}\Delta_2^+(f)(x,y)\Lambda^2(s)(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}s\\ &+\mu\int_0^t\int_{\mathbb{N}^2}(f(0,0)-f(x,y))\left(\Lambda^1(s)-\Lambda^2(s)\right)(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}s\\ &+\mu\int_0^t\int_{\mathbb{N}^2} y\Delta^{\pm}(f)(x,y)\left(\Lambda^1(s)-\Lambda^2(s)\right)(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}s.\end{aligned}$$

For any signed measure m on $\mathbb{N}^2$, denote

$$\|m\|_{TV} = \sup\left\{\int_{\mathbb{N}^2} f(x,y)m(\mathrm{d}x,\mathrm{d}y), f:\mathbb{N}^2\to\mathbb{R} \text{ with } \|f\|_\infty\le 1\right\}.$$

Therefore, for every f on $\mathbb{N}^2$ with $\|f\|_\infty\le 1$ and every $K>0$, we have

$$\begin{aligned}\left|\Lambda^1(t)(f)-\Lambda^2(t)(f)\right| &\le (6\lambda+2\mu+2\mu K)\int_0^t \|\Lambda^1(s)-\Lambda^2(s)\|_{TV}\,\mathrm{d}s\\ &\quad +2\mu\int_0^t\int_{\mathbb{N}^2} y\mathbb{I}_{\{y\ge K\}}\Lambda^1(s)(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}s+2\mu\int_0^t\int_{\mathbb{N}^2} y\mathbb{I}_{\{y\ge K\}}\Lambda^2(s)(\mathrm{d}x,\mathrm{d}y)\,\mathrm{d}s.\end{aligned}$$

Now using (3.27) of Lemma 3.2, and taking the supremum over all functions f on $\mathbb{N}^2$ with $\|f\|_\infty\le 1$, we have

$$\|\Lambda^1(t)-\Lambda^2(t)\|_{TV}\le 4\mu C_t e^{-K/2}+(6\lambda+2\mu+2\mu K)\int_0^t\|\Lambda^1(s)-\Lambda^2(s)\|_{TV}\,\mathrm{d}s.$$

Therefore, by another application of Grönwall's inequality,

$$\|\Lambda^1(t)-\Lambda^2(t)\|_{TV}\le 4\mu C_T e^{-K/2}e^{(6\lambda+2\mu+2\mu K)t}.$$

For $t<1/(4\mu)$, by letting K go to infinity in the above relation, one gets that $\Lambda^1(t)=\Lambda^2(t)$. By repeating the same argument on successive time intervals of width less than $1/(4\mu)$, one obtains the uniqueness result. □

Now we can conclude the proof of Theorem 3.2.

Proof of Theorem 3.2. Let $(x,y)\in N^2$ and $\Lambda_0=\delta_{(x,y)}$, then if $(\overline{R}_1(t),\overline{R}_2(t))$ is the unique solution of Equation (3.14) and the measure valued process $(\Lambda^1(t))$ is defined by, if f is a function with finite support on $\mathbb{N}^2$,

$$\Lambda^1(t)(f)\overset{\text{def.}}{=}\mathbb{E}\left(f(\overline{R}_1(t),\overline{R}_2(t))\right),$$

it is straightforward to check that this is a solution of Equation (3.25). The convergence of $(\Lambda^N(t))$ follows from Propositions 3.1 and 3.2. The last assertion is a simple consequence of Proposition 2.2 in Sznitman [158]. □

3.5 An Asymptotic Bound on the Decay of the Network

The asymptotic process $(\overline{R}(t))=(\overline{R}_1(t),\overline{R}_2(t))$ of Theorem 3.1 is an inhomogeneous Markov process with the following transitions: if $(\overline{R}(t))$ is in state $r=(r_1,r_2)$ at time t, the next possible state and the corresponding rates are given by

$$r\mapsto\begin{cases}(0,0) & \text{with rate } \mu\\ r+e_2 & \text{with rate } \lambda p(t)\end{cases}\quad\text{and}\quad r\mapsto\begin{cases}r-e_1+e_2 & \text{with rate } \lambda\mathbb{I}_{\{r_1>0\}}\\ r+e_1-e_2 & \text{with rate } \mu r_2,\end{cases}\tag{3.29}$$

where $p(t)=\mathbb{P}\left(\overline{R}_1(t)>0\right)$ is the non-linear part of the dynamic. A simple feature of this process is that it resets to the state $(0,0)$ at the epoch times of a Poisson process

with rate μ, and between two consecutive epoch times, the sum of its coordinates grows according to an inhomogeneous Poisson process with rate $p(\cdot)$. With this observation, the following proposition gives a representation of the distribution of the total number of copies with the function $(p(t))$.

Proposition 3.3. *If the initial state of $(\overline{R}_1(t), \overline{R}_2(t))$ is $(0, r_2)$ with $r_2 \in \mathbb{N}$ then, for $u \in [0,1]$ and $t \geq 0$,*

$$\mathbb{E}\left(u^{\overline{R}_1(t)+\overline{R}_2(t)}\right) = e^{-\mu t} u^{r_2} \exp\left(-\lambda(1-u)\int_0^t p(z)\,\mathrm{d}z\right) + \int_0^t \exp\left(-\lambda(1-u)\int_0^s p(t-z)\,\mathrm{d}z\right)\mu e^{-\mu s}\,\mathrm{d}s.$$

Proof. From Relation (3.29), one obtains that the transition rates of the process $(\overline{R}_1(t)+\overline{R}_2(t))$ are given by

$$r \mapsto r+1, \text{ at rate } \lambda p(t) \text{ and } r \mapsto 0, \text{ at rate } \mu.$$

The Fokker-Planck equation associated to this process yields the relation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\left(u^{\overline{R}_1(t)+\overline{R}_2(t)}\right) = \mu + (\lambda p(t)(u-1)-\mu)\,\mathbb{E}\left(u^{\overline{R}_1(t)+\overline{R}_2(t)}\right).$$

It is then easy to conclude. □

The problem with the above formula is that the function $t \mapsto p(t)$ is unknown. In the following, we obtain a lower bound on the asymptotic rate of decay of the network, i.e. the exponential rate of convergence of the process $(\overline{R}_1(t), \overline{R}_2(t))$ to $(0,0)$.

Recall that $R^N_{i,1}(t)$ and $R^N_{i,2}(t)$ are the number of files on server i with one and two copies, respectively. Therefore, the quantity

$$L^N(t) = \sum_{i=1}^N R^N_{i,1}(t) + \frac{1}{2}\sum_{i=1}^N R^N_{i,2}(t)$$

is the total number of distinct files in the system at time t. By Theorem 3.2, Equation (3.20) of Lemma 3.1, and an application of the dominated convergence theorem, we have

$$(L(t)) \stackrel{\text{def.}}{=} \lim_{N\to\infty}\left(\frac{L^N(t)}{N}\right) = \left(\mathbb{E}\left(\overline{R}_1(t)\right) + \frac{1}{2}\mathbb{E}\left(\overline{R}_2(t)\right)\right).$$

The following proposition gives therefore a lower bound on the exponential rate of decay $(L(t))$.

Proposition 3.4. *If $(\overline{R}(t)) = (\overline{R}_1(t), \overline{R}_2(t))$ is the solution of Equation (3.14), then*

$$\mathbb{E}\left(\overline{R}_1(t) + \frac{1}{2}\overline{R}_2(t)\right) \leq \mathbb{E}\left(\overline{R}_1(0) + \frac{1}{2}\overline{R}_2(0)\right) e^{-\kappa_2^+(\rho)\mu t}, \tag{3.30}$$

where

$$\kappa_2^+(x) = \frac{(3+x) - \sqrt{(3+x)^2-8}}{2}, \qquad x \in \mathbb{R},$$

and $\rho = \lambda/\mu$.

The quantity $\kappa_2^+(\rho)\mu$ is thus a lower bound for the exponential rate of decay. When there is no duplication capacity, i.e. $\lambda=0$, $\kappa_2^+(\rho)=1$ and the lower bound becomes μ, the failure rate of servers, as expected. On the other hand, when the duplication capacity goes to infinity, the lower bound goes to 0.

Proof of Proposition 3.4. Let $m_1(t)=\mathbb{E}\left(\overline{R}_1(t)\right)$ and $m_2(t)=\mathbb{E}\left(\overline{R}_2(t)\right)$. Taking expectation from both sides of Equations (3.14), we conclude that the pair (m_1, m_2) satisfy the following set of Ordinary Differential Equations (ODEs):

$$\begin{cases} \dot{m}_1(t) = -\lambda p(t) + \mu(m_2(t) - m_1(t)), \\ \dot{m}_2(t) = 2\lambda p(t) - 2\mu m_2(t). \end{cases} \tag{3.31}$$

Defining $g(t)=m_1(t)-p(t)$, then clearly $0\leq g(t)\leq m_1(t)$. The ODEs (3.31) can be rewritten as

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} m_1(t) \\ m_2(t) \end{pmatrix} = A\begin{pmatrix} m_1(t) \\ m_2(t) \end{pmatrix} + \lambda\begin{pmatrix} g(t) \\ -2g(t) \end{pmatrix}, \tag{3.32}$$

where A is the matrix

$$A = \begin{pmatrix} -(\lambda+\mu) & \mu \\ 2\lambda & -2\mu \end{pmatrix}.$$

It has two negative eigenvalues, $-\mu\kappa_2^+(\rho)$ and $-\mu\kappa_2^-(\rho)$ with

$$\kappa_2^+(\rho) = \frac{(3+\rho) - \sqrt{(3+\rho)^2 - 8}}{2}, \quad \kappa_2^-(\rho) = \frac{(3+\rho) + \sqrt{(3+\rho)^2 - 8}}{2}.$$

Defining the constants $y_1=(-\mu\kappa_2^+(\rho)+\lambda+\mu)/\mu$, $y_2=(-\mu\kappa_2^-(\rho)+\lambda+\mu)/\mu$,

$$h_1 = \frac{1}{y_1-y_2}(-y_2 m_1(0)+m_2(0)), \quad \text{and} \quad h_2 = \frac{1}{y_1-y_2}(y_1 m_1(0)-m_2(0)),$$

the standard formula for explicit solution of the linear ODE (3.32), with g regarded as an external force, gives

$$\begin{aligned} m_1(t) &= h_1 e^{-\mu\kappa_2^+(\rho)t} + h_2 e^{-\mu\kappa_2^-(\rho)t} \\ &\quad - \frac{\lambda}{y_1-y_2}\int_0^t g(s)\left[(y_2+2)e^{-\mu\kappa_2^+(\rho)(t-s)} - (y_1+2)e^{-\mu\kappa_2^-(\rho)(t-s)}\right] \mathrm{d}s, \end{aligned} \tag{3.33}$$

$$\begin{aligned} m_2(t) &= y_1 h_1 e^{-\mu\kappa_2^+(\rho)t} + y_2 h_2 e^{-\mu\kappa_2^-(\rho)t} \\ &\quad - \frac{\lambda}{y_1-y_2}\int_0^t g(s)\left[(y_2+2)y_1 e^{-\mu\kappa_2^+(\rho)(t-s)} - (y_1+2)y_2 e^{-\mu\kappa_2^-(\rho)(t-s)}\right] \mathrm{d}s. \end{aligned} \tag{3.34}$$

Therefore, using the fact $g(s)\geq 0$ in the first inequality, and the relations $y_1>y_2\geq -2$ and $\kappa_2^+(\rho)<\kappa_2^-(\rho)$ in the second inequality below, we conclude

$$\begin{aligned} m_1(t) + \frac{1}{2}m_2(t) &= \left(1+\frac{y_1}{2}\right)h_1 e^{-\mu\kappa_2^+(\rho)t} + \left(1+\frac{y_2}{2}\right)h_2 e^{-\mu\kappa_2^-(\rho)t} \\ &\quad - \frac{\lambda}{y_1-y_2}\frac{1}{2}(y_1+2)(y_2+2)\int_0^t g(s)\left(e^{-\mu\kappa_2^+(\rho)(t-s)} - e^{-\mu\kappa_2^-(\rho)(t-s)}\right)\mathrm{d}s \\ &\leq \left(1+\frac{y_1}{2}\right)h_1 e^{-\mu\kappa_2^+(\rho)t} + \left(1+\frac{y_2}{2}\right)h_2 e^{-\mu\kappa_2^-(\rho)t} \\ &\leq \left(m_1(0) + \frac{1}{2}m_2(0)\right)e^{-\mu\kappa_2^+(\rho)t}, \end{aligned}$$

This completes the proof. □

3.6 The Case of Multiple Copies

In this section, we consider the general case where each file has a maximum number of d copies in the system. We now describe the algorithm without too much formalism for sake of simplicity. The duplication capacity of a given node is used for one of its copies corresponding to a file with the least number of copies in the network. Provided that this number is strictly less than d, a new copy is done at rate λ at random on a node. See Section 3 of Sun at al. [156] for a quick description of how this kind of mechanism can be implemented in practice. The node receives copies from other nodes from this duplication mechanism. As before, at rate μ all copies of the node are removed.

As it will be seen in the next section, the model does not seem to be mathematically tractable. To understand the effect of the maximum number of copies d on the performance of the file system, we study the asymptotic behavior of a stochastic model which is dominating (in some sense) our network. We study the decay rate of this new model.

The initial condition of our system are given by the following assumption.

Assumptions 3.2. *(Initial State) There is a set $\mathcal{F}_N$ of F_N initial files, and for $f \in \mathcal{F}_N$, a subset of d nodes of $\{1, \ldots, N\}$ is taken at random and on each of them a copy of f is done.*

It should be noted that, with the duplication mechanism described above, a copy can be made on a node which has already a copy of the same file. But, with a similar approach as in the proof of Lemma 3.1, it can be shown that on any finite time interval, with probability 1, there is only a finite number of files which have at least two copies on a server. In particular, this assumption has no influence on the asymptotic results obtained in this section since they are concerning asymptotic growth in N of the number of files alive at time t.

With these assumptions, if f is a file, $f \in \mathcal{F}_N$, one denotes by $\mathcal{A}_f^N(t) \subset \{1, \ldots, N\}$ the subset of nodes which have a copy of f at time t. The cardinality of the set $\mathcal{A}_f^N(t)$ is denoted as $c_f^N(t)$, it is at most d. The process

$$(\mathcal{A}^N(t)) \stackrel{\text{def.}}{=} (\mathcal{A}_f^N(t), f \in \mathcal{F}_N)$$

gives a (Markovian) representation of the time evolution of the state of the network.

3.6.1 The Additional Complexity of the Model

A analogous Markovian description as for the case $d=2$ can be done in the following way. If $\mathcal{S}$ is the set of non-empty subsets of $\{1, \ldots, N\}$ whose cardinality is less or equal to d and, for $A \in \mathcal{S}$, if $X_A^N(t)$ is the number of files with a copy only in the nodes whose index is in A,

$$X_A^N(t) = \sum_{f \in \mathcal{F}_N} \mathbb{I}_{\left\{\mathcal{A}_f^N(t) = A\right\}}$$

then it is not difficult to show that $(X^N(t))=(X^N_A(t), A \in \mathcal{S})$ is a Markov process, even if its transitions are not so easy to write formally. Following the analysis done for the case $d=2$, it is natural to introduce, for $1\leq i\leq N$, and $1\leq k\leq d$, the quantity

$$R^N_{i,k}(t) = \sum_{\substack{A\in\mathcal{S}, i\in A\\ \mathrm{card}(A)=k}} X^N_A(t) = \sum_{f\in\mathcal{F}_N} \mathbb{I}_{\left\{i\in\mathcal{A}^N_f(t), c^N_f(t)=k\right\}}$$

is the number of files having k copies in the whole network, with a copy on server i. It is the equivalent of the variables $R^N_1(t)$ and $R^N_2(t)$ of the case $d=2$.

The vector $R^N_i(t) = (R^N_{i,k}(t), 1 \leq k \leq d)$ gives also a reduced representation of the state of the node i at time t. It turns out that the evolution equations of this model are much more involved. To observe why our method cannot be worked out for general $d>2$, let us try, as in Section 3.3, to heuristically obtain the transition rates of a possible asymptotic limit process for this model. Fix $1\leq k<d$ and $1\leq i\leq N$, e_k is the kth unit vector of $\mathbb{N}^d$, and $r=(r_j)=(R^N_i(t-))$, then the process $(R^N_i(t)$ jumps from r to $r-e_k+e_{k+1}$ at time t according with two types of events:

a. due to the duplication capacity at node i, at rate λ, if $r_1=r_2=\cdots=r_{k-1}=0$ and $r_k>0$. Recall that only files with the least number of copies are duplicated.

b. If a file present at i in k copies is duplicated on one of the other $k-1$ servers having a copy of this file, conditionally on the past before t, it occurs at rate

$$\lambda \sum_{j\neq i} \mathbb{I}_{\left\{R^N_{j,\ell}(t-)=0, 1\leq\ell<k, R^N_{j,k}(t-)>0\right\}} \frac{1}{R^N_{j,k}(t-)} \sum_{\substack{A\in\mathcal{S}: i,j\in A\\ \mathrm{card}(A)=k}} X^N_A(t-).$$

The first event is similar as in the case $d=2$, the jump rate can be expressed in terms of the vector r. This is not the case for the second event. The last sum of the above expression does not seem to have an expression in terms of the components of the vector r. It requires a much more detailed description. The information provided by r is not enough, even in the limit when N goes to infinity as it is the case when $d=2$. Consequently, it does not seem that one can derive autonomous equations describing the asymptotic dynamics of $(R^N(t))$.

3.6.2 Introduction of a Dominating Process

We now consider the following related Markov process. We first describe it without too much formalism for sake of simplicity in terms of a duplication system. Note however that this is *not* an alternative algorithm but merely a way of having a mathematically tractable stochastic process that will give us a lower bound of the exponential decay rate of the initial system. For convenience, we will use nevertheless the terminology of "files", "copies", "duplication" and "servers" to describe this new process.

The initial condition is also given by Assumption 3.2. If f is one of the files of the network, as long as the total number of copies of f is strictly less that d, each of the nodes having one of these copies generates a new copy of f at rate λ at random on a node. The case of multiple copies on the same server is taken care of as for the process

$(\mathcal{A}^N(t))$. The failures of a given node occur according to a Poisson process with rate μ and, as for our algorithm, all copies are lost.

The system works the same as the original model, except that each file on a server i with strictly less than d copies in the network, has a dedicated duplication rate λ, regardless of any other copy on that server. Consequently, if, for $1 \leq k \leq d$, a node has r_k files, each of them with a total of k copies, at a given time, the "duplication capacity" of this node for $(Y^N(t))$ is given

$$\lambda\,(r_1 + r_2 + \cdots + r_{d-1})$$

instead of λ in our algorithm. Remember nevertheless that such system is not possible in practice, it is used only to estimate the performances of the algorithm introduced at the beginning of this section.

For $f \in \mathcal{F}_N$, one denotes by $\mathcal{B}_f^N(t)$ the finite subset of $\{1, \ldots, N\}$ of nodes having a copy of f at time t and its cardinality is denoted by $d_f^N(t)$. We define

$$(\mathcal{B}^N(t)) \stackrel{\text{def.}}{=} (\mathcal{B}_f^N(t), f \in \mathcal{F}_N).$$

In this model, for $k \in \{1, \ldots, d\}$ and $1 \leq i \leq N$, we will denote by $T_{i,k}^N(t)$ the number of files of type k and with one copy on node i at time $t \geq 0$, this is the analogue of the variable $R_{i,k}^N(t)$ defined above,

$$T_{i,k}^N(t) = \sum_{f \in \mathcal{F}_N} \mathbb{I}_{\left\{i \in \mathcal{B}_f^N(t), d_f^N(t) = k\right\}}.$$

For a given node i, if $(T_{i,k}^N(t-), 1 \leq k \leq d) = r = (r_k)$, provided that there are no multiple copies on the same server just before time t, the transition rates of this process at time t are given by

$$r \mapsto \begin{cases} r - e_k + e_{k+1}, & \text{at rate } \lambda k r_k, \qquad 1 \leq k < d, \\ r + e_{k-1} - e_k, & \qquad \mu(k-1) r_k, \qquad 1 < k \leq d. \end{cases}$$

and

$$r \mapsto \begin{cases} (0,0), & \text{at rate } \mu, \\ r + e_k, & \qquad \lambda \dfrac{1}{N-k+1} \displaystyle\sum_{f \in \mathcal{F}_N} (k-1) \mathbb{I}_{\left\{i \notin \mathcal{B}_f^N(t), d_f^N(t-) = k-1\right\}}. \end{cases}$$

Note that the last sum is the sum of the terms $T_{j,k-1}^N(t-)$, $j = 1, \ldots, N$ minus some term which is less that $(k-1) T_{i,k-1}^N(t-)$. The term $T_{i,k-1}^N(t-)$, with appropriate estimates as in Section 3.4, will vanish in the limit when divided by $N - k + 1$. Consequently, asymptotically, the transitions of the vector $(T_{i,k}^N(t))$ can be expressed as a functional of its coordinates. With the same methods as for the original model for $d = 2$ in Section 3.4, it is not difficult to show that an analogue of Theorem 3.2 holds.

Theorem 3.3. *(Mean-Field Convergence Theorem) Suppose the process $(\mathcal{A}^N(t))$ is initialized according to Assumption 3.2. The process of the empirical distribution*

$$(\Lambda^N(t)) = \left(\frac{1}{N} \sum_{i=1}^{N} \delta_{(T_{i,k}^N(t), 1 \leq k \leq d)} \right)$$

converges in distribution to a process $(\Lambda(t) \in \mathcal{D}(\mathbb{R}_+, \mathcal{P}(\mathbb{N}^2))$ *such that : for* f *with finite support on* $\mathbb{N}^d$,

$$\Lambda(t)(f) \stackrel{def.}{=} \mathbb{E}\left(f\left(\overline{T}_k(t)\right)\right),$$

where $(\overline{T}(t))=(\overline{T}_k(t))$ *is a nonlinear Markov process with the following transition rates: if* $(\overline{T}(t))$ *is in state* $r=(r_k)$ *just before time* t*, the next possible state and the corresponding rates are given by*

$$r \mapsto \begin{cases} (0,0), & \mu \\ r+e_k, & \lambda\mathbb{E}\left(\overline{T}_{k-1}(t)\right),\ 1 \le k \le d \end{cases} \text{ and } \begin{cases} r-e_k+e_{k+1}, & \lambda k r_k,\ 1 \le k < d \\ r+e_{k-1}-e_k, & \mu(k-1)r_k,\ 1 < k \le d. \end{cases}$$

An argument similar to that in the proof of Theorem 3.1 shows the existence and uniqueness of the Markov process $(\overline{T}(t))$. Note that the nonlinear component is now given by the vector of the mean values $\mathbb{E}(\overline{T}_k(t))$, $k=1, \ldots, d$.

The limiting Markov process $(\overline{T}(t))=(\overline{T}_k(t))$ can also be seen as the solution of the following SDEs, for $t \ge 0$

$$\begin{aligned} \mathrm{d}\overline{T}_1(t) = -\overline{T}_1(t-)\mathcal{N}_\mu(\mathrm{d}t) - \int_{\mathbb{R}_+} \mathbb{I}_{\left\{0 \le h \le \overline{T}_1(t-)\right\}} \overline{\mathcal{N}}_\lambda(\mathrm{d}t, \mathrm{d}h) \\ + \int_{\mathbb{R}_+} \mathbb{I}_{\left\{0 \le h \le \overline{T}_2(t-)\right\}} \overline{\mathcal{N}}_\mu(\mathrm{d}t, \mathrm{d}h), \end{aligned} \tag{3.35}$$

for $1<k<d$,

$$\begin{aligned} \mathrm{d}\overline{T}_k(t) &= -\overline{T}_k(t-)\mathcal{N}_\mu(\mathrm{d}t) \\ &- \int_{\mathbb{R}_+} \mathbb{I}_{\left\{0 \le h \le \overline{T}_k(t-)\right\}} \overline{\mathcal{N}}_{k\lambda}(\mathrm{d}t, \mathrm{d}h) + \int_{\mathbb{R}_+} \mathbb{I}_{\left\{0 \le h \le \overline{T}_{k-1}(t-)\right\}} \overline{\mathcal{N}}_{(k-1)\lambda}(\mathrm{d}t, \mathrm{d}h) \\ &- \int_{\mathbb{R}_+} \mathbb{I}_{\left\{0 \le h \le \overline{T}_k(t-)\right\}} \overline{\mathcal{N}}_{(k-1)\mu}(\mathrm{d}t, \mathrm{d}h) + \int_{\mathbb{R}_+} \mathbb{I}_{\left\{0 \le h \le \overline{T}_{k+1}(t-)\right\}} \overline{\mathcal{N}}_{k\mu}(\mathrm{d}t, \mathrm{d}h) \\ &+ \int_{\mathbb{R}_+} \mathbb{I}_{\left\{0 \le h \le \mathbb{E}(\overline{T}_{k-1}(t))\right\}} \overline{\mathcal{N}}_{\lambda,k-1}(\mathrm{d}t, \mathrm{d}h), \end{aligned} \tag{3.36}$$

$$\begin{aligned} \mathrm{d}\overline{T}_d(t) &= -\overline{T}_d(t-)\mathcal{N}_\mu(\mathrm{d}t) + \int_{\mathbb{R}_+} \mathbb{I}_{\left\{0 \le h \le \mathbb{E}(\overline{T}_{d-1}(t))\right\}} \overline{\mathcal{N}}_{\lambda,d-1}(\mathrm{d}t, \mathrm{d}h) \\ &+ \int_{\mathbb{R}_+} \mathbb{I}_{\left\{0 \le h \le \overline{T}_{d-1}(t-)\right\}} \overline{\mathcal{N}}_{(d-1)\lambda}(\mathrm{d}t, \mathrm{d}h) - \int_{\mathbb{R}_+} \mathbb{I}_{\left\{0 \le h \le \overline{T}_d(t-)\right\}} \overline{\mathcal{N}}_{(d-1)\mu}(\mathrm{d}t, \mathrm{d}h). \end{aligned} \tag{3.37}$$

The interesting property of these SDEs is that the vector of expected values can be expressed as the solution of a classical ODE, as the next proposition states.

Proposition 3.5. *For* $t \ge 0$*, the function* $V(t)=\mathbb{E}[\overline{T}_k(t/\mu)]$ *satisfies*

$$\frac{\mathrm{d}}{\mathrm{d}t} V(t) = M_\rho \cdot V(t), \tag{3.38}$$

with

$$M_\rho = \begin{pmatrix} -(\rho+1) & 1 & 0 & & & 0 \\ 2\rho & -2(\rho+1) & 2 & 0 & & 0 \\ 0 & 3\rho & -3(\rho+1) & 3 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & k\rho & -k(\rho+1) & k & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & & & 0 & d\rho & -d \end{pmatrix}$$

and $\rho=\lambda/\mu$. Moreover, the matrix M_ρ has d distinct negative eigenvalues and the largest of them, $-\kappa_d^+(\rho)$, satisfies

$$0 < \kappa_d^+(\rho) \leq \overline{\kappa}_d(\rho) \stackrel{\text{def.}}{=} \left(\sum_{k=1}^{d} \frac{\rho^{k-1}}{k}\right)^{-1} < 1. \tag{3.39}$$

Finally, there exists a positive constant K_0 such that, for all $1 \leq k \leq d$ and $t \geq 0$,

$$\mathbb{E}\left(\overline{T}_k(t)\right) \leq K_0 e^{-\mu\kappa_d^+(\rho)t}. \tag{3.40}$$

Proof. Equation (3.38) can be obtained by taking the expected value of both sides of the integral version of Equations (3.35), (3.36) and (3.37). For the next claim, since the matrix M_ρ is a tridiagonal matrix, it has d distinct real eigenvalues (see e.g. Chapter 1 of Fallat and Johnson [60]). If D is the $d\times d$ diagonal matrix whose kth diagonal component is $1/\sqrt{k\rho^{k-1}}$, then

$$DM_\rho D^{-1}=\begin{pmatrix} -(\rho+1) & \sqrt{2\rho} & 0 & & & 0 \\ \sqrt{2\rho} & -2(\rho+1) & \sqrt{6\rho} & 0 & & 0 \\ 0 & \sqrt{6\rho} & -3(\rho+1) & \sqrt{12\rho} & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \sqrt{k(k-1)\rho} & -k(\rho+1) & \sqrt{k(k+1)\rho} & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & & & 0 & \sqrt{d(d-1)\rho} & -d \end{pmatrix}$$

is a symmetric matrix with the same eigenvalues as M_ρ. A straightforward calculation shows that its associated quadratic form is given by

$$q(x_1,...,x_d) \stackrel{\text{def.}}{=} -\sum_{i=1}^{d-1}\left(\sqrt{k\rho}x_k - \sqrt{(k+1)}x_{k+1}\right)^2 - x_1^2, \quad (x_1,...,x_d) \in \mathbb{R}^d,$$

which implies that all eigenvalues of M_ρ are negative. The maximal eigenvalue of the symmetric matrix can be expressed as

$$\sup\left(q(y) : y = (y_1,...,y_d) \in \mathbb{R}^d, \|y\| = 1\right),$$

with $\|y\|^2=y_1^2+\cdots+y_d^2$, see e.g. page 176 of Horn and Johnson [76]. Taking the vector $x=(x_1,...,x_d)$ such that

$$x_k=x_{k-1}\sqrt{\frac{k-1}{k}\rho}, \quad 1 < k \leq d,$$

and x_1 is chosen so that $\|x\|=1$, one gets the upper bound (3.39).

Finally, Equation (3.38) shows that the components of $V(\cdot)$ can be expressed as a linear combination of the functions $(\exp(\lambda_k\mu t))$, $1\leq k\leq d$. Since all eigenvalues of M_ρ are negative, $-\kappa_d^+$ is the largest eigenvalue, Relation (3.40) follows. □

Remarks

1. We have not been able to get a closed form expression for the actual exponential decay rate $\kappa_d^+(\rho)$ associated to the process $(\overline{T}_k(t))$. However, the upper bound $\overline{\kappa}_d(\rho)$ defined in Equation (3.39) gives a lower bound for the decay rate. In Figure 3.1, we plot the ration $\bar{\kappa}_d(\rho)/\kappa_d^+(\rho)$ for different values of ρ and d.
2. Note that if the duplication rate λ is larger than μ, i.e. $\rho>1$, then

$$\lim_{d\to+\infty} \kappa_d^+(\rho) = 0.$$

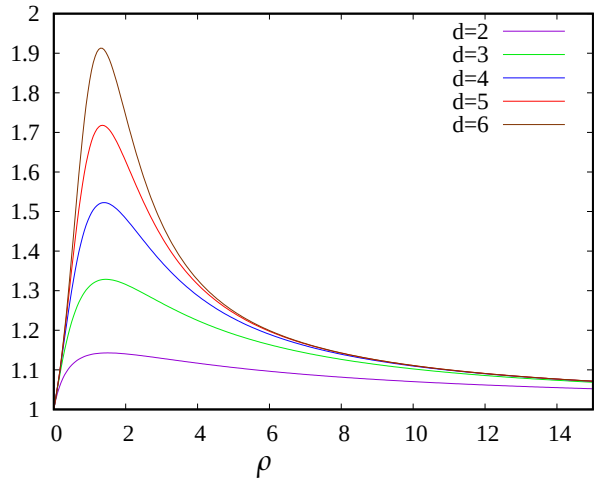


Figure 3.1 – Accuracy of the upper bound of Relation (3.39): The ratio $\overline{\kappa}_d(\rho)/\kappa_d^+(\rho)$ for various values of ρ and d.

Finally, for the case d=2, we can compare our result on the decay rate of file system with the decay rate of the process $(\mathbb{E}(\overline{T}_k(t)))$.

Corollary 3.1. *For d=2, if $(\overline{T}_1(0), \overline{T}_2(0))=(0, r_2)$, we have*

$$\mathbb{E}\left(\overline{T}_1(t)\right) = \frac{r_2}{\kappa_2^+(\rho) - \kappa_2^-(\rho)} \left(e^{-\mu\kappa_2^+(\rho)t} - e^{-\mu\kappa_2^-(\rho)t}\right)$$

$$\mathbb{E}\left(\overline{T}_2(t)\right) = \frac{r_2}{\kappa_2^+(\rho) - \kappa_2^-(\rho)} \left(y_+ e^{-\mu\kappa_2^+(\rho)t} - y_- e^{-\mu\kappa_2^-(\rho)t}\right),$$

where

$$\kappa_2^\pm(\rho) \stackrel{\text{def.}}{=} \frac{(3+\rho) \pm \sqrt{(3+\rho)^2-8}}{2}$$

and $y_\pm \stackrel{\text{def.}}{=} -\kappa_2^\pm+\rho+1$.

The definitions with $\pm$ just indicate that the identities are taken for $+$ and $-$ separately. Note that the definition of $\kappa_2^+(\rho)$ is consistent with the definition of Proposition 3.5.

Proof of Corollary 3.1. The above proposition can be used but the work has already been done to prove Proposition 3.4. It is not difficult to show that

$$(m_1(t), m_2(t)) = \left(\mathbb{E}\left(\overline{T}_1(t)\right), \mathbb{E}\left(\overline{T}_2(t)\right)\right)$$

satisfies Relation (3.31) with $p(t)=m_1(t)$, i.e. so that $g(t)=0$ with the notations of the proof of Proposition 3.4. One concludes by using Relations (3.33) and (3.34). □

A Bound on the Exponential Decay Rate of the Algorithm

Since the duplication mechanism associated to the process $(\mathcal{B}_f^N(t))$ is more active than for our algorithm, intuitively the decay rate of our system should be faster that the decay rate of the process $(\mathcal{B}_f^N(t))$. The following lemma will be used to establish rigorously this relation.

Lemma 3.3. *There exists a coupling of the processes $(\mathcal{A}^N(t))$ and $(\mathcal{B}^N(t))$ such that, almost surely, for all $f\in\mathcal{F}_N$ and $t\geq 0$, $\mathcal{A}_f^N(t)\subset\mathcal{B}_f^N(t)$.*

Proof. This is done by induction on the number of jumps of the process $(\mathcal{B}^N(t))$. By assumption one can take $\mathcal{A}^N(0)=\mathcal{B}^N(0)$. If the relation $\mathcal{A}_f^N(t)\subset\mathcal{B}_f^N(t)$, at the instant $t=\tau_n$ of the nth jump of $(\mathcal{A}^N(t))$ and $(\mathcal{B}^N(t))$. We review the different scenarios for the next jump after time τ_n, at time τ_{n+1},

1. if some node $i_0\in\{1,\ldots,N\}$ fails, then, for $f\in\mathcal{F}_N$,

$$\mathcal{A}_f^N(\tau_{n+1})=\mathcal{A}_f^N(\tau_n)\setminus\{i_0\} \text{ if } i_0\in\mathcal{A}_f^N(\tau_n), \qquad \mathcal{A}_f^N(\tau_{n+1})=\mathcal{A}_f^N(\tau_n) \text{ otherwise,}$$

   and a similar relation holds for $\mathcal{B}_f^N(\tau_{n+1})$. The relation $\mathcal{A}_f^N(t)\subset\mathcal{B}_f^N(t)$ still holds for $t=\tau_{n+1}$ since it is true at time τ_n.

2. If a duplication occurs for the process $(\mathcal{A}^N(t))$ at time τ_{n+1} at some node $i_0\in\{1,\ldots,N\}$ and for file a $f\in\mathcal{F}_N$, In particular $i_0\in\mathcal{A}_f^N(\tau_n)$ and therefore, by induction hypothesis, $i_0\in\mathcal{B}_f^N(\tau_n)$, so that we can couple both the duplication process for both processes $(\mathcal{A}^N(t))$ and $(\mathcal{B}^N(t))$ as follows
   — If $\text{card}(\mathcal{B}_f^N(\tau_n))<d$. A copy is made on the same node for both processes $(\mathcal{A}^N(t))$ and $(\mathcal{B}^N(t))$.
   — If $\text{card}(\mathcal{B}_f^N(\tau_n))=d$. There exists some node i_0 such that $i_0\notin\mathcal{A}_f^N(\tau_n)$ and $i_0\in\mathcal{B}_f^N(\tau_n)$. We can then set $\mathcal{A}_f^N(\tau_{n+1})=\mathcal{A}_f^N(\tau_n)\cup\{i_0\}$, remember that for the process $(\mathcal{B}^N(t))$ the servers where to make a copy are also chosen at random.
   In both cases the relation $\mathcal{A}_f^N(\tau_{n+1})\subset\mathcal{B}_f^N(\tau_{n+1})$ will hold.

3. If a duplication occurs for the process $(\mathcal{B}^N(t))$ at time τ_{n+1} but not for the process $(\mathcal{A}^N(t))$ then, clearly, the desired relation will then also hold at time τ_{n+1}.

□

The following proposition gives an estimation of the rate of decay of the network.

Proposition 3.6. *If $L^N(t)$ is the number of files alive at time t,*

$$L^N(t) = \sum_{f \in \mathcal{F}_N} \mathbb{I}_{\left\{\mathcal{A}_f^N(t) \neq \emptyset\right\}},$$

then there exists a constant $K_1 > 0$ such that, for all $t \geq 0$,

$$\limsup_{N \to +\infty} \mathbb{E}\left(\frac{L^N(t)}{N}\right) \leq K_1 e^{-\mu \kappa_d^+(\rho) t},$$

where $\kappa_d^+(\rho)$ is the constant defined in Proposition 3.5.

Proof. By using the coupling of the last proposition, one gets

$$\begin{aligned}\mathbb{E}\left(L^N(t)\right) &\leq \sum_{f \in \mathcal{F}_N} \mathbb{P}\left(\mathcal{B}_f^N(t) \neq \emptyset\right) \\ &\leq \sum_{f \in \mathcal{F}_N} \mathbb{E}\left(\mathcal{B}_f^N(t)\right) = \sum_{i=1}^{N} \sum_{k=1}^{d} \frac{1}{k} \mathbb{E}\left(T_{i,k}^N(t)\right) = N \sum_{k=1}^{d} \frac{1}{k} \mathbb{E}\left(T_{1,k}^N(t)\right).\end{aligned}$$

Theorem 3.3 gives the convergence of the sequence of processes $(T_{1,k}^N(t))$ to the solution of the EDS (3.35), (3.37) and (3.37). It is not difficult to obtain an analogue of Lemma 3.1 which guarantee the boundedness of the second moments of the variables $T_{1,k}^N(t)$, $k = 1,\ldots,d$, which gives the convergence of the first moments. One has obtained the relation,

$$\limsup_{N \to +\infty} \mathbb{E}\left(\frac{L^N(t)}{N}\right) \leq \sum_{k=1}^{d} \frac{1}{k} \mathbb{E}\left(\overline{T}_k(t)\right),$$

one concludes with Inequality (3.40). The proposition is proved. □

Chapter 4

Analysis of Large Urn Models with Local Mean-Field Interactions

Contents

4.1 Motivations: allocation in large storage distributed system 102

4.1.1 Introduction 102

4.1.2 Placement policies 104

4.1.3 Simulations 105

4.1.4 Mathematical Models 108

4.1.5 Conclusion 118

4.2 Large Urn Models with Local Mean-Field Interactions 123

4.2.1 Introduction 123

4.2.2 The Stochastic Model 127

4.2.3 Evolution Equations 133

4.2.4 Mean-Field Convergence Results 138

4.2.5 Equilibrium of the McKean-Vlasov Process 148

This chapter consists of two sections.

The fist section is devoted to the analysis of the efficiency of allocation algorithms in a large storage distributed system. Three allocation algorithms are tested against simulations in the context of a real implementation of a DHT network: Random, Least Loaded and Power of Choice. A mathematical model is presented to explain the phenomenon observed in the simulations.

In the second section, we have developed a large urn model with local mean-field interactions to study general local allocation algorithms on a abstract symmetrical graph. We present a thorough study on the large scale, long time and heavy load behaviors of this model. Results in the first section can be seen as an application of results in the second section.

4.1 Motivations: allocation in large storage distributed system

Abstract Distributed storage systems such as Hadoop File System or Google File System (GFS) ensure data availability and durability using replication. Persistence is achieved by replicating the same data block on several nodes, and ensuring that a minimum number of copies are available on the system at any time. Whenever the contents of a node are lost, for instance due to a hard disk crash, the system regenerates the data blocks stored before the failure by transferring them from the remaining replicas. This section is focused on the analysis of the efficiency of replication mechanism that determines the location of the copies of a given file at some server. The variability of the loads of the nodes of the network is investigated for several policies. Three replication mechanisms are tested against simulations in the context of a real implementation of a DHT network: Random, Least Loaded and Power of Choice.

The simulations show that some of these policies may lead to quite unbalanced situations: if β is the average number of copies per node it turns out that, at equilibrium, the load of the nodes may exhibit a high variability. It is shown in this section that a simple variant of a power of choice type algorithm has a striking effect on the loads of the nodes: at equilibrium, almost 100% of the nodes are less than 2β. Furthermore, the load of a given node has in fact a uniform distribution between 0 and 2β. In particular the load of a random node of the network for this algorithm is bounded by 2β which is an interesting property for the design of the storage space of these systems.

A mathematical model is presented to explain this unusual, quite surprising, phenomenon. The analysis of these systems turns out to be quite complicated mainly because of the large dimensionality of the state spaces involved. Our study relies on probabilistic methods to analyze the asymptotic behavior of an arbitrary node of the network when the total number of nodes gets large. An additional ingredient is the use of stochastic calculus with marked Poisson point processes to establish some of our results.

4.1.1 Introduction

For scalability, performance or for fault-tolerance concerns in distributed storage systems, the pieces of data are spread among many distributed nodes. Most famous distributed data stores include Google File System (GFS) [66], Hadoop Distributed File System (HDFS) [21], Cassandra [101], Dynamo [43], Bigtable [34], PAST [144] or DHASH [40].

Distributed data storage permits to enhance access performance by spreading the load among many nodes and by placing conveniently pieces of data. It can also improve fault tolerance by maintaining multiple copies of each piece of data. However, while implementing a distributed data store, many problems have to be tackled. For instance, it is necessary to efficiently locate a given piece of data: to balance the storage load evenly among nodes, to maintain consistency and the fault-tolerance level. While consistency and fault-tolerance in replicated data stores are widely studied, the storage load balance has received little importance despite of its importance.

The distribution of the storage load among the storing nodes is a critical issue. On a daily basis, new pieces of data have to be stored and while a failure occurs, maintenance mechanisms are supposed to create and store new copies to replace the lost ones. A key

feature of these systems is that the storage infrastructure itself is dynamic: nodes may crash and new nodes may be added. If the placement policy used does not balance the storage load evenly among nodes, the imbalance will become harmful. The overloaded nodes may have to serve much more requests than the others, and in case of failure, the recovery procedure will take more time, increasing the probability to lose data.

To circumvent this issue most systems rely on data redistribution. They use a hash function in the case for distributed hash tables (DHTs) [144, 40]. However, as shown in previous studies, these systems imply many data movements and may lose data under churn [104]. Rodrigues and Blake have shown that classical DHTs storing large amounts are usable only if the node lifetime is of the order of several days [140].

In this section we study data placement policies avoiding data redistribution: once a piece of data is assigned to a node, it will remain on it until the node crashes. We focus specifically on the evaluation of the impact of several placement strategies on the storage load balance on a long term. To the best of our knowledge, there are few papers devoted to the analysis of the evolution of the storage load of the nodes of a DHT system on such a long term period. Our investigation has been done in two complementary steps.

1. A simulation environment of a real system to simulate several years of evolution of this system for three placement policies which are defined below: Random, Least Loaded and Power of Choice. See Figures 4.1 and 4.2.
2. Simplified mathematical models to explain some of the surprising results obtained through simulations for the Random and Power of Choice Policies. We show that for a large network, if β is the average load per node and $\overline{X}^R_\beta$, [resp. $\overline{X}^P_\beta$] is the load of a node at equilibrium for the random policy [resp. power of choice policy] then, for $x \geq 0$,

$$\lim_{\beta\to+\infty} \mathbb{P}\left(\frac{\overline{X}^R_\beta}{\beta} \geq x\right) = e^{-x}$$
$$\lim_{\beta\to+\infty} \mathbb{P}\left(\frac{\overline{X}^P_\beta}{\beta} \geq x\right) = \begin{cases} 1-x/2 & \text{if } x < 2 \\ 0 & \text{if } x \geq 2. \end{cases} \tag{4.1}$$

   See Theorems 4.2 and 4.3 below. The striking feature is that, for the power of choice policy, the load of a node has, in the limit, a *finite support* $[0, 2\beta]$ for a large average load per node β. This is an important and desirable property for the design of such systems, to dimension the storage of the nodes in particular. Note that this is not the case for the random policy. The simulations of a real system exhibit this surprising phenomenon, even for moderately large loads, see Figure 4.2.

   It should be noted that, usually, power of choice policies used in computer science and communication networks are associated with $\log\log N$ loads instead of $\log N$ loads, see Mitzenmacher [112], or with double exponential decay for tail distributions of the load instead of simple exponential decay, see Vvedenskaya et AI. [165]. Here the phenomenon is the asymptotic finite support instead of an exponential decay.

   A mathematical model is presented to explain this unusual, quite surprising, phenomenon. The analysis of these systems turns out to be quite complicated

mainly because of the large dimensionality of the state spaces involved. Our study relies on probabilistic methods to analyze the asymptotic behavior of an arbitrary node of the network when the total number of nodes gets large. An additional ingredient is the use of stochastic calculus with marked Poisson point processes to establish some of our results.

The section is organized as follows. The main placement policies are introduced in Section 4.1.2. Section 4.1.3 describes the simulation model and presents the results obtained with the simulator. Concerning mathematical models, the Random policy is analyzed in Section 4.1.4 and Power of Choice policy in Section 4.1.4. All (quite) technical details of the proofs of the results for the random policy are included. This is not the case for the power of choice policy, for sake of simplicity and due to the much more complex framework of general mean-field results, convergence results of the sample paths (Proposition 4.4) and of the invariant distributions (Proposition 4.6) are stated without proof. A reference is provided. The complete proof of the important convergence (4.1) is provided in both cases nevertheless.

4.1.2 Placement policies

To each data block is associated a *root node*, a node having a copy of the block in charge of its duplication if necessary. During the recovery process to replace a lost copy, the root node has to choose a new storage node within a dedicated set of nodes, *the selection range* of the node. Any node of this subset that does not already store a copy of the same data block may be chosen. Three policies of placement are investigated.

Least Loaded Policy

For this algorithm the root node of the data block selects the least loaded node, in terms of storage load, of its selection range not already storing a copy of the same data block. This strategy clearly aims at reducing the variation of storage loads among nodes. As it has been seen in earlier studies, this policy has a bad impact on the systems reliability,see [147]. Indeed, a node having a small storage load will be chosen by all its neighbors in the ring. Furthermore, this policy implies for a root node to monitor the load of all the nodes of its selection range, which may be costly. It is nevertheless in terms of placement an optimal policy. It is used in this section as a reference for comparison with the other policies.

Random Policy

The node chooses uniformly at random a new storage node within its selection range among nodes not already hosting a copy of the same data block.

Power of Choice Policy

For this algorithm, the root node chooses, uniformly at random, two nodes of its selection range not storing a copy of the data block. It selects the least loaded among the two.

It is inspired of algorithms studied by Mitzenmacher and others in the context of static allocation schemes of balls into bins in computer science, see [112] for a survey. In queueing theory, a similar algorithm has been investigated by the seminal Vvedenskaya et al. [165] in 1996. There is a huge literature on these algorithms in this context. Our framework is quite different, the placement is dynamic, data blocks have to move because of crashes, and the number of requests is constant in the system contrary to the queueing models. The idea is nevertheless the same: reducing the load by just checking some finite subset instead of all possibilities.

4.1.3 Simulations

Our simulator is based on PeerSim [82], see also [114]. It simulates a real distributed storage system. Every node, every piece of data, and every transfer is represented. Each piece of data is replicated and each copy is assigned to a different storage node.

System model We have simulated N nodes, storing F_N^* data blocks with a fixed size s and replicated k times. The nodes and the data blocks are assigned unique identifiers (*id*). The nodes are organized according to their identifiers, forming a virtual ring, as it is usual in distributed hash tables (DHTs) [144, 40]. To each data block is associated a *root node*, a node having a copy of the block in charge of its duplication if necessary. See below.

Failure model Failures in the systems occur according to a Poisson process with a fixed mean of seven days. The failures are crashes: a node acts correctly until it fails. After a crash it stops and never comes back again (fail-stop model). All the copies stored become unavailable at that time. To maintain the number of nodes constant equal to N, each time a node fails, an empty node with a new *id* joins the system at the same position in the ring of nodes.

Simulation parameters In the simulations, based on PeerSim, the parameters have been fixed as follows:

— The number of nodes N=200,
— the number of data blocks F_N^*=10000,
— the block size s=10MB,
— the replication degree of data blocks k=3,
— the mean time between failures (MTBF) is 7 days.

The network latency is fixed to $0.1s$ and the bandwidth is $5.5Mbps$.

At the beginning of each simulation, the N nodes are empty and the F_N^* blocks are placed using the corresponding policy and the system is simulated for a period of 2 years. We have studied the storage load distribution and its time evolution. With these parameters, the average load is $\beta = k \times F_N^* / N = 150$ blocks per node. The optimal placement from the point of view of load balancing would consist of having 150 blocks at every node. We will investigate the deviation from this scenario for the three policies.

Network simulation The impact of policies on bandwidth management has been carefully monitored. In case of failure, many data blocks have to be transferred among a subset of nodes to repair the system. Taking into account bandwidth limitation and network contention is crucial since a delayed recovery may lead to the loss of additional blocks because of additional crashes in the meantime.

Recovery mechanism Concerning the data blocks it is root for, a node is responsible:

— to regularly check the presence of all copies;
— In case of failure of a node having a copy of a data block, it has to select a node. This node has to transfer the data block from one of the remaining copies in the system.

Note that in case of failure of a node, lost data blocks will involve, in general, different root nodes. Each root node is only responsible for the recovery of the data blocks it is root for. The mechanism in charge of the failure of root nodes (each node is both root for many data blocks and stores copies for many data blocks) is beyond the scope of this section, see, for example, the work on RelaxDHT [104]. What is important here is that, each time a node fails it is eventually detected, and for each data block it stores, a new node is chosen to store a new copy replacing the lost one. See Section 4.1.4 for a discussion of this assumption from the point of view of the mathematical model.

We have also observed that the wider the selection range, the faster the recovery process is. Intuitively, this can be easily explained as follows: if the number of sources available for transfers is large then the local traffic at each of the nodes will be reduced thereby reducing congestion and therefore delays and, potentially additional losses. We now focus on the placement policies within the selection range.

Simulation results

Figure 4.1 shows the evolution of the average load of a node with respect to the duration of its lifetime within the network. One can conclude that:

— For the *Least Loaded* strategy, the load remains almost constant and equal to the optimal value 150 . By systematically choosing the least loaded node within the selection range to store a data block copy, the storage load tends to be constant among nodes.
  As observed in simulations, this policy has however two main drawbacks. First, it requires that nodes maintain an up-to-date knowledge of the load of all the nodes within their selection range. Second, it is more likely to generate network contention for the following reason: If one of the nodes is really underloaded, it will receives most of the transfer requests of its neighborhood. See [147].
— For the *Random* strategy, the load increases linearly until the failure of the node.This is an undesired feature since it implies that the failure of "old" nodes will imply in this case a lot of transfers to recover the large number of blocks lost.
— The growth of the *Power of Choice* policy is slow as it can be seen from the figure. It should be noted that, contrary to the Least Loaded Policy, the required information to allocate data blocks is limited. Furthermore, the random choices

of nodes for allocation spread the load from the point of view of the network contention.

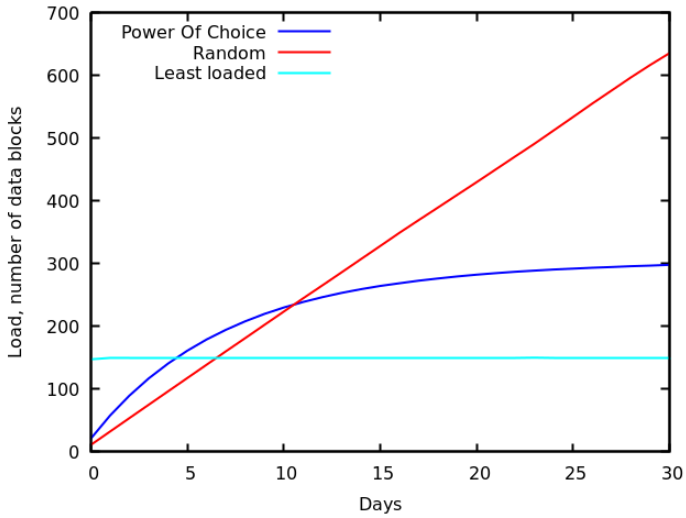


Figure 4.1 – Evolution of the Average Load of a Node with Respect to its Age in the System.

The distribution function of the storage loads after two simulated years is presented in Figure 4.2 . For clarity, the figure has been truncated. Each point of each policy has been obtained with 210 runs. At the beginning, the data block copies are placed using the corresponding strategy. After two years of failures and reparations, one gets that:

— The *Random* strategy presents a highly non-uniform distribution profile, note that more 10% of the nodes have a loaded greater than 350. This is consistent with our previous remark on the fact that old nodes are overloaded.
— For the *Least Loaded* strategy, as expected, the load is highly concentrated around 150.
— The striking feature concerning the *Power of Choice* policy is that the load of a node seems to follow a uniform distribution between 0 and 300. In particular almost all nodes have a load bounded by 300 which is absolutely remarkable.

Table 4.1 gives the maximum loads that have been observed for each strategy over 132 090 samples: starting from day 100, the maximal load has been measured and recorded every day, this for the 210 runs. We can see that the mean maximum load of the random strategy is already high (more than five times the average), and furthermore, the load varies a lot, the maximum measured load being 2188 data blocks. This implies that, with the random strategy, the storage devices for each node has to be over-sized, recall that the average load is 150 data blocks.

As a conclusion, the simulations show that, with a limited cost in terms of complexity, the power of choice policy has remarkable properties. The load of each node is bounded by 300. It may be remarked that each possible load between 0 and 300 is represented by the same amount of nodes in average. Figure 4.2 shows that there is approximately the same number of nodes having 0 data blocks, than nodes having

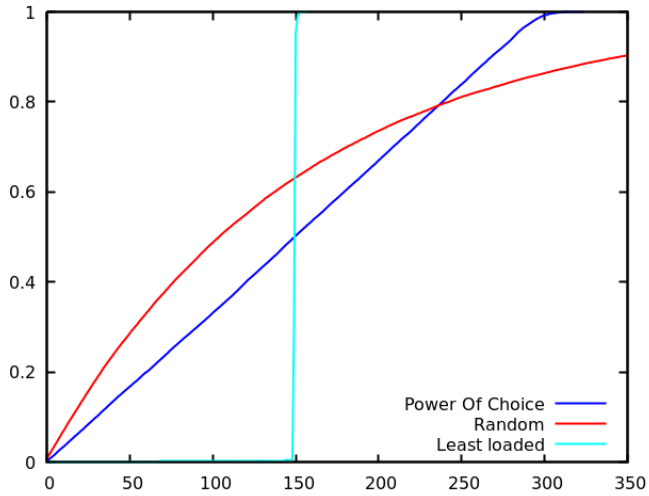


Figure 4.2 – Distribution function of load distribution of a random node after 729 days.

Strategy	Mean of Max	Min	Max
Least Loaded	153	150	165
Random	864	465	2188
Power of Choice	300	269	328

Table 4.1 – Statistics of Extremal Loads

150 data block or nodes having 300 data blocks. Note that this is a stationary state. Additionally, the variation is low, we can observe in Table 4.1 that upon the 132,090 samples, the most loaded node was never above 328. From a practical point of view, it means that, at the cost of a slightly oversized storage device at each node, a bit more than twice the average, is enough to guarantee the durability of the system.

In the following sections we investigate simplified mathematical models of two placement policies: Random and Power of Choice. The goal is of explaining these striking qualitative properties of these policies.

4.1.4 Mathematical Models

The main goal of the section is of investigating the performances of duplication algorithms in terms of the overhead for the loads of the nodes of the network. For simplicity we will assume that the replication rate is 2, each data block has at most two copies. Without loss of generality, we will assume that the breakdown of each server occurs according to a Poisson process with rate 1. After a breakdown, a server restarts empty (in fact a new server replaces it). A complete Markovian description of such a system is quite complex. Indeed, if there are N servers and F_N^* initial data blocks, for $1 \leq i \leq F_N^*$, the locations of the ith data block are either the index of two distinct servers if there are two copies, or the index of one server if only one copy is present, or 0 if the data block has been lost for good. A simple calculation shows that the size of the state space of the Markov process is at least of the order of $(N^2/2)^{F_N^*}$ which is huge if

it is remembered that F_N^* is of the order of N. For this reason, we shall simplify the mathematical model.

Assumption on Duplication Rate In order to focus specifically on the efficiency of the replacement strategy from the point of view of the distribution of the load of an arbitrary node, we will study the system with the assumption that it does not loose files. We will only track the location of the node of each copy of a data block with a simplifying assumption: just before when a node fails, all the copies it contains are allocated to the other nodes with respect to the algorithm of placement investigated. In this way, every data block has always two copies in the system.

Note that this system is like the original one by assuming that the time to make a new copy is instantaneous. Once a server has failed, a copy of each of the data blocks it contains is produced immediately with the remaining copy in the network. With this model, a copy could be done on the same node as the other copy, but this occurs with probability $1/(N-1)$, it can be shown that this has a negligible effect at the level of the network, in the same way as in Proposition 4.1 below. This approximation is intuitively reasonable to describe the original evolution of the system when few data blocks are lost. As we will see, qualitatively, its behavior is close to the observed simulations of the real system, few files were lost after two years.

Now $F_N = 2F_N^*$ denotes the total number of copies of files, it is assumed that, for some $\beta > 0$,

$$\lim_{N\to+\infty} F_N/N = \beta.$$

β is therefore the average load per server. For $1 \leq i \leq N$, $L_i^N(0)$ is the initial number of copies on server i. Throughout the section, it is assumed that that the distribution of the variables $(L_i^N(0))$ of the initial state are invariant by any permutation of indices, i.e. it is an exchangeable vector, and that

$$\sup_{N\geq 1} \frac{\mathbb{E}\left(L_1^N(0)^2\right)}{N} < +\infty. \tag{4.2}$$

Note that this condition is satisfied if we start with an optimal exchangeable allocation, i.e. for which, for all $1\leq i\leq N$, $L_i^N(0)\in\{k_N-1, k_N\}$ with $k_N=\lceil F_N/N\rceil$.

The Random Allocation

For $1 \leq i \leq N$, we denote by $\overline{\mathcal{N}}_i = (t_n^i, \overline{U}_n^i)$ the marked Poisson point process defined as follows:

— (t_n^i) is a Poisson process on $\mathbb{R}_+$ with rate 1;
— $\overline{U}_n^i=(U_p^{i,n})$ is an i.i.d. sequence of uniform random variables on the subset $\{1,\ldots,N\}\setminus\{i\}$.

For $1 \leq i \leq N$ and $n \geq 1$, t_n^i is the instant of the nth breakdown of server i. For $p \geq 1$, $U_p^{i,n}$ is the server where the pth copy present on node i is allocated after this breakdown. The random variables $\overline{\mathcal{N}}_i$, $1\leq i\leq N$ are assumed to be independent. Concerning marked Poisson point processes, see Page 130 in next subsection.

One will use an integral representation for these processes, if $\mathcal{M}_U=\{1,\dots,N\}^{\mathbb{N}}$ and $f:\mathbb{R}_+\times\mathcal{M}_U\to\mathbb{R}_+$,

$$\sum_{n\geq 1} f(t_n^i,(U_p^{i,n})) = \int_{t=0}^{+\infty}\int_{\overline{u}=(u_p)\in\mathcal{M}_U} f(t,\overline{u})\overline{\mathcal{N}}_i(\mathrm{d}t,\mathrm{d}\overline{u}).$$

Equations of Evolution For $1\leq i\leq N$ and $t\geq 0$, $L_i^N(t)$ is the number of copies on server i at time t. The dynamics of the random allocation algorithm is represented by the following stochastic differential equation, for $1\leq i\leq N$,

$$\mathrm{d}L_i^N(t) \stackrel{\text{def.}}{=} L_i^N(t)-L_i^N(t-) = -L_i^N(t-)\overline{\mathcal{N}}_i(\mathrm{d}t,\mathcal{M}_U) + \sum_{m=2}^{N} z_i(L_m^N(t-),\overline{u})\overline{\mathcal{N}}_m(\mathrm{d}t,\mathrm{d}\overline{u}) \quad (4.3)$$

where $z_i:\mathbb{N}\times\mathcal{M}_U\mapsto\mathbb{N}$ is the function

$$z_i(\ell,\overline{u}) = \mathbb{I}_{\{u_1=i\}}+\mathbb{I}_{\{u_2=i\}}+\cdots+\mathbb{I}_{\{u_\ell=i\}}. \quad (4.4)$$

The first term of the right hand side of Relation (4.3) corresponds to a breakdown of node i, all files are removed from the node. The second concerns the files added to node i when other servers break down and send copies to node i. Note that the ith term of the sum is always 0.

Denote $L^N(t)=(L_i^N(t),1\leq i\leq N)\in\mathcal{S}=\mathbb{N}^N$, then clearly $(L^N(t))$ is a Markov process. Not that because of the symmetry of the initial state and of the dynamics of the system, the variables $L_i^N(t)$ have the same distribution and since the sum of these variables is F_N, one has in particular $\mathbb{E}(L_i^N(t))=F_N/N$ for all N, i and $t\geq 0$. In the following, we will only consider the process $(L_1^N(t))$ due to this property of symmetry.

The integrand in the second term of the right hand side of Equation (4.3) has a binomial distribution with parameter $L_m^N(t)$ and $1/(N-1)$ and the sum of these terms is $F_N/(N-1)$ which is converging to β. Hence, this suggests that this second term could be a Poisson process with rate β. The process $(L_1^N(t))$ should be in the limit, a jump process with a Poissonnian input and return to 0 at rate 1. This is what we are going to prove now.

By integrating Equation (4.3) on gets the relation

$$L_1^N(t) = L_1^N(0) - \int_0^t L_1^N(s)\,\mathrm{d}s + \frac{1}{N-1}\sum_{m=2}^{N}\int_0^t L_m^N(s)\,\mathrm{d}s + M_1^N(t), \quad (4.5)$$

where $(M_1^N(t))$ is the martingale

$$M_1^N(t) = -\int_0^t L_1^N(s-)\left[\overline{\mathcal{N}}_1(\mathrm{d}t,\mathcal{M}_U)-\mathrm{d}s\right] + \sum_{m=2}^{N}\int_0^t\left[z_1\left(L_m^N(s-),\overline{u}\right)\overline{\mathcal{N}}_m(\mathrm{d}s,\mathrm{d}\overline{u}) - \frac{L_m^N(s)}{N-1}\,\mathrm{d}s\right].$$

The following proposition shows that the process $(L_1^N(t))$ does not have jumps of size ≥ 2 on a finite time interval with high probability.

Proposition 4.1. *For $T > 0$ then*

$$\lim_{N\to+\infty} \mathbb{P}\left(\sup_{t\in[0,T]} \mathrm{d}L_1^N(t) \geq 2\right) = 0.$$

Proof. For $0<\varepsilon<1$, from Equation (4.15) in the Appendix, there exists $K>0$ such that $\mathbb{P}(\mathcal{E}_{N,K})\geq 1-\varepsilon$ holds for all $N\geq 2$, if

$$\mathcal{E}_{N,K} = \left\{\sup(L_1^N(t) : 0 \leq t \leq T) \leq K\right\}.$$

On the event $\mathcal{E}_{N,K}$ the probability that a failure of some node will send more than 2 new copies on node 1 is upper bounded by $(K/N)^2$. Since the total number of failures on the time interval $[0, T]$ affecting node 1 has a Poisson distribution with parameter $N-1$, one obtains that the probability that $(L_1^N(t))$ has a jump of size at least 2 on $[0, T]$ is bounded by K^2/N hence goes to 0 as N gets large. The proposition is proved. □

Convergence to a Simple Jump Process Define

$$\mathcal{P}^N[0,t] = \int_0^t \sum_{m=2}^N z_i \left(L_m^N(s-), \overline{u}\right) \overline{\mathcal{N}}_m(\mathrm{d}s, \mathrm{d}\overline{u}),$$

this is a counting process with jumps of size 1. Define

$$C_P^N(t) = \frac{1}{N-1} \sum_{m=2}^N \int_0^t L_m^N(s)\ \mathrm{d}s$$

then $(C_P^N(t))$ is the compensator of $(\mathcal{P}^N[0,t])$ in the sense that it is a previsible process and that $(\mathcal{P}^N[0,t]-C_P^N(t))$ is a martingale. The proof is analogous to the proof of Proposition 4.7 in the Appendix.

Proposition 4.2. *If the initial distribution of $(L_i^N(t))$ satisfies Condition (4.2) then, for the convergence in distribution of processes,*

$$\lim_{N\to+\infty} (C_P^N(t)) = (\beta t).$$

Proof. We first prove that the sequence $(C_P^N(t))$ is tight for the convergence in distribution with the topology of the uniform norm. By using that the sum of the L_m's is less than F_N, for $0 \leq s \leq t \leq T$,

$$|C_P^N(t) - C_P^N(s)| \leq \frac{1}{N-1} \int_s^t \sum_{m=2}^N L_m^N(u)\ \mathrm{d}u \leq \frac{F_N}{N-1}(t-s).$$

Hence for any $\eta>0$ and $\varepsilon>0$, there exists some $\delta > 0$ such that, for all $N \geq 1$,

$$\mathbb{P}\left(\sup_{\substack{0\leq s,t\leq T\\ |t-s|\leq\delta}} \left|C_P^N(t) - C_P^N(s)\right| \geq \eta\right) \leq \varepsilon.$$

The sequence $(C_P^N(t))$ satisfies the criterion of the modulus of continuity, see Theorem 7.2 page 81 of Billingsley [19]. The property of tightness has been proved.

The symmetry of the variables $(L_m(t))$ and the fact that their sum is F_N give that

$$\mathbb{E}(C_P^N(t)) = \int_0^t \mathbb{E}\left(L_1^N(s)\right) \mathrm{d}s = \frac{F_N}{N} t.$$

Hence, the sequence $(\mathbb{E}(C_P^N(t)))$ is converging to βt.

By using again the same arguments, one has

$$\begin{aligned}\mathbb{E}\left(C_P^N(t)^2\right) =& \mathbb{E}\left(\int_{[0,t]^2}\left[\frac{1}{N-1}\sum_{1\le m\le N} L_m^N(s)\right]\left[\frac{1}{N-1}\sum_{1\le m\le N} L_m^N(s')\right]\mathrm{d}s\,\mathrm{d}s'\right)\\ &- 2\mathbb{E}\left(\int_{[0,t]^2}\frac{L_1^N(s)}{N-1}\left[\frac{1}{N-1}\sum_{1\le m\le N} L_m^N(s')\right]\mathrm{d}s\,\mathrm{d}s'\right) + \mathbb{E}\left(\int_{[0,t]}\frac{L_1^N(s)}{N-1}\,\mathrm{d}s\right)^2\\ =& \left(\frac{F_N}{N-1}t\right)^2 - 2\int_0^t \frac{\mathbb{E}(L_1^N(s))F_N t}{(N-1)^2}\,\mathrm{d}s + \mathbb{E}\left(\int_{[0,t]}\frac{L_1^N(s)}{N-1}\,\mathrm{d}s\right)^2. \quad (4.6)\end{aligned}$$

With Lemma 4.1 of Appendix and Cauchy-Shwartz's Inequality, one obtains therefore that the second moment of $C_P^N(t)$ is converging to $(\beta t)^2$, hence

$$\lim_{N\to+\infty} \mathbb{E}\left(\left(C_P^N(t)^2 - \beta t\right)^2\right) = 0.$$

One concludes that finite marginals of the process $(C_P^N(t))$ converge to the corresponding marginals of (βt). Consequently, (βt) is the only limiting point of the sequence $(C_P^N(t))$ for the convergence in distribution. The tightness property gives therefore the desired convergence. The proposition is proved. □

Theorem 4.1. *If the initial distribution of $(L_i^N(t))$ satisfies Condition (4.2) and converges to some distribution π_0, then, for the convergence in distribution,*

$$\lim_{N\to+\infty} (L_1^N(t)) = (X_\beta^R(t)),$$

where $(X_\beta^R(t))$ is a jump process on $\mathbb{N}$ with initial distribution π_0 whose Q-matrix $Q=(q(x,y))$ is given by, for $x\in\mathbb{N}$, $q(x,x+1)=\beta$ and $q(x,0)=1$.

See also Proposition 4.14 below.

Proof. By using Proposition 4.1, Proposition 4.2 of the Appendix and Theorem 5.1 of [88], one concludes that the sequence of point processes $(\mathcal{P}^N[0,t])$ is converging in distribution to a Poisson process $\mathcal{N}_\beta$ with rate β.

Recall that $\overline{\mathcal{N}}_1(\mathrm{d}t, \mathcal{M}_U) = (t_n^1)$, from SDE (4.3), one has

$$\mathrm{d}L_i^N(t) = -L_i^N(t-)\overline{\mathcal{N}}_i(\mathrm{d}t, \mathcal{M}_U) + \mathcal{P}^N(\mathrm{d}t),$$

thus, for $t>0$,

$$L_1^N(t) = \mathcal{P}^N(t_n^1, t] \text{ if } t_n^1 \le t < t_{n+1}^1.$$

The convergence we have obtained shows that $(L_1^N(t))$ is converging in distribution to $(\overline{L}_1(t))$ where

$$\overline{L}_1(t) = \mathcal{N}_\beta(t_n^1, t] \text{ if } t_n^1 \leq t < t_{n+1}^1.$$

This is the desired result. □

Proposition 4.3. *The equilibrium distribution of $(L_1^N(t))$ is converging in distribution to $\overline{X}_\beta^R$, a geometrically distributed random variable with parameter $\beta/(1+\beta)$.*

Proof. Denote by π_β^N the invariant distribution of the process $(L_1^N(t))$. By symmetry, we know that

$$\mathbb{E}_{\pi_\beta^N}\left(L_1^N(0)\right) = \frac{F_N}{N},$$

hence the sequence of probability distributions (π_β^N) is tight. Let π be some limiting point of this sequence for some subsequence (N_k). If f is some function on $\mathbb{N}$ with finite support, then the cycle formula for the invariant distribution of the ergodic Markov process $(L_1^N(t))$ gives

$$\mathbb{E}_{\pi_\beta^N}(f(L_1)) = \mathbb{E}_{\widehat{\pi}_\beta^N}\left(\int_0^{t_1^1} f(L_1^N(s))\ \mathrm{d}s\right),$$

where $\widehat{\pi}_\beta^N$ is the distribution of $(L_i^N(t))$ at the instants of jumps of breakdowns of node 1. In particular,

$$\widehat{\pi}_\beta^N\left(\ell=(\ell_i)\in\mathbb{N}^N,\ \ell_1=0\right) = 1.$$

By Proposition 4.8 of Appendix, Theorem 4.1 is also true when the initial distribution is $\widehat{\pi}_\beta^N$ hence, for the convergence in distribution,

$$\lim_{N\to+\infty}\left(\int_0^{t_1^1} f(L_1^N(s))\ \mathrm{d}s\right) = \left(\int_0^{t_1^1} f(X_\beta^R(s))\ \mathrm{d}s\right)$$

when the process $(X_\beta^R(t))$ has initial point 0. Consequently, by Lebesgue's Theorem,

$$\mathbb{E}_\pi(f) = \lim_{N\to+\infty}\mathbb{E}_{\pi_\beta^N}(f(L_1)) = \mathbb{E}_0\left(\int_0^{t_1^1} f(X_\beta^R(s))\ \mathrm{d}s\right).$$

The last term of this equation is precisely the invariant distribution of $(X_\beta^R(t))$, again with the cycle formula for ergodic Markov processes. The probability π is necessarily the invariant distribution of $(X_\beta^R(t))$, hence the sequence (π_β^N) is converging to π. It is easily checked that π is a geometric distribution with parameter $\beta/(1+\beta)$. □

By using the fact that $\mathbb{P}(\overline{X}_\beta^R \geq n) = (\beta/(1+\beta))^n$, it is then easy to get the following result.

Theorem 4.2 (Equilibrium at High Load)**.** *The convergence in distribution*

$$\lim_{\beta\to+\infty}\frac{\overline{X}_\beta^R}{\beta} = E_1,$$

holds, with E_1 an exponential random variable with parameter 1.

In particular the probability that, at equilibrium, the load of a given node has more than twice the average load is

$$\lim_{\beta\to+\infty} \mathbb{P}\left(\overline{X}^R_\beta \geq 2\beta\right) = \exp(-2) \sim 0.135,$$

which is consistent with the simulations, see Figure 4.2.

The Power of Choice Algorithm

Similarly as before, for $1 \leq i \leq N$, $\overline{\mathcal{N}}_i = (t^i_n, (\overline{V}^i_n))$ denotes the marked Poisson point process defined as follows:

— (t^i_n) is a Poisson process on $\mathbb{R}_+$ with rate 1;

— $\overline{V}^i_n = (V^{i,n}_p) = ((V^{i,n}_{0,p}, V^{i,n}_{1,p}, B^{i,n}_p))$ where $(V^{i,n}_{0,p}, V^{i,n}_{1,p})$ is an i.i.d. sequence with common distribution (V_0, V_1) is uniform on the set of pairs of distinct elements of $\{1, \ldots, N\} \backslash \{i\}$. Finally, $(B^{i,n}_p)$ is i.i.d. Bernoulli sequence of random variables with parameter $1/2$.

The set of marks $\mathcal{M}_V$ is defined as

$$\mathcal{M}_V = \left\{\overline{v} = (v_p) = (v_{0,p}, v_{1,p}, b_p) \in \{1, \ldots, N\}^2 \times \{0, 1\} : v_{0,p} \neq v_{1,p}\right\}$$

For $1 \leq i \leq N$ and $n \geq 1$, t^i_n is the instant of the nth breakdown of server i. For $p \geq 1$, $V^{i,n}_{0,p}$ and $V^{i,n}_{1,p}$ are the servers where the pth copy present on node i may be allocated after this breakdown, depending on their respective loads of course. If the two loads are equal, the Bernoulli random variable $B^{i,n}_p$ is then used.

Equations of Evolution For $1 \leq i \leq N$ and $t \geq 0$, $Q^N_i(t)$ is the number of copies on server i at time t for this policy and $(Q^N(t)) = (Q^N_i(t))$ The dynamics of the power of choice algorithm is represented by the following stochastic differential equation, for $1 \leq i \leq N$,

$$\mathrm{d}Q^N_i(t) = -Q^N_i(t-)\overline{\mathcal{N}}_i(\mathrm{d}t, \mathcal{M}_V) + \sum_{m=1, m\neq i}^{N} R^N_{mi}(Q^N(t-), \overline{v})\, \overline{\mathcal{N}}_m(\mathrm{d}t, \mathrm{d}\overline{v}) \tag{4.7}$$

where $R^N_{mi} : \mathbb{N}^N \times \mathcal{M}_V \mapsto \mathbb{N}$ is the function, for $\ell = (\ell_k)$ and $\overline{v} = (v_{0,p}, v_{1,p}, b_p) \in \mathcal{M}_V$

$$R^N_{mi}(\ell, \overline{v}) = \sum_{k=1}^{\ell_m} \mathbb{I}_{\{i \in \{v_{0,k}, v_{1,k}\}\}} \times \left(\mathbb{I}_{\left\{\ell_i < \ell_{v_{0,k}} \vee \ell_{v_{1,k}}\right\}} + \mathbb{I}_{\left\{\ell_{v_{0,k}} = \ell_{v_{1,k}}, i = v_{b_k,k}\right\}}\right).$$

As it can be seen, when node m breaks down while the network is in state ℓ, $R^N_{mi}(\ell, \overline{v})$ is the number of copies sent to node i by the power of choice policy if $\overline{v} = V^m_\cdot$ is the corresponding mark associated to this instant.

Contrary to the random policy, the allocation depends on the state $(Q^N(t))$, for this reason it is convenient to introduce the empirical distribution $\Lambda^N(t)$ as follows, if f is some real-valued function on $\mathbb{N}$,

$$\left\langle \Lambda^N(t), f \right\rangle = \int_{\mathbb{N}} f(\ell)\Lambda^N(t)(\mathrm{d}\ell) = \frac{1}{N}\sum_{i=1}^{N} f\left(Q^N_i(t)\right).$$

If $0 \leq a \leq b$, $\langle \Lambda^N(t), [a,b] \rangle$ denotes $\Lambda^N(t)$ applied to the indication function of $[a,b]$. In the same way as in the proof of Proposition 4.1, it can be proved that with high probability, uniformly on any finite time interval, the positive jumps of all processes are $+1$. By using Equation (4.7) and the definition of $\Lambda^N(t)$, one gets that, for a finite support function f, with high probability

$$\begin{aligned} \mathrm{d}\left\langle \Lambda^N(t), f \right\rangle &= \mathrm{d}M_f^N(t) + \left\langle \Lambda^N(t), f(0) - f(\cdot) \right\rangle \mathrm{d}t + \\ &\sum_{\ell \in \mathbb{N}} \left[f(\ell+1) - f(\ell) \right] \sum_m Q_m^N(t) \frac{1}{(N-1)(N-2)} \times \Bigg[\sum_{\substack{j \neq j' \\ j,j' \neq m}} \mathbb{I}_{\left\{ Q_j^N(t) \geq \ell \right\}} \mathbb{I}_{\left\{ Q_{j'}^N(t) \geq \ell \right\}} \\ &\qquad - \sum_{\substack{j \neq j' \\ j,j' \neq m}} \mathbb{I}_{\left\{ Q_j^N(t) \geq \ell+1 \right\}} \mathbb{I}_{\left\{ Q_{j'}^N(t) \geq \ell+1 \right\}} \Bigg] \mathrm{d}t, \end{aligned}$$

where $M_f(t)$ is a martingale. Note that the terms inside the brackets in the last equation is simply the number of pair of nodes whose state is greater than ℓ and the state of at least one of them is ℓ. By integrating, this gives the relation

$$\begin{aligned} \langle \Lambda^N(t), f \rangle = \langle \Lambda^N(0), f \rangle + M_f^N(t) + \int_0^t \langle \Lambda^N(s), f(0) - f \rangle \, \mathrm{d}s \\ + \beta \int_0^t \langle \Lambda^N(s), g \rangle \, \mathrm{d}s + O(1/N), \end{aligned} \tag{4.8}$$

with

$$g(\ell) = (f(\ell+1) - f(\ell)) \times \frac{\left[\Lambda^N(s)([\ell, +\infty))^2 - \Lambda^N(s)([\ell+1, +\infty))^2 \right]}{\Lambda^N(s)(\{\ell\})}.$$

See the derivation of Ψ_{pc} page 130 for example. The term $O(1/N)$ is coming from the martingale part which is vanishing by a standard argument using Doob's Inequality.

Proposition 4.4 (Mean-Field Convergence)**.**

1. *The distribution of $(Q_1^N(t))$ is converging in distribution to $(X_\beta^P(t))$, a non-homogeneous Markov process whose Q matrix $(q(t)(x,y))$ is given by, for $x \in \mathbb{N}$, $q(t)(x,0) = 1$ and*

$$q(t)(x, x+1) = \beta \frac{\mathbb{P}\left(X_\beta^P(t) \geq x\right)^2 - \mathbb{P}\left(X_\beta^P(t) \geq x+1\right)^2}{\mathbb{P}\left(X_\beta^P(t) = x\right)}$$

2. *For the convergence in distribution, if f has finite support,*

$$\lim_{N \to +\infty} \left(\langle \Lambda^N(t), f \rangle \right) = \left(\mathbb{E}\left[f\left(X_\beta^P(t) \right) \right] \right)$$

The proof is developed in a more general setting in Section 4.2.4. It is based on the proof of the convergence of the process $(\Lambda^N(t))$ by using Equation (4.8). Additionally, Theorem 4.6 shows that a propagation of chaos also holds, i.e. in the limit, as N gets large, the finite marginals of the processes $(Q_i^N(t))$ converge in distribution to the law of independent processes.

The Invariant Distribution In this part, we study the asymptotic behavior of the invariant distribution of the load of a node at equilibrium.

Proposition 4.5. *The process $(X_\beta^P(t))$ of Proposition 4.4 has a unique invariant distribution π_β^P on $\mathbb{N}$, which can be defined by induction as*

$$\pi_\beta^P([x+1,+\infty))=\frac{-1+\sqrt{1+4\beta^2\pi_\beta^P([x,+\infty))^2}}{2\beta}, \qquad x\in\mathbb{N},$$

with $\pi_\beta^P([0,+\infty))=1$.

It should be noted that, due to the non-homogeneity of the Markov process, the uniqueness property is not clear in principle.

Proof. Let π a possible invariant probability of the process. If we start from this initial distribution, obviously the coefficients of the Q-matrix do not depend of time, the invariant equations can be written as

$$\begin{cases}\pi(x)(1+q(x,x+1)) = \pi(x-1)q(x-1,x),\ x>0,\\ \pi(0)(1+q(0,1)) = 1.\end{cases} \tag{4.9}$$

Define, for $x\geq 1$, $\xi(x)=\pi(x-1)q(x-1,x)$, then

$$\pi(x)=\xi(x)-\xi(x+1),$$

in particular $\mathbb{P}_\pi(X_\beta^P\geq x)=\xi(x)$, hence by definition of the Q-matrix

$$\xi(x+1)=\beta(\xi(x)^2-\xi(x+1)^2), \tag{4.10}$$

hence, necessarily

$$\xi(x+1)=\frac{-1+\sqrt{1+4\beta^2\xi(x)^2}}{2\beta},$$

with initial value $\xi(0)=1$. It is easily seen that the sequence $(\xi(x))$ is converging to 0. Hence the positive measure π defined by Relation (4.9) is indeed a probability distribution on $\mathbb{N}$. Such a probability is of course invariant with respect to the McKean-Vlasov process. The proposition is proved. □

Proposition 4.6 (Invariant Distributions)**.** *The invariant distribution of $(Q_1^N(t))$ is converging to the unique invariant distribution of $(X_\beta^P(t))$.*

See the proof of Theorem 4.7 below. It shows that it is enough to analyze the invariant distribution π_β^P of the limiting process we have just obtained. We can now state the main result of this section which explains the phenomenon observed in the simulations, see Figure 4.2.

Theorem 4.3 (Equilibrium with High Load)**.** *If $\overline{X}_\beta^P$ is a random variable with distribution π_β^P, then, for the convergence in distribution,*

$$\lim_{\beta\to+\infty}\frac{\overline{X}_\beta^P}{\beta}=U,$$

where U is a uniformly distributed random variable on $[0,2]$.

Theorem 4.8 below gives an analogous result in a more general setting.

Proof. In the proof of Proposition 4.5, we have seen that, by Equation (4.10), for $k \geq 0$,

$$\mathbb{P}\left(X_\beta^P \geq k+1\right) = \beta\left(\mathbb{P}\left(X_\beta^P \geq k\right)^2 - \mathbb{P}\left(X_\beta^P \geq k+1\right)^2\right), \tag{4.11}$$

by summing these equations, one obtains for all x,

$$\mathbb{E}\left(X_\beta^P \wedge x\right) = \beta\left(1 - \mathbb{P}\left(X_\beta^P \geq x\right)^2\right),$$

where $a \wedge b = \min(a, b)$. Hence, as expected, $\mathbb{E}(X_\beta^P) = \beta$, and therefore

$$\mathbb{P}\left(X_\beta^P \geq x\right)^2 = \frac{1}{\beta}\left(\mathbb{E}\left(X_\beta^P\right) - \mathbb{E}\left(X_\beta^P \wedge x\right)\right) = \frac{1}{\beta}\mathbb{E}\left(\left(X_\beta^P - x\right)^+\right). \tag{4.12}$$

By multiplying Equation (4.11) by $k+1$ and by summing up, one gets

$$\sum_{k=1}^{x} k\,\mathbb{P}\left(X_\beta^P \geq k\right) = \beta \sum_{k=0}^{x-1} \mathbb{P}\left(X_\beta^P \geq k\right)^2 + \beta\left(1 - \mathbb{P}\left(X_\beta^P \geq x\right)^2\right).$$

The right hand side of this relation is bounded by

$$\beta\left(\sum_{k=0}^{+\infty} \mathbb{P}\left(X_\beta^P \geq k\right) + 1\right) = \beta(\beta + 2)$$

hence, by using Fubini's Theorem on the left hand side,

$$\mathbb{E}\left(\left(X_\beta^P\right)^2\right) \leq 2\beta(\beta+2),$$

so that

$$\sup_{\beta > 0} \mathbb{E}\left(\left(\frac{X_\beta^P}{\beta}\right)^2\right) < +\infty$$

holds. In particular the family of random variables

$$(Y_\beta) \stackrel{\text{def.}}{=} \left(\frac{X_\beta^P}{\beta}\right)$$

is tight when β goes to infinity. Let Y be one of its limiting points,

$$\mathbb{P}\left(Y_\beta \geq x\right)^2 = \mathbb{E}\left(\left(Y_\beta - \frac{\lceil x\beta \rceil}{\beta}\right)^+\right).$$

The uniform integrability property of (Y_β), consequence of the boundedness of the second moments, gives that Y satisfies necessarily the relation

$$\mathbb{P}\left(Y \geq x\right)^2 = \mathbb{E}\left((Y - x)^+\right) = \int_x^{+\infty} \mathbb{P}(Y > s)\, \mathrm{d}s.$$

The function $f(x)=\mathbb{P}(Y \geq x)$ is thus differentiable and satisfies the differential equation

$$2f'(x)f(x)=-f(x),$$

for $x\geq 0$, so that $f'(x)=-1/2$ when $f(x)\neq 0$. One obtains the solution

$$\mathbb{P}(Y \geq x)=(2-x)^+/2, x\geq 0,$$

with $a^+ = \max(a,0)$, Y is a uniformly distributed random variable on the interval $[0,2]$. The family of random variables (Y_β) has therefore a unique limiting point when β goes to infinity. One deduces the convergence in distribution. The theorem is proved.

□

4.1.5 Conclusion

Our investigations through simulations and mathematical models have shown that

— a simple, random placement strategy may lead to heavily unbalanced situations;
— Classical load balancing techniques, like choosing the least loaded nodes are optimal from the point of view of placement. They have the drawback of requiring a detailed information on the state of the network, hence a significant cost in terms of complexity and bandwidth.
— the *power of two random choices* policy has the advantage of having good performances with a limited cost in terms of storage space and of complexity.

Appendix: Convergence Results

The technical results of this section concern the random allocation scheme. The notations of the corresponding section are used.

Proposition 4.7. *The previsible increasing process of the martingale* $(M_1^N(t))$ *is*

$$\left\langle M_1^N \right\rangle (t) = \int_0^t L_1^N(s)^2 \, \mathrm{d}s + \sum_{m=2}^{N} \int_0^t \left[\frac{1}{(N-1)^2} L_m^N(s)^2 + \frac{N-2}{(N-1)^2} L_m^N(s) \right] \mathrm{d}s. \tag{4.13}$$

Concerning previsible increasing process of martingales, see Section VI-34 page 377 of Rogers and Williams [142].

Proof. The proof is not difficult, it is included for the sake of completeness for readers not familiar with the properties of martingales associated to marked Poisson point processes. The previsible increasing process of the martingale

$$\left(\int_0^t L_1^N(s-) \left[\overline{\mathcal{N}}_1(\mathrm{d}t, \mathcal{M}_U) - \mathrm{d}s \right] \right) \text{ is } \left(\int_0^t L_1^N(s)^2 \, \mathrm{d}s \right),$$

see Theorem (28.1) page 50 of [142]. By independence of the Poisson processes, it is enough to calculate the previsible increasing process of the martingale, for $2\leq m\leq N$,

$$M_{1,m}^N(t) \stackrel{\text{def.}}{=} \int_0^t \left[z_1 \left(L_m^N(s-), \overline{u} \right) \overline{\mathcal{N}}_m(\mathrm{d}s, \mathrm{d}\overline{u}) - \frac{L_m^N(s)}{N-1} \, \mathrm{d}s \right]$$
$$= \sum_{t_n^m \leq t} \sum_{p=1}^{L_m^N(t_n^m-)} \mathbb{I}_{\{U_p^{m,n}=1\}} - \int_0^t \frac{L_m^N(s)}{N-1} \, \mathrm{d}s.$$

It is enough in fact to show that the second moment of this martingale is such that

$$\mathbb{E}\left(M_{1,m}^N(t)^2\right) = \int_0^t \left[\frac{1}{(N-1)^2}\mathbb{E}\left(L_m^N(s)^2\right) + \frac{N-2}{(N-1)^2}\mathbb{E}\left(L_m^N(s)\right)\right] \mathrm{d}s,$$

the property of independent increments of Poisson processes will then give the martingale property of $M_{1,m}^N(t)^2$ minus this term. By integrating with respect to the values of $(U_p^{m,n})$, one has

$$\mathbb{E}\left(\left(\sum_{p=1}^{L_m^N(t_n^m-)} \mathbb{I}_{\{U_p^{m,n}=1\}}\right)^2\right) = \frac{N-2}{(N-1)^2}\mathbb{E}\left(L_m^N(t_n^m-)\right) + \frac{1}{(N-1)^2}\mathbb{E}\left(L_m^N(t_n^m-)^2\right),$$

which gives the relation

$$\mathbb{E}\left(\left(\sum_{t_n^m \le t} \sum_{p=1}^{L_m^N(t_n^m-)} \mathbb{I}_{\{U_p^{m,n}=1\}}\right)^2\right)$$
$$= \frac{1}{(N-1)^2}\mathbb{E}\left(\left(\int_0^t L_m^N(s-)\,\mathcal{N}_m(\mathrm{d}s)\right)^2\right) + \frac{N-2}{(N-1)^2}\mathbb{E}\left(\int_0^t L_m^N(s-)\,\mathcal{N}_m(\mathrm{d}s)\right).$$

In the same way, by integrating with respect to the values of $(U_p^{m,n})$, with the notation $\mathcal{N}_m(\mathrm{d}s)=\overline{\mathcal{N}}_m(\mathrm{d}s, \mathcal{M}_U)$,

$$\mathbb{E}\left(\int_0^t \frac{L_m^N(s)}{N-1}\,\mathrm{d}s \sum_{t_n^m \le t} \sum_{p=1}^{L_m^N(t_n^m-)} \mathbb{I}_{\{U_p^{m,n}=1\}}\right) = \mathbb{E}\left(\int_0^t \frac{L_m^N(s)}{N-1}\,\mathrm{d}s \int_0^t \frac{L_m^N(s)}{N-1}\,\mathcal{N}_m(\mathrm{d}s)\right).$$

By using the last two relations one gets

$$\mathbb{E}\left(M_{1,m}^N(t)^2\right) =$$
$$\frac{1}{(N-1)^2}\mathbb{E}\left(\left(\int_0^t L_m^N(s)\,[\mathcal{N}_m(\mathrm{d}s) - \mathrm{d}s]\right)^2\right) + \frac{N-2}{(N-1)^2}\mathbb{E}\left(\int_0^t L_m^N(s-)\,\mathcal{N}_m(\mathrm{d}s)\right).$$

Since the the martingale $(\mathcal{N}_m([0,t]-t)$ has the increasing previsible process (t), one gets

$$\mathbb{E}\left(M_{1,m}^N(t)^2\right) = \frac{1}{(N-1)^2}\int_0^t \mathbb{E}\left(L_m^N(s)^2\right)\mathrm{d}s + \frac{N-2}{(N-1)^2}\int_0^t \mathbb{E}\left(L_m^N(s-)\right)\mathrm{d}s.$$

The proposition is proved. □

Lemma 4.1. *If the initial distribution of $(L_i^N(t))$ satisfies Condition (4.2) then, for any $T>0$,*

$$\sup_{N\ge1}\sup_{0\le t\le T} \mathbb{E}\left((L_1^N(t))^2\right) < +\infty. \tag{4.14}$$

and

$$\sup_{N\ge1} \mathbb{E}\left(\sup_{0\le s\le T} L_1^N(s)\right) < +\infty. \tag{4.15}$$

Proof. With Relation (4.3), by writing the SDE satisfied by $(L_1^N(t)^2)$,

$$L_1^N(t)^2 = L_1^N(0)^2 - \int_0^t L_1^N(s-)^2 \mathcal{N}_1(\mathrm{d}s, \mathcal{M}_U) + \sum_{m=2}^N \int_0^t z_1\left(L_m^N(s-), \overline{u}\right) \times \left[2L_1^N(s-) + z_1\left(L_m^N(s-), \overline{u}\right)\right] \overline{\mathcal{N}}_m(\mathrm{d}s, \mathrm{d}\overline{u})$$

by taking the expectation, one obtains

$$\mathbb{E}\left(L_1^N(t)^2\right) = \mathbb{E}\left(L_1^N(0)^2\right) - \int_0^t \mathbb{E}\left(L_1^N(s)^2\right) \mathrm{d}s + \int_0^t \mathbb{E} \sum_{m=2}^N 2\frac{L_m^N(s)}{N-1} L_1^N(s)\ \mathrm{d}s + \int_0^t \sum_{m=2}^N \mathbb{E}\left(\frac{L_m^N(s)(L_m^N(s)-1)}{(N-1)^2} + \frac{L_m^N(s)}{N-1}\right) \mathrm{d}s$$

By using the fact that the $L_m^N(t)$'s have the same distribution and their sum is F_N, if $f_N(t)=\mathbb{E}(L_1^N(t)^2)$, Equation (4.5) gives that, for $0 \le t \le T$,

$$f_N(t) \le (2T+1)\left[\frac{F_N}{N}\right]^2 + \frac{1}{N-1}\int_0^t f_N(s)\ \mathrm{d}s.$$

If $p \in \mathbb{N}$ such that $F_N/N \le p$ for all N, then, by Gronwall's Inequality, see Ethier and Kurtz [57] p.498,

$$f_N(t) \le p^2(1+2T)e^{T/(N-1)}, \quad \forall N \ge 2.$$

Relation (4.14) is proved.
Denote by

$$S_m^N(t) = \sup(L_m^N(s) : 0 \le s \le t)$$

then, by Equation (4.5), for $t \le T$,

$$S_1^N(t) \le L_1^N(0) + \frac{1}{N-1}\sum_{m=2}^N \int_0^t S_m^N(s)\ \mathrm{d}s + \sup_{0\le s\le T} |M_1^N(s)|.$$

with the help of Doob's Inequality, see Theorem (52.6) of Rogers and Williams [141], one gets

$$\mathbb{E}\left(\sup_{0\le s\le T} M_1^N(s)^2\right) \le 2\mathbb{E}(M_1^N(T)^2) = 2\mathbb{E}(\left\langle M_1^N\right\rangle(T))$$

and this last quantity is bounded with respect to $N \ge 2$ by Relations (4.13) and (4.14). Hence, by using the previous inequality, one can find a constant K_0 such that, for any $N \ge 2$ and $t \le T$,

$$\mathbb{E}(S_1^N(t)) \le K_0 + \int_0^t \mathbb{E}\left(S_1^N(s)\right) \mathrm{d}s,$$

one concludes again with Gronwall's Inequality. The lemma is proved. □

Lemma 4.2. *If the initial condition $(L_j^N(0))$ is such that the variables $L_j^N(0)$, $j \ge 2$ are exchangeable and that*

$$\sup_{N\ge 1} \mathbb{E}(L_1^N(0)^2) + \mathbb{E}(L_2^N(0)^2) < +\infty$$

holds, then, for all $T \geq 0$,

$$\sup_{N\geq 1}\sup_{0\leq t\leq T} \mathbb{E}(L_1^N(t)^2) + \mathbb{E}(L_2^N(t)^2) < +\infty,$$

Proof. The proof is similar to the proof of Lemma 4.1. One has to introduce the functions $f_N^1(t)=\mathbb{E}(L_1^N(t)^2)$ and $f_N^2(t)=\mathbb{E}(L_2^N(t)^2)$, by using an integral equation for $(L_1^N(t)^2)$ and $(L_1^N(t)^2)$ and the symmetry properties of the vector $(L_j^N(t), j\geq 2)$, one obtains the relations

$$\begin{cases} f_N^1(t) \leq C_1 + A_1 \displaystyle\int_0^t f_N^1(s)\ \mathrm{d}s + B_1 \int_0^t f_N^2(s)\ \mathrm{d}s \\ \\ f_N^2(t) \leq C_2 + A_2 \displaystyle\int_0^t f_N^1(s)\ \mathrm{d}s + B_2 \int_0^t f_N^2(s), \end{cases}$$

for convenient positive constants A_i, B_i, C_i, $i = 1, 2$ independent of N. One uses Gronwall's Inequality for the first relation to get an upper bound on f_N^1,

$$f_N^1(t) \leq \left(C_1 + B_1 \int_0^t f_N^2(s)\ \mathrm{d}s\right) e^{A_1 t}$$

and Gronwall's Inequality is again used after injecting this relation in the second inequality. □

The next result is a technical extension of Proposition 4.2 used to prove Proposition 4.3.

Proposition 4.8. *If $\widehat{\pi}_\beta^N$ is the invariant distribution of the state of the network at the instants of failures of node 1, then, with the notations of Section 4.1.4, for the convergence in distribution,*

$$\lim_{N\to+\infty} \left(C_N^P(t)\right) = (\beta t)$$

if the initial distribution of $(L^N(t))$ is $\widehat{\pi}_\beta^N$.

Proof. Let $\widetilde{\pi}_\beta^N$ be the invariant distribution of the process $(L_1^N(t))$ at the instants of failures on nodes, not only of node 1. The sequence of states of the corresponding Markov chain is denoted as $(\widetilde{L}_n^N) = (\widetilde{L}_{n,j}^N, 2\leq j\leq N)$ where $\widetilde{L}_{n,j}^N$, $2\leq j\leq N$ is the state of the nodes at the instant of the nth failure, i.e. the state of network reordered but with the failed node is excluded. If

$$W_n = \left(\widetilde{L}_{n,2}^N\right)^2 + \left(\widetilde{L}_{n,3}^N\right)^2 + \cdots + \left(\widetilde{L}_{n,N}^N\right)^2,$$

by invariance one has

$$\mathbb{E}_{\widetilde{\pi}_\beta^N}(W_0) = \mathbb{E}_{\widetilde{\pi}_\beta^N}(W_1),$$

after some trite calculations, one obtains

$$\mathbb{E}_{\widetilde{\pi}_\beta^N}\left(\left(\widetilde{L}_{0,2}^N\right)^2\right) = \frac{N-1}{N}\frac{F_N^2}{N^2} + \frac{F_N}{N}\frac{N-2}{N}$$

hence

$$\sup_{N\geq 2} \mathbb{E}_{\widetilde{\pi}_\beta^N}\left(\left(\widetilde{L}_{0,2}^N\right)^2\right) < +\infty.$$

The same property will hold when one considers only the instants of failures of node 1 since, recall that t_1^1 is the first of these instants,

$$\widehat{\pi}_\beta^N \overset{\text{dist.}}{=} (\widetilde{L}_i^N(t_1^1), i\geq 2) \text{ if } (L_i^N(0)) \overset{\text{dist.}}{=} \widetilde{\pi}_\beta^N.$$

By proceeding as in the proof of Lemma 4.1, but by stopping at time t_1^1 instead of a fixed time t, one obtains that

$$\sup_{N\geq 2} \mathbb{E}_{\widehat{\pi}_\beta^N}\left(\left(L_2^N(0)\right)^2\right) = \sup_{N\geq 2} \mathbb{E}_{\widetilde{\pi}_\beta^N}\left(\left(\widetilde{L}_2^N(t_1^1)\right)^2\right) < +\infty.$$

Lemma 4.2 implies therefore that

$$\sup_{N\geq 2}\sup_{0\leq t\leq T} \mathbb{E}_{\widehat{\pi}_\beta^N}\left(\left(L_1^N(t)\right)^2\right) + \mathbb{E}\left(\left(L_2^N(t)\right)^2\right) < +\infty.$$

One can now proceed as in the proof of Proposition 4.2 by noting that the crucial argument is the fact that the two last terms of the right hand side of Equation (4.6) vanish when N gets large. □

4.2 Large Urn Models with Local Mean-Field Interactions

Abstract The stochastic models investigated in this section describe the evolution of a set of F_N identical balls scattered into N urns connected by an underlying symmetrical graph with constant degree h_N. After some exponentially distributed amount of time *all the balls* of one of the urns are redistributed locally, among the h_N urns of its neighborhood. The allocation of balls is done at random according to a set of weights which depend on the state of the system. The main original features of this context is that the cardinality h_N of the range of interaction is not necessarily linear with respect to N as in a classical mean-field context and, also, that the number of simultaneous jumps of the process is not bounded due to the redistribution of all balls of an urn at the same time. The approach relies on the analysis of the evolution of the local empirical distributions associated to the state of urns in the neighborhood of a given urn. Under some convenient conditions, by taking an appropriate Wasserstein distance and by establishing appropriate technical estimates for local empirical distributions, we are able to establish a mean-field convergence result. Convergence results of the corresponding invariant distributions are obtained for several allocation policies.

For the class of power of d choices policies for the allocations of balls, we show that the invariant measure of the corresponding McKean-Vlasov process has an asymptotic finite support property when the average load per urn gets large. This result differs somewhat from the classical double exponential decay property usually encountered in the literature for power of d choices policies. This finite support property has interesting consequences in practice.

4.2.1 Introduction

The stochastic models investigated in this paper describe the evolution of a set of N urns indexed by $i \in \{1, \ldots, N\}$ with F_N identical balls. There is an underlying deterministic symmetrical graph structure connecting the urns. The system evolves as follows. After some exponentially distributed amount of time *all the balls* of an urn, index $i \in \{1, \ldots, N\}$ say, are redistributed among a subset $\mathcal{H}^N(i)$ of urns "near" urn i. Urn i is then empty after that moment except if some of its balls are re-allocated to it. An important feature of the model is that the allocation of the balls into urns of $\mathcal{H}^N(i)$ is done at random according to a set of weights depending on the vector of the number of balls in each of these urns.

Quite general allocation schemes are investigated but two policies stand out because of their importance. Their specific equilibrium properties are analyzed in detail in the section. We describe them quickly. Assume that a ball of urn i has to be allocated.

1. Random Policy.
   The ball is allocated uniformly at random in one of the neighboring urns, i.e. in one of the urns whose index is in $\mathcal{H}^N(i)$, this occurs with probability $1/\mathrm{card}(\mathcal{H}^N(i))$.

2. Power of d-Choices
   For this scheme, a subset of d urns whose indices are in $\mathcal{H}^N(i)$ is taken at random, the ball is allocated to the urn having the least number of balls among these d urns. Ties are broken with coin tossing.

Under some weak symmetry assumption and supposing that the state of the system can be represented by an irreducible finite state Markov process, at equilibrium the average number of balls per urn is F_N/N whatever the allocation procedure is. The main problem considered in this section concerns distribution of the number of balls in a given urn: how likely is the event that, at equilibrium, an urn has a large number of balls such a context? It may be expected that if the set of weights to perform allocation is chosen conveniently, then there are few heavily loaded urns, i.e. the probability of such event is significantly small. An additional desirable feature is that only a limited information is available for the allocation of the balls. In our case this will be the knowledge of the occupancy's of few of the neighboring urns.

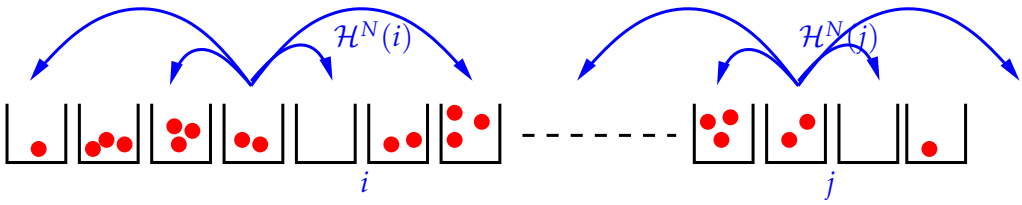


Figure 4.3 – Urn Model with Neighborhoods

Urn Models and Allocation Algorithms in the Literature

These problems have important applications in several domains, in physics to describe interactions of particles and their non-equilibrium dynamics, see Ehrenfest [54], Godrèche and Luck [69] and Evans and Hanney [59] or Godrèche [68] for recent overviews. They are also used in theoretical computer science to evaluate the efficiency of algorithms to allocate tasks to processors in a large scale computing network, see Karthik et al. [87], Maguluri et al. [107]; to place copies of files in the servers of a large distributed system, see Lakshman et al. [101] and Sun et al. [156] or to design hash tables, see Azar et al. [14] and Mitzenmacher [112]. We review some of their aspects.

Urn models have been used in statistical mechanics as toy models to understand conceptual problems related to non-equilibrium properties of particle systems. As a model they describe the evolution a set of particles moving from one urn to another one. Ehrenfest [54] is one the most famous of these models: in continuous time models, each particle moves at random to another urn after an exponentially distributed amount of time. There may also be underlying structure for the urns, a ball can be allocated to the connected urns at random according to some weights on the corresponding edges or, via a metropolis algorithm, associated to some energy functional. See Evans and Hanney [59] and Godrèche [68]. Due to their importance and simplicity, these stochastic models have been thoroughly investigated over the years: reversibility properties, precise estimates of fluctuations, hitting times, convergence rate to equilibrium, ... See, for example, Bingham [20], Karlin and McGregor [86], Diaconis [47] to name but a few. For these models balls are moved one by one. Closely related to these models is the zero range process for which the corresponding rate is $f(x)$ for some general function f on $\mathbb{N}$. When $f(\cdot)$ is constant the jumps are in fact associated with the urns instead of the balls, as it is our case. See Evans and Hanney [59].

A classical problem in theoretical computer science concerns the assignments of N balls into N urns. Balls are assumed to be allocated one by one. The constraints in this setting are of minimizing the maximum of the number of balls in the urns with a reduced information on the state of the system. When balls are distributed at random, it has been proved that this variable is, with high probability, of the order of $\log N / \log\log N$. See Kolchin et al. [97] for example. By using an algorithm of the type power of d-choices as above, Azar et al. [14] has shown that the maximum is of the order of $\log\log N / \log d$. Hence, with a limited information on the system, only the state of d urns is required, the improvement over the random policy is striking. See also Mitzenmacher [112].

A related problem is of assigning the jobs to the queues of N processing units working at rate 1. The jobs are assumed to be arriving according to a Poisson process with rate λN. When the natural stability condition $\lambda<1$ holds, if the jobs are allocated at random, then, at equilibrium, it is easily shown that the tail distribution of the number of jobs at a given unit decreases exponentially. If the allocation follows a power of d choices, Vvedenskaya et al. [165] has shown that the corresponding tail distribution has a double exponential decay, i.e. of the type $x \to \exp(-\alpha \exp(\delta x))$ for some positive constants α and δ. See Bramson et al. [23] for more delicate asymptotic mean-field results at equilibrium in a non-Markovian framework. See also Alanyali and Dashouk [2], Alistarh et al. [7] for similar results on related models.

The initial motivation of this work is coming from a collaboration with computer scientists to study the efficiency of duplication algorithms in the context of a large distributed system. The urns are the disks of a set of servers and the balls are copies of files on these disks. The redistribution of balls of an urn correspond to a disk crash, in this case, the copies of its lost files are retrieved on other neighboring disks. See Sun et al. [156] for more details on the modelling aspects.

Results

Mean-Field Convergence

For the stochastic models investigated in this section there are N urns and a total of F_N balls with $F_N \sim \beta N$ for some $\beta>0$. The underlying graph is symmetrical with degree $h_N = \text{card}(\mathcal{H}^N(i))$, $1 \le i \le N$. The sequence (h_N) is assumed to converge to infinity. The state of the system is represented by a vector $\ell = (\ell_i, 1 \le i \le N)$ describing the number of balls in the urns. The distinctive features of the stochastic model investigated in this section can be described as follows.

1. Multiple Simultaneous Jumps.
   When the exponential clock associated to urn i rings, then all its balls are redistributed to other urns of the system. For this reason, there are ℓ_i jumps occurring simultaneously.

2. Local Search.
   Let

$$\Lambda_i^N = \frac{1}{h_N} \sum_{k \in \mathcal{H}^N(i)} \delta_{\ell_k}$$

be *the local empirical distribution* around urn i, where δ_a is the Dirac mass at $a \in \mathbb{N}$. When urn i has to allocate a ball, it is sent to one of h_N neighboring urns, $j \in \mathcal{H}^N(i)$, with a probability of the order of $\Psi(\Lambda_i^N, \ell_j)/h_N$, where Ψ is a some functional on $M_1(\mathbb{N}) \times \mathbb{N}$ and $M_1(\mathbb{N})$ is the set of probability distributions on $\mathbb{N}$.

This interaction with only local neighborhoods and the unbounded number of simultaneous jumps are at the origin of the main technical difficulties to establish a mean-field convergence theorem. See the quite intricate evolution equation (4.39) for these local empirical distributions below. More specifically, this is, partially, due to a factor ℓ_i/h_N which has to be controlled in several integrands of the evolution equations. This is where unbounded jumps play a role. See Relation (4.50) for example. A convenient Wasserstein distance (4.32) is introduced for this reason.

It should be noted a classical mean-field analysis can be achieved when the sequences (h_N) has a linear growth, this is in fact close to a classical mean-field framework with full interaction. In this case the term ℓ_i/h_N is not anymore a problem since ℓ_i is bounded by $F_N \sim \beta N$, The same is true if the simultaneous jumps feature is removed by assuming for example that only a ball is transferred for each event. In this case a mean-field result can be established with standard methods for the model with neighborhoods.

Literature

For an introduction to the classical mean-field approach, see Sznitman [158]. Specific results for power of d choices policies when $h_N = N$ are presented in Graham [70], see also Luczak and McDiarmid [106]. One of the earliest works on local mean-field models seems to be Comets [37]. Budhiraja et al. [28] considers also such a local interaction for a queueing network with an underlying graph which is, possibly, random, with jumps of size 1. In this setting when the graph is deterministic, the mean-field analysis can be carried out by using standard arguments. Andreis et al. [10] investigates mean-field of jump processes associated to neural networks with a large number of simultaneous jumps with an infinitesimal amplitude. Luçon and Stannat [105] establishes mean-field results in a diffusion context when the mean-field interaction involves particles in a box whose size is linear with respect to N and a spatial component with possible singularities. A related setting is also considered in Müller [116].

Asymptotic Finite Support Property

The mean field results show that, for a fixed asymptotic load β per urn and under appropriate conditions, the evolution of the state of the occupancy of a given urn is converging in distribution to some non-linear Markov process $(L_\beta(t))$.

We show that, for all classes of allocations considered in this section, when it exists, an invariant distribution of $(L_\beta(t)/\beta)$ has at a tail which is upper-bounded by an exponentially decreasing function. Additionally, for the random algorithm this is an exact exponentially decreasing tail distribution in fact. It shows, as it can be expected, that the performances of random algorithms are weak in terms of occupancy of urns. Recall that these algorithms do not use any information to allocate balls into urns.

For power of d choices algorithms, contrary to the model analyzed in Vvedenskaya et al. [165], we do not have an explicit expression for the invariant distribution of $(L_\beta(t))$. We show that, for any $\beta > 0$, this distribution exists and is unique and that it can be

expressed via a recursion. See Proposition 4.15. We prove the following behavior for high loads. The invariant distribution of $(L_\beta(t)/\beta)$ for power of d choices is converging in distribution to a distribution with a continuous density on the compact interval $[0, d/(d-1)]$. When $d=2$, this is a uniform distribution on $[0,2]$. The striking feature is that, for an average load of β per urn, at equilibrium, the occupancy of a given urn is at most $\beta d/(d-1)$ with probability arbitrarily close to 1. In some way this can be seen as the equivalent of the double exponential decay property in this context. It should be noted that this is an asymptotic picture for N large and also β large. Experiments seem to show nevertheless that, in practice, this is an accurate description. See for example Figure 2 of Sun et al. [156] where $d=2$, $N=200$ and $\beta=150$ and the uniform distribution on $[0,300]$ is quite neat. Explicit bounds on error terms would be of interest but seem to be out of reach for the moment.

Outline of the Section

The stochastic model is presented in Subsection 4.2.2. Subsection 4.2.3 introduces the main evolution equations for the local empirical distributions and establishes the existence and uniqueness properties of the corresponding asymptotic McKean-Vlasov process. The main convergence results are proved in Subsection 4.2.4 via several technical estimates. Subsection 4.2.5 is devoted to the analysis of the invariant distribution of the McKean-Vlasov process. The finite support property is proved in this section.

4.2.2 The Stochastic Model

We give a precise mathematical description of our system with N urns and F_N balls with the following scaling assumption

$$\beta = \lim_{N\to+\infty} \frac{F_N}{N}, \tag{4.16}$$

for some $\beta>0$. An index N will be added on the important quantities describing the system to stress the dependence on N but not systematically to make the various equations involved more readable.

Graph Structure

One considers a graph with N vertices labeled from 1 to N and $\mathcal{H}^N \subset \{2,\dots,N\}$ is assumed to satisfy the following property of symmetry

$$k\in\mathcal{H}^N \Leftrightarrow (N+2-k \bmod N)\in\mathcal{H}^N, \tag{4.17}$$

with the convention used throughout the section that $(0 \bmod N)$ is N.

The set $\mathcal{H}^N$ is the set of neighbors of node 1. The set $\mathcal{H}^N(i)$ of neighbors of a node $i\in\{1,\dots,N\}$ is defined by translation in the following way

$$\mathcal{H}^N(i) = \left\{(i+j-1 \bmod N) : j\in\mathcal{H}^N\right\}.$$

In particular $\mathcal{H}^N(1)=\mathcal{H}^N$, one denotes by h_N the cardinality of $\mathcal{H}^N$. Relation (4.17) gives the property that the associated graph is non-oriented that is, if $i\in\mathcal{H}^N(j)$ then $j\in\mathcal{H}^N(i)$ and regular, its constant degree being $\mathrm{card}(\mathcal{H}^N)$. Here are three examples.

— Full Graph: $\mathcal{H}^N_{\mathrm{cc}}\stackrel{\mathrm{def.}}{=}\{2,\dots,N\}$.

— Torus: for $\alpha\in(0,1)$,

$$\mathcal{H}^N_\alpha\stackrel{\mathrm{def.}}{=}\{(1+j \bmod N) : j\in[-\alpha N,\alpha N]\cap\mathbb{Z}\}\setminus\{1\}.$$

— Log-Torus: for $\delta>0$,

$$\mathcal{H}^N_{\log}\stackrel{\mathrm{def.}}{=}\{(1+j \bmod N) : j\in[-\lfloor\delta\log N\rfloor,\lfloor\delta\log N\rfloor]\cap\mathbb{Z}\}\setminus\{1\}.$$

We have chosen that $1\notin\mathcal{H}^N$ and, consequently, $i\notin\mathcal{H}^N(i)$. Our results in the following do not need this assumption in fact. It just simplifies some steps of the proofs, to compute the previsible increasing processes of some martingales in particular.

Assumptions 4.1. (Topology).

1. *The sequence of the degrees of nodes* $(h_N)\stackrel{\mathrm{def.}}{=}(\mathrm{card}(\mathcal{H}^N))$ *is converging to infinity.*
2. *The interaction set of node i,* $1\le i\le N$, *is defined as*

$$\mathcal{A}^N(i)\stackrel{\mathrm{def.}}{=}\Big\{j\in\{1,\dots,N\} : \mathcal{H}^N(j)\cap\mathcal{H}^N(i)\neq\emptyset\Big\} \tag{4.18}$$

and its cardinality satisfies

$$\lim_{N\to+\infty}\frac{\mathrm{card}(\mathcal{A}^N(1))}{h_N^2}=0.$$

Remark.

a) The associated graph can be seen as a torus and the set of neighbors of each node are defined by a translation of the neighbors of node 1. *We have chosen it with convenience. In fact, as long as symmetry properties hold, so that the invariant distributions of the number of balls into the urns are identically distributed and that Assumptions 4.1 holds, then it is not difficult to see that the main convergence results of Section 4.2.4 still hold.*

b) The set $\mathcal{A}^N(i)$ *is the set of nodes at distance* 2 *from i. Assumption 4.1 is clearly satisfied for the full graph, torus graphs described above, in these cases* $\mathrm{card}(\mathcal{A}^N(1))\le Ch_N$, *for some constant C.*

Allocation Policies

A set of F_N balls are scattered in the nodes, the state of the system is thus described by a vector $\ell=(\ell_i)\in\mathcal{S}_N$, with

$$\mathcal{S}_N=\{(\ell_i)\in\mathbb{N}^N : \ell_1+\ell_2+\cdots+\ell_N=F_N\},$$

for $i\in\{1,\dots,N\}$, ℓ_i is the number of balls in urn i. For each $\ell\in\mathcal{S}_N$, one associates a probability vector $P^N(\ell)=(p_i^N(\ell),1\le i\le N)$ with support on $\mathcal{H}^N$, i.e. $p_i^N=0$ if $i\notin\mathcal{H}^N$. The

vector $P^N(\ell)$ is in fact the set of weights associated to urn 1 in state ℓ and $p_i^N(\ell)$ is the probability that a given ball of urn 1 is allocated to urn i. As for the topology, we define the vector of weights for the other urns by translation.

For $i\in\{1,\dots\}$ a probability vector $P_i^N(\ell)=(p_{ij}^N(\ell))$ with support on $\mathcal{H}^N(i)$ is defined by

$$p_{ij}^N(\ell)=p_a^N(y), \tag{4.19}$$

with $a=(j-i+1 \bmod N)$ and $y=(\ell_{(i+k-1 \bmod N)}, 1\le k\le N)$, for any $j\in\{1,\dots,N\}$. In particular $P_1^N(\ell)=P^N(\ell)$ and $P_i^N(\cdot)$ is the vector $P^N(\cdot)$ "centered" around node i. The quantity $p_{ij}^N(\ell)$ is the probability that, in state ℓ, a ball of urn i is allocated to urn j.

The dynamics are as follows: After an exponentially distributed amount of time with mean 1, the balls of an urn i, $1\le i\le N$, are distributed into the neighboring urns in the subset $\mathcal{H}^N(i)$ one by according to some policy depending on the state $\ell=(\ell_j)$ of the system just before the event. Each of the ℓ_i balls of the ith urn is distributed on $\mathcal{H}^N(i)$ according to the probability vector $P_i(\ell)$ and the corresponding ℓ_i random variables used are assumed to be independent. As it will be seen our asymptotic results hold under general assumptions, the following cases will be discussed due to their practical importance.

— Random Algorithm.
Balls of a given urn $i\in\{1,\dots,N\}$ are sent at random into an urn with index in $\mathcal{H}^N(i)$. This is the simplest policy which is used in large distributed systems. See Sun et al. [156] for example. this corresponds to the case where

$$p_j^N(\ell)=\frac{1}{h_N}, \quad j\in\mathcal{H}^N. \tag{4.20}$$

— Random Weighted Algorithm.
Each ball is sent into urn $j\in\mathcal{H}^N(i)$ with a probability proportional to $W(\ell_j)$, where W is some function on $\mathbb{N}$, that is

$$p_j^N(\ell) = W(\ell_j)\Big/\sum_{k\in\mathcal{H}^N} W(\ell_k), \quad j\in\mathcal{H}^N. \tag{4.21}$$

— Power of d Choices, $d\ge 2$.
For each ball, d urns are chosen at random in $\mathcal{H}^N(i)$, the ball is allocated to the urn of this subset having the minimum number of balls. Ties are broken with coin tossing. A simple combinatorial argument gives that, for $j\in\mathcal{H}^N(i)$,

$$p_j^N(\ell) = \frac{1}{\binom{h_N}{d}\sum_{k\in\mathcal{H}^N}\mathbb{I}_{\{\ell_k=\ell_j\}}}\left[\binom{\sum_{k\in\mathcal{H}^N}\mathbb{I}_{\{\ell_k\ge\ell_j\}}}{d}-\binom{\sum_{k\in\mathcal{H}^N}\mathbb{I}_{\{\ell_k>\ell_j\}}}{d}\right], \tag{4.22}$$

with the convention that $\binom{n}{d}=0$ if $n<d$.

Remark.

1. *For simplicity we have chosen to consider only the state of the system just before the jump for all the balls which have to be moved. The state of the system could be also updated after each ball allocation and, consequently the vector $p_i^N(\cdot)$, the dynamical description would then be more intricate to express. Using an argument similar as in the proof Proposition 4.1*

of Sun et al. [156], it is not difficult to see that the two corresponding processes describing the number of balls in the urns has the same asymptotic behavior on any finite time interval as N gets large.

2. *It should be noted that the exponential clock associated to the ith urn gives simultaneous jumps of ℓ_j coordinates with high probability if N is large. In particular the magnitude of a possible jump is not bounded which leads to significant technical complications to prove a mean-field result as it will be seen.*

Assumptions 4.2. (Allocations).

1. The sequence $(P^N(\cdot))=(p_i^N(\cdot))$ satisfies the relation

$$\lim_{N\to+\infty} \sup_{\substack{i\in\mathcal{H}^N \\ \ell\in\mathcal{S}_N}} \left| h_N p_i^N(\ell) - \Psi\left(\frac{1}{h_N}\sum_{j\in\mathcal{H}^N} \delta_{\ell_j}, \ell_i\right)\right| = 0, \tag{4.23}$$

where Ψ is a non-negative bounded function on $M_1(\mathbb{N})\times\mathbb{N}$ such that, for any $\sigma\in M_1(\mathbb{N})$,

$$\int_{\mathbb{N}} \Psi(\sigma, x)\,\sigma(\mathrm{d}x) = 1. \tag{4.24}$$

2. There exist constants C_Ψ, $D_\Psi>0$ such that the relations

$$\begin{cases} |\Psi(\sigma,l)-\Psi(\sigma,l')| \le C_\Psi\,|l-l'|\,, \\ |\Psi(\sigma,l)-\Psi(\sigma',l)| \le D_\Psi\,\|\sigma-\sigma'\|_{\mathrm{tv}} \end{cases} \tag{4.25}$$

hold for any (σ,l), $(\sigma',l')\in M_1(\mathbb{N})\times\mathbb{N}$, where $\|\cdot\|_{\mathrm{tv}}$ is the total variation norm.

Relation (4.24) is a conservation of mass condition, all balls are reallocated.

— Random Weighted Algorithm.
Assumptions 4.2 are satisfied with the function Ψ given by

$$\Psi_{\mathrm{cc}}(\sigma,l) = W(l) \Big/ \int W(x)\,\sigma(\mathrm{d}x)$$

if the range of W is in $[c,C]$, for some positive constants c and C.

— Power of d Choices.
It is not difficult to check that Assumptions 4.2 are satisfied with

$$\Psi_{\mathrm{pc}}(\sigma,l)=\frac{(\sigma([l,+\infty))^d-\sigma((l,+\infty))^d}{\sigma(\{l\})},$$

with the convention that $0/0=0$.

Marked Poisson Point Processes

In order to use a convenient stochastic calculus to study these allocation algorithms, one has to introduce marked Poisson processes. See Kingman [96] and Robert [138] for an introduction on marked Poisson point processes.

We define the space of marks $\mathcal{M}=[0,1]^{\mathbb{N}}$, a mark $u=(u_k)\in\mathcal{M}$ associated to an urn $i\in\{1,\dots,N\}$ describes how the balls of this urn are allocated in the system: If the state is $\ell=(\ell_j)$, assuming that the balls are indexed by $1\leq k\leq \ell_i$, the kth ball is allocated to urn $j\in\{1,\dots,N\}$ if

$$u_k \in I_{ij}(\ell) \stackrel{\text{def.}}{=} \left[\sum_{n=1}^{j-1} p_{in}^N(\ell), \sum_{n=1}^{j} p_{in}^N(\ell)\right), \tag{4.26}$$

if u_k is a uniform random variable on $[0,1]$, this occurs with probability $p_{ij}^N(\ell)$. For i, $j\in\{1,\dots,N\}$, we introduce the family of mappings Z_{ij} on $\mathcal{M}\times\mathcal{S}_N$ defined by, for $u\in\mathcal{M}$ and $\ell\in\mathcal{S}_N$,

$$Z_{ij}(u,\ell) \stackrel{\text{def.}}{=} \sum_{k=1}^{\ell_i} \mathbb{I}_{\{u_k\in I_{ij}(\ell)\}}, \tag{4.27}$$

the quantity $Z_{ij}(u,\ell)$ is the number of balls of urn i which are allocated to urn j if the ith urn is emptied when the system is in state ℓ and with mark u. If $U=(U_k)$ is an i.i.d. sequence of uniform random variables on $[0,1]$ then, clearly, for $i\in\{1,\dots,N\}$ and $j\in\mathcal{H}^N(i)$, and $\ell\in\mathcal{S}_N$,

$$(Z_{ij}(U,\ell), j\in\{1,\dots,N\}) \stackrel{\text{dist.}}{=} B_i(\ell)(\mathrm{d}z_1,\dots,\mathrm{d}z_N), \tag{4.28}$$

where $B_i(\ell)$ is a multinomial distribution with parameters ℓ_i and $p_{i1}^N(\ell)$, ..., $p_{iN}^N(\ell)$, in particular

$$Z_{ij}(U,\ell) \stackrel{\text{dist.}}{=} B_{ij}(\ell)(\mathrm{d}z), \tag{4.29}$$

$B_{ij}(\ell)$ is a binomial distribution with parameter ℓ_i and $p_{ij}^N(\ell)$.

Let $\overline{\mathcal{N}}$ be a marked Poisson point process on $\mathbb{R}_+\times\mathcal{M}$ with intensity measure $\mathrm{d}t\otimes\prod_1^{+\infty}\mathrm{d}u_i$ on $\mathbb{R}_+\times\mathcal{M}$. Such a process can be represented as follows. If $\mathcal{N}=(t_n)$ is a standard Poisson process on $\mathbb{R}_+$ with rate 1 and $((U_k^n), n\in\mathbb{N})$ is a sequence of i.i.d. sequences of uniform random variables on $[0,1]$, then the point process $\overline{\mathcal{N}}$ on $\mathbb{R}_+\times[0,1]^{\mathbb{N}}$ can be defined by

$$\overline{\mathcal{N}}=\sum_{n\geq 1}\delta_{(t_n,(U_k^n))},$$

where $\delta_{(a,b)}$ is the Dirac mass at (a,b). If $A\in\mathcal{B}(\mathbb{R}_+\times\mathcal{M})$ is a Borelian subset of $\mathbb{R}_+\times\mathcal{M}$,

$$\overline{\mathcal{N}}(A) = \int_A \overline{\mathcal{N}}(\mathrm{d}t,\mathrm{d}u)$$

is the number of points of $\overline{\mathcal{N}}$ in A. We denote by $\mathcal{N}$ the point process on $\mathbb{R}_+$ defined by the first coordinates of the points of $\overline{\mathcal{N}}$, i.e. $\mathcal{N}(\mathrm{d}t)=\overline{\mathcal{N}}(\mathrm{d}t,\mathcal{M})$, $\mathcal{N}$ is a Poisson process on $\mathbb{R}_+$ with rate 1.

We denote by $\overline{\mathcal{N}}_i$, $i\in\mathbb{N}$, i.i.d. marked Poisson point processes with the same distribution as $\overline{\mathcal{N}}$. The martingale property mentioned in the following is associated to the natural filtration $(\mathcal{F}_t)$ of these marked Poisson point processes, for $t\geq 0$,

$$\mathcal{F}_t = \sigma\left\langle \overline{\mathcal{N}}_i([0,s]\times B) : i\in\mathbb{N}, s\leq t, B\in\mathcal{B}(\mathcal{M})\right\rangle,$$

We recall an elementary result concerning the martingales associated to marked Poisson point processes. It is used throughout the section. See Subsection 4.5 of Jacobsen [79], see also Last and Brandt [102] for more details.

Proposition 4.9. *For $1 \leq i \leq N$, if h is a Borelian function on $\mathbb{R}_+ \times \mathcal{M}$ depending on a finite number of coordinates of $u \in \mathcal{M}$, càdlàg on the first coordinate and such that*

$$\int_{[0,t]\times\mathcal{M}} h(s,u)^2 \, \mathrm{d}s \, \mathrm{d}u < +\infty, \quad \forall t \geq 0,$$

where $\mathrm{d}u$ denotes the product of Lebesgue measures on $\mathcal{M} = [0,1]^{\mathbb{N}}$, then the process

$$(M(t)) \stackrel{\text{def.}}{=} \left(\int_{[0,t]\times\mathcal{M}} h(s-,u)) \overline{\mathcal{N}}_i(\mathrm{d}s, \mathrm{d}u) - \int_{[0,t]\times\mathcal{M}} h(s,u) \, \mathrm{d}s \, \mathrm{d}u \right)$$

is a square integrable martingale with respect to the filtration $(\mathcal{F}_t)$, its previsible increasing process is given by

$$(\langle M \rangle (t)) = \left(\int_{[0,t]\times\mathcal{M}} h(s,u)^2 \, \mathrm{d}s \, \mathrm{d}u \right).$$

We conclude this subsection with some notations which will be used throughout the section.

Technical Framework

We denote by $M_1(\mathbb{N})$ the set of probability distributions on $\mathbb{N}$. If $\mu \in M_1(\mathbb{N})$ and $f : \mathbb{N} \to \mathbb{R}$,

$$\langle \mu, f \rangle \stackrel{\text{def.}}{=} \int_{\mathbb{N}} f(x) \, \mu(\mathrm{d}x) = \sum_{k \in \mathbb{N}} f(k) \mu(\{k\}),$$

provided that the later term is well defined. If μ_1, $\mu_2 \in M_1(\mathbb{N})$, the total variation norm of $\mu_1 - \mu_2$ is defined by

$$\|\mu_1 - \mu_2\|_{\mathrm{tv}} = \sup_{f : \mathbb{N} \to \{0,1\}} |\, \langle \mu_1, f \rangle - \langle \mu_2, f \rangle \,|. \tag{4.30}$$

The space $M_1(\mathbb{N})$ endowed with the total variation norm is a separable Banach space. For $T \geq 0$, we will denote by $\mathcal{C}([0,T], M_1(\mathbb{N}))$ (resp. $\mathcal{D}([0,T], M_1(\mathbb{N}))$) the space of continuous (resp. càdlàg) functions with values in $M_1(\mathbb{N})$. We denote by $d_T(\cdot,\cdot)$ a distance associated to the topology of $\mathcal{D}_T \stackrel{\text{def.}}{=} \mathcal{D}([0,T], M_1(\mathbb{N}))$.

We introduce the Wasserstein metric W_T on the corresponding stochastic process, on $M_1(\mathcal{D}_T)$, the space of càdlàg process with values in $M_1(\mathbb{N})$. Let Π_1 and $\Pi_2 \in M_1(\mathcal{D}_T)$,

$$W_T(\Pi_1, \Pi_2) \stackrel{\text{def.}}{=} \inf_{\Pi \in [\Pi_1, \Pi_2]} \int_{\mathcal{D}_T^2} [d_T(\Lambda_1, \Lambda_2) \wedge 1] \; \Pi(\mathrm{d}\Lambda_1, \mathrm{d}\Lambda_2), \tag{4.31}$$

where $[\Pi_1, \Pi_2]$ is the set of couplings of distributions Π_1 and Π_2, i.e. $\Pi \in [\Pi_1, \Pi_2]$ is an element of $M_1(\mathcal{D}_T^2)$ such that the marginals are Π_1 and Π_2 respectively. The metric space $(M_1(\mathcal{D}_T), W_T)$ is complete and separable. See Billingsley [19] for example. See also Dawson [42] for a more specific presentation of measure-valued stochastic processes.

We will use the stronger Wasserstein distance $\overline{W}_T$ to establish our convergence results,

$$\overline{W}_T(\Pi_1, \Pi_2) \stackrel{\text{def.}}{=} \inf_{\Pi \in [\Pi_1, \Pi_2]} \int_{\mathcal{D}_T^2} \sup_{s \leq T} \|\Lambda_1(s) - \Lambda_2(s)\|_{\mathrm{tv}} \, \Pi(\mathrm{d}\Lambda_1, \mathrm{d}\Lambda_2). \tag{4.32}$$

We will use the notation C_0 or C_T for several different constants to avoid the multiplication of notations for constants in the text.

4.2.3 Evolution Equations

The state of the system at time t is denoted by a càdlàg process $(L^N(t))$, with

$$(L^N(t)) = \left(\left(L_i^N(t), 1 \leq i \leq N\right)\right) \in \mathcal{S}_N,$$

$L_i^N(t)$ is the number of balls in urn i at time t. One defines the *local empirical distribution* at i, $1 \leq i \leq N$, at time $t \geq 0$ by

$$\Lambda_i^N(t) \stackrel{\text{def.}}{=} \frac{1}{h_N} \sum_{j \in \mathcal{H}^N(i)} \delta_{L_j^N(t)}, \tag{4.33}$$

the global empirical distribution is

$$\Lambda^N(t) \stackrel{\text{def.}}{=} \frac{1}{N} \sum_{j=1}^{N} \delta_{L_j^N(t)} = \frac{1}{N} \sum_{j=1}^{N} \Lambda_j^N(t). \tag{4.34}$$

Assumptions 4.3 (Initial State)**.**

— The distribution of the vector $L^N(0) \stackrel{\text{def.}}{=} (L_i^N(0)) \in \mathcal{S}_N$ satisfies the relation

$$\left(L_1^N(0), \ldots, L_N^N(0)\right) \stackrel{\text{dist.}}{=} \left(L_2^N(0), \ldots, L_N^N(0), L_1^N(0)\right).$$

— The local empirical distribution of the initial state converges in distribution to a probability distribution π_0 on $\mathbb{N}$ for the total variation distance

$$\lim_{N \to +\infty} \mathbb{E}\left(\|\Lambda_1^N(0) - \pi_0\|_{\text{tv}}\right) = 0. \tag{4.35}$$

— There exists some $\eta > 0$ such that

$$\sup_{N \in \mathbb{N}} \mathbb{E}\left(e^{\eta L_1^N(0)}\right) < +\infty. \tag{4.36}$$

Remark. *Let us assume for the moment that the above condition holds. If f is a bounded function on $\mathbb{N}$ with $\|f\|_\infty \leq 1$, by the symmetry property,*

$$\left|\mathbb{E}\left(f(L_1^N(0))\right) - \langle \pi_0, f \rangle\right| = \left|\mathbb{E}\left(\left\langle \Lambda_1^N(0), f \right\rangle\right) - \langle \pi_0, f \rangle\right| \leq 2\mathbb{E}\left(\|\Lambda_1^N(0) - \pi_0\|_{\text{tv}}\right),$$

the sequence $(L_1^N(0))$ is thus converging in distribution to π_0. For $\eta_0 < \eta$, Relation (4.36) gives the existence of a constant C_0 such that

$$\mathbb{P}(L_1^N(0) \geq x) \leq C_0 e^{-\eta_0 x}, \quad \forall x \geq 0. \tag{4.37}$$

In particular all moments of $L_1^N(0)$ are converging to the corresponding moments of π_0. The same property holds for exponential moments of order $\eta_0 < \eta$. Since $L^N(0) \in \mathcal{S}_N$, implies that $\mathbb{E}(L_1^N(0)) = F_N/N$. we obtain that

$$\int x\, \pi_0(\mathrm{d}x) = \lim_{N \to +\infty} \mathbb{E}\left(L_1^N(0)\right) = \beta,$$

by Relation (4.16). Let I be the identity function on $\mathbb{N}$, for $K>0$, the relation

$$\mathbb{E}\left(\left|\left\langle \Lambda_1^N(0), I\right\rangle - \langle \pi_0, I\rangle\right|\right) \leq \mathbb{E}\left(\left|\left\langle \Lambda_1^N(0) - \pi_0, I\mathbb{1}_{[0,K]}\right\rangle\right|\right) + \mathbb{E}\left(L_1^N(0)\mathbb{I}_{\{L_1^N(0)\geq K\}}\right) + \left\langle \pi_0, I\mathbb{1}_{[K,+\infty)}\right\rangle,$$

the convergence in distribution and the exponential estimate (4.37) show that the right hand side of this inequality can be made arbitrarily small as N gets large. Therefore, under Condition 4.3, we have the convergence

$$\lim_{N\to+\infty} \mathbb{E}\left(\left|\left\langle \Lambda_1^N(0)-\pi_0, I\right\rangle\right|\right) = \mathbb{E}\left(\left|\left\langle \Lambda_1^N(0), I\right\rangle - \beta\right|\right) = 0.$$

The process $(L_i^N(t))$ can be represented as the solution of the following stochastic differential equation (SDE), for $1\leq i\leq N$,

$$\mathrm{d}L_i^N(t) = \sum_{j\in\mathcal{H}^N(i)} \int_{\mathcal{M}} Z_{ji}(u, L^N(t-))\,\overline{\mathcal{N}}_j(\mathrm{d}t, \mathrm{d}u) - L_i^N(t-)\,\mathcal{N}_i(\mathrm{d}t), \tag{4.38}$$

where $f(t-)$ denotes the left limit of the function f at $t>0$. For $i\in\{1,\ldots,N\}$, the points of the process $\mathcal{N}_i(\mathrm{d}t)$ correspond to the instants when the ith urn is emptied. Recall the notation $\mathcal{N}_i(\mathrm{d}t)=\overline{\mathcal{N}}_i(\mathrm{d}t, \mathcal{M})$. If time t is one of these instants, due to the uniform distribution assumption of the variables (U_k^i), Relation (4.26) gives that, conditionally on $\mathcal{F}_{t-}$, a ball from urn i is allocated to urn j with probability $p_{ij}^N(L^N(t-))$. This shows that the solution of Equation (4.38) does represent our allocation process of balls into the urns.

Evolution Equations for Local Empirical Distributions

Recall that, for any function f with finite support on $\mathbb{N}$,

$$\left\langle \Lambda_i^N(t), f\right\rangle \stackrel{\text{def.}}{=} \int_{\mathbb{N}} f(x)\Lambda_i^N(t)(\mathrm{d}x),$$

The SDE (4.38) can then be rewritten in the following way, for all $1\leq i\leq N$,

$$\left\langle \Lambda_i^N(t), f\right\rangle = \left\langle \Lambda_i^N(0), f\right\rangle + \frac{1}{h_N}\sum_{j\in\mathcal{H}^N(i)} \int_0^t \left[f(0)-f(L_j^N(s-))\right]\mathcal{N}_j(\mathrm{d}s)$$
$$+\frac{1}{h_N}\sum_{\substack{j\in\mathcal{A}^N(i)\\ k\in\mathcal{H}^N(i)\cap\mathcal{H}^N(j)}} \int_{[0,t]\times\mathcal{M}} \left[f\left(L_k^N(s-)+Z_{jk}(u, L^N(s-))\right) - f(L_k^N(s-))\right]\overline{\mathcal{N}}_j(\mathrm{d}s, \mathrm{d}u), \tag{4.39}$$

where the variables $(Z_{jk}(\cdot,\cdot))$ are defined by Relation (4.27) and $\mathcal{A}^N(i)$ by Relation (4.18).

A Heuristic Asymptotic Description

As it will be seen, the proof of the main convergence result of this section in Subsection 4.2.4 is quite involved. We first present an informal, hopefully intuitive,

motivation for the asymptotic SDE satisfied by the time evolution of the number of balls in a given urn. It should be noted that we will not establish our mean-field result in the same way. The method can be used in a simpler setting, see Sun [156]. It does not seem to be possible for our current model. The integration of Equation (4.38) and the use of Proposition 4.9 lead to the relation

$$L_i^N(t) = L_i^N(0) + C_i^N(t) - \int_0^t L_i^N(s-)\,\overline{\mathcal{N}}_i(\mathrm{d}s, \mathcal{M}) \tag{4.40}$$

with $(C_i^N(t))$ is the process associated to the interaction of the nodes in the neighborhood of i,

$$C_i^N(t) = \sum_{j\in\mathcal{H}^N(i)} \int_{[0,t]\times\mathcal{M}} Z_{ji}(u, L^N(s-))\,\overline{\mathcal{N}}_j(\mathrm{d}s, \mathrm{d}u).$$

As it will be seen, under Assumptions 4.2, with high probability the process

$$(Z_{ji}(u, L^N(s-)), u\in\mathcal{M}, s\leq t)$$

is either 0 or 1 and, consequently, $(C_i^N(t))$ is a counting process. Additionally the process $(C_i^N(t)-\widehat{C}_i^N(t))$ is a martingale, where

$$\left(\widehat{C}_i^N(t)\right) = \left(\sum_{j\in\mathcal{H}^N(i)} \int_0^t L_j^N(s) p_{ji}^N(L^N(s))\,\mathrm{d}s\right)$$

is the *compensator* of $(C_i^N(t))$. See Jacobsen [79]. With the definition of $(p_{ji}^N(\ell))$, Relation (4.23) of Assumptions 4.2 gives the equivalence

$$\left(h_N p_{ji}^N(L^N(t))\right) \sim \left(\Psi(\Lambda_j^N(t), L_i^N(t))\right)$$

with high probability on finite time intervals, then

$$\left(\widehat{C}_i^N(t)\right) \sim \left(\int_0^t \frac{1}{h_N} \sum_{j\in\mathcal{H}^N(i)} L_j^N(s)\Psi(\Lambda_j^N(t), L_i^N(s))\,\mathrm{d}s\right).$$

Assuming a convenient mean-field convergence, i.e. that the local empirical measures $(\Lambda_j^N(t))$ [resp. the processes $(L_j^N(t))$] are converging in distribution to a continuous deterministic process $(\Lambda(t))$ [resp. to a process $(L(t))$]. In particular $(\Lambda(t)(f))=(\mathbb{E}(f(L(t))))$ and, due to the fact that $L^N(t)\in\mathcal{S}_N$ and the scaling assumption (4.16),

$$\langle\Lambda(t), I\rangle = \mathbb{E}(L(t)) = \beta, \quad \forall t\geq 0,$$

where $I(x)=x$ is the identity. Under this hypothesis, one would have the equivalence in distribution for the compensator of $(C_i^N(t))$,

$$\left(\widehat{C}_i^N(t)\right) \sim \left(\int_0^t \Psi\left(\Lambda(s), L_i^N(s)\right) \frac{1}{h_N} \sum_{j\in\mathcal{H}^N(i)} L_j^N(s)\,\mathrm{d}s\right)$$
$$\sim \left(\widehat{C}(t)\right) \overset{\text{def.}}{=} \left(\beta \int_0^t \Psi(\Lambda(s), L(s))\,\mathrm{d}s\right).$$

This suggest that the sequence of counting processes $(C_i^N(t))$ is converging in distribution to a counting process $(C(t))$ given by

$$(C(t)) = \left(\int_0^t \overline{\mathcal{P}} \left(\mathrm{d}s \times [0, \beta\Psi(\Lambda(s), L(s-))] \right) \right),$$

where $\overline{\mathcal{P}}$ is an homogeneous Poisson point process on $\mathbb{R}_+^2$. In view of Relation (4.40), one can now introduce a potential candidate for the asymptotic process $(L(t))$.

Existence and Uniqueness of the McKean-Vlasov process

We first establish the existence and uniqueness of such a process. The proof of the convergence in distribution of $(L_i^N(t))$, $1 \leq i \leq N$, to this asymptotic process is achieved in the next subsection.

Theorem 4.4. *If the functional Ψ satisfies Assumptions 4.2, there exists a unique càdlàg process $(L(t))$ with an initial condition $L(0)$ and such that the SDE*

$$\mathrm{d}L(t) = \overline{\mathcal{P}} \left(\mathrm{d}t \times \left[0, \beta\, \Psi(\pi(t), L(t-)) \right] \right) - L(t-)\mathcal{P}(\mathrm{d}t) \tag{4.41}$$

holds, where, for $t>0$, $\pi(t)$ is the distribution of $L(t)$ on $\mathbb{N}$ and $\overline{\mathcal{P}}$ [resp. $\mathcal{P}$] is an homogeneous Poisson point process on $\mathbb{R}_+^2$ [resp. $\mathbb{R}_+$] with rate 1 and the random processes $\overline{\mathcal{P}}$ and $\mathcal{P}$ are independent .

This will be referred to as the McKean-Vlasov process associated to this model. This non-linear Markov process can be also be seen analytically as the solution of the following Fokker-Planck equation, i.e. find $(\pi(t))$ a continuous $M_1(\mathbb{N})$-valued function such that, for any function f with finite support on $\mathbb{N}$,

$$\langle \pi(t), f \rangle = \langle \pi(0), f \rangle + \int_0^t \left\langle \pi(s), \Omega_{\pi(s)} f \right\rangle \mathrm{d}s, \quad \forall t \geq 0, \tag{4.42}$$

where the generator Ω is defined by, for $\sigma \in M_1(\mathbb{N})$,

$$\Omega_\sigma(f)(x) = \beta\Psi(\sigma, x)\,(f(x+1) - f(x)) + (f(0) - f(x)), \tag{4.43}$$

for $x \in \mathbb{N}$. See Sznitman [158].

Proof. Recall that $\mathcal{C}(\mathbb{R}_+, M_1(\mathbb{N}))$ is the set of continuous mappings on $\mathbb{R}_+$ with values in $M_1(\mathbb{N})$ endowed with the topology associated to the total variation distance. A coupling approach is used. For any $\pi = (\pi(t)) \in \mathcal{C}(\mathbb{R}_+, M_1(\mathbb{N}))$, we can construct the solution $(L_\pi(t))$ of the SDE (4.41) with $L_\pi(0) = L_0 \in \mathbb{N}$, with $L_0 \overset{\text{dist.}}{=} \pi_0$, on the same probability space. All we have to show is that the mapping

$$\begin{aligned} \Phi : \mathcal{C}(\mathbb{R}_+, M_1(\mathbb{N})) &\to \mathcal{C}(\mathbb{R}_+, M_1(\mathbb{N})) \\ (\pi(t)) &\longrightarrow (\mathbb{P}(L_\pi(t) \in \cdot)) \end{aligned}$$

has a unique fixed point. Let f be a real-valued function on $\mathbb{N}$ then, for $t\geq 0$,

$$f(L_\pi(t)) = f(L_0) + \int_0^t \Delta_1(f)(L_\pi(s-))\overline{\mathcal{P}}\left(\mathrm{d}s\times\left[0,\beta\,\Psi(\pi(s),L_\pi(s-))\right]\right) + \int_0^t (f(0)-f(L_\pi(s-)))\mathcal{P}(\mathrm{d}s), \quad (4.44)$$

where, for $x\geq 0$, $\Delta_1(f)(x)=f(x+1)-f(x)$. By using standard properties of Poisson processes, one gets the relation

$$\mathbb{E}(f(L_\pi(t))) = \mathbb{E}(f(L_0)) + \beta\int_0^t \mathbb{E}(\Delta_1(f)(L_\pi(s))\Psi(\pi(s),L_\pi(s)))\,\mathrm{d}s + \int_0^t \mathbb{E}[f(0)-f(L_\pi(s))]\,\mathrm{d}s,$$

In particular $(\mathbb{E}(f(L_\pi(t))))$ is a continuous function, so that Φ is indeed well defined.

For $\pi_1=(\pi_1(t))$ and $\pi_2=(\pi_2(t))\in\mathcal{C}(\mathbb{R}_+,M_1(\mathbb{N}))$, and $t\geq 0$, let $m(t)$ [resp. $M(t)$] be the minimum [resp. maximum] of $\Psi(\pi_1(t),L_{\pi_1}(t))$ and $\Psi(\pi_2(t),L_{\pi_2}(t))$, then Equation (4.44) gives the relation

$$|L_{\pi_1}(t)-L_{\pi_1}(t)| \leq \int_0^t \overline{\mathcal{P}}\,(\mathrm{d}s\times[\beta m(s-),\beta M(s-)]) + \int_0^t |L_{\pi_1}(s)-L_{\pi_2}(s)|\mathcal{P}(\mathrm{d}s).$$

By integrating this inequality, we get

$$h(t)\stackrel{\text{def.}}{=}\mathbb{E}\left(|L_{\pi_1}(t)-L_{\pi_1}(t)|\right) \leq \beta\int_0^t \mathbb{E}\left(|\Psi(\pi_1(s),L_{\pi_1}(s))-\Psi(\pi_2(s),L_{\pi_2}(s))|\right)\,\mathrm{d}s + \int_0^t \mathbb{E}\left(|L_{\pi_1}(s)-L_{\pi_2}(s)|\right)\,\mathrm{d}s.$$

From the Lipschitz conditions (4.25) we get the relation

$$|\Psi(\pi_1(s),L_{\pi_1}(s))-\Psi(\pi_2(s),L_{\pi_2}(s))| \leq D_\Psi\|\pi_1(s)-\pi_2(s)\|_{\mathrm{tv}} + C_\Psi|L_{\pi_1}(s)-L_{\pi_1}(s)|,$$

and therefore the existence of constants a and b such that

$$h(t) \leq a\int_0^t h(s)\,\mathrm{d}s + b\int_0^t\|\pi_1(s)-\pi_2(s)\|_{\mathrm{tv}}\,\mathrm{d}s.$$

From Grönwall's Inequality, we get that

$$h(t) \leq be^{at}\int_0^t \|\pi_1(s)-\pi_2(s)\|_{\mathrm{tv}}\,\mathrm{d}s.$$

Since

$$\|\Phi(\pi_1)(t)-\Phi(\pi_2)(t)\|_{\mathrm{tv}}=\frac{1}{2}\sup_{f:\|f\|_\infty\leq 1}|\mathbb{E}(f(L_{\pi_1}(t)))-\mathbb{E}(f(L_{\pi_2}(t)))| \leq \mathbb{E}(|L_{\pi_1}(t)-L_{\pi_2}(t)|) = h(t),$$

for $T>0$, there is some constant $C_T\geq 0$ such that for any $t\leq T$, the relation

$$\|\Phi(\pi_1)(t)-\Phi(\pi_2)(t)\|_{\mathrm{tv}} \leq C_T\int_0^t \|\pi_1(s)-\pi_2(s)\|_{\mathrm{tv}}\,\mathrm{d}s$$

holds. From there, the existence and uniqueness result of a fixed point for Φ is established with standard arguments. The theorem is proved. □

Proposition 4.10. *If there exists some η_0 such that $\mathbb{E}(\exp(\eta_0 L(0)))<+\infty$, then, for $T>0$, there exists some $\eta_T>0$ such that*

$$\mathbb{E}\left(\sup_{t\leq T} e^{\eta_T L(t)}\right) < +\infty,$$

where $(L(t))$ is the solution of SDE (4.41).

Proof. With the notations of Theorem 4.4, SDE (4.41) gives the inequality, for all $t\leq T$,

$$L(t) \leq L(0)+\overline{\mathcal{P}}\left([0,T]\times[0,\beta\|\Psi\|_\infty]\right),$$

which leads to the desired result. □

We conclude the subsection with the following straightforward lemma.

Lemma 4.3. *If $\Omega_\cdot$ is the operator defined by Relation (4.43), then under Assumptions 4.2, for any function f on $\mathbb{N}$ with $\|\Delta_1(f)\|_\infty<+\infty$, the relation*

$$\left|\langle\sigma,\Omega_\sigma(f)\rangle-\langle\sigma',\Omega_{\sigma'}(f)\rangle\right| \leq 2\beta\|\Delta_1(f)\|_\infty\left[\|\Psi\|_\infty+D_\Psi\right]\|\sigma-\sigma'\|_{\mathrm{tv}}+\left|\langle\sigma-\sigma',f\rangle\right|$$

holds for any σ, $\sigma'\in M_1(\mathbb{N})$ such that f is integrable with respect to σ and σ'.

Proof. This is a simple consequence of the inequality

$$\left|\langle\sigma,\Omega_\sigma(f)\rangle-\langle\sigma',\Omega_{\sigma'}(f)\rangle\right| \leq \left|\langle\sigma,\Omega_\sigma(f)-\Omega_{\sigma'}(f)\rangle\right|+\left|\langle\sigma-\sigma',\Omega_{\sigma'}(f)\rangle\right|$$

and Relation (4.25). □

4.2.4 Mean-Field Convergence Results

The general strategy to establish the mean-field convergence is to use Relation (4.39) and decompose it in a convenient way with the help of (careful) stochastic calculus for Poisson process and of several technical estimates. We begin by recalling and introducing some notations which will be used throughout this subsection.

— As before, if f is some function on $\mathbb{N}$, for $y\in\mathbb{N}$, Δ_y is the operator defined by, for $x\in\mathbb{N}$,

$$\Delta_y(f)(x)\stackrel{\text{def.}}{=}f(x+y)-f(x).$$

— The set of 1-Lipschitz functions on N is denoted as $\mathrm{Lip}(1)$,

$$\mathrm{Lip}(1) = \{f:\mathbb{N}\to\mathbb{R} \mid \|\Delta_1(f)\|_\infty\leq 1\},$$

and, for $K\geq 0$,

$$I_K(x)\stackrel{\text{def.}}{=}x\mathbb{I}_{\{x\geq K\}} \quad \text{and } I(x)\stackrel{\text{def.}}{=}x. \tag{4.45}$$

Relation (4.39) is rewritten by compensating the stochastic integrals with respect to the Poisson processes then, by using Proposition 4.9, one gets that, for a bounded function f on $\mathbb{N}$ and $t\geq 0$,

$$\begin{aligned}\left\langle\Lambda_i^N(t),f\right\rangle &= \left\langle\Lambda_i^N(0),f\right\rangle+\int_0^t\left\langle\Lambda_i^N(s),f(0)-f(\cdot)\right\rangle\mathrm{d}s\\ &+\sum_{z\geq 1}\int_0^t\sum_{j\in\mathcal{H}^N(i)}\frac{1}{h_N}\quad\Delta_z(f)\left(L_j^N(s)\right)\left(\sum_{k\in\mathcal{H}^N(j)}B_{kj}(L^N(s))(\mathrm{d}z)\right)\mathrm{d}s+M_{f,i}^N(t),\end{aligned} \tag{4.46}$$

$B_{kj}(\cdot)$ is the binomial distribution defined by Relation (4.29) and $(M^N_{f,i}(t))$ is a martingale whose previsible increasing process $(\langle M^N_{f,i}(t)\rangle)$ is given, via some simple calculations, by

$$\left\langle M^N_{f,i}(t)\right\rangle = \frac{1}{h_N^2}\sum_{j\in\mathcal{A}^N(i)}\int_0^t\int_{(z_i)\in\mathbb{N}^N}\Bigg(\left[f(0)-f(L^N_j(s))\right]\mathbb{I}_{\{j\in\mathcal{H}^N(i)\}} + \sum_{k\in\mathcal{H}^N(i)\cap\mathcal{H}^N(j)}\Delta_{z_k}(f)(L^N_k(s))\Bigg)^2 B_j(L^N(s))(\mathrm{d}z_1,\ldots,\mathrm{d}z_N)\,\mathrm{d}s. \quad (4.47)$$

where $\mathcal{A}^N(i)$ is defined by Relation (4.18) and $B_j(\cdot)$ is the multinomial distribution defined by Relation (4.28).

We introduce the potential asymptotic process of Theorem 4.4 in this relation. A careful (and somewhat cumbersome) rewriting of Relation (4.46) gives the identity

$$\left\langle \Lambda^N_i(t), f\right\rangle = \left\langle \Lambda^N_i(0), f\right\rangle + \int_0^t \left\langle \Lambda^N_i(s), \Omega_{\Lambda^N_i(s)}(f)\right\rangle \mathrm{d}s + X^N_{f,i}(t) + Y^N_{f,i}(t) + Z^N_{f,i}(t) + M^N_{f,i}(t), \quad (4.48)$$

where $\Omega_{\cdot}$ is the operator defined by Relation (4.43) and I is the identity function. Note that the relation $\langle\pi(s), I\rangle = \beta$, $s\geq 0$, has been used in this derivation. The other terms are

$$X^N_{f,i}(t) \stackrel{\text{def.}}{=} \frac{1}{h_N}\int_0^t \sum_{j\in\mathcal{H}^N(i)}\Delta_1(f)(L^N_j(s))\left\langle \Lambda^N_j(s)-\pi(s), I\right\rangle \Psi\left(\Lambda^N_i(s), L^N_j(s)\right)\mathrm{d}s, \quad (4.49)$$

where $(\pi(s))$ is defined by Theorem 4.4, and

$$Y^N_{f,i}(t) \stackrel{\text{def.}}{=} \frac{1}{h_N}\int_0^t \sum_{\substack{j\in\mathcal{H}^N(i)\\ k\in\mathcal{H}^N(j)}}\Delta_1(f)(L^N_j(s))\frac{L^N_k(s)}{h_N}\left(\Psi\left(\Lambda^N_k(s), L^N_j(s)\right) - \Psi\left(\Lambda^N_i(s), L^N_j(s)\right)\right)\mathrm{d}s. \quad (4.50)$$

The term $L^N_k(s)/h_N$ in the expression of $(Y^N_{f,i}(t))$ is the main source of difficulty to prove the mean-field convergence. When the sequence (h_N) grows linearly with N then, since $|L^N_k|\leq F_N\sim\beta N$, this term is bounded and the usual contraction methods, via Grönwall's Inequality, can be used without too much difficulty. A more careful approach has to be considered if the growth of (h_N) is sublinear. Finally,

$$\begin{aligned} Z^N_{f,i}(t) \stackrel{\text{def.}}{=} & \int_0^t \sum_{j\in\mathcal{H}^N(i)}\frac{1}{h_N}\sum_{z\geq 2}\Delta_z(f)\left(L^N_j(s)\right)\left(\sum_{k\in\mathcal{H}^N(j)} B_{kj}(L^N(s))(\mathrm{d}z)\right)\mathrm{d}s \\ & + \int_0^t \sum_{\substack{j\in\mathcal{H}^N(i)\\ k\in\mathcal{H}^N(j)}}\frac{1}{h_N}\Delta_1(f)\left(L^N_j(s)\right)L^N_k(s)p^N_{kj}\left(L^N(s)\right)\left(\left(1-p^N_{kj}\left(L^N(s)\right)\right)^{L^N_k(s)-1}-1\right)\mathrm{d}s \\ & + \int_0^t \sum_{\substack{j\in\mathcal{H}^N(i)\\ k\in\mathcal{H}^N(j)}}\frac{1}{h_N}\Delta_1(f)\left(L^N_j(s)\right)\frac{L^N_k(s)}{h_N}\left(h_N p^N_{kj}\left(L^N(s)\right) - \Psi\left(\Lambda^N_k(s), L^N_j(s)\right)\right)\mathrm{d}s \end{aligned} \quad (4.51)$$

where $(B_{kj}(\ell))$ are the binomial distributions defined by Relation (4.29).

We first consider the last four terms of Relation (4.48) via three technical lemmas.

Lemma 4.4. *Under Condition (4.3), for any $T\geq 0$, there exists a finite constant C_T such that*

$$\sup_{N\geq 1}\sup_{t\leq T}\mathbb{E}\left(L_1^N(t)^2\right)\leq C_T.$$

Proof. Condition (4.3) shows that the quantity

$$q_0\overset{\text{def.}}{=}\sup_{N\geq 1}\mathbb{E}\left(L_1^N(0)^2\right)$$

is finite. From Relation (4.23) and the boundedness of the functional Ψ, we get the existence of a constant C_0 and $N_0\in\mathbb{N}$, such that, for $N\geq N_0$, the relation

$$\sup_{i\in\mathcal{H}^N,\ell\in\mathcal{S}_N}p_i^N(\ell)\leq\frac{C_0}{h_N}<1 \tag{4.52}$$

holds. Proposition 4.9 and Relation (4.38) give the identity

$$\begin{aligned}\mathbb{E}\left(L_1^N(t)^2\right)=&\mathbb{E}\left(L_1^N(0)^2\right)\\&+\sum_{j\in\mathcal{H}^N}\int_0^t\mathbb{E}\left(\int_{\mathbb{N}}\left(2L_1^N(s)z+z^2\right)B_{j1(L^N(s))}(\mathrm{d}z)\right)\mathrm{d}s-\int_0^t\mathbb{E}\left(L_1^N(s)^2\right)\mathrm{d}s.\end{aligned} \tag{4.53}$$

For $s\geq 0$, by using Relation (4.52) and the symmetry of the model, we get

$$\begin{aligned}\mathbb{E}\left(L_1^N(s)\int_{\mathbb{N}}zB_{j1(L^N(s))}(\mathrm{d}z)\right)&=\mathbb{E}\left(L_1^N(s)L_j^N(s)p_{j1}^N\left(L^N(s)\right)\right)\\&\leq\frac{C_0}{h_N}\mathbb{E}\left(L_1^N(s)L_j^N(s)\right)\leq\frac{C_0}{h_N}\mathbb{E}\left(L_1^N(s)^2\right),\end{aligned}$$

by Cauchy-Shwartz's Inequality. Similarly, by using the expression of the second moment of a binomial variable,

$$\begin{aligned}\mathbb{E}\left(\int_{\mathbb{N}}z^2B_{j1(L^N(s))}(\mathrm{d}z)\right)&\leq\mathbb{E}\left(L_1^N(s)^2p_{j1}^N\left(L^N(s)\right)^2\right)+\mathbb{E}\left(L_1^N(s)p_{j1}^N\left(L^N(s)\right)\right)\\&\leq\frac{C_0}{h_N}\mathbb{E}\left(L_1^N(s)^2\right)\left(1+\frac{C_0}{h_N}\right).\end{aligned}$$

By plugging these estimates in Equation (4.53), we obtain the following inequality, for all $N\geq 1$,

$$\mathbb{E}\left(L_1^N(t)^2\right)\leq q_0+(3C_0-1)\int_0^t\mathbb{E}\left(L_1^N(s)^2\right)\mathrm{d}s.$$

A straightforward use of Grönwall's Inequality gives then directly the estimation since the constants q_0 and C_0 do not depend on N. The lemma is proved. □

Lemma 4.5. *Under Conditions (4.2) and (4.3), if $(Z^N_{f,i}(t))$ is the process defined by Relation (4.51), for $T\geq 0$,*

$$\|Z^N\|_T \stackrel{\text{def.}}{=} \mathbb{E}\left(\sup_{t\leq T}\sup_{f\in \mathrm{Lip}(1)}\left|Z^N_{f,i}(t)\right|\right)$$

then the sequence $(\|Z^N\|_T)$ converges to 0.

Proof. We denote by $\delta^N_1(f,t)$, $\delta^N_2(f,t)$ and $\delta^N_3(f,t)$ the three terms of the right hand side of Definition (4.51) of $(Z^N_{f,i}(t))$.

Let, for $l\in\mathbb{N}$, $\overline{B}(l)$ be binomial distribution with parameter l and $p_0 \stackrel{\text{def.}}{=} C_0/h_N$ defined in Relation (4.52), then

$$\mathbb{E}\left(\overline{B}(l)\mathbb{I}_{\{\overline{B}(l)\geq 2\}}\right) = \mathbb{E}\left(\overline{B}(l)\right) - \mathbb{P}\left(\overline{B}(l)=1\right) = lp_0\left(1-(1-p_0)^{l-1}\right) \leq (lp_0)^2. \tag{4.54}$$

By using this inequality and the fact that if $f\in\mathrm{Lip}(1)$, then $\|\Delta_z(f)\|_\infty\leq z$ holds, for $z\geq 1$. For $T>0$,

$$\mathbb{E}\left(\sup_{t\leq T}\sup_{f\in\mathrm{Lip}(1)}\left|\delta^N_1(f,t)\right|\right)$$
$$\leq \frac{1}{h_N}\int_0^T \sum_{\substack{j\in\mathcal{H}^N(i)\\ k\in\mathcal{H}^N(j)}} \mathbb{E}\left(\overline{B}(L^N_k(s))\mathbb{I}_{\{\overline{B}(L^N_k(s))\geq 2\}}\right)\mathrm{d}s \leq \frac{C_0^2}{2h_N}\int_0^T \mathbb{E}\left(L^N_k(s)^2\right)\mathrm{d}s.$$

From Lemma 4.4 we deduce that the right hand side of the relation is converging to 0 as N gets large. A similar argument can also be used for the term $(\delta^N_2(f,t))$. For the last term of $(Z^N_{f,i}(t))$

$$\mathbb{E}\left(\sup_{t\leq T}\sup_{f\in\mathrm{Lip}(1)}\left|\delta^N_3(f,t)\right|\right) \leq \sup_{\substack{i\in\mathcal{H}^N\\ \ell\in\mathcal{S}_N}}\left|h_N p^N_i(\ell) - \Psi\left(\frac{1}{h_N}\sum_{j\in\mathcal{H}^N}\delta_{\ell_j},\ell_i\right)\right|\int_0^T \mathbb{E}\left(L^N_k(s)\right)\mathrm{d}s.$$

Relation (4.23) of Condition (4.2) shows that this term is converging to 0 when N goes to infinity. The lemma is proved. □

Lemma 4.6. *Under Conditions (4.1) and (4.3), the relation*

$$\lim_{N\to+\infty}\sup_{f\in\mathrm{Lip}(1)}\mathbb{E}\left(\sup_{t\leq T}\left(M^N_{f,i}(t)\right)^2\right) = 0,$$

holds for any $T\geq 0$, where, for $f\in\mathrm{Lip}(1)$, $(M^N_{f,i}(t))$ is the martingale of SDE (4.46).

Proof. By using Relation (4.47), we get the inequality

$$\mathbb{E}\left(\sup_{f\in\mathrm{Lip}(1)}\left\langle M^N_{f,i}\right\rangle(T)\right) \leq \frac{2}{h_N^2}\sum_{j\in\mathcal{A}^N(i)}\int_0^T\Bigg[\mathbb{E}\left(L^N_j(s)^2\right)$$
$$+\mathbb{E}\left(\int_{(z_i)\in\mathbb{N}}\left(\sum_{k\in\mathcal{H}^N(i)} z_k\right)^2 B_j(L^N(s))(\mathrm{d}z_1,\ldots,\mathrm{d}z_N)\right)\Bigg]\mathrm{d}s.$$

Note that, for $\ell\in\mathcal{S}_N$, the multinomial distribution $B_j(\ell)$ on $\mathbb{N}^N$ has the support $\{z\in\mathbb{N}^N{:}z_1+z_2+\cdots+z_N=\ell_j\}$, which gives the relation

$$\mathbb{E}\left(\sup_{f\in\mathrm{Lip}(1)}\left\langle M^N_{f,i}\right\rangle(T)\right)\leq 4\frac{\mathrm{card}(\mathcal{A}^N(i))}{h_N^2}\int_0^T\mathbb{E}\left(L_1^N(s)^2\right)\mathrm{d}s.$$

We conclude the proof by using again Lemma 4.4, Condition (4.1) and Doob's Inequality. □

Now we can turn to the remaining terms of Relation (4.48) to establish the convergence results.

Proposition 4.11. *Under Condition (4.2), for $T\geq 0$ and $t\leq T$,*

$$\|X^N\|_t\stackrel{\mathrm{def.}}{=}\mathbb{E}\left(\sup_{f\in\mathrm{Lip}(1)}\sup_{s\leq t}\left|X^N_{f,1}(s)\right|\right)\leq\|\Psi\|_\infty\int_0^t\mathbb{E}\left(\left|\left\langle\Lambda_1^N(s)-\pi(s),I\right\rangle\right|\right)\mathrm{d}s,\tag{4.55}$$

and for $K>0$,

$$\begin{aligned}\|Y^N\|_t\stackrel{\mathrm{def.}}{=}\mathbb{E}\left(\sup_{f\in\mathrm{Lip}(1)}\sup_{s\leq t}\left|Y^N_{f,1}(s)\right|\right)&\leq 4D_\Psi K\int_0^t\mathbb{E}\left(\left\|\Lambda_1^N(s)-\pi(s)\right\|_{\mathrm{tv}}\right)\mathrm{d}s\\&+D_\Psi\int_0^t\mathbb{E}\left(\left|\left\langle\Lambda_1^N(s)-\pi(s),I\right\rangle\right|\right)\mathrm{d}s+D_\Psi\int_0^t\mathbb{E}\left(\langle\pi(s),I_K\rangle\right)\mathrm{d}s\end{aligned}\tag{4.56}$$

where I is the identity function, and $(\pi(s))$ is defined by Theorem 4.4.

Proof. The first Inequality is straightforward to derive from Relation (4.49).

Let $f\in\mathrm{Lip}(1)$ and $t\leq T$, by the Lipschitz property of Relation (4.25) we get that

$$\begin{aligned}\sup_{s\leq t}\left|Y^N_{f,1}(s)\right|&\leq\frac{1}{h_N^2}\int_0^t\sum_{\substack{j\in\mathcal{H}^N(1)\\k\in\mathcal{H}^N(j)}}L_k^N(s)\left|\Psi\left(\Lambda_k^N(s),L_j^N(s)\right)-\Psi\left(\Lambda_1^N(s),L_j^N(s)\right)\right|\mathrm{d}s\\&\leq\frac{D_\Psi}{h_N^2}\int_0^t\sum_{j\in\mathcal{H}^N(1)}\sum_{k\in\mathcal{H}^N(j)}L_k^N(s)\left\|\Lambda_k^N(s)-\Lambda_1^N(s)\right\|_{\mathrm{tv}}\mathrm{d}s.\end{aligned}\tag{4.57}$$

For $s\geq 0$,

$$\begin{aligned}&\frac{1}{h_N^2}\sum_{j\in\mathcal{H}^N(1)}\sum_{k\in\mathcal{H}^N(j)}L_k^N(s)\left\|\Lambda_k^N(s)-\Lambda_1^N(s)\right\|_{\mathrm{tv}}\\&\qquad\leq\frac{1}{h_N^2}\sum_{j\in\mathcal{H}^N(1)}\sum_{k\in\mathcal{H}^N(j)}\left[K\left\|\Lambda_k^N(s)-\Lambda_1^N(s)\right\|_{\mathrm{tv}}+L_k^N(s)\mathbb{I}_{\left\{L_k^N(s)\geq K\right\}}\right],\end{aligned}$$

since $\|\Lambda_k^N(s)-\Lambda_1^N(s)\|_{\mathrm{tv}}\leq 1$. We get therefore, by symmetry, that

$$\begin{aligned}&\frac{1}{h_N^2}\sum_{j\in\mathcal{H}^N(1)}\sum_{k\in\mathcal{H}^N(j)}\mathbb{E}\left(L_k^N(s)\left\|\Lambda_k^N(s)-\Lambda_1^N(s)\right\|_{\mathrm{tv}}\right)\\&\quad\leq\frac{1}{h_N^2}\sum_{j\in\mathcal{H}^N(1)}\sum_{k\in\mathcal{H}^N(j)}\left[2K\mathbb{E}\left(\left\|\Lambda_1^N(s)-\pi(s)\right\|_{\mathrm{tv}}\right)+\mathbb{E}\left(L_1^N(s)\mathbb{I}_{\left\{L_1^N(s)\geq K\right\}}\right)\right]\\&\quad\leq 2K\mathbb{E}\left(\left\|\Lambda_1^N(s)-\pi(s)\right\|_{\mathrm{tv}}\right)+\mathbb{E}\left(L_1^N(s)\mathbb{I}_{\left\{L_1^N(s)\geq K\right\}}\right),\end{aligned}$$

and this term is smaller than

$$2K\mathbb{E}\left(\left\|\Lambda_1^N(s)-\pi(s)\right\|_{\mathrm{tv}}\right)+\mathbb{E}\left(\left|\left\langle\Lambda_1^N(s)-\pi(s),I_K\right\rangle\right|\right)+\langle\pi(s),I_K\rangle$$
$$\leq 4K\mathbb{E}\left(\left\|\Lambda_1^N(s)-\pi(s)\right\|_{\mathrm{tv}}\right)+\mathbb{E}\left(\left|\left\langle\Lambda_1^N(s)-\pi(s),I\right\rangle\right|\right)+\langle\pi(s),I_K\rangle\,,$$

which gives the desired result. □

For the next step to prove the main mean-field result, one has to estimate the deviations of the local mean,

$$\mathbb{E}\left(\sup_{s\leq t}\left|\left\langle\Lambda_1^N(s)-\pi(s),I\right\rangle\right|\right)=\mathbb{E}\left(\sup_{s\leq t}\left|\left\langle\Lambda_1^N(s),I\right\rangle-\beta\right|\right),$$

this is the next proposition. We define, for $t\geq 0$, $d^N(t)\stackrel{\text{def.}}{=}d_1^N(t)+d_2^N(t)$, with

$$\begin{cases} d_1^N(t)\stackrel{\text{def.}}{=}\mathbb{E}\left(\sup_{s\leq t}\left\|\Lambda_1^N(s)-\pi(s)\right\|_{\mathrm{tv}}\right), \\ d_2^N(t)\stackrel{\text{def.}}{=}\mathbb{E}\left(\sup_{s\leq t}\left|\left\langle\Lambda_1^N(s)-\pi(s),I\right\rangle\right|\right). \end{cases} \tag{4.58}$$

We are going to show that, for $T>0$, the sequence $(d^N(T))$ is converging to 0. Let

$$\overline{d}(t)=\limsup_{N\to+\infty}d^N(t).$$

Proposition 4.12. *Under Conditions (4.1), (4.2) and (4.3), for any $T>0$, there exists a constant $C_0>0$ such that the relation*

$$\overline{d}(t)\leq C_0K\int_0^t\overline{d}(s)\,\mathrm{d}s+C_0TK\sup_{s\leq T}\mathbb{E}\left(\langle\pi(s),I_K\rangle\right). \tag{4.59}$$

holds for all $t\leq T$ and $K\geq 1$.

Proof. Noting that $I\in\mathrm{Lip}(1)$, Relations (4.42) and (4.48) give the inequality, for $T>0$ and $t\leq T$,

$$\mathbb{E}\left(\sup_{s\leq t}\left|\left\langle\Lambda_1^N(s)-\pi(s),I\right\rangle\right|\right)\leq\mathbb{E}\left(\left|\left\langle\Lambda_1^N(0)-\pi(0),I\right\rangle\right|\right)$$
$$+\int_0^t\mathbb{E}\left(\left|\left\langle\Lambda_1^N(s),\Omega_{\Lambda_1^N(s)}(I)\right\rangle-\left\langle\pi(s),\Omega_{\pi(s)}(I)\right\rangle\right|\right)\mathrm{d}s$$
$$+\mathbb{E}\left(\sup_{s\leq T}\left|M_{I,1}^N(s)\right|\right)+\left\|X^N\right\|_t+\left\|Y^N\right\|_t+\left\|Z^N\right\|_T,$$

with the notations of Proposition 4.11 and Lemma 4.5. From Lemma 4.3 we get the relation, for $s\geq 0$,

$$\left|\left\langle\Lambda_1^N(s),\Omega_{\Lambda_1^N(s)}(I)\right\rangle-\left\langle\pi(s),\Omega_{\pi(s)}(I)\right\rangle\right|$$
$$\leq(2\|\Psi\|_\infty+D_\Psi)\beta\|\Lambda_1^N(s)-\pi(s)\|_{\mathrm{tv}}+\left|\left\langle\Lambda_1^N(s)-\pi(s),I\right\rangle\right|.$$

By using Proposition 4.11 and Lemma 4.5, we get that

$$\begin{aligned} d_2^N(t) = \mathbb{E}\left(\sup_{s\leq t}\left|\left\langle \Lambda_1^N(s)-\pi(s), I\right\rangle\right|\right) &\leq \mathbb{E}\left(\left|\left\langle \Lambda_1^N(0)-\pi(0), I\right\rangle\right|\right) \\ &+ [(2\|\Psi\|_\infty+D_\Psi)\beta+4D_\Psi K]\int_0^t \|\Lambda_1^N(s)-\pi(s)\|_{\mathrm{tv}}\,\mathrm{d}s \\ &+ [\|\Psi\|_\infty+D_\Psi+1]\int_0^t \mathbb{E}\left(\left|\left\langle \Lambda_1^N(s)-\pi(s), I\right\rangle\right|\right)\,\mathrm{d}s \\ &+ D_\Psi\int_0^t \mathbb{E}\left(\langle \pi(s), I_K\rangle\right)\,\mathrm{d}s + \mathbb{E}\left(\sup_{s\leq T}\left|M_{I,1}^N(s)\right|\right) + \left\|Z^N\right\|_T. \end{aligned} \tag{4.60}$$

We now turn to the estimation of $d_1^N(t)$. For $f{:}\mathbb{N}\to\{0,1\}$ and $T>0$, by Lemma 4.3, if $t\leq T$,

$$\int_0^t \left|\left\langle \Lambda_1^N(s), \Omega_{\Lambda_1^N(s)}(f)\right\rangle - \left\langle \pi(s), \Omega_{\pi(s)}(f)\right\rangle\right|\mathrm{d}s \leq C_0\int_0^t \|\Lambda_1^N(s)-\pi(s)\|_{\mathrm{tv}}\,\mathrm{d}s$$

with $C_0{=}1{+}\beta(2\|\Psi\|_\infty{+}D_\Psi)$, Relations (4.42) and (4.48) give then the inequality

$$\begin{aligned} \left|\left\langle \Lambda_1^N(t)-\pi(t), f\right\rangle\right| &\leq C_0\int_0^t \|\Lambda_1^N(s)-\pi(s)\|_{\mathrm{tv}}\,\mathrm{d}s \\ &+ \left\|\Lambda_1^N(0)-\pi(0)\right\|_{\mathrm{tv}} + \left|M_{f,i}^N(t)\right| + |X_{f,1}^N(t)|+|Y_{f,1}^N(t)|+|Z_{f,1}^N(t)| \end{aligned} \tag{4.61}$$

By definition of the total variation norm, see Relation (4.30), we have

$$\left\|\Lambda_1^N(t)-\pi(t)\right\|_{\mathrm{tv}} = \sup_{f:\mathbb{N}\to\{0,1\}}\left|\left\langle \Lambda_1^N(t)-\pi(t), f\right\rangle\right| \leq R_1^N(t)+R_2^N(t), \tag{4.62}$$

with

$$R_1^N(t) \stackrel{\text{def.}}{=} \sup_{F\subset(K,+\infty)}\left|\left\langle \Lambda_1^N(t)-\pi(t), \mathbb{1}_F\right\rangle\right| \quad \text{and} \quad R_2^N(t) \stackrel{\text{def.}}{=} \sup_{F\subset[0,K]}\left|\left\langle \Lambda_1^N(t)-\pi(t), \mathbb{1}_F\right\rangle\right|.$$

With the same argument as before, we get

$$R_1^N(t) \leq \left\langle \Lambda_1^N(t)+\pi(t), (K,+\infty)\right\rangle \leq \left|\left\langle \Lambda_1^N(t)-\pi(t), \mathbb{1}_{[0,K]}\right\rangle\right| + 2\left\langle \pi(t), I_K\right\rangle \tag{4.63}$$

and therefore

$$R_1^N(t) \leq R_2^N(t) + 2\sup_{s\leq t}\left\langle \pi(s), I_K\right\rangle.$$

Denote by $\mathcal{E}_K{=}\{f{=}\mathbb{1}_F{:}F\subset[0,K]\}$, By taking successively the supremum on all $f\in\mathcal{E}_K$ and $s\leq t$ for Relation (4.61), we obtain the inequality, for $t\leq T$,

$$\begin{aligned} &\mathbb{E}\left(\sup_{s\leq t} R_2^N(s)\right) \leq C_0\int_0^t \mathbb{E}\left(\sup_{u\leq s}\|\Lambda_1^N(u)-\pi(u)\|_{\mathrm{tv}}\right)\mathrm{d}s \\ &+\mathbb{E}\left(\left\|\Lambda_1^N(0)-\pi(0)\right\|_{\mathrm{tv}}\right) + 2^{K+1}\sup_{f\in\mathcal{E}_K}\mathbb{E}\left(\sup_{s\leq t}\left|M_{f,i}^N(s)\right|\right) + \|X^N\|_t + \|Y^N\|_t + \|Z^N\|_T. \end{aligned} \tag{4.64}$$

Again the quantity $\|X^N\|_t+\|Y^N\|_t$ is upper bounded with the help of Proposition 4.11. If we gather the estimates (4.60), (4.62), (4.63) and (4.64), we obtain that, for $T\geq 0$, there exists a constant C_0 independent of $K\geq 1$ and T such that, for any $t\leq T$,

$$\frac{1}{C_0}d^N(t) \leq K\int_0^t d^N(s)\ \mathrm{d}s + KT\sup_{s\leq T}\langle \pi(s), I_K\rangle + \mathbb{E}\left(\left|\left\langle \Lambda_1^N(0)-\pi(0), I\right\rangle\right|\right)$$
$$+\mathbb{E}\left(\left\|\Lambda_1^N(0)-\pi(0)\right\|_{\mathrm{tv}}\right) + 2^K \sup_{f\in\mathcal{E}_K\cup\{I\}} \mathbb{E}\left(\sup_{s\leq t}\left|M_{f,i}^N(s)\right|\right) + \left\|Z^N\right\|_T.$$

Note that, for $s\geq 0$, $|\langle \Lambda_1^N(s)-\pi(s), I\rangle| \leq F_N/N+\beta$, the mapping $s\mapsto d^N(s)$ is therefore bounded by a constant. By using Condition (4.3), Lemmas 4.5 and 4.6, and by applying Fatou's Lemma, we obtain the relation,

$$\overline{d}(t) \leq C_0K\int_0^t \overline{d}(s)\ \mathrm{d}s + KC_0T\sup_{s\leq t}\langle \pi(s), I_K\rangle.$$

The proposition is proved. □

We can now formulate our main convergence result.

Theorem 4.5 (Convergence of Local Empirical Distributions)**.** *If Conditions* (4.1), (4.2) *and* (4.3) *hold, then, for any* $1\leq i\leq N$*, the local empirical distribution process at node* i*,* $(\Lambda_i^N(t))$*,*

$$(\Lambda_i^N(t)) = \left(\frac{1}{\mathrm{card}(\mathcal{H}^N(i))}\sum_{j\in\mathcal{H}^N(i)}\delta_{L_j^N(t)}\right)$$

is converging in distribution to $(\pi(t))$*, where, if* $(L(t))$ *is the solution of McKean-Vlasov SDE* (4.41) *with initial distribution* $\pi(0)$*, for* $t\geq 0$*,* $\pi(t)$ *is the distribution of* $L(t)$*.*

Proof. The notations of the proof of the previous proposition are used. With Grönwall's Inequality and Relation (4.59), we get the relation

$$\overline{d}(t) \leq C_0TK\sup_{s\leq T}\mathbb{E}\left(\langle \pi(s), I_K\rangle\right)e^{C_0Kt}$$

for all $t\leq T$. Proposition 4.10 gives the existence of some $\eta>0$ and $C_1>0$ such that

$$\sup_{s\leq T}\mathbb{E}\left(\langle \pi(s), I_K\rangle\right) \leq C_1e^{-\eta K}, \quad \forall K>0.$$

By letting K go to infinity, we obtain that $\overline{d}(t)=0$ for all $t\leq t_1 \overset{\text{def.}}{=} \eta/(2C_0)\wedge T$ and therefore the desired convergence on the time interval $[0, t_1]$. In particular if we make a time shift at t_1, it is easy to check that Condition (4.3) are also satisfied for this initial state and we can then repeat the same procedure until the time T is reached. The theorem is proved. □

Theorem 4.6 (Mean-Field Convergence)**.** *Under the conditions of Theorem* 4.5*, then, for the convergence in distribution of processes,*

$$\lim_{N\to+\infty}\left(\frac{1}{N}\sum_{i=1}^N\delta_{L_i^N(t)}\right) = (\pi(t)),$$

and, for any $p\geq 2$*, and* $1\leq n_1<n_2<\cdots<n_p$*,*

$$\lim_{N\to+\infty}\left(\left(L^N_{n_1}(t)\right),\left(L^N_{n_2}(t)\right),\ldots,\left(L^N_{n_p}(t)\right)\right)\stackrel{\text{dist.}}{=}\Pi^{\otimes p}$$

where Π *is the distribution of the stochastic process* $(L(t))$*, the solution of McKean-Vlasov SDE* (4.41) *and, for* $s\geq 0$*,* $\pi(s)$ *is the distribution of the marginal of* Π *at* s*, i.e. it is the distribution of* $L(s)$*.*

Proof. By using relation (4.34), one has

$$\sup_{t\leq T}\left\|\Lambda^N(t)-\pi(t)\right\|_{\text{tv}}\leq\frac{1}{N}\sum_{i=1}^N\sup_{t\leq T}\left\|\Lambda^N_i(t)-\pi(t)\right\|_{\text{tv}}.$$

The first claim of the theorem follows directly from Theorem 4.5 and the fact that the processes $(\Lambda^N_i(t))$, $1\leq i\leq N$, have the same distribution. The last assertion is a simple consequence of Proposition 2.2 p. 177 of Sznitman [158]. □

We conclude this subsection with a mean-field convergence result for the invariant distributions when the sequence of the sizes (h_N) of the neighborhoods grows at linearly with respect N. For a fixed N, the finite state space Markov process $(L^N_i(t))$ has a unique invariant distribution. We will denote by $(L^N_i(\infty))$ a random variable whose distribution is invariant for $(L^N_i(t))$.

Theorem 4.7. *Under the conditions of Theorem 4.5, and if*

$$\liminf_{N\to+\infty}h_N/N>0,$$

and π_N *is the distribution at equilibrium of* $(L^N_1(t))$*, then the sequence* (π_N) *is tight and any of its limiting points is an invariant distribution of the McKean-Vlasov process defined by SDE* (4.41)*.*

Proof. Since $(L^N(t))$ is an irreducible finite state Markov chain, it has therefore a unique invariant distribution. We denote by $L^N(\infty)\stackrel{\text{def.}}{=}(L^N_1(\infty),\ldots,L^N_N(\infty))$ a random variable with such a distribution.

We first show that some exponential moments of the variables $L^N_1(\infty)$, $N\geq 1$, are bounded. For $\eta>0$, the balance equation for the function $f(\ell)=\exp(\eta\ell_1)$, $\ell\in\mathcal{S}_N$, gives the relation

$$\mathbb{E}\left(e^{\eta L^N_1(\infty)}\right)-1=\mathbb{E}\left(e^{\eta L^N_1(\infty)}\left[(e^\eta-1)\sum_{i\in\mathcal{H}^N(1)}L^N_i(\infty)p^N_{i1}\left(L^N(\infty)\right)+U_N\left(L^N(\infty)\right)\right]\right),$$

where, for $\ell\in\mathcal{S}_N$,

$$U_N(\ell)=\sum_{i\in\mathcal{H}^N(1)}\left[1+p^N_{i1}(\ell)(e^\eta-1)\right]^{\ell_i}-1-\ell_ip^N_{i1}(\ell)(e^\eta-1).$$

By using Relation (4.52), we have $p_{i1}^N(\ell)\leq C_0/h_N$, we get then the estimation

$$|U_N(\ell)| \leq \sum_{i\in\mathcal{H}^N(1)} (e^\eta-1)^2 \frac{C_0^2\ell_i^2}{2h_N^2}\left(1+\frac{C_0}{h_N}(e^\eta-1)\right)^{\ell_i-2}$$
$$\leq (e^\eta-1)^2\frac{C_0^2}{2}\left(1+\frac{C_0}{h_N}(e^\eta-1)\right)^{F_N}\frac{F_N^2}{h_N^2},$$

since $\ell\in\mathcal{S}_N$. The assumption of the sequence (h_N) shows that there exists some $\eta_1>0$ such that if $\eta<\eta_1$ then the last term is bounded by $D_0(e^\eta-1)^2$ for some constant $D_0\geq 0$. By using this inequality, we get the relation

$$\mathbb{E}\left(e^{\eta L_1^N(\infty)}\right)-1 \leq \mathbb{E}\left(e^{\eta L_1^N(\infty)}\left(\sum_{i\in\mathcal{H}^N(1)} L_i\frac{C_0}{h_N}(e^\eta-1)+D_0(e^\eta-1)^2\right)\right)$$
$$\leq \mathbb{E}\left(e^{\eta L_1^N(\infty)}\left(D_1(e^\eta-1)+D_0(e^\eta-1)^2\right)\right),$$

for some constant $D_1\geq 0$. If $\eta_0<\eta_1$ is such that $D_1(e^{\eta_0}-1)+D_0(e^{\eta_0}-1)^2<1$, then we get that the exponential moments of order η_0 of L_1^N are bounded,

$$\sup_{N\geq 1}\mathbb{E}\left(e^{\eta_0 L_1^N(\infty)}\right) < +\infty. \tag{4.65}$$

This shows in particular that the sequence (π_N) is tight. Let $\overline{\pi}$ be one of its limits. For $K>0$,

$$\mathbb{E}\left(\Lambda_1^N(\infty)([K,+\infty))\right) = \mathbb{P}\left(L_1^N\geq K\right), \tag{4.66}$$

from Lemma 3.2.8 of Dawson [42], we deduce that the sequence of local empirical distribution at equilibrium $(\Lambda_1^N(\infty))$ is tight for the convergence in distribution. We take a subsequence (N_k) so that (π_{N_k}) and $(\Lambda_1^{N_k}(\infty))$ converge.

By Skorohod's representation theorem, one can construct a probability space where the sequence of random probabilities $(\Lambda_1^{N_k}(\infty))$ on $\mathbb{N}$ converges almost surely towards some random probability $\overline{\Lambda}$. In particular, for any bounded function f on $\mathbb{N}$,

$$\mathbb{E}\left(\overline{\Lambda}(f)\right) = \lim_{k\to+\infty}\mathbb{E}\left(\Lambda_1^{N_k}(\infty)(f)\right) = \lim_{k\to+\infty}\int f(x)\,\mathrm{d}\pi_{N_k}(\mathrm{d}x) = \int_{\mathbb{N}} f(x)\,\overline{\pi}(\mathrm{d}x).$$

From Relations (4.65) and (4.66), we get that

$$\lim_{N\to+\infty}\mathbb{E}\left(\left\|\Lambda_1^N(\infty)-\overline{\Lambda}\right\|_{\mathrm{tv}}\right) = 0.$$

Condition (4.3) is therefore satisfied for the initial vector $L^N(\infty)$.

If $(\widetilde{L}^N(t))$ be the solution of SDE (4.38) with initial state $(L^N(\infty))$ and $(\widetilde{\Lambda}_i^N(t))$, $1\leq i\leq N$, are the corresponding local empirical distributions. Theorem 4.5 gives the relation

$$\lim_{N\to+\infty}\mathbb{E}\left(\Lambda_1^N(t)(f)\right) = \mathbb{E}\left(f(\widetilde{L}(t))\right),$$

where $(\widetilde{L}(t))$ is the solution of McKean-Vlasov SDE (4.41) with initial distribution $\overline{\pi}$. By invariance, $\mathbb{E}\left(f(\widetilde{L}(t))\right)$ is constant with respect to t, hence $\overline{\pi}$ is an invariant distribution of the McKean-Vlasov process. The theorem is proved. □

The following subsection is devoted to the properties of the invariant distributions of the McKean-Vlasov process.

4.2.5 Equilibrium of the McKean-Vlasov Process

Recall that β is the average load of an urn at any time. The purpose of this subsection is to investigate the properties of the tail distribution of the number of balls in an urn at equilibrium when β get large. Recall that in our asymptotic picture, the time evolution of the number of balls in a given urn can be seen as the solution of the SDE defining the McKean-Vlasov process

$$\mathrm{d}L_\beta(t) = \overline{\mathcal{P}}\left(\mathrm{d}t\times\left[0,\beta\,\Psi(\pi(t),L_\beta(t-))\right]\right) - L_\beta(t-)\mathcal{P}(\mathrm{d}t), \tag{4.67}$$

where $\overline{\mathcal{P}}$ [resp. $\mathcal{P}$] is an homogeneous Poisson point process on $\mathbb{R}_+^2$ [resp. $\mathbb{R}_+$] with rate 1.

If the process $(L_\beta(t))$ is at equilibrium, then the function $(\pi(t))$ is constant and is equal to some $\pi_\beta\in M_1(\mathbb{N})$. In this case, the process $(L_\beta(t))$ is a simple Markov jump process on $\mathbb{N}$. It is easily checked that any invariant distribution π_β satisfies the following fixed point equation $F_\beta(\pi_\beta)=\pi_\beta$ in $M_1(\mathbb{N})$ where F_β is defined by, for $\pi\in M_1(\mathbb{N})$,

$$F_\beta(\pi)=\big(F_\beta(\pi)(n)\big) \stackrel{\text{def.}}{=} \left(\frac{1}{1+\beta\Psi(\pi,n)}\prod_{k=0}^{n-1}\frac{\beta\Psi(\pi,k)}{1+\beta\Psi(\pi,k)}\right), \tag{4.68}$$

with the convention that $\prod_0^{-1}=1$. Clearly $F_\beta(\pi)\in M_1(\mathbb{N})$ and, due to Relation (4.24), the quantity

$$\sum_{n=0}^{+\infty}\prod_{k=0}^{n}\frac{\beta\Psi(\pi,k)}{1+\beta\Psi(\pi,k)}$$

is β when π is a fixed point of F_β.

The following proposition shows that, for a large class of functionals Ψ, a potential invariant distribution is always concentrated around values proportional to its average β. We will give more precise results later for power of d-choices algorithms. Additionally, an existence and uniqueness result is proved when the average load per urn is small enough.

Proposition 4.13. *If Assumptions 4.2 holds for the functional Ψ,*

1. *if, for $\beta>0$, there exists and invariant distribution π_β for the McKean-Vlasov process (4.41), then it is stochastically bounded by a geometric distribution with parameter $\beta\|\Psi\|_\infty/(1+\beta\|\Psi\|_\infty)$, in particular, for any $x>0$,*

$$\pi_\beta([x,+\infty))\leq \exp\left(-\frac{\beta}{1+\beta\|\Psi\|_\infty}x\left(1-\frac{1}{2(1+\beta\|\Psi\|_\infty)}\right)\right). \tag{4.69}$$

2. *There exists $\beta_0>0$ such that, if $\beta<\beta_0$, then the McKean-Vlasov process (4.41) has a unique invariant distribution.*

Proof. If the initial distribution of the McKean-Vlasov process is π_β, then $(L_\beta(t))$ is a simple Markov process which jumps from n to $n{+}1$ at rate $\beta\Psi(\pi_\beta, n)$ and returns at 0 with rate 1. Denote by $(\widetilde{L}_\beta(t))$ a Markov process on $\mathbb{N}$ with the same characteristics except that the jumps from n to $n{+}1$ occur at rate $\beta\|\Psi\|_\infty$. It is easy to construct a coupling such that $\widetilde{L}_\beta(0){=}L_\beta(0){=}Y_\beta$ and that $L_\beta(t){\le}\widetilde{L}_\beta(t)$ holds for all $t{\ge}0$. By assumption, the distribution of $L_\beta(t)$ is constant with respect to t. It is easy to check that the invariant distribution of $(\widetilde{L}_\beta(t))$ is a geometric distribution with parameter $\beta\|\Psi\|_\infty/(1{+}\beta\|\Psi\|_\infty)$. This proves the first part of the proposition.

For $k{\in}\mathbb{N}$ and $\pi{\in}M_1(\mathbb{N})$, one defines

$$a_\pi(k) \stackrel{\text{def.}}{=} \frac{\beta\Psi(\pi,k)}{1{+}\beta\Psi(\pi,k)},$$

the function F_β can be expressed as, for $n{\in}\mathbb{N}$,

$$F_\beta(\pi)(n) = \prod_0^{n-1} a_\pi(k) - \prod_0^{n} a_\pi(k).$$

From Relations (4.25) and the boundedness of Ψ, we get that

$$\|a_\pi - a_{\pi'}\|_\infty \le \beta D_\Psi \|\pi - \pi'\|_{\text{tv}} \quad \text{and} \quad a_\pi(k) \le \delta \stackrel{\text{def.}}{=} \frac{\beta\|\Psi\|_\infty}{1{+}\beta\|\Psi\|_\infty}$$

hold for any π, $\pi'{\in}M_1(\mathbb{N})$ and $k{\in}\mathbb{N}$. Note that, for $n{\in}\mathbb{N}$,

$$\prod_0^n a_{\pi'}(k) - \prod_0^n a_\pi(k) = \sum_{m=0}^n \prod_0^m a_\pi(k) \prod_{m+1}^n a_{\pi'}(k) - \prod_0^{m-1} a_\pi(k) \prod_m^n a_{\pi'}(k),$$

and therefore that, for $n{\ge}1$,

$$\left|\prod_0^n a_{\pi'}(k) - \prod_0^n a_\pi(k)\right| \le \sum_{m=0}^n |a_\pi(m) - a_{\pi'}(m)|\,\delta^n \le (n{+}1)\delta^n \beta D_\Psi \|\pi - \pi'\|_{\text{tv}}.$$

This gives the Lipschitz property for the mapping F_β for the total variation norm,

$$\begin{aligned}\|F_\beta(\pi') - F_\beta(\pi)\|_{\text{tv}} &= \frac{1}{2}\sum_{n=0}^{+\infty} |F_\beta(\pi')(n) - F_\beta(\pi)(n)| \\ &\le \sum_{n=0}^{+\infty} \left|\prod_0^n a_{\pi'}(k) - \prod_0^n a_\pi(k)\right| \le \frac{\beta D_\Psi}{(1{-}\delta)^2}\|\pi' - \pi\|_{\text{tv}}.\end{aligned}$$

Hence if $\beta D_\Psi(1{+}\beta\|\Psi\|_\infty)^2{<}1$, F_β is a contracting application for the total variation norm. The proposition is proved. □

We have not been able to improve significantly this result. As it will be seen for specific functionals Ψ much more can be said.

Random Weighted Algorithm

For $\sigma \in M_1(\mathbb{N})$, the function Ψ is, in this case, defined by

$$\Psi_{\mathrm{cc}}(\sigma, l) = \frac{W(l)}{\langle \sigma, W \rangle},$$

and the range of W is assumed to be in $[c, C]$, for some positive constants c and C.

It is not difficult to see that there exists a unique invariant distribution π_β. If $\pi \in M_1(\mathbb{N})$ is invariant if and only if it has the representation

$$(\pi(n)) = \left(\frac{\gamma}{\gamma + \beta W(n)} \prod_{k=0}^{n-1} \frac{\beta W(k)}{\gamma + \beta W(k)} \right),$$

where $\gamma = \langle \pi, W \rangle$, i.e.

$$\beta = \sum_{n=0}^{+\infty} \prod_{k=0}^{n} \frac{\beta W(k)}{\gamma + \beta W(k)}. \tag{4.70}$$

It is easy to see that this equation has a unique solution γ_β which gives the existence and the uniqueness of the invariant distribution in this case.

When $(W(k))$ is constant equal to $w > 0$, then balls are placed at random in the neighboring urns, independently of their loads in particular. In this case $\gamma_\beta = w$ and the corresponding invariant measure is the geometric distribution with parameter $\beta/(1+\beta)$. From Theorem 4.7, we get the following proposition.

Proposition 4.14 (Random Algorithm)**.** *Under Assumptions 4.1, for any $\beta > 0$, when N goes to infinity, the equilibrium distribution of the process $(L_1^N(t))$ is converging in distribution to a random variable Y_β with a geometric distribution with parameter $\beta/(1+\beta)$. In particular, for the convergence in distribution,*

$$\lim_{\beta \to +\infty} \frac{Y_\beta}{\beta} = E_1,$$

where E_1 is an exponentially distributed random variable with parameter 1.

It turns out that the random algorithm behaves poorly in terms of the load of a given urn. For a large β, the asymptotic tail distribution of the occupancy of an urn at equilibrium is, as expected, the upper bound (4.69). The simple consequence of this result is that, if on average, there are β balls per urn, there is a significant fraction of urns with an arbitrarily large number of balls. We will see that the situation is completely different for the power of d choices algorithm.

Power of d-choices

For $\sigma \in M_1(\mathbb{N})$, the function Ψ is, in this case, defined by

$$\Psi_{\mathrm{pc}}(\sigma, l) = \frac{(\sigma([l, +\infty))^d - \sigma((l, +\infty))^d}{\sigma(\{l\})},$$

Proposition 4.15. *There exists a unique invariant distribution π_β of $(L_\beta(t))$, it is given by*

$$\pi_\beta(n) = \xi_n - \xi_{n+1}, \quad n \geq 0, \tag{4.71}$$

where (ξ_n) is the non-increasing sequence defined by induction by $\xi_0 = 1$ and

$$\xi_n = \beta\left(\xi_{n-1}^d - \xi_n^d\right), \quad n \geq 1. \tag{4.72}$$

Proof. The existence and uniqueness of such a sequence (ξ_n) is clear. Let π be an invariant distribution of $(L_\beta(t))$, it satisfies the balance equation

$$\pi(n)(1 + \beta\Psi(\pi, n)) = \pi(n-1)\beta\Psi(\pi, n-1), \quad n \geq 1.$$

and $\pi(0)\beta\Psi(\pi, 0) = 1 - \pi(0)$. Define $\xi_n = \pi([n, +\infty))$. The balance equation for $n=0$ shows that Relation (4.72) is clearly true for $n=1$. If we assume that Relation (4.71) holds up to p, the balance equation can be rewritten as

$$(\xi_p - \xi_{p+1}) + \beta\left(\xi_p^d - \xi_{p+1}^d\right) = \beta\left(\xi_{p-1}^d - \xi_p^d\right) = \xi_p,$$

hence Relation (4.72) is valid for $p+1$. The proposition is proved. □

The following theorem shows that the power of d-choices policy is efficient in terms of the load of an arbitrary urn, the invariant distribution of this load is asymptotically concentrated on the finite interval $[0, d/(d-1)\beta]$. Only an extra capacity $\beta/(d-1)$ has to be added to the minimal capacity β for any urn in order to handle properly this allocation policy. This has important algorithmic consequences in some contexts. See Sun et al. [156].

Theorem 4.8 (Power of d choices)**.** *Under Assumptions 4.1, for any $\beta > 0$, when N goes to infinity, the equilibrium distribution of the process $(L_1^N(t))$ is converging in law to a random variable Y_β whose distribution is the unique invariant measure of the McKean-Vlasov process $(L_\beta(t))$ defined by SDE (4.41).*

Furthermore, for the convergence in distribution,

$$\lim_{\beta \to +\infty} \frac{Y_\beta}{\beta} = \frac{d}{d-1}\left(1 - U^{d-1}\right),$$

where U is a uniform random variable on $[0, 1]$.

Proof. The first part of the theorem is a direct consequence of Theorem 4.7 and Proposition 4.15.

For $k \geq 1$, by summing up Equation (4.72) for 1 to $k-1$, one gets the relation

$$\beta\left(1 - \mathbb{P}(Y_\beta \geq k)^d\right) = \mathbb{E}\left(Y_\beta \wedge k\right),$$

since $\mathbb{E}(Y_\beta) = \beta$, this gives

$$\mathbb{P}(Y_\beta \geq k)^d = \frac{1}{\beta}\left(\mathbb{E}(Y_\beta) - \mathbb{E}\left(Y_\beta \wedge k\right)\right) = \mathbb{E}\left(\left(\frac{Y_\beta}{\beta} - \frac{k}{\beta}\right)^+\right),$$

hence, if $Z_\beta \stackrel{\text{def.}}{=} Y_\beta/\beta$, for $x>0$,

$$\mathbb{P}\left(Z_\beta \geq x\right)^d = \mathbb{E}\left(\left(Z_\beta - \frac{\lceil \beta x \rceil}{\beta}\right)^+\right).$$

Relation (4.69) shows that, for β_1 sufficiently large the family of random variables Z_β, $\beta \geq \beta_1$ is tight and that the corresponding second moments are bounded. Let Z be a limiting point when β gets large and $h(x) \stackrel{\text{def.}}{=} \mathbb{P}(Z \geq x)$, $x \geq 0$. The previous relation gives the identity, for $x \geq 0$,

$$h(x)^d = \int_x^{+\infty} h(u)\ \mathrm{d}u.$$

It is easy to see that this relation determines h and therefore gives the desired convergence in distribution. The theorem is proved. □

Part II

Variations in polymerization models

Chapter 5

On the Asymptotic Distribution of Nucleation Times of Polymerization Processes

Contents

5.1 **Introduction** 156
5.2 **Stochastic Model** 161
5.3 **The First Instant of Nucleation** 167
5.4 **Fast Growth of Stable Polymers** 176
5.5 **Asymptotic Behavior of Lag Time** 178
5.A **Appendix: Biological background** 181
5.B **Appendix: Technical Details** 182
5.B.1 A Reminder on Marked Poisson Point Processes 182
5.B.2 Additional Calculations for the Proof of Proposition 5.2 183
5.B.3 Proof of Proposition 5.3 184

Abstract In this chapter, we investigate a stochastic model describing the time evolution of a polymerization process. A polymer is a macromolecule resulting from the aggregation of several elementary subunits called monomers. Polymers can grow by addition of monomers or can be split into several polymers. The initial state of the system consists of isolated monomers. We study the *lag time* of the polymerization process, that is, the first instant when the fraction of monomers used in polymers is above some threshold. The mathematical model includes *a nucleation property*: If n_c is defined as the size of the nucleus, polymers with a size smaller than n_c are quickly fragmented into smaller polymers. For polymers of size greater than n_c, the fragmentation still occurs but at a smaller rate. A scaling approach is used, by taking the volume N of the system as a scaling parameter. If $n_c \geq 3$ and under quite general assumptions on the way polymers are fragmented, if T^N is the instant of creation of the first "stable" polymer, i.e. a polymer of size n_c, then it is proved that (T^N/N^{n_c-3}) converges in distribution. We also show that, if $n_c \geq 4$, then the lag time has the same order of magnitude as T^N and, if $n_c = 3$, it is of the order of $\log N$. An original

feature proved for this model is the significant variability of T^N. This is a well known phenomenon observed in the experiments in biology but the previous mathematical models used up to now did not exhibit this magnitude of variability. The results are proved via a series of technical estimates for occupations measures on fast time scales. Stochastic calculus with Poisson processes, coupling arguments and branching processes are the main ingredients of the proofs.

5.1 Introduction

The protein polymerization processes is an important phenomenon occurring in many biological processes. Macromolecules proteins, also called monomers, may be assembled into several aggregated states called polymers. A polymer of size k is an assembly of k monomers linked by some chemical bonds. See Appendix 5.A for a more detailed description of the biological background.

In this chapter, we introduce a stochastic model describing the time evolution of the set of polymers. It is assumed that, initially, there are only single monomers. Polymers are formed by the successive additions of monomers to an existing polymer, these events are essentially caused by random fluctuations inside the cell due to thermal noise for example. The same fluctuations may also cause the breakage of polymers into several polymers of smaller sizes and also monomers.

The main quantity of interest is the time evolution of the polymerized mass of monomers, that is the the mass of the polymers of size greater or equal to 2. In practice, this quantity can be measured in the experiments.

Using the classical notations for chemical reactions, if $\mathcal{X}_k$ is the set of polymers of size k, the polymerization process analyzed in this chapter can be represented as

$$\begin{cases} \mathcal{X}_1+\mathcal{X}_k \xrightarrow{\kappa^k_{\text{on}}} \mathcal{X}_{k+1}, \\ \qquad \mathcal{X}_k \xrightarrow{\kappa^k_{\text{off}}} \mathcal{X}_{a_1}+\mathcal{X}_{a_2}+\cdots+\mathcal{X}_{a_p}, \quad p\geq 2, a_1+\cdots+a_p=k. \end{cases} \tag{5.1}$$

The quantity κ^k_{on} [resp. κ^k_{off}] is the chemical rate at which a polymer of size k is bound with a monomer [resp. at which it is broken].

The Large Variability of Polymerization Processes

Starting with only monomers, the system will stay for some time with a negligible polymerized mass, i.e. few monomers are polymerized at a given instant. The main phenomenon observed in the experiments is that there is an instant when a sharp phase transition occurs, the polymerized mass goes then from 0 to reach very quickly its final value. This explosive behavior is a constant feature of polymerization processes observed in experiments. Another key property from a biological point of view is that the instant when this event occurs, the *lag time*, varies significantly from an experiment to another. See Figure 5.1 in the appendix. This phenomenon is believed to be at the origin of the long delays observed in the development of several neuro-degenerative diseases such as the Bovine Spongiform Encephalopathy (Mad Cow) or Alzheimer's disease. See Appendix 5.A for a quick review and references for the biological aspects. The main goal of mathematical models in this domain is of representing this phenomenon via a

simple model exhibiting this phenomenon and so that the variability of the lag time can analyzed quantitatively.

A natural description involves an infinite dimensional state space, the kth coordinate giving the number of polymers of size k. The growth of polymer of size k is described by the interaction of the kth and first coordinates. The fragmentation of polymer gives a more intricate transition since a polymer can be fragmented into a subset of polymers of smaller sizes and also monomers. We begin with the description of a classical mathematical model of polymerization.

Literature: Becker-Döring Equations

The corresponding sequence of chemical reactions are given by

$$\mathcal{X}_1+\mathcal{X}_k \underset{b_{k+1}}{\overset{a_k}{\rightleftarrows}} \mathcal{X}_{k+1},\ k\geq 1. \tag{5.2}$$

They have been quite often used to represent polymerization processes. It amounts to assume that polymers can evolve only by the addition/removal of one monomer. In a deterministic setting, there is an associated system of ODEs called the *Becker-Döring Dynamical System* $(c_k(t))$ solution of the system of differential equations

$$\begin{cases} \dfrac{\mathrm{d}c_1}{\mathrm{d}t}(t) = -2J_1(c(t)) - \displaystyle\sum_{k\geq 2} J_k(c(t)), \\ \dfrac{\mathrm{d}c_k}{\mathrm{d}t}(t) = J_{k-1}(c(t)) - J_k(c(t)),\ k>1, \end{cases} \tag{BD}$$

with $J_k(c)=a_kc_1c_k-b_{k+1}c_{k+1}$ if $c=(c_k)\in\mathbb{R}_+^{\mathbb{N}}$ and with a convenient initial condition. This is a first order description of the process, the term $c_k(t)$ should be thought of the concentration of polymers of size $k\geq 1$. The conservation of mass holds, provided that the system is well-behaved, i.e. the quantity

$$c_1(t)+2c_2(t)+\cdots+kc_k(t)+\cdots$$

is constant with respect to t. See Becker and Döring [16] for the original paper. The conditions on existence/uniqueness of solutions have been extensively investigated. See Ball et al. [15] and Penrose [125, 126]. For the rate of convergence to equilibrium, see Jabin and Niethammer [78] and for the so-called gelation property when the conservation of mass is not satisfied, see Jeon [83]. In the literature of biological models it has been extensively used, see the various classes of ODEs in Table 2 of Morris et al. [115] and Hingant and Yvinec [75] for example, see also Wu and Shea [167] and Yvinec et al. [170].

With convenient parameters estimations, these ODEs can describe first order characteristics such as the mean concentration of polymers with a given size. See Davis and Sindi [41], Rubenstein et al. [75] for a survey of these results. Due to their deterministic formulation, they cannot really be used to investigate the fluctuations of the lag time of the polymerization process. Recall that this is one of the most important aspects of these phenomena.

There is a past and recent interest in *growth-fragmentation models* which are generalizations of the polymerization process described by Relation (5.1). For these

models, the growth can occur also by coagulation and not only by addition of a single particle/monomer at a time. Additionally, the rates of occurrence of the growth or fragmentation events can be state dependent. See Aldous [5] for a general survey. See also Bertoin [18] and Calvez et al. [29] and subsequent references for recent work on these topics. These processes are also used to study some population processes. These studies focus mainly on the existence of such processes, on their scaling properties (like self-similarity) or their branching process representation. It should be noted that the term "polymer" is also used for stochastic processes defined in terms of a Gibbs distribution on finite sample paths in $\mathbb{Z}^d$, these processes are related to several classes of random walks in random environment, see den Hollander [44] and Comets [38] for example. The problems investigated in this chapter are described differently and they focus on transient properties rather than equilibrium properties.

Central Limit Theorems vs Rare Events Asymptotics

There are several analyses of stochastic versions of the Becker-Döring equations. Jeon [83] shows, that starting from N particles, and with Poisson processes governing the dynamics of the transitions of monomer additions and deletions then, under appropriate conditions, a convergence result holds for the coordinates scaled by N and the limit is the solution of Becker-Döring equations. The corresponding functional central limit theorem has been proved in Sun [153]. A stochastic analysis of the lag time of the polymerization processes based on Becker-Döring equation can then be done in terms of a central limit theorem (CLT) in Szavits et al. [157]. A mathematical analysis of a stochastic model with only two species, monomers and polymerized monomers is achieved in this way in Szavits et al. [157], Eugène et al. [58], see also Doumic et al. [51]. It should be pointed out that these mathematical models fail to explain the large variability of the lag time which is observed in practice. For these models the ratio of the variance and the mean of the lag time is converging to 0 in the asymptotic regimes considered. In a small volume, theses estimations can nevertheless be reasonably accurate.

We believe that the variability cannot be explained only by a central limit theorem (CLT), see Szavits et al. [157] and Eugène et al. [58]. The main result of this chapter states that the fluctuations are due to the occurrence of a set of rare events. This has been suggested in earlier works, see for example Yvinec et al [171], but to the best of our knowledge it has never been proved rigorously for a convenient mathematical model. We also think that the Becker-Döring model itself cannot be a convenient model to describe polymerization processes. The fragmentation mechanism of the model, the removal of monomers of a polymer one by one, is too limited to give the appropriate explosive behavior mentioned above. On this matter, Condition A-3 of Assumption A* below is key to get this property.

The Nucleation Phenomenon

One has first to introduce the notion of nucleation which is a popular assumption in the biological literature but, to the best of our knowledge, does not seem to have been properly included in the previous mathematical models of polymerization. See Morris et al. [115] and Gillam and MacPhee [67], and the references therein, Kashchiev [91, 90] for example. The assumption is that polymers with small sizes

are quite unstable. They are very quickly broken by the random fluctuations of the environment. There is nevertheless a critical size n_c above which a polymer is more stable. It can be still broken due to random events but at a much smaller rate. See also the discussion on Appendix 5.A. The quantity n_c is called the *nucleus size*, in this way polymerization can also be seen as a nucleation process as in the literature in physics, see Kashchiev [89]. These assumptions are generally based on considerations of statistical mechanics expressed in terms of the free energy of assemblies of proteins. See Andrews and Roberts [11], Firestone et al. [62] and Zhang and Muthukumar [172] for example. We will assume in this chapter that the size of the nucleus is at least 3, $n_c \geq 3$. See the discussion in Section 5.3 concerning the case $n_c=2$.

The assumption of the existence of a nucleus size can be translated as follows in the chemical reactions: for $k<n_c$ then $\kappa^k_{\text{off}} \gg \kappa^k_{\text{on}}$, otherwise, if $k \geq n_c$, κ^k_{off} is of the same order of magnitude as κ^k_{on} or much smaller. In our model, with the scaling parameter N, the number of initial monomers, we will assume that, for $k<n_c$, then $\kappa^k_{\text{off}}/\kappa^k_{\text{on}}=N\alpha$.

In a first step in the analysis of these models, we have chosen a linear dependence on the scaling parameter N for the ratio of the chemical rates. A more general dependence of the type $\kappa^k_{\text{off}}/\kappa^k_{\text{on}}=\Phi(N)\mu_k$, $2 \leq k < n_c$, for a non-decreasing function Φ will be investigated in a future work.

The Main Results

Assuming that there are only N initial monomers, the mathematical results we obtain is that if T^N is the first instant when a polymer of size n_c is created, under appropriate conditions, the following convergence in distribution holds

$$\lim_{N \to +\infty} \frac{T^N}{N^{n_c-3}} = E_\rho, \tag{5.3}$$

where E_ρ is an exponential random variable with parameter ρ defined by Relation(5.27), a constant depending on the polymerization and fragmentation rates before nucleation. In particular the variability of the lag time is the variability of an exponential distribution with a large average and, therefore, gives a large variance.

Furthermore, it is also shown that, for a sufficiently small δ, the *lag time* L^N_δ, i.e. the first instant when a fraction δ of the initial monomers is polymerized, then, with high probability,

$$L^N_\delta = T^N + O(\log N). \tag{5.4}$$

This relation shows that, for $n_c \geq 4$, the variable T^N has the same order of magnitude in N as the lag time L^N_δ. See Proposition 5.8 and Theorem 5.2 for a precise formulation of this result.

The polymerization process can thus be decomposed into two steps.

1. The evolution of the process until a stable polymer (a polymer whose size is n_c) is created. With our scaling assumptions, this is a rare event as mentioned above. The convergence (5.3) is natural in probability theory. For stochastic processes converging quickly to equilibrium, rare events are, generally, reached after an exponentially distributed amount of time with a large average. See Keilson [92] and Aldous [3] for example. Nevertheless, getting a precise limiting result for this convergence turns out to be quite challenging. The main reason is that our

assumptions on the way the polymers are fragmented are somewhat minimal, the transitions of the associated Markov processes are thus quite complicated. See Section 5.3 and Appendix 5.B.

2. With an initial state with only one stable polymer, it is shown that the number of stable polymers is lower bounded by a super critical branching process, in particular with positive probability it is growing exponentially fast or it dies out. This is the explosion phase of the polymerization process.

The picture is then as follows: Stable polymers are created individually from the aggregation of monomers according to a Poisson process with a small parameter proportional to $1/N^{n_c-3}$. One of them will be eventually successful to produce a significant fraction of stable polymer in a very small amount of time (with respect to N^{n_c-3}). The lag time is then essentially the amount of time it takes to have a successful stable polymer that generates sufficiently many stable polymers. When $n_c \geq 4$, the duration of the growth period is negligible with respect to N^{n_c-3}, because of the exponential growth during this phase.

Our model has thus the two main characteristics observed in the experiments in biology: a take-off phase with a large variability (the asymptotic exponential random variable) and a steep growth (the super-critical branching process). See Figure 5.1 in the appendix. The results are proved via a series of technical estimates for occupations measures on fast time scales associated to the first n_c coordinates of the corresponding Markov process. Stochastic calculus with Poisson processes, coupling arguments and branching processes are the main ingredients of the proofs. A related model (in a quite different context) with a Becker-Döring flavor is analyzed in Sun et al. [154] with a different perspective since the goal is of analyzing the asymptotic behavior of a transient multi-dimensional Markov process. Outside the fragmentation feature which does not appear in the models of [154], the transition rates are linear with respect to the state in this reference, instead of a quadratic dependence for the present chapter. In both cases, several estimates on fast time scales have nevertheless to be derived.

Outline of the Chapter

Section 5.2 introduces the stochastic model used to investigate the polymerization process, together with the notations and definitions used throughout this chapter. In Section 5.3 limiting results are proved for the first instant when there is a stable polymer, see Theorem 5.1 and Proposition 5.5. For that purpose, several estimates on the time scale $t \mapsto N^{n_c-2}t$ are derived via stochastic calculus and coupling methods. The most technical parts of the proofs of these estimates are postponed in Appendix 5.B. Section 5.4 considers the dynamic of the polymerization for the polymers whose sizes are above the level of nucleation n_c. Under appropriate conditions on the fragmentation process, a coupling with a super-critical branching process is obtained, see Proposition 5.6. In Section 5.5 the main results for the asymptotic behavior of the distribution of the lag time are proved, see Theorem 5.2. Appendix 5.A gives a quick overview of the biological aspects of these processes.

Acknowledgments

The authors are grateful to Marie Doumic and Wei-Feng Xue for numerous conversations on these topics. They have been very helpful for our understanding of these quite delicate phenomena.

5.2 Stochastic Model

In this section we introduce the stochastic model describing the polymerization processes with a nucleation phenomenon. The later feature is introduced via a multiplicative scaling parameter for the rate of fragmentation when a polymer has a size strictly less than $n_c \geq 3$ the size of the nucleus. In this way, polymers of small size are unstable and quickly break into monomers.

We define, for $p \geq 1$

$$\mathcal{S}_p \stackrel{\text{def.}}{=} \left\{ (u_i) \in \mathbb{N}_+^{\mathbb{N}} : u_1 + 2u_2 + \cdots + \ell u_\ell + \cdots = p \right\}.$$

The state space of the process is thus given by $\mathcal{S}_N$, if $u = (u_i) \in \mathcal{S}_N$, u_i is the number of polymers of size i. Since for $u \in \mathcal{S}_p$, all components with index strictly greater than $p+1$ are null, we will occasionally use the slight abuse of notation $u = (u_1, \ldots, u_p)$. We denote by

$$\mathcal{S}_\infty = \bigcup_{k \geq 0} \mathcal{S}_k, \tag{5.5}$$

the set of states with finite mass. For $p \in \mathbb{N}$, e_p will denote the pth unit vector of $\mathcal{S}_\infty$.

One starts from N monomers, i.e. N polymers of size 1, as the initial state

$$u(0) = (N, 0, \ldots, 0, \ldots).$$

The parameter N is the original mass of monomers of the system, it will be also interpreted as a volume parameter. When the system is in state $u = (u_k) \in \mathcal{S}_N$ then, for $k \geq 1$, the quantity u_k / N is defined as the concentration of polymers of size k. In the following, N is used, classically, as a scaling parameter.

In the special case of Becker-Döring chemical reactions (5.1), the law of mass action gives the Michaelis Menten's kinetics equations for the concentration of the various polymers. If $[\mathcal{X}_k]$ is the concentration of polymers of size k, one has for $k \geq 2$,

$$\frac{\mathrm{d}}{\mathrm{d}t}[\mathcal{X}_k] = \kappa_{\text{on}}^{k-1}[\mathcal{X}_1][\mathcal{X}_{k-1}] - \kappa_{\text{on}}^{k}[\mathcal{X}_1][\mathcal{X}_k] - \kappa_{\text{off}}^{k}[\mathcal{X}_k] + \kappa_{\text{off}}^{k+1}[\mathcal{X}_{k+1}] \tag{5.6}$$

See Chapter 6 of Murray [117] for example. The fragmentation processes analyzed in this chapter are more general. In a stochastic context, because of the different time scales of the phenomena of aggregation and fragmentation, we will study the evolution of the numbers of polymers rather than their concentration.

Growth of Polymers

We make the, classical, assumption that the growth of polymers occurs only through the successive additions of monomers. The rate at which a given monomer, a polymer

of size k, $k<n_c$, collides with a polymer of size 1 is classically taken as proportional to the concentration u_1/N of these polymers. See Anderson and Kurtz [8]. Hence the total rate of production of polymers of size $k+1$ via this kind of reaction is given by

$$\lambda_k u_k \frac{u_1}{N},$$

for some constant $\lambda_k>0$. which is consistent with the two first terms of the right-hand side of Relation (5.6).

Fragmentation

This is where the nucleation phenomenon is introduced in the stochastic model. Fragmentation occurs because, as long as the size of the polymer is below the nucleation size n_c, it is quite unstable, the thermal noise breaks it after some time. For $k\geq 2$, define

$$\mu_k^N = \begin{cases} N\mu_k & \text{if } k<n_c, \\ \mu & \text{if } k\geq n_c, \end{cases} \tag{5.7}$$

where $(\mu_k, 2\leq k\leq n_c-1)$ and μ are positive real numbers. Polymers with size greater than n_c will be qualified as stable to stress that its fragmentation rate is not large.

In the literature for chemical reactions of the type (5.6), the ratio $\kappa_{\text{off}}^k/\kappa_{\text{off}}^k$ is assumed to be large for $k<n_c$ and small otherwise. In our case we used the scaling parameter N to stress the difference before and after nucleation.

In this setting, a quite general fragmentation process is considered. A polymer of size k splits at rate μ_k^N according to a fragmentation distribution ν_k on the state space $\mathcal{S}_k$. A polymer of size k is broken into Y_i^k polymers of size i, $1\leq i\leq k$ where $Y^k=(Y_1^k,\ldots,Y_k^k)$ are random variables such that

$$\mathbb{E}\left(f(Y_1^k,\ldots,Y_n^k)\right) = \int_{\mathcal{S}_k} f(y_1,\ldots,y_k)\,\nu_k(\mathrm{d}y).$$

The case where the fragmentation of k gives one item of size k is a priori not excluded, i.e. $Y^k=e_k=(0,\ldots,0,1)$ with positive probability. We will assume in the following that the fragmentation measure is not singular from this point of view, i.e. that $\nu_k(e_k)<1$ holds for all $k\geq 1$. Under this assumption, ultimately, a polymer of size $k\geq 2$ will be decomposed into polymers of size less or equal to $k-1$. In the case $\nu_k(e_k)>0$, we allow then transitions with no effect on the state. This is not a restriction since one could use an equivalent model where the fragmentation rate μ_k is replaced by $\mu_k(1-\nu_k(e_k))$ and the measure $\nu_k(\mathrm{d}y)$ by

$$\frac{1}{1-\nu_k(e_k)}\mathbb{I}_{\{y\neq e_k\}}\nu_k(\mathrm{d}y).$$

For $p<k$, we denote by $I_p(y)$ the pth coordinate of $y\in\mathcal{S}_k$ and

$$\langle \nu_k, I_p\rangle \overset{\text{def.}}{=} \int_{\mathcal{S}_k} I_p(y)\,\nu_p(\mathrm{d}y) = \int_{\mathcal{S}_k} y_p\,\nu_p(\mathrm{d}y)$$

is the average number of polymers of size p for ν_k, in particular

$$\sum_{p=1}^{k} p\,\langle \nu_k, I_p\rangle = k.$$

When fragmented, a polymer of size k is decomposed into a set of Y_i^k polymers of size i for $1 \leq i < k$, where (Y_i^k) is a random vector of integers with distribution ν_k.

Remark on the Nucleus Size

If the nucleus size is 2, in state $u \in \mathcal{S}_N$ stable polymers are created at rate $\lambda_1 u_1^2/N$. Hence, initially, in a time interval of duration $\varepsilon > 0$, $\lambda_1 \varepsilon N$ stable polymers are created directly with monomers. Polymerization starts right away initially. When $n_c \geq 3$, it will be seen that that the stable polymers are essentially produced by an autocatalytic process: under appropriate conditions on the parameters, a stable polymer grows for some time and then is fragmented into multiple polymers and, with positive probability, the size of several of them will be larger than n_c and therefore a stable polymer can create stable polymers with positive probability. It turns out that this production scheme is much faster than the creation of stable polymers by the successive addition of monomers. In this way the case $n_c = 2$ stands out. It is in fact used in most of the mathematical models of polymerization, implicitly sometimes. In our view, for this reason it cannot be used to explain the large variability observed for these processes.

Examples of Fragmentation Measures.
Fragmentation is associated to the decomposition of integers. An important literature is devoted to this topic: from the point of view of combinatorics as well as for statistical aspects. See, for example, Fristed [64], Pitman [131] or Ercolani et al. [56]. We now give some classical examples. Binary fragmentations are single out due to their importance.

For $k \geq 2$, a polymer of size k is split according to the distribution $\nu_k(\cdot)$.

1. Binary Fragmentations.
   — UF: Uniform Binary Fragmentation,
   for $0 \leq \ell \leq k$,
   $$\nu_k^{\mathrm{UF}}(e_\ell + e_{k-\ell}) = \begin{cases} \frac{2}{k+1}, & & k \text{ odd} \\ \frac{2}{k+1} & \ell \neq k/2 & k \text{ even} \\ \frac{1}{k+1} & \ell = k/2 & k \text{ even} \end{cases}$$
   where e_ℓ is the ℓth unit vector of $\mathbb{N}^{\mathbb{N}}$, e_ℓ is representing a single polymer of size ℓ. We take the convention that $e_0 = 0$ when $\ell \in \{0, k\}$, in this case the corresponding transition is empty.
   — BF: Binomial Fragmentation, for $p \in (0,1)$,
   $$\nu_k^{\mathrm{BF}}(e_\ell + e_{k-\ell}) = 2\binom{k}{\ell} p^\ell (1-p)^{k-\ell}, \quad 0 \leq \ell \leq k,$$
   if k is odd. If k is even, the case $\ell = k/2$ has to be singled out as before.

2. Multiple Fragmentations.
   — MF: Multinomial Fragmentation.
   For $m \geq 2$ and $p = (p_i) \in (0,1)^m$ with $p_1 + p_2 + \cdots + p_m = 1$, a polymer of size $k \geq m$ is fragmented into at most m polymers according to a multinomial distribution,

if $\ell = e_{\ell_1} + e_{\ell_2} + \cdots + e_{\ell_m} \in \mathcal{S}_k$

$$\nu_{k,m}^{\mathrm{MF}}(\ell) = \sum_{\substack{0 \le n_1, \ldots, n_m \le k \\ \{n_1, \ldots, n_m\} = \{\ell_1, \ldots, \ell_m\}}} \frac{k!}{n_1! n_2! \cdots n_m!} \prod_{i=1}^{m} p_i^{n_i}.$$

Recall the convention $e_0 = 0$.

— SF: Symmetrical Fragmentation. This is similar to the previous case except that the degree of splitting is a random variable. If $(q_i, i \ge 2)$ is a probability distribution on integers greater than 2, for $k \ge 2$,

$$\nu_k^{\mathrm{SF}}(\ell) = q_m \nu_{k,m}^{\mathrm{MF}}(e_{\ell_1} + e_{\ell_2} + \cdots + e_{\ell_m}),$$

with if $\ell = e_{\ell_1} + e_{\ell_2} + \cdots + e_{\ell_m} \in \mathcal{S}_k$

The family of probability distributions to describe fragmentation have various definitions depending on the context investigated. For continuous fragmentation, i.e. when the state space is $\mathbb{R}_+$ instead of $\mathbb{N}$, the corresponding quantity is the *dislocation measure*. In the self-similar case, the fragmentation process is described by a measure associated with the outcome of the breaking of a particle of size 1. There are generalizations in this continuous setting with fragmentation kernels $K(x, \mathrm{d}y)$ depending on the initial size x to decompose. See Bertoin [17] for example. For fragmentation of integers, as it is our case, Mohamed and Robert [113] describes the fragmentation also via a single measure, *the splitting measure* for the analysis of first order quantities. See [113] for an overview of this literature in this framework. In our case, we have to use instead a family depending of the size of initial polymer.

The following conditions will be assumed to hold throughout this chapter.

Assumptions 5.1.

A-1) Non-Degeneracy of Fragmentation.

$$\sup_{k \ge 2} \nu_k(e_k) < 1.$$

For simplicity it will be assumed in the rest of the chapter that $\nu_k(e_k) = 0$, for all $k \ge 2$. The above discussion shows that, up to a change of the fragmentation rate, there is no loss of generality. It is used in several proofs in Section 5.3.

A-2) Lower Bound for Polymerization Rates.

$$\underline{\lambda} \stackrel{\text{def.}}{=} \inf_{k \ge 1} \lambda_k > 0. \tag{5.8}$$

A-3) Fragmentation of Large Polymers. The quantity

$$\liminf_{k \to +\infty} \nu_k \left(y = (y_i) \in \mathcal{S}_k : \sum_{i \ge n_c} y_i \ge a \right)$$

is 1 for $a = 1$ and is positive for $a = 2$.

A-4) Monotonicity Property: If $k \leq k'$, then, for all $a \in \mathbb{N}$ and $\ell \geq 1$, the relation

$$\nu_k\left(y \in \mathcal{S}_k \colon \sum_{i=1}^{+\infty} \mathbb{I}_{\{y_i \geq \ell\}} \geq a\right) \leq \nu_{k'}\left(y \in \mathcal{S}_{k'} \colon \sum_{i=1}^{+\infty} \mathbb{I}_{\{y_i \geq \ell\}} \geq a\right)$$

holds.

Assumption A-2 states that growth of polymers occurs at least at a minimal positive rate. Assumption A-3 states that, with probability close to 1, the fragmentation of a large polymer gives at least a stable polymer, i.e. whose size is greater than n_c, and, with positive probability, at least two stable polymers are produced. Under this condition a large polymer cannot in particular be broken into monomers with probability 1 in particular. The monotonicity property A-4 gives that the sizes of fragments are stochastically increasing with respect to the size of the initial polymer.

Proposition 5.1. *The fragmentation measures* UB, UF, MB *and* MF *satisfy Assumption A^*.*

Proof. The proof is done for the multinomial fragmentation, the proof for the symmetrical fragmentation is similar. The fragmentation as the distribution of k balls into m urns so that the probability the ith urn is chosen is p_i, if A_i^k the number of balls in this urn, then ν_k is simply the empirical distribution of these variables. A simple coupling can be constructed so that $A_i^k \leq A_i^{k+1}$ holds for all $i \in \{1, \ldots, m\}$. This gives right away the monotonicity property. The law of large numbers gives the convergence in distribution

$$\lim_{k \to +\infty} \frac{1}{k}(A_1^k, \ldots, A_m^k) = (p_1, \ldots, p_m),$$

Condition A-3 holds since p_i is positive for all i. □

The Infinitesimal Generator of the Stochastic Evolution

We define by $U_k^N(t)$ the number of polymers of size $k \geq 1$ at time $t \geq 0$. Clearly, the process $(U^N(t)) \stackrel{\text{def.}}{=} (U_k^N(t))$ has the Markov property on the countable state space $\mathcal{S}_N$. The generator Ω_N of this process is given by, for $u \in \mathcal{S}_N$,

$$\Omega_N(f)(u) = \sum_{k=1}^{+\infty} \lambda_k u_k \frac{u_1}{N}\left[f(u+e_{k+1}-e_k-e_1)-f(u)\right] + \sum_{k=2}^{+\infty} \mu_k^N u_k \int_{\mathcal{S}_k} \left[f(u+y-e_k)-f(u)\right] \nu_k(\mathrm{d}y) \quad (5.9)$$

where, for $i \geq 1$, e_i is the ith unit vector of $\mathbb{N}^{\mathbb{N}}$ and f is a function on $\mathcal{S}_N$ with finite support.

The growth mechanism involves the scaled term u_1/N, the concentration of monomers. This is not the case for the rates of fragmentation which do not depend on the concentration, in the state $u = (u_i) \in \mathcal{S}_N$, polymers of size k are fragmented at rate $\mu_k^N u_k$.

Stochastic Differential Equations

Throughout the chapter, we use the following notations for Poisson processes. For $\xi>0$, $\mathcal{N}_\xi$ is a Poisson point process on $\mathbb{R}_+$ with parameter ξ and $(\mathcal{N}_{\xi,i})$ is an i.i.d. sequence of such processes.

For $k\geq 2$, $\overline{\mathcal{N}}^k_\xi$ is a marked Poisson point process on $\mathbb{R}_+\times\mathcal{S}_k$ with intensity measure $\xi\,\mathrm{d}t\otimes\nu_k(\mathrm{d}y)$ and $(\overline{\mathcal{N}}^k_{\xi,i})$ is an i.i.d. sequence of such processes. Such a process can be represented as follows. If $\mathcal{N}_\xi=(t^k_n)$ is a Poisson process on $\mathbb{R}_+$ with rate ξ and (Y^k_n) is an i.i.d. sequence of random variables on $\mathcal{S}_k$ with distribution ν_k, then $\overline{\mathcal{N}}^k_\xi$ has the same distribution as the point process on $\mathbb{R}_+\times\mathcal{S}_k$ given by

$$\sum_{n\geq 1}\delta_{(t^k_n,Y^k_n)},$$

where $\delta_{(a,b)}$ is the Dirac mass at (a,b). All Poisson processes used are assumed to be independent.

In this context, the Markov process $(U^N(t))$ can also be seen as the stochastic process solution of the system of stochastic differential equations

$$\begin{aligned}\mathrm{d}U^N_k(t) = &\sum_{\ell=1}^{U^N_1U^N_{k-1}(t-)}\mathcal{N}_{\lambda_{k-1}/N,\ell}(\mathrm{d}t)+\sum_{m=k+1}^{+\infty}\sum_{\ell=1}^{U^N_m(t-)}\int_{\mathcal{S}_m}y_k\overline{\mathcal{N}}^m_{\mu^N_m,\ell}(\mathrm{d}t,\mathrm{d}y)\\ &-\sum_{\ell=1}^{U^N_k(t-)}\overline{\mathcal{N}}^k_{\mu^N_k,\ell}(\mathrm{d}t,\mathcal{S}_k)-\sum_{\ell=1}^{U^N_1U^N_k(t-)}\mathcal{N}_{\lambda_k/N,\ell}(\mathrm{d}t),\quad k\geq 2.\end{aligned}\tag{5.10}$$

The conservation of mass condition defines the evolution of the first coordinate $(U^N_1(t))$, the process of the number of monomers,

$$U^N_1(t)+2U^N_2(t)+3U^N_3(t)+\cdots+\ell U^N_\ell(t)+\cdots=N.\tag{5.11}$$

The initial condition is given by

$$(U^N(0))=(N,0,\ldots,0,\ldots).\tag{5.12}$$

The process $(U^N(t))$ is càdlàg, i.e. right continuous with left limits at every point of $\mathbb{R}_+$, and $\mathrm{d}U^N_k(t)$ is $U^N_k(t)-U^N_k(t-)$, where $f(t-)$ denotes the left limit of a function f at t.

Nucleation Times

We can now introduce the lag time of the polymerisation process, for $\delta\in(0,1)$,

$$L^N_\delta=\inf\left\{t\geq 0: U_{n_c}(t)+U_{n_c+1}(t)+\cdots+U_k(t)+\cdots\geq\delta N\right\}.\tag{5.13}$$

This is the main quantity of interest in this chapter, the first instant when there is a fraction δ of monomers in stable polymers, i.e. of size greater than n_c. The first nucleation time is defined as

$$T^N=\inf\{t\geq 0: U_{n_c}(t)=1\},\tag{5.14}$$

it is the first time a stable polymer appears.

5.3 The First Instant of Nucleation

In this section, one introduces an auxiliary Markov process $(X^N(t)) \stackrel{\text{def.}}{=} (X_k^N(t))$, the solution of the system of stochastic differential equations

$$\mathrm{d}X_k^N(t) = \sum_{\ell=1}^{X_1^N X_{k-1}^N(t-)} \mathcal{N}_{\frac{\lambda_{k-1}}{N^2},\ell}(\mathrm{d}t) + \sum_{m=k+1}^{+\infty} \sum_{\ell=1}^{X_m^N(t-)} \int_{\mathcal{S}_m} y_k \overline{\mathcal{N}}^m_{\mu_m,\ell}(\mathrm{d}t,\mathrm{d}y)$$
$$- \sum_{\ell=1}^{X_k^N(t-)} \overline{\mathcal{N}}^k_{\mu_k,\ell}(\mathrm{d}t,\mathcal{S}_k) - \sum_{\ell=1}^{X_1^N X_k^N(t-)} \mathcal{N}_{\frac{\lambda_k}{N^2},\ell}(\mathrm{d}t), \quad k\geq 2, \tag{5.15}$$

with initial condition $(X^N(0))=(N,0,\ldots,0,\ldots)$. By an abuse of notation, here we assume that for the process $(X^N(t))$, $\lambda_k=\mu_k=0$, for all $k\geq n_c$. Therefore, $(X_k^N(t))\equiv(0)$ for all $k>n_c$. As before, the mass conservation condition on any finite time interval defines the evolution of the first coordinate $(X_1^N(t))$. Up to a change of time scale, the process $(X_k^N(t))$ is closely related to the polymerization process, see the lemma below. The main difference is that polymers with size n_c cannot be fragmented and therefore do not produce polymers with smaller size. This process is used to investigate the first phase of the polymerization process, until the first nucleus is created. Another auxiliary process will be introduced in section 5.4 to investigate the second phase during which polymers of size greater than n_c have a total mass of the order of N.

Lemma 5.1. *If $\tau^N=\inf\{t : X_{n_c}^N(t)=1\}$ and $((U^N(t))$ is the process defined by Relations (5.10), (5.11) and (5.12), then the following identity in distribution holds*

$$\left(U^N(t), t \leq T^N\right) \stackrel{\text{dist.}}{=} \left(X^N(Nt), t \leq \tau^N/N\right) \tag{5.16}$$

with T^N defined by Relation (5.14). In particular $T^N \stackrel{\text{dist.}}{=} \tau^N/N$.

Proof. The arguments are quite simple. Up to time T^N, the size of any polymer is at most n_c-1, hence its fragmentation rate is proportional to N. Since the time scale $t\mapsto Nt$ adds a multiplicative factor to the transition rates of $(X_k^N(t))$, we easily check from the SDEs (5.10) and (5.15) that the two processes on both sides of Relation (5.16) satisfy the same SDE. Since they have the same initial point, we get the desired result. □

The strategy to derive the limiting behavior of the distribution of the variable τ^N is explained briefly. One will first prove, via a coupling argument, that, on the fast time scale $t\mapsto N^{n_c-1}t$, the values of the coordinates with index between 2 and n_c-1 are essentially negligible with respect to N. The key results are the asymptotic balance equations (5.24) of Corollary (5.1), they will finally give the order of magnitude in N of τ^N as well as the corresponding limit convergence in distribution to an exponentially distributed random variable, see Proposition 5.5 at the end of the section. The derivation of these balance equations is achieved through a sequence of estimations of occupation measures associated to the coordinates of the process. Several of the technical details are postponed in Appendix 5.B.

Lemma 5.2. *If* $(X_k^N(t))$ *is the solution of the SDE* (5.15) *with initial state* (5.12) *then, for* $2 \leq k \leq n_c - 1$ *and* $\gamma > 0$*, the convergence in distribution of processes*

$$\lim_{N \to +\infty} \left(\frac{X_k^N(N^{n_c-1}t)}{N^\gamma} \right) = (0)$$

holds.

Proof. Define $\lambda \stackrel{\text{def.}}{=} \max(\lambda_k, 1 \leq k < n_c)$ and

$$\left(R^N(t)\right) \stackrel{\text{def.}}{=} \left(\sum_{i=2}^{n_c-1} (i-1) X_i^N(t) \right),$$

Definition (5.9) of the infinitesimal generator gives that the process $(X^N(t))$ has positive jumps of size 1 only. In state $x \in \mathcal{S}_N$ they occur at rate

$$\frac{x_1}{N} \sum_{i=1}^{n_c-2} \lambda_i \frac{x_i}{N} \leq \lambda.$$

Remember that a polymer of size n_c is in a cemetery state from the point of view of the dynamic of the process $(X_k^N(t))$. The size of negative jumps of $(R^N(t))$ is less than -1 and they occur at rate at least

$$\sum_{i=2}^{n_c-1} \frac{\mu_i}{i-1} (i-1) x_i \geq \mu R^N(t),$$

with $\mu \stackrel{\text{def.}}{=} \min(\mu_i/(i-1), 2 \leq i \leq n_c - 1)$.

With a simple coupling, one can therefore construct the process $(L(t))$ of jobs of an $M/M/\infty$ queue, see Chapter 6 of Robert [138], with arrival rate λ and service rate μ such that $L(0) = 0$ and the relation $R^N(t) \leq L(t)$ holds for all $t \geq 0$.

Proposition 6.10 of [138] gives that, for $\varepsilon > 0$, the hitting time of $K = \varepsilon N^\gamma$ by $(L(t))$ is of the order of $(K-1)!(\mu/\lambda)^K$ as N gets large. In particular this does not happen on the time scale $t \mapsto N^{n_c-1}t$. The lemma is proved. □

The integration of Equations (5.3) gives the following representation, for $k \geq 2$,

$$X_k^N(t) = \lambda_{k-1} \int_0^t \frac{X_1^N(s)}{N} \frac{X_{k-1}^N(s)}{N} \, \mathrm{d}s + \sum_{\ell=k+1}^{+\infty} \mu_\ell \langle \nu_\ell, I_k \rangle \int_0^t X_\ell^N(s) \, \mathrm{d}s$$
$$- \mu_k \int_0^t X_k^N(s) \, \mathrm{d}s - \lambda_k \int_0^t \frac{X_1^N(s)}{N} \frac{X_k^N(s)}{N} \, \mathrm{d}s + M_k^N(t), \quad (5.17)$$

where $(M_k^N(t))$ is a martingale, obtained by the compensation of the Poisson processes of the dynamics. Stochastic calculus, see Section 5.B.1 of the Appendix for example, gives that its previsible increasing process is

$$\left\langle M_k^N \right\rangle (t) = \lambda_{k-1} \int_0^t \frac{X_1^N(s)}{N} \frac{X_{k-1}^N(s)}{N} \, \mathrm{d}s + \sum_{\ell=k+1}^{+\infty} \mu_\ell \langle \nu_\ell, I_k \rangle \int_0^t X_\ell^N(s) \, \mathrm{d}s$$
$$+ \mu_k \int_0^t X_k^N(s) \, \mathrm{d}s + \lambda_k \int_0^t \frac{X_1^N(s)}{N} \frac{X_k^N(s)}{N} \, \mathrm{d}s. \quad (5.18)$$

Proposition 5.2. *For $n_c \geq r \geq 2$, $k \leq r-1$ and $2 \leq h \leq n_c - 1$ then, for the convergence in distribution of continuous processes,*

$$\lim_{N\to\infty} \left(\frac{1}{N^r} \int_0^{N^{n_c-2}t} X^N_{n_c-k}(u) X^N_h(u)\, \mathrm{d}u \right) = (0).$$

Specially, for all $1 \leq k \leq n_c - 2$,

$$\lim_{N\to\infty} \left(\frac{1}{N^{k+1}} \int_0^{N^{n_c-2}t} X^N_{n_c-k}(u)^2\, \mathrm{d}u \right) = (0).$$

Proof. We prove the lemma by induction on r, from n_c to 2.
When $r = n_c$ and $k \leq n_c - 1$, one has

$$\left(\frac{1}{N^{n_c}} \int_0^{N^{n_c-2}t} X^N_{n_c-k}(u) X^N_h(u)\, \mathrm{d}u \right) = \left(\int_0^t \frac{X^N_{n_c-k}(N^{n_c-2}u)}{N} \frac{X^N_h(N^{n_c-2}u)}{N}\, \mathrm{d}u \right),$$

Lemma 5.2 shows that this process converges in distribution to 0 when N goes to infinity for all $k = 1, \ldots, n_c - 1$ and $2 \leq h \leq n_c - 1$.

Now suppose, by induction, that for all $\ell > r$ and $1 \leq k < \ell$, $2 \leq h \leq n_c - 1$, one has the convergence in distribution

$$\lim_{N\to\infty} \left(\frac{1}{N^{\ell}} \int_0^{N^{n_c-2}t} X^N_{n_c-k}(u) X^N_h(u)\, \mathrm{d}u \right) = 0. \tag{5.19}$$

On will prove that this property holds for $\ell = r$. Take $k = 1$ and $h = n_c - 1$. Relations (5.15), via stochastic calculus with Poisson processes, see Section 5.B.1 of the appendix for example, give the identity

$$\begin{aligned} X^N_{n_c-1}(t)^2 = M^N_{n_c-1,n_c-1}(t) &+ \lambda_{n_c-2} \int_0^t \left(2X^N_{n_c-1}(u) + 1 \right) \frac{X^N_1(u) X^N_{n_c-2}(u)}{N^2}\, \mathrm{d}u \\ &+ \int_0^t \left(-2X^N_{n_c-1}(u) + 1 \right) \left(\mu_{n_c-1} X^N_{n_c-1}(u) + \lambda_{n_c-1} \frac{X^N_1(u) X^N_{n_c-1}(u)}{N^2} \right) \mathrm{d}u \end{aligned} \tag{5.20}$$

where $(M^N_{n_c-1,n_c-1}(t))$ is a martingale whose previsible increasing process is given by

$$\begin{aligned} \left\langle M^N_{n_c-1,n_c-1} \right\rangle (t) = \lambda_{n_c-2} &\int_0^t \left(2X^N_{n_c-1}(u) + 1 \right)^2 \frac{X^N_1(u) X^N_{n_c-2}(u)}{N^2}\, \mathrm{d}u \\ &+ \int_0^t \left(-2X^N_{n_c-1}(u) + 1 \right)^2 \left(\mu_{n_c-1} X^N_{n_c-1}(u) + \lambda_{n_c-1} \frac{X^N_1(u) X^N_{n_c-1}(u)}{N^2} \right) \mathrm{d}u. \end{aligned}$$

Since $X^N_i(t) \leq N$ for all $i \geq 1$, for $2 \leq h \leq n_c - 1$,

$$\left(\frac{1}{N^r} \right)^2 \int_0^{N^{n_c-2}t} X^N_{n_c-1}(u)^2 \frac{X^N_1(u) X^N_h(u)}{N^2}\, \mathrm{d}u \leq \frac{1}{N^{r-1}} \frac{1}{N^{r+1}} \int_0^{N^{n_c-2}t} X^N_{n_c-1}(u) X^N_h(u)\, \mathrm{d}u,$$

as a process, the last term converges to 0 in distribution by the induction assumption. Similarly, since $r \geq 2$, the term

$$\left(\frac{1}{N^r}\right)^2 \int_0^{N^{n_c-2}t} X^N_{n_c-1}(u)^3 \, \mathrm{d}u \leq \frac{1}{N^{r-2}} \frac{1}{N^{r+1}} \int_0^{N^{n_c-2}t} X^N_{n_c-1}(u)^2 \, \mathrm{d}u$$

converges to 0 in distribution as a process by induction. By considering in the same way each of the terms defining $\left\langle M^N_{n_c-1,n_c-1} \right\rangle (N^{n_c-2}t)$, one gets the previsible increasing process of the martingale

$$\left(\frac{1}{N^r} M^N_{n_c-1,n_c-1}(N^{n_c-2}t)\right)$$

is converging in distribution to 0 as N goes to infinity. By using Lemma I.3.30 of Jacod and Shiryaev [81], this martingale is thus vanishing in distribution for N large. Similarly, the relation

$$\frac{1}{N^r} \int_0^{N^{n_c-2}t} X^N_{n_c-1}(u) \frac{X^N_1(u) X^N_h(u)}{N^2} \, \mathrm{d}u \leq \frac{1}{N^{r+1}} \int_0^{N^{n_c-2}t} X^N_{n_c-1}(u) X^N_h(u) \, \mathrm{d}u$$

and the induction assumption give the convergence in distribution

$$\lim_{N \to +\infty} \left(\frac{1}{N^r} \int_0^{N^{n_c-2}t} X^N_{n_c-1}(u) \frac{X^N_1(u) X^N_h(u)}{N^2} \, \mathrm{d}u \right) = (0).$$

From Lemma 5.2 we get that, for the convergence in distribution,

$$\lim_{N \to +\infty} \left(\frac{1}{N^r} X^N_{n_c-1}(N^{n_c-2}t)^2 \right) = (0).$$

By gathering these results in Equation (5.20), one finally gets that

$$\lim_{N \to \infty} \left(\frac{1}{N^r} \int_0^{N^{n_c-2}t} \left(2X^N_{n_c-1}(u) - 1\right) X^N_{n_c-1}(u) \, \mathrm{d}u \right) = (0).$$

For $x \in \mathbb{N}$ the relation $x^2 \leq (2x-1)x$ gives therefore that

$$\lim_{N \to \infty} \left(\frac{1}{N^r} \int_0^{N^{n_c-2}t} X^N_{n_c-1}(u) X^N_h(u) \, \mathrm{d}u \right) = (0), \tag{5.21}$$

for $h = n_c - 1$. Now, one proves this convergence by induction on h, from $n_c - 1$ to 2. Assume it holds for all $h \in \{h'+1, \ldots, n_c - 1\}$. By using Relation (5.15), one gets, again

with stochastic calculus with Poisson processes, for $h = h' \leq n_c - 2$,

$$\begin{aligned}
X^N_{n_c-1}(t)X^N_h(t) &= M^N_{n_c-1,h}(t) \\
&+ \int_0^t X^N_h(u)\left(\lambda_{n_c-2}\frac{X^N_{n_c-2}(u)X^N_1(u)}{N^2} - \lambda_{n_c-1}\frac{X^N_{n_c-1}(u)X^N_1(u)}{N^2}\right)\mathrm{d}u \\
&+ \mu_{n_c-1}\int_0^t \left[-X^N_h(u) + \langle \nu_{n_c-1}, I_h\rangle \left(X^N_{n_c-1}(u) - 1\right)\right] X^N_{n_c-1}(u)\,\mathrm{d}u \\
&+ \int_0^t X^N_{n_c-1}(u)\left(\lambda_{h-1}\frac{X^N_{h-1}(u)X^N_1(u)}{N^2} - \lambda_h\frac{X^N_h(u)X^N_1(u)}{N^2}\right. \\
&\qquad \left. - \mu_h X^N_h(u) + \sum_{i=h+1}^{n_c-2}\mu_i X^N_i(u)\langle \nu_i, I_h\rangle\right)\mathrm{d}u \\
&\qquad - \mathbb{I}_{\{h=n_c-2\}}\lambda_{n_c-2}\int_0^t \frac{X^N_{n_c-2}(u)X^N_1(u)}{N^2}\,\mathrm{d}u, \qquad (5.22)
\end{aligned}$$

where $(M^N_{n_c-1,h}(t))$ is a martingale.

We now show that in this identity, when t is replaced by $N^{n_c-2}t$ and it is multiplied by $1/N^r$, then several of its terms vanish in the limit. They are examined one by one.

a) By lemma 5.2 and the fact that $r \geq 2$, one has

$$\lim_{N\to+\infty}\left(\frac{1}{N^r}X^N_{n_c-1}(N^{n_c-2}t)X^N_h(N^{n_c-2}t)\right) = (0).$$

b) For $h_0 = h$, $h-1$,

$$\frac{1}{N^r}\int_0^{N^{n_c-2}t} X^N_{h_0}(u)\frac{X^N_{n_c-1}(u)X^N_1(u)}{N^2}\,\mathrm{d}u \leq \frac{1}{N^r}\int_0^{N^{n_c-2}t} X^N_{n_c-1}(u)\,\mathrm{d}u.$$

Equation (5.21) obtained for $h = n_c - 1$ shows that the last term of this inequality converges in distribution to 0 when N gets large.

c) For $h_0 = h$ and $n_c - 2$,

$$\frac{1}{N^r}\int_0^{N^{n_c-2}t} X^N_{h_0}(u)\frac{X^N_{n_c-2}(u)X^N_1(u)}{N^2}\,\mathrm{d}u \quad \leq \quad \frac{1}{N^{r+1}}\int_0^{N^{n_c-2}t} X^N_{n_c-2}(u)X^N_{h_0}(u)\,\mathrm{d}u,$$

the recurrence relation (5.19) gives that the last term of this relation vanishes as N goes to infinity.

d) For all $n_c - 1 \geq i \geq h+1$, the recurrence assumption (5.21) gives

$$\lim_{N\to+\infty}\left(\frac{1}{N^r}\int_0^{N^{n_c-2}t} X^N_i(u)X^N_{n_c-1}(u)\,\mathrm{d}u\right) = (0).$$

e) The martingale. The relation

$$\left\langle \frac{1}{N^r}M^N_{n_c-1,h}(N^{n_c-2}\cdot)\right\rangle(t) = \frac{1}{N^{2r}}\left\langle M^N_{n_c-1,h}\right\rangle(N^{n_c-2}t)$$

and, in the same way as before, by checking each term of the expression (5.33) of $\langle M^N_{n_c-1,h}\rangle(t)$ in the appendix, one also gets the convergence in distribution

$$\lim_{N\to+\infty}\left(\frac{1}{N^r}M^N_{n_c-1,h}\left(N^{n_c-2}t\right)\right)=(0).$$

One gets finally that the remaining term of Relation (5.22) is also vanishing, the convergence in distribution

$$\lim_{N\to\infty}\left(\frac{1}{N^r}\int_0^{N^{n_c-2}t}X^N_{n_c-1}(u)X^N_h(u)\,\mathrm{d}u\right)=(0)$$

holds, i.e. Relation (5.21) is true for h. This gives the proof of this recurrence scheme for $k=1$ and all $n_c-1\geq h\geq 2$.

To proceed the induction on k, from 1 to $r-1$, one uses the SDE (5.34) and Relation (5.35) for $(X^N_{n_c-k}(t)^2)$ and (5.36) for $(X^N_{n_c-k}(t)X^N_h(t))$, $2\leq h\leq n_c-1$ of the Appendix, and, with the same method which has been used for the first step one gets that, for $1\leq k\leq r-1$ and $2\leq h\leq n_c-1$,

$$\lim_{N\to\infty}\left(\frac{1}{N^r}\int_0^{N^{n_c-2}t}X^N_{n_c-k}(u)^2\,\mathrm{d}u\right)=0,$$

and

$$\lim_{N\to\infty}\left(\frac{1}{N^r}\int_0^{N^{n_c-2}t}X^N_{n_c-k}(u)X^N_h(u)\,\mathrm{d}u\right)=0.$$

The proposition is proved.

□

Proposition 5.3. *For $k=1,\ldots,n_c-2$, the relation*

$$X^N_1(t)^kX^N_{n_c-k}(t)=H^N_k(t)$$
$$+\lambda_{n_c-k-1}\int_0^t\frac{X^N_1(u)^{k+1}X^N_{n_c-k-1}(u)}{N^2}\,\mathrm{d}u-\mu_{n_c-k}\int_0^t X^N_1(u)^kX^N_{n_c-k}(u)\,\mathrm{d}u$$

holds where $(H^N_k(t))$ has the property that, for the convergence in distribution,

$$\lim_{N\to+\infty}\left(\frac{1}{N^{2k}}H^N_k(N^{n_c-2}t)\right)=(0).$$

Proof. By a careful use of the SDE (5.15), one gets that, for $t\geq 0$ and $m=n_c-k$, where

$1 \leq k \leq n_c - 2$

$$
\begin{aligned}
X_1^N(t)^k X_m^N(t) = & M_{(k)}^N(t) \\
& + \lambda_{m-1} \int_0^t \left(\left(X_1^N(u) - 1 \right)^k \left(X_m^N(u) + 1 \right) - X_1^N(u)^k X_m^N(u) \right) \frac{X_1^N(u) X_{m-1}^N(u)}{N^2} \, \mathrm{d}u \Big\} A_1^N(t) \\
& + \mu_m \int_0^t \int_{\mathcal{S}_m} \left(\left(X_1^N(u) + y_1 \right)^k \left(X_m^N(u) - 1 \right) - X_1^N(u)^k X_m^N(u) \right) \nu_m(\mathrm{d}y) X_m^N(u) \, \mathrm{d}u \Big\} A_2^N(t) \\
& + \sum_{i \neq m-1, m} \lambda_i \int_0^t \left(\left(X_1^N(u) - 1 - \mathbb{I}_{\{i=1\}} \right)^k - X_1^N(u)^k \right) X_m^N(u) \frac{X_1^N(u) X_i^N(u)}{N^2} \, \mathrm{d}u \Big\} A_3^N(t) \\
& + \lambda_m \int_0^t \left(\left(X_1^N(u) - 1 \right)^k \left(X_m^N(u) - 1 \right) - X_1^N(u)^k X_m^N(u) \right) \frac{X_1^N(u) X_m^N(u)}{N^2} \, \mathrm{d}u \Big\} A_4^N(t) \\
& + \sum_{i=2}^{m-1} \mu_i \int_0^t \int_{\mathcal{S}_i} \left(\left(X_1^N(u) + y_1 \right)^k - X_1^N(u)^k \right) X_m^N(u) X_i^N(u) \, \nu_i(\mathrm{d}y) \, \mathrm{d}u \Big\} A_5^N(t) \\
& + \sum_{i=m+1}^{n_c-1} \mu_i \int_0^t \int_{\mathcal{S}_i} \left(\left(X_1^N(u) + y_1 \right)^k \left(X_m^N(u) + y_m \right) - X_1^N(u)^k X_m^N(u) \right) \nu_i(\mathrm{d}y) X_i^N(u) \, \mathrm{d}u. \Big\} A_6^N(t)
\end{aligned}
\tag{5.23}
$$

As usual $(M_{(k)}^N(t))$ is a martingale. By looking at all the terms of its previsible increasing process, in the same way as in the proof of Proposition 5.2, it can be shown that, for the convergence in distribution,

$$
\lim_{N \to +\infty} \left(\frac{1}{N^{2k}} M_{(k)}^N \left(N^{n_c - 2} t \right) \right) = (0).
$$

Denote respectively by $A_j^N(t)$, $1 \leq i \leq 6$ the five remaining terms of the right hand side of Relation (5.23). One has

$$
\begin{aligned}
H_k^N(t) = & \left(A_1^N(t) - \lambda_{m-1} \int_0^t \frac{X_1^N(u)^{k+1} X_{m-1}^N(u)}{N^2} \, \mathrm{d}u \right) + \\
& \left(A_2^N(t) + \mu_m \int_0^t X_1^N(u)^k X_m^N(u) \, \mathrm{d}u \right) + \sum_{i=3}^{6} A_i^N(t).
\end{aligned}
$$

One rewrites

$$
\begin{aligned}
A_1^N(t) = & \frac{\lambda_{m-1}}{N^2} \int_0^t X_1^N(u)^{k+1} X_{m-1}^N(u) \, \mathrm{d}u \\
& + \lambda_{m-1} \int_0^t B_1(X_1^N(u), X_{m-1}^N(u)) \frac{X_1^N(u) X_{m-1}^N(u)}{N^2} \, \mathrm{d}u,
\end{aligned}
$$

with $B_1(x,z) \stackrel{\text{def.}}{=} ((x-1)^k - x^k)(z+1)$. Since $X_1^N(t) \leq N$ and that $B_1(x,z)$ is a polynomial of degree $k-1$ in x, by taking $m = n_c - k$ with $1 \leq k \leq n_c - 2$,

$$
\begin{aligned}
& \frac{1}{N^{2k}} \int_0^{N^{n_c-2}t} B_1 \left(X_1^N(u), X_{n_c-k}^N(u) \right) \frac{X_1^N(u) X_{n_c-k-1}^N(u)}{N^2} \, \mathrm{d}u \\
& \qquad \leq c_1 \frac{1}{N^{k+2}} \int_0^{N^{n_c-2}t} (X_{n_c-k}^N(u) + 1) X_{n_c-k-1}^N(u) \, \mathrm{d}u,
\end{aligned}
$$

for some constant c_1. By using Proposition 5.2 and the fact $X_1^N(t)\leq N$ (for the case $k=n_c-2$), one gets that the process associated to this last term converges in distribution to 0 as N goes to infinity.

The second term $(A_2^N(t))$ can be written as

$$A_2^N(t) = -\mu_m \int_0^t X_1^N(u)^k X_m^N(u)\ \mathrm{d}u + \mu_m \int_0^t \int_{\mathcal{S}_m} B_2(X_1^N(u), X_m^N(u))\nu_m(\mathrm{d}y) X_m^N(u)\ \mathrm{d}u,$$

with $B_2(x,z)=((x+y_1)^k-x^k)(z-1)$. Again, by taking $m=n_c-k$ with $1\leq k\leq n_c-2$,

$$\frac{1}{N^{2k}} \int_0^{N^{n_c-2}t} \int_{\mathcal{S}_{n_c-k}} B_2(X_1^N(u), X_{n_c-k}^N(u))\nu_{n_c-k}(\mathrm{d}y) X_{n_c-k}^N(u)\ \mathrm{d}u$$
$$\leq c_2 \frac{1}{N^{k+1}} \int_0^{N^{n_c-2}t} \left(X_{n_c-k}^N(u)-1\right) X_{n_c-k}^N(u)\ \mathrm{d}u,$$

for some constant c_2. Proposition 5.2, again, gives the convergence in distribution to 0 of the process associated to this last term.

For the four remaining terms $(A_3^N(t))$, $(A_4^N(t))$, $(A_5^N(t))$ and $(A_6^N(t))$, with the same type of method, one can show, that, for the convergence in distribution,

$$\lim_{N\to+\infty} \left(\frac{1}{N^{2k}} A_i^N\left(N^{n_c-2}t\right)\right) = (0), \quad i=3,4,5,6.$$

See Section 5.B.3 of the Appendix for the list of inequalities used to get these convergences. By gathering all these results in Equation (5.23) with $m=n_c-k$, one gets the desired identity. □

Corollary 5.1 (Balance Equations). *For $1\leq k\leq n_c-2$ and $t>0$,*

$$\lim_{N\to\infty} \left(\int_0^{N^{n_c-2}t} \left[\frac{\lambda_{n_c-k-1}}{\mu_{n_c-k}} \frac{X_1^N(u)^{k+1}}{N^{2(k+1)}} X_{n_c-k-1}^N(u) - \frac{X_1^N(u)^k}{N^{2k}} X_{n_c-k}^N(u) \right] \mathrm{d}u \right) = (0), \quad (5.24)$$

and, consequently,

$$\lim_{N\to\infty} \left(\int_0^{N^{n_c-2}t} \left[\prod_{k=2}^{n_c-1} \frac{\lambda_{k-1}}{\mu_k} \frac{X_1^N(u)^{n_c}}{N^{2n_c-2}} - \frac{X_1^N(u) X_{n_c-1}^N(u)}{N^2} \right] \mathrm{d}u \right) = (0).$$

Proof. Lemma 5.2 gives the convergence in distribution

$$\lim_{N\to+\infty} \left(\frac{1}{N^{2k}} X_1^N(N^{n_c-2}t)^k X_{n_c-k}^N(N^{n_c-2}t) \right) = (0),$$

the first identity is then a consequence of Proposition 5.3.

□

The next proposition states that, on the time scale $t\mapsto N^{n_c-2}t$, most of the polymers are of size 1.

Proposition 5.4. *The convergence in distribution*

$$\lim_{N\to\infty}\left(\frac{1}{N}X^N\left(N^{n_c-2}t\right)\right)=(1,0,\ldots,0)$$

holds.

Proof. By using Lemma 5.2 and the fact that $(X^N(t))\in\mathcal{S}_N$, we just have to prove the convergence result for the last coordinate $X^N_{n_c}$ of X^N. By using SDE (5.15) and Relation (5.18), we have

$$X^N_{n_c}(t)=\lambda_{n_c-1}\int_0^t\frac{X^N_1(u)X^N_{n_c-1}(u)}{N^2}\,\mathrm{d}u+M^N_{n_c}(t), \tag{5.25}$$

where $(M^N_{n_c}(t))$ whose previsible increasing previsible process is

$$\left\langle M^N_{n_c}\right\rangle(t)=\lambda_{n_c-1}\int_0^t\frac{X^N_1(s)X^N_{n_c-1}(s)}{N^2}\,\mathrm{d}s.$$

The inequality

$$\frac{1}{N^2}\int_0^{N^{n_c-2}t}\frac{X^N_1(s)X^N_{n_c-1}(s)}{N^2}\,\mathrm{d}s\le\frac{1}{N^3}\int_0^{N^{n_c-2}t}X^N_{n_c-1}(s)^2\,\mathrm{d}s,$$

Proposition 5.2 and Lemma I.3.30 of Jacod and Shiryaev [81] give the convergence in distribution

$$\lim_{N\to+\infty}\left(\frac{1}{N}M^N_{n_c}(N^{n_c-2}t)\right)=(0).$$

From Corollary 5.1, one has

$$\lim_{N\to\infty}\left(\int_0^{N^{n_c-2}t}\prod_{k=2}^{n_c-1}\frac{\lambda_{k-1}}{\mu_k}\frac{X^N_1(u)^{n_c}}{N^{2n_c-2}}-\frac{X^N_1(u)X^N_{n_c-1}(u)}{N^2}\,\mathrm{d}u\right)=(0). \tag{5.26}$$

and, since

$$\frac{1}{N}\int_0^{N^{n_c-2}t}\frac{X^N_1(u)^{n_c}}{N^{2n_c-2}}\,\mathrm{d}u=\frac{1}{N}\int_0^t\frac{X^N_1(N^{n_c-2}u)^{n_c}}{N^{n_c}}\,\mathrm{d}u\le\frac{t}{N},$$

we obtain therefore the convergence in distribution

$$\lim_{N\to+\infty}\left(\frac{1}{N}\int_0^{N^{n_c-2}t}\frac{X^N_1(u)X^N_{n_c-1}(u)}{N^2}\,\mathrm{d}u\right)=(0).$$

One concludes the proof of the proposition with Equation (5.25). □

Theorem 5.1. *The sequence of processes* $(X^N_{n_c}(N^{n_c-2}t))$ *is converging in distribution to a Poisson process on* $\mathbb{R}_+$ *with rate*

$$\rho\overset{\text{def.}}{=}\lambda_1\prod_{k=2}^{n_c-1}\frac{\lambda_k}{\mu_k}. \tag{5.27}$$

Proof. Indeed Clearly $(X_{n_c}^N(N^{n_c-2}t))$ is a counting process, i.e. a non-decreasing integer valued process with jumps of size one. By Equation (5.25), its compensator is given by

$$\left(\lambda_{n_c-1}\int_0^{N^{n_c-2}t}\frac{X_1^N(u)X_{n_c-1}^N(u)}{N^2}\,\mathrm{d}u\right).$$

Relation (5.26) shows that the later process has the same limit as the process

$$\left(\lambda_{n_c-1}\prod_{k=2}^{n_c-1}\frac{\lambda_{k-1}}{\mu_k}\int_0^{N^{n_c-2}t}\frac{X_1^N(u)^{n_c}}{N^{2n_c-2}}\,\mathrm{d}u\right)=\left(\lambda_{n_c-1}\prod_{k=2}^{n_c-1}\frac{\lambda_{k-1}}{\mu_k}\int_0^t\left(\frac{X_1^N(N^{n_c-2}u)}{N^{n_c}}\right)^{n_c}\mathrm{d}u\right).$$

From Proposition 5.4, the right hand side of this relation converges in distribution to (t). We have thus shown that the compensator of $(X_{n_c}^N(N^{n_c-2}t))$ is converging in distribution to

$$\left(\lambda_1\prod_{k=2}^{n_c-1}\frac{\lambda_k}{\mu_k}t\right),$$

Theorem 5.1 of Kasahara and Watanabe [88], see also Brown [26], gives the desired convergence in distribution to a Poisson process. The theorem is proved. □

Proposition 5.5 (Asymptotic of the First Nucleation Time)**.** *If $((U^N(t))$ is the process defined by Relations (5.10), (5.11) and (5.12) and*

$$T^N=\inf\{t\geq 0: U_{n_c}^N(t)=1\},$$

then, for the convergence in distribution,

$$\lim_{N\to+\infty}\frac{T^N}{N^{n_c-3}}=E_\rho,$$

where E_ρ is an exponential random variable with parameter ρ defined by Relation (5.27).

5.4 Fast Growth of Stable Polymers

In this section, the evolution of the polymerization process after nucleation is investigated. The main difference with Section 5.3 lies in the fact that the polymers become more stable after nucleation: a polymer of size $k\geq n_c$ is degraded at rate μ instead of the rate $\mu_k N$ when $k<n_c$ as in Section 5.3, where, as before, N is the scaling parameter. See Relation (5.7). As before we will introduce an auxiliary process to study this phase.

We will be interested by the evolution of the number of polymers whose size are greater than n_c. For $\alpha>0$, a Markov process $(Z^\alpha(t))=(Z_i^\alpha(t), i\geq n_c)$ will be introduced for this purpose. With a slight abuse of notations, we will consider $(Z^\alpha(t))$ as a process in the state space $\mathcal{S}_\infty$. Formally, it can be done by assuming that the n_c-1 first coordinates are null. We denote by $\|z\|$ the sum of the components of the vector $z=(z_i)$ of the set $\mathcal{S}_\infty$ of states with finite mass defined by Relation (5.5),

$$\|z\|=\sum_{i\geq n_c}z_i.$$

For $z \in \mathcal{S}_\infty$, then the generator Ω_{Z^α} of the process is given by

$$\Omega_{Z^\alpha}(f)(z) = \sum_{k=n_c}^{+\infty} [f(z+e_{k+1}-e_k)-f(z)]\, \alpha z_k + \sum_{k=n_c}^{+\infty} \int_{\mathcal{S}_k} [f(z+y-e_k)-f(z)]\, \mu z_k \nu_k(\mathrm{d}y) \quad (5.28)$$

where, as before, for $i \geq 1$, e_i is the ith unit vector of $\mathcal{S}_\infty$ and f is a function on $\mathcal{S}_\infty$ with finite support. Since the first n_c-1 first coordinates of $(Z^\alpha(t))$ are implicitly 0, in Relation (5.28) we use the convention that $f(z+y)=f(z+\pi_{n_c}(y))$ where $\pi_{n_c}(y)=(0,\ldots,0,y_{n_c},y_{n_c+1},\ldots)$ for $y \in \mathcal{S}_\infty$, π_{n_c} is the projection on the coordinates with index greater or equal to n_c.

We give a quick, informal, motivation for the introduction of the Markov process $(Z^\alpha(t))$. It describes in fact the evolution of stable polymers. Assume for the moment that the polymerization rates are independent of the sizes of polymers, i.e. $\lambda_k=\alpha$ for all $k \geq n_c$. A monotonicity argument will be used to have a more general framework. The initial state is assumed to be $Z^\alpha(0)=e_{n_c}$ with only one polymer of size n_c present at time 0, as it is the case just after the first nucleation instant. Proposition 5.4 gives that, at this instant, the number of polymers is small with respect to N, i.e. the fraction of monomer is close to 1. A polymer of size greater that n_c grows therefore at a rate close α as in Relation (5.28). If the fragmentation of a polymer of size $k \geq n_c$ gives a polymer of size $<n_c$, then, due to the fast fragmentation rates below n_c, this last one is fragmented quickly into monomers and thus vanishes as in Relation (5.28) since coordinates with index less than n_c are assumed to be 0 for the process $(Z^\alpha(t))$.

Proposition 5.6. *Under Assumption A-3, if $Z^\alpha(0)=e_m$ for some $m \geq n_c$, there exist $\kappa_0 \geq 0$, $a_0>0$ and $\eta>0$ such for any α and $\mu>0$ such that $\alpha/\mu \geq \kappa_0$, then the event*

$$\mathcal{F}_{Z^\alpha} = \left\{ \liminf_{t \to +\infty} e^{-a_0 t} \|Z^\alpha(t)\| > \eta \right\}$$

has a positive probability.

Proof. The evolution of $(Z^\alpha(t))$ describes the the population of stable polymers generated by an initial polymer with size $m \geq n_c$. By condition A-3, for $\varepsilon>0$ sufficiently small, there exists some $k_0 \geq n_c$ such that if $k \geq k_0$,

$$\nu_k\left(y : \sum_{i \geq n_c} y_i \geq 1\right) > 1-\varepsilon \text{ and } \nu_k\left(y : \sum_{i \geq n_c} y_i \geq 2\right) > 2\varepsilon$$

A stable polymer of size $m \geq n_c$ is fragmented after an exponentially distributed amount of time E_μ with parameter μ. Just before this instant its size has the same distribution as $M=m+\mathcal{N}_\alpha([0,E_\mu])$. Recall that only the fragments with size greater than n_c are considered for $(Z^\alpha(t))$. The process $(Z^\alpha(t))$ can then be also seen as a multi-type branching process with the type of an individual being the size of the corresponding polymer. The average number of stable polymers generated by the fragmentation of the

polymer is greater than

$$\mathbb{E}\left(\nu_M\left(y:\sum_{i\geq n_c} y_i\geq 1\right)\right)+\mathbb{E}\left(\nu_M\left(y:\sum_{i\geq n_c} y_i\geq 2\right)\right)$$
$$\geq \mathbb{P}(M\geq k_0)((1-\varepsilon)+2\varepsilon)=(1+\varepsilon)\left(\frac{\alpha}{\alpha+\mu}\right)^{k_0-n_c}.$$

By choosing α/μ sufficiently large, the last quantity is strictly greater than 1. We have thus shown that the process $(\|Z^\alpha(t)\|)$ is lower bounded by a continuous time supercritical branching process. The proposition is then a simple consequence of a classical result in this domain, see Chapter V of Athreya and Ney [13] for example. □

5.5 Asymptotic Behavior of Lag Time

We now return to the original polymerization process $(U_k^N(t))$ with values in the state space $\mathcal{S}_N$ defined by the SDE (5.10). The goal is of establishing our main result on the asymptotic behavior of the associated lag time L_δ^N defined by Relation (5.13) for a small δ which will be determined later. The variable T^N defined by Relation (5.14) is a stopping time clearly satisfying $T^N\leq L_\delta^N$. Let

$$u^N=(u_k^N)\overset{\text{def.}}{=}U^N(T^N), \tag{5.29}$$

in particular $u_{n_c}^N=1$ and $u_k^N=0$ for $k>n_c$. For the other coordinates of u^N, from Lemmas 5.1 and 5.2 and Proposition 5.5, one gets the following convergence in distribution,

$$\lim_{N\to+\infty}\frac{1}{\sqrt{N}}\sum_{k=2}^{n_c-1}u_k^N=0.$$

In the rest of this section, one denotes by $(\widehat{U}^N(t))$ the solution of the SDE (5.10) with initial point u^N and $\widehat{L}_\delta^N$ the corresponding lag time, in particular $L_\delta^N=T^N+\widehat{L}_\delta^N$.

Fix some $\delta_0>\delta$ and define

$$\mathcal{E}_N\overset{\text{def.}}{=}\left\{\widehat{U}_1^N(t)\geq(1-\delta_0)N,\forall t\leq\widehat{L}_\delta^N\wedge N\right\}, \tag{5.30}$$

with $a\wedge b=\min(a,b)$. If N is sufficiently large so that $(\delta_0-\delta)\sqrt{N}>1$, then

$$\left\{\frac{1}{\sqrt{N}}\sum_{k=2}^{n_c-1}\widehat{U}_i^N(t)\leq 1,\forall t\leq\widehat{L}_\delta^N\wedge N\right\}\subset\mathcal{E}_N$$

By using again Lemmas 5.1 and 5.2 and the strong Markov property, one gets that

$$\lim_{N\to+\infty}\mathbb{P}(\mathcal{E}_N)=1. \tag{5.31}$$

We have to study the order of magnitude of $\widehat{L}_\delta^N$. We will prove that, with a positive probability, the mass of stable polymers hits the value $\lfloor\delta N\rfloor$ in a duration of time of the order of $\log N$.

A Coupling

Let us introduce a process $(Z^{\alpha_0}(t))$ with the initial state $Z^{\alpha_0}(0)=e_{n_c}$ and generator (5.28) of Section 5.4 with $\alpha_0=\underline{\lambda}(1-\delta_0)$.

Proposition 5.7 (Coupling). *Under Assumptions A-2 and A-4, one can construct a version of the processes $(Z^{\alpha_0}(t), t\geq 0)$ and $(\widehat{U}^N(t), t\geq 0)$ such that, $Z^{\alpha_0}(0)=e_{n_c}$ and, on the event $\mathcal{E}_N$, the relation*

$$\sum_{k\geq n} Z_k^{\alpha_0}(t) \leq \sum_{k\geq n} \widehat{U}_k^N(t) \tag{5.32}$$

holds for all $0\leq t\leq \widehat{L}_\delta^N \wedge N$ and $n\geq n_c$.

Proof. At time 0, there is at least a polymer of size n_c for $(\widehat{U}^N(t))$ and a unique one for $(Z^{\alpha_0}(t))$. For $t\geq 0$, recall that

$$\|Z^{\alpha_0}(t)\| = \sum_{n\geq n_c} Z_n^{\alpha_0}(t)$$

is the number of polymers for $(Z^{\alpha_0}(t))$ and $A_1(t), \ldots, A_{\|Z^{\alpha_0}(t)\|}(t)$, their respective sizes. One will show, by induction on the number of jumps of $(Z^{\alpha_0}(t))$ and $(\widehat{U}^N(t))$ after time 0, that there are at least $\|Z^{\alpha_0}(t)\|$ of polymers of $\widehat{U}^N(t)$ whose sizes are given respectively by $B_1(t), \ldots, B_{\|Z^{\alpha_0}(t)\|}(t)$ with $A_k(t) \leq B_k(t)$ holds for all $1\leq k\leq \|Z^{\alpha_0}(t)\|$. Assume that this relation holds at some fixed time $0\leq t\leq \widehat{L}_\delta^N$, we will show that one can construct a version of the two process after that time so that the relation will also hold after the next jump of $(Z^{\alpha_0}(t))$ and $(\widehat{U}^N(t))$. We give the construction of the next jump.

1. A monomer addition to the polymer of size $A_i(t-)$ occurs at rate α. A monomer addition to the polymer of size $B_i(t)$ occurs also at that time for the process $(\widehat{U}^N(t))$.
2. A monomer addition to the polymer of size $B_i(t-)=k$ occurs at rate given by $\lambda_k\widehat{U}_1^N(t-)-\underline{\lambda}(1-\delta_0)N$. This last quantity is non-negative because, by definition, $\lambda_k\geq\underline{\lambda}$ and, since $t<\widehat{L}_\delta^N$, one has the relation $\widehat{U}_1^N(t-)\geq\underline{\lambda}(1-\delta_0)N$. There is no change for the process $(Z^{\alpha_0}(t))$ with this event.
3. If $A_i(t-)=k$ and $B_i(t-)=k'$, at rate μ the polymer of size k [resp. k'] is fragmented as $A_1^1, \ldots A_{n_A}^1$ [resp. $B_1^1, \ldots B_{n_B}^1$] according to the distribution ν_k [resp. ν_k']. By Assumption A-4, the random variables (A_i^1) and (B_i^1) can be chosen so that, for any $1\leq i\leq n_A$, there exists some $1\leq m_i\leq n_B$ such that $A_i^1\leq B_{m_i}^1$ and all indices m_i, $i = 1, \ldots, n_A$ are distinct.

With this construction, it is easily seen that $(\widehat{U}^N(t))$ and $(Z^{\alpha_0}(t))$ are indeed Markov processes with the generator (5.9) and (5.28) respectively. Moreover each of the three transitions described above preserve the desired relation. Equation (5.32) is a simple consequence of this relation. □

Proposition 5.8. *Under Assumptions* A*, *if $\underline{\lambda}>\kappa_0\mu$, where κ_0 is the constant of Proposition 5.6 and $\underline{\lambda}$ is defined by Relation (5.8), then for any $\delta_0<1-\kappa_0\mu/\underline{\lambda}$, there exist positive constants K and p_0 such that*

$$\liminf_{N\to+\infty} \mathbb{P}\left(\frac{\widehat{L}_{\delta_0}^N}{\log N} \leq K\right) > p_0.$$

Proof. In the following, all statements are understood on the event $\mathcal{E}_N$ defined by Equation (5.30). Relation (5.31) gives that this event has a probability close to 1 as N gets large.

From the coupling proved in Proposition 5.7 and with the same notations, it can be assumed that there exists a Markov process $(Z^{\alpha_0}(t))$ with generator defined by Relation (5.28) and initial point e_{n_c} such that the relation

$$\|Z^{\alpha_0}(t)\| = \sum_{k\geq n_c} Z_k^{\alpha_0}(t) \leq \sum_{k\geq n_c} \widehat{U}_k^N(t),$$

holds for all $t\leq \widehat{L}_{\delta_0}^N$. For $K>0$,

$$\mathcal{H}_N \stackrel{\text{def.}}{=} \left\{\widehat{L}_{\delta_0}^N > K\log N\right\} \subset \{\|Z^{\alpha_0}(K\log N)\| \leq \lfloor \delta_0 N\rfloor\} \subset \{\|Z^{\alpha_0}(K\log N)\| \leq \lfloor \delta_0 N\rfloor\},$$

Since $\underline{\lambda}(1-\delta_0)>\mu\kappa_0$, with the notations of Proposition 5.6, one can take $K=2/a_0$ then

$$\mathcal{F}_{Z^{\alpha_0}}\cap\mathcal{H}_N=\emptyset,$$

as soon as $N>2\delta_0/\eta$, where the event $\mathcal{F}_{Z^{\alpha_0}}$ is defined in Proposition 5.6, hence

$$\lim_{N\to+\infty} \mathbb{P}\left(\mathcal{E}_N\cap\mathcal{F}_{Z^{\alpha_0}}\cap\left\{\widehat{L}_{\delta_0}^N \leq A\log N\right\}\right) = \mathbb{P}(\mathcal{F}_{Z^{\alpha_0}})>0.$$

The proposition is proved. □

Theorem 5.2 (Asymptotics for the Lag Time). *Under Assumptions A^*, if $\underline{\lambda}>\kappa_0\mu$, where κ_0 is the constant of Proposition 5.6 and $\underline{\lambda}$ is defined by Relation (5.8), for all $0<\delta<1-\kappa_0\mu/\underline{\lambda}$, then, for any $\varepsilon>0$, there exists K_1 and K_2, $0<K_1<K_2$, such that*

1. *If $n_c=3$,*
$$\liminf_{N\to+\infty} \mathbb{P}\left(\frac{L_\delta^N}{\log N} \leq K_2\right) > 1-\varepsilon.$$
2. *If $n_c\geq 4$,*
$$\liminf_{N\to+\infty} \mathbb{P}\left(K_1 \leq \frac{L_\delta^N}{N^{n_c-3}} \leq K_2\right) > 1-\varepsilon.$$

Proof. By Proposition 5.5, at time T^N a polymer of size n_c is created. According to Proposition 5.8, for N sufficiently large, with probability at least p_0 there is a fraction δ of monomers polymerized as stable polymers before time $R_1^N\stackrel{\text{def.}}{=}T^N+K\log N$. If this does not happen, there are two possibilities:

1. There is at least one stable polymer at time R_1^N. One can construct another coupling with an independent process $(\widetilde{Z}^{\alpha_0}(t))$ with the same distribution as $(Z^{\alpha_0}(t))$. Again there is a probability p_0 that a fraction δ of monomers is polymerized into stable polymers before time $R_1^N+K\log N$. If this does not happen, there are again the same two possibilities.
2. There are no stable polymers at time R_1^N. From Lemma 5.1 and 5.2, the fraction of monomers is 1 and thus the systems starts afresh: a polymer of size n_c is created at a time $R_1^N + T_1^N$, with T_1^N independent of T^N and with the same distribution. We can then repeat our argument.

This decomposition shows that the lag time L_δ^N can be stochastically upper-bounded by the random variable

$$\sum_{i=1}^{1+G_{p_0}} \left(T_i^N + K \log N\right),$$

where (T_i) is an i.i.d. sequence of random variables with the same distribution as T^N and G is a geometrically distributed random variable with parameter p_0 which is independent of the sequence (T_i^N). The theorem is then a simple consequence of Proposition 5.5. The theorem is proved. □

5.A Appendix: Biological background

The protein polymerization processes investigated in this chapter are believed to be the main phenomena at the origin of several neuro-degenerative diseases such as Alzheimer's, Parkinson's and Huntington's diseases for example. The general picture of this setting is the following. At some moment, for some reasons, within a neural cell a fraction of the proteins of a given type are produced in a anomalous state, defined as misfolded state. Recall that if a protein is a sequence of amino-acids, its three-dimensional structure determines also its functional properties.

A misfolded protein has the same sequence of amino-acids but a different spatial architecture. It turns out that misfolded proteins tend to aggregate to form fibrils, also called polymers. These fibrils are believed to have a toxic impact in the cell, on its membrane in particular, leading to its death. The prion protein PrP^C is an example of such protein that can be polymerized when it is in the state PrP^{SC}. The corresponding disease is the Bovine Spongiform Encephalopathy (BSE), also known as the mad cow disease. See Dobson [50, 49] and Ross and Poirier [143]. There are mechanisms within the cell, via other proteins, to "correct" misfolded proteins, but they seem to be inefficient when the phenomenon has really started. See Bozaykut et al. [22] or Smith et al. [150]. This (rough) description is not completely accurate or complete, moreover some aspects are disputed, but it is used in a large part of the current literature. See also the interesting historical surveys of Sipe and Cohen [148] and Pujo-Menjouet [133]. Other biological processes such as actin filamentation, or yet industrial processes exhibit similar mechanisms, see McManus et al. [110], Wegner and Engel [166], Voter and Erickson [164] and Ow and Dustan [123].

The Variability of the Polymerization Process

Neuro-degenerative diseases are quite diverse. They can be infectious, like the BSE, others are not, like Alzheimer (apparently). Nevertheless they all exhibit large, variable, time spans for the development of the disease, from several years to 10 years.

When experiments are done in vitro with convenient types of proteins/monomers and with no initial polymers, a related phenomenon is observed. The fraction of monomers consumed by the polymerization process exhibit an S-curve behavior: it stays at 0 for several hours, and quickly reaches 1, the state where most of monomers are polymerized. The other key feature of these experiments concerns the variability of the instant of the take-off phase of the S-curve from an experiment to another. See

Szavits-Nossan et al. [157] and Xue et al. [168]. See Figure 5.1 where twelve experiments are represented, the instant when half of the proteins are polymerized varies from 7.3h. to 10.1h. with an average of 8.75h. See also Pigolotti et al. [129] and Eden et al. [53].

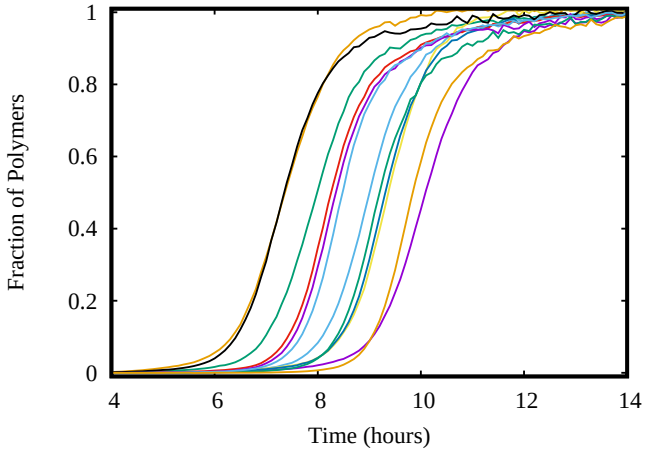


Figure 5.1 – Twelve experiments for the evolution of polymerised mass. From data published in Xue et al. [168], see also Eugene et al. [58].

The initial step of the chain reactions giving rise to polymers consists in the spontaneous formation of a so-called *nucleus*, that is, the simplest possible polymer able to ignite the reaction of polymerization. This early phase is called *nucleation*, and is still far from being understood. See Philo and Arakawa [127].

5.B Appendix: Technical Details

5.B.1 A Reminder on Marked Poisson Point Processes

We first recall briefly some elementary aspects of stochastic calculus with marked Poisson point processes. They are used throughout this chapter. See Jacobsen [79] and Last and Brandt [102] for more details. Let $\mathcal{N}_\lambda = (t_n)$ be a Poisson point process on $\mathbb{R}_+$ with parameter λ and an independent sequence (U_n) of i.i.d. random variables on some locally compact space H, μ denotes the common distribution of these variables. The marked Poisson point process $\mathcal{N}_\lambda^U$ is defined as a point process on $\mathbb{R}_+ \times H$, by

$$\mathcal{N}_\lambda^U = \sum_{n\in\mathbb{N}} \delta_{(t_n, U_n)},$$

if f is a non-negative measurable function on $\mathbb{R}_+ \times H$, one defines, for $t \geq 0$,

$$\int_0^t f(s,u)\mathcal{N}_\lambda^U(\mathrm{d}s, \mathrm{d}u) = \sum_{n\in\mathbb{N}} f(t_n, U_n)\mathbb{I}_{\{t_n \leq t\}}$$

and, if $F \in \mathcal{B}(H)$ is a Borelian subset of H,

$$\mathcal{N}_\lambda^U([0,t] \times F) = \int_0^t \mathbb{I}_{\{u\in F\}}\mathcal{N}_\lambda^U(\mathrm{d}s, \mathrm{d}u) = \sum_{n\in\mathbb{N}} \mathbb{I}_{\{t_n \leq t, U_n \in F\}}.$$

The natural filtration associated to $\mathcal{N}_\lambda^U$ is $(\mathcal{F}_t)$, with, for $t\geq 0$,

$$\mathcal{F}_t = \sigma\left(\mathcal{N}_\lambda^U([0,s]\times F) : s\leq t, F\in\mathcal{B}(H)\right).$$

Proposition 5.9. *If g is a càdlàg function on $\mathbb{R}_+$ and h is Borelian on H such that*

$$\int_0^t g(s)^2\,\mathrm{d}s<+\infty, \quad \forall t\geq 0 \text{ and } \int_H h(u)^2\nu(\mathrm{d}u)<+\infty,$$

then the process

$$(M(t)) \stackrel{\text{def.}}{=} \left(\int_0^t g(s-)h(u)\mathcal{N}_\lambda^U(\mathrm{d}s,\mathrm{d}u) - \lambda\int_H h(u)\nu(\mathrm{d}u)\int_0^t g(s)\,\mathrm{d}s\right)$$

is a square integrable martingale with respect to the filtration $(\mathcal{F}_t)$, its previsible increasing process is given by

$$(\langle M\rangle(t) = \lambda\int_H h^2(u)\nu(\mathrm{d}u)\int_0^t f(s)^2\,\mathrm{d}s$$

In this chapter, since we are dealing with several Poisson point processes, the (implicit) definition of the filtration $(\mathcal{F}_t)$ is extended so that it includes all of them.

5.B.2 Additional Calculations for the Proof of Proposition 5.2

After some careful calculations, for $2\leq h\leq n_c-2$, the previsible increasing process of the martingale of Relation (5.22) can be expressed by

$$\begin{aligned}
&\left\langle M_{n_c-1,h}^N\right\rangle(t)\\
&\quad=\mathbb{I}_{\{h=n_c-2\}}\lambda_{n_c-2}\int_0^t\left(X_{n_c-2}^N(u)-X_{n_c-1}^N(u)-1\right)^2\frac{X_{n_c-2}^N(u)X_1^N(u)}{N^2}\,\mathrm{d}u\\
&\quad+\int_0^t X_h^N(u)^2\left(\lambda_{n_c-2}\frac{X_{n_c-2}^N(u)X_1^N(u)}{N^2}\mathbb{I}_{\{h<n_c-2\}}+\lambda_{n_c-1}\frac{X_{n_c-1}^N(u)X_1^N(u)}{N^2}\right)\mathrm{d}u\\
&\quad+\mu_{n_c-1}\int_0^t\int_{y\in\mathcal{S}_{n_c-1}}\left(X_h^N(u)-y_h(X_{n_c-1}^N(u)-1)\right)^2X_{n_c-1}^N(u)\,\nu_{n_c-1}(\mathrm{d}y)\,\mathrm{d}u\\
&\quad+\int_0^t X_{n_c-1}^N(u)^2\left(\lambda_{h-1}\frac{X_{h-1}^N(u)X_1^N(u)}{N^2}+\lambda_h\frac{X_h^N(u)X_1^N(u)}{N^2}\mathbb{I}_{\{h<n_c-2\}}+\mu_hX_h^N(u)\right)\mathrm{d}u\\
&\quad+\sum_{i=h+1}^{n_c-2}\mu_i\int_{\mathcal{S}_i}y_h^2\nu_i(\mathrm{d}y)\int_0^t X_{n_c-1}^N(u)^2X_i^N(u)\,\mathrm{d}u. \qquad (5.33)
\end{aligned}$$

The SDE (5.15) gives for, $1<m<n_c-1$,

$$\begin{aligned}
X_m^N(t)^2=&M_{m,m}^N(t)+\lambda_{m-1}\int_0^t\left(2X_m^N(u)+1\right)\frac{X_1^N(u)X_{m-1}^N(u)}{N^2}\,\mathrm{d}u\\
&+\sum_{i=m+1}^{n_c-1}\mu_i\int_0^t\int_{y\in\mathcal{S}_i}\left(2X_m^N(u)y_m+y_m^2\right)\nu_i(\mathrm{d}y)X_i^N(u)\,\mathrm{d}u\\
&+\int_0^t\left(1-2X_m^N(u)\right)\left(\mu_mX_m^N(u)+\lambda_m\frac{X_1^N(u)X_m^N(u)}{N^2}\right)\mathrm{d}u, \qquad (5.34)
\end{aligned}$$

$(M^N_{m,m}(t))$ is a martingale whose previsible increasing process is given by

$$\begin{aligned}\left\langle M^N_{m,m}\right\rangle(t)=&\lambda_{m-1}\int_0^t\left(2X^N_m(u)+1\right)^2\frac{X^N_1(u)X^N_{m-1}(u)}{N^2}\,\mathrm{d}u\\&+\sum_{i=m+1}^{n_c-1}\mu_i\int_0^t\int_{y\in\mathcal{S}_i}\left(2X^N_m(u)y_m+y_m^2\right)^2\nu_i(\mathrm{d}y)X^N_i(u)\,\mathrm{d}u\\&+\int_0^t\left(1-2X^N_m(u)\right)^2\left(\mu_mX^N_m(u)+\lambda_m\frac{X^N_1(u)X^N_m(u)}{N^2}\right)\mathrm{d}u.\end{aligned}\tag{5.35}$$

For $1<h\leq n_c-1$,

$$\begin{aligned}X^N_1(t)X^N_h(t)=&M^N_{1,h}(t)+\lambda_{h-1}\int_0^t\left(X^N_1(u)-1\right)\frac{X^N_{h-1}(u)X^N_1(u)}{N^2}\,\mathrm{d}u\\&+\lambda_h\int_0^t\left(1-X^N_1(u)\right)\frac{X^N_h(u)X^N_1(u)}{N^2}\,\mathrm{d}u\\&-\sum_{j=2}^{n_c-1}\lambda_j\int_0^tX^N_h(u)\frac{X^N_j(u)X^N_1(u)}{N^2}\,\mathrm{d}u-2\lambda_1\int_0^tX^N_h(u)\frac{X^N_1(u)^2}{N^2}\,\mathrm{d}u\\&+\sum_{i=2}^{n_c-1}\mu_i\int_0^t\int\left[y_1X^N_h(u)+y_h\left(X^N_1(u)+y_1\right)\mathbb{I}_{\{i>h\}}-\left(X^N_1(u)+y_1\right)\mathbb{I}_{\{i=h\}}\right]\\&\qquad\nu_i(\mathrm{d}y)X^N_i(u)\,\mathrm{d}u.\end{aligned}\tag{5.36}$$

5.B.3 Proof of Proposition 5.3

The End of the Proof of Proposition 5.3. With the same notations as at the beginning of the proof of Proposition 5.3, we review the components of the A^N_j of Equation (5.23), for $j=3,\ldots,6$. Recall that $m=n_c-k$ with $1\leq k\leq n_c-2$.

The generic scaled expression of $A^N_3(t)$ is, for $1\leq i\leq n_c-1$,

$$\begin{aligned}&\frac{1}{N^{2k}}\int_0^{N^{n_c-2}t}\left(\left(X^N_1(u)-1-\mathbb{I}_{\{i=1\}}\right)^k-\left(X^N_1(u)\right)^k\right)X^N_m(u)\frac{\lambda_iX^N_1(u)X^N_i(u)}{N^2}\,\mathrm{d}u\\&\qquad\leq c_3\frac{1}{N^{k+1}}\int_0^{N^{n_c-2}t}X^N_m(u)\,\mathrm{d}u\leq\frac{c_3}{N^{k+1}}\int_0^{N^{n_c-2}t}X^N_m(u)^2\,\mathrm{d}u,\end{aligned}$$

for some constant c_3.

For $A^N_4(t)$ the analogous term is, for a convenient constant c_5,

$$\begin{aligned}&\frac{1}{N^{2k}}\int_0^{N^{n_c-2}t}\left(\left(X^N_1(u)-1\right)^k\left(X^N_m(u)-1\right)-X^N_1(u)^kX^N_m(u)\right)\frac{X^N_1(u)X^N_m(u)}{N^2}\,\mathrm{d}u\\&\qquad\leq c_4\frac{1}{N^{k+1}}\int_0^{N^{n_c-2}t}X^N_m(u)^2\,\mathrm{d}u.\end{aligned}$$

For $A_5^N(t)$, if $i{=}n_c{-}k{-}1,\ldots,2$, this is

$$\frac{1}{N^{2k}}\int_0^{N^{n_c-2}t}\int_{\mathcal{S}_i}\left(\left(X_1^N(u){+}y_1\right)^k-X_1^N(u)^k\right)X_m^N(u)X_i^N(u)\nu_i(\mathrm{d}y)\ \mathrm{d}u$$
$$\leq c_5\frac{1}{N^{k+1}}\int_0^{N^{n_c-2}t}X_m^N(u)X_i^N(u)\ \mathrm{d}u$$

For $A_6^N(t)$ it is, if $i{=}n_c{-}k{+}1,\ldots,n_c{-}1$, and $n_c{-}2{\geq}k{\geq}2$,

$$\frac{1}{N^{2k}}\int_0^{N^{n_c-2}t}\int_{\mathcal{S}_i}\left(\left(X_1^N(u){+}y_1\right)^k\left(X_m^N(u){+}y_m\right)-X_1^N(u)^kX_m^N(u)\right)\nu_i(\mathrm{d}y)X_i^N(u)\,\mathrm{d}u$$
$$\leq c_6\frac{1}{N^k}\int_0^{N^{n_c-2}t}X_i^N(u)\ \mathrm{d}u\leq c_6\frac{1}{N^k}\int_0^{N^{n_c-2}t}X_i^N(u)^2\ \mathrm{d}u.$$

The right-hand-side of the four previous inequalities satisfy the assumptions of Lemma 5.2 and therefore each of them, as a process, converges in distribution to 0. This completes the proof of the proposition.

□

Chapter 6

A Functional Central Limit Theorem in the Becker-Döring Model

Contents

6.1 Introduction 187
6.2 The Stochastic Model 192
6.3 SDE in Hilbert space 195
6.3.1 Remark on the assumptions on the coefficients (a_i) and (b_i) . . . 197
6.4 Convergence of the Fluctuation Processes 201

abstract We investigate the fluctuations of the stochastic Becker-Döring model of polymerization when the initial size of the system converges to infinity. A functional central limit problem is proved for the vector of the number of polymers of a given size. It is shown that the stochastic process associated to fluctuations is converging to the strong solution of an infinite dimensional stochastic differential equation (SDE) in a Hilbert space. We also prove that, at equilibrium, the solution of this SDE is a Gaussian process. The proofs are based on a specific representation of the evolution equations, the introduction of a convenient Hilbert space and several technical estimates to control the fluctuations, especially of the first coordinate which interacts with all components of the infinite dimensional vector representing the state of the process.

6.1 Introduction

Polymerization is a key phenomenon in several important biological processes. Macro-molecules proteins, also called monomers, may be assembled randomly into several aggregated states called polymers or clusters. These clusters can themselves be fragmented into monomers and polymers at some random instants. The fluctuations of the number of polymerized monomers analyzed in this chapter is an important characteristic of polymerization processes in general.

The Becker-Döring model

We investigate the fluctuations of a stochastic version of the Becker-Döring model which is a classical mathematical model to study polymerization. The Becker-Döring model describes the time evolution of the distribution of cluster sizes in a system where only additions (coagulation mechanism) or removals (fragmentation) of one monomer from a cluster are possible. A cluster of size 1 is a monomer and clusters of size greater than 2 are polymers. Under Becker-Döring model, coagulation and fragmentation are simple synthesis and decomposition reactions: a polymer of size k may react with a monomer to form a polymer of size $k+1$ at kinetic rate a_k; a polymer of size $k+1$ may break down into a polymer of size k and a monomer at kinetic rate b_{k+1}, i.e.,

$$(1)+(k) \underset{b_{k+1}}{\overset{a_k}{\rightleftharpoons}} (k+1).$$

The ODEs associated to the deterministic version of the Becker-Döring model have been widely studied in physics since 1935, see Becker and Döring [16]. This is an infinite system of ordinary differential equations of $c(t)=(c_k(t), k\in\mathbb{N}^+)$, given by

$$\begin{cases} \dfrac{\mathrm{d}c_1}{\mathrm{d}t}(t)=-2J_1(c(t))-\sum\limits_{k\geq 2} J_k(c(t)), \\ \dfrac{\mathrm{d}c_k}{\mathrm{d}t}(t)=J_{k-1}(c(t))-J_k(c(t)),\ k>1, \end{cases} \tag{BD}$$

with $J_k(c)=a_kc_1c_k-b_{k+1}c_{k+1}$ if $c=(c_k)\in\mathbb{R}_+^{\mathbb{N}^+}$. For $k\geq 1$, $c_k(t)$ represents the concentration of clusters of size k at time t. The conditions on existence/uniqueness of solutions for the Becker-Döring equations (BD) have been extensively investigated. See Ball et al. [15], Niethammer [118] and Penrose [125, 126].

The evolution equations satisfied by $(c(t))$ can be rewritten under a more compact form as,

$$\mathrm{d}c(t)=\tau\circ s(c(t))\,\mathrm{d}t, \tag{6.1}$$

where s is a mapping from $\mathbb{R}_+^{\mathbb{N}^+}\to\mathbb{R}_+^{\mathbb{N}^+}$: for any $k\in\mathbb{N}^+$ and $c\in\mathbb{R}_+^{\mathbb{N}^+}$

$$s_{2k-1}(c)=a_kc_1c_k \quad \text{and} \quad s_{2k}(c)=b_{k+1}c_{k+1}; \tag{6.2}$$

and τ is a linear mapping from $\mathbb{R}_+^{\mathbb{N}^+}\to\mathbb{R}_+^{\mathbb{N}^+}$: for any $z\in\mathbb{R}_+^{\mathbb{N}^+}$ and $k\geq 2$,

$$\begin{cases} \tau_1(z)=-\sum\limits_{i\geq 1}(1+\mathbb{I}_{\{i=1\}})z_{2i-1}+\sum\limits_{i\geq 2}(1+\mathbb{I}_{\{i=2\}})z_{2i-2}, \\ \tau_k(z)=z_{2k-3}-z_{2k-2}-z_{2k-1}+z_{2k}. \end{cases} \tag{6.3}$$

As it will be seen this representation will turn out to be very useful to derive the main results concerning fluctuations.

Becker-Döring ODEs and Polymerization Processes

This set of ODEs is used to describe the evolution of the concentration $c_i(t)$, $i\geq 1$, of polymers of size i. The classical framework assumes an initial state with only

polymers of size 1, monomers. In a biological context, experiments show that the concentration of polymers of size greater than 2 stays at 0 until some instant, defined as the *lag time* , when the polymerized mass grows very quickly to reach its stationary value. With convenient parameters estimations, these ODEs can be used to describe first order characteristics such as mean concentration of polymers of a given size. The use of systems of ODEs to describe the evolution of polymerization processes started with Oosawa's pioneering work in 1962, see Oosawa and Asakura [122] for example. Morris et al. [115] presents a quite detailed review of the classical sets of ODEs used for polymerization processes. As it can be seen and also expected, the basic dynamics of the Becker-Döring model of adding/removing a monomer to/from a polymer occupy a central role in most of these mathematical models. See also Prigent et al. [132] and Hingant and Yvinec [75] for recent developments in this domain.

Outside the rapid growth of polymerized mass at the lag time, the other important aspect observed in the experiments is the high variability of the instant when it occurs, the lag time, from an experiment to another. This is believed to explain, partially, the variability of the starting point of diseases associated to these phenomena in neural cells, like Alzheimer's disease for example. See Xue et al. [168]. Hence if the deterministic Becker-Döring ODEs describes the first order of polymerization through a limiting curve, the fluctuations around these solutions will give a characterization of the variability of the processes itself. Up to now the mathematical studies of these fluctuations are quite scarce, the stochastic models analyzed include generally only a finite number of possible sizes for the polymers. See Szavits et al. [157], Xue et al. [168] and Eugène et al. [58] for example. To study these important aspects, we have to introduce a stochastic version of the Becker-Döring model.

The stochastic Becker-Döring model

The polymerization process is described as a Markov process $(X^N(t)):=(X_k^N(t), 1\leq k\leq N)$, where $X_k^N(t)$ is the total number of clusters of size k at time t, it takes values in the state space,

$$\mathcal{S}^N:=\left\{\vec{x}\in\mathbb{N}^N \,\middle|\, \sum_{k=1}^N kx_k=N\right\},$$

of configurations of polymers with mass N.

For any $\vec{x}\in\mathcal{S}^N$ and $1\leq k<N$, the associated jump matrix $Q^N=(q^N(\cdot,\cdot))$ is given by

$$\begin{cases} q^N(\vec{x},\vec{x}-\vec{e}_1-\vec{e}_k+\vec{e}_{k+1})=a_k x_1(x_k-\mathbb{I}_{\{k=1\}})/N, \\ q^N(\vec{x},\vec{x}+\vec{e}_1+\vec{e}_k-\vec{e}_{k+1})=b_{k+1}x_{k+1}, \end{cases}$$

where $(\vec{e}_k, k\in\mathbb{N}^+)$ is the standard orthonormal basis of $\mathbb{N}^{\mathbb{N}^+}$. In other words, a monomer is added to a given polymer of size k at rate $a_k x_1/N$ and a monomer is detached from a polymer of size k at rate b_k. Note that if N is interpreted as a "volume", the quantity $X_1^N(t)/N$ can be seen as the concentration of monomers.

There are few studies of this important stochastic process. The large scale behavior of the stochastic Becker-Döring model, when N gets large, is an interesting and challenging problem. Given the transition rates, one can expect that the deterministic Becker-Döring

equations (BD) give the limiting equations for the concentration of the different species of polymers, i.e. for the convergence in distribution

$$\lim_{N\to+\infty}(X_k^N(t)/N, k\geq 1)=(c_k(t), k\geq 1).$$

Such a first order analysis is achieved in Jeon [83]. This result is in fact proved for a more general model, the Smoluchowski coagulation-fragmentation model. The Becker-Döring model is a special case of Smoluchowski model, see Aldous [5] for a survey on the coalescence models. The stochastic approximation of the pure Smoluchowski coalescence equation through Marcus-Lushnikov process is investigated in Norris [119]. There are also results on the relation of stochastic Becker-Döring model and the deterministic Lifshitz-Slyozov equation, see Deschamps et al. [45] for example.

The Main Contributions

In this chapter, we investigate the fluctuations of the Becker-Döring model, i.e. the limiting behavior of the $\mathbb{R}^N$-valued process

$$\left(W^N(t)\right) := \left(\frac{1}{\sqrt{N}}\left(X^N(t)-Nc(t)\right)\right) \tag{6.4}$$

is analyzed. We prove that, under appropriate conditions, the fluctuation process $(W^N(t))$ converges for the Skorohod topology to a $L_2(w)$-valued process $(W(t))$ which is the strong solution of the SDE

$$\mathrm{d}W(t)=\tau\Big(\nabla s(c(t))\cdot W(t)\Big)\,\mathrm{d}t+\tau\left(\mathrm{Diag}\left(\sqrt{s(c(t))}\right)\cdot \mathrm{d}\beta(t)\right), \tag{6.5}$$

where

1. $L_2(w)$ is a Hilbert subspace of $\mathbb{R}^{\mathbb{N}^+}$, see definition 6.2 of Section 6.3;
2. The operators $s(\cdot)$ and $\tau(\cdot)$ are defined respectively by relations (6.2) and (6.3);
3. $\nabla s(c)$ is the Jacobian matrix

$$\nabla s(c)=\left(\frac{\partial s_i}{\partial c_j}(c)\right), c\in\mathbb{R}_+^{\mathbb{N}^+}$$

4. $\beta(t)=(\beta_k(t), k\in\mathbb{N}^+)$ is a sequence of independent standard Brownian motions in $\mathbb{R}$;
5. $\mathrm{Diag}(\vec{v})$ is a diagonal matrix whose diagonal entries are the coordinates of the vector $\vec{v}\in\mathbb{R}^{\mathbb{N}^+}$.

See Theorem 6.3 in Section 6.3 for a precise formulation of this result. Note that the drift part of (6.5) is the gradient of the Becker-Döring equation (6.1).

Under appropriate assumptions, see Ball et al. [15], the Becker-Döring equations (6.1) have a unique fixed point $\widetilde{c}$. If $c(0)=\widetilde{c}$, the asymptotic fluctuations around this stationary state are then given by $(\widetilde{W}(t))$. It is shown in Proposition 6.2 that this process, the solution of SDE (6.5), can be represented as

$$\widetilde{W}(t)=\mathcal{T}(t)\widetilde{W}(0)+\int_0^t \mathcal{T}(t-s)\tau\Big(\mathrm{d}\mathcal{B}(s)\Big),$$

where $\mathcal{T}$ is the semi-group associated with linear operator $\tau \circ \nabla s(c)$ and $(\mathcal{B}(t))$ is a $\widetilde{Q}$-Wiener process in $L_2(w)$ where $\widetilde{Q}$=Diag $(w_n s_n(c), n \geq 1)$. In particular if the initial state $\widetilde{W}(0)$ is deterministic, then $(\widetilde{W}(t))$ is a Gaussian process. For the definitions and properties of $\widetilde{Q}$-Wiener process and Gaussian processes in Hilbert spaces, see Section 4.1 and Section 3.6 in Da Prato and Zabczyk [39] for example.

Literature

For the fluctuation problems in the models with finite chemical reactions, results are well known, see Kurtz [99, 100] for example. However, in the Becker-Döring model, there are countable many species and reactions, where the results for the finite reactions are not directly applicable. See the Open Problem 9 in Aldous [5] for the description of fluctuation problem in the general Smoluchowski coagulation model. For this reason, the studies of fluctuations of the Becker-Döring models are quite scarce. Ranjbar and Rezakhanlou [136] investigated the fluctuations at equilibrium of a family of coagulation-fragmentation models with a spatial component. They proved that, at equilibrium, the limiting fluctuations can be described by an Ornstein-Uhlenbeck process in some abstract space. Their results rely on balance equations which hold because of the stationary framework. Durrett et al. [52] gave the stationary distributions for all reversible coagulation-fragmentation processes. Then they provided the limits of the mean values, variances and covariances of the stationary densities of particles of any given sizes when the total mass tends to infinity.

Concerning fluctuations of infinite dimensional Markov processes, several examples have received some attention, in statistical physics mainly. In the classical Vlasov model, the first result on the central limit theorem seems to be given by Braun and Hepp [24] in 1977. They investigated the fluctuations around the trajectory of a test particle. The central limit theorem for the general McKean-Vlasov model when the initial measures of the system are products of i.i.d. measures is proved by Sznitman [160] in 1984. For the Ginzburg-Landau model on $\mathbb{Z}$, where the evolution of the state at a site i depends only on the state of its nearest neighbors $i \pm 1$. The independent (Brownian) stochastic fluctuations at each site do not depend of the state of the process. The hydrodynamic limit, i.e. a first order limit, is given by Guo, Papanicolao and Varadhan [72]. The fluctuations around this hydrodynamic limit live in an infinite dimensional space. Zhu [173] proved that, at equilibrium, the limiting fluctuations converge to a stationary Ornstein-Uhlenbeck process taking values in the dual space of a nuclear space. In the non-equilibrium case, the fluctuations have been investigated by Chang and Yau [33], the limiting fluctuations can be described as an Ornstein-Uhlenbeck in a negative Sobolev space. For more results on related fluctuation problems in statistical mechanics, see Spohn [151, 152]. Fluctuations of an infinite dimensional Markov process associated to load balancing mechanisms in large stochastic networks have been investigated in Graham [71], Budhiraja and Friedlander [27].

One of the difficulties of our model is the fact that the first coordinate $(X_1^N(t))$ of our Markov process, the number of monomers, interact with all non-null coordinates of the state process. Recall that monomers can react with all other kinds of polymers. This feature has implications on the choice of the Hilbert space $L_2(w)$ chosen to formulate the SDE (6.10) and, additionally, several estimates have to be derived to control the stochastic fluctuations of the first coordinate. This situation is different from the examples of the

Ginzburg-Landau model in [33, 151, 173], since each site only have interactions with a finite number of sites, or in the stochastic network example [27] where the interaction range is also finite. It should be also noted that our evolution equations are driven by a set of independent Poisson processes whose intensity is state dependent which is not the case in the Ginzburg-Landau models for which the diffusion coefficients are constant. For this reason Lipschitz properties have to be established for an appropriate norm for several functionals, it complicates the already quite technical framework of these problems.

Outline of the chapter

Section 6.2 introduces the notations, assumptions and the stochastic model as well as the evolution equations. Section 6.3 investigates the problem of existence and uniqueness of the solution of the SDE (6.5). The proof of the fluctuations at equilibrium is also given. Section 6.4 gives the proof of the main result, Theorem 6.3, the convergence of the fluctuation processes.

6.2 The Stochastic Model

In this section we introduce the notations and assumptions used throughout this chapter. The stochastic differential equations describing the evolution of the model are introduced.

For any $h\in\mathbb{R}^+$, $(\mathcal{N}_h^i)_{i=1}^\infty$ denotes a sequence of independent Poisson processes with intensity h. The stochastic Becker-Döring equation with aggregation rates $(a_k, k\in\mathbb{N}^+)$ and fragmentation rates $(b_k, k\in\mathbb{N}^+)$ can be expressed as the solution of the SDEs

$$\begin{aligned}
\mathrm{d}X_1^N(t) &= -\sum_{k\geq 1}(1+\mathbb{I}_{\{k=1\}})\sum_{i=1}^{X_1^N(t-)(X_k^N(t-)-\mathbb{I}_{\{k=1\}})}\mathcal{N}^i_{a_k/N}(\mathrm{d}t)+\sum_{k\geq 2}(1+\mathbb{I}_{\{k=2\}})\sum_{i=1}^{X_k^N(t-)}\mathcal{N}^i_{b_k}(\mathrm{d}t),\\
\mathrm{d}X_k^N(t) &= \sum_{i=1}^{X_1^N(t-)(X_{k-1}^N(t-)-\mathbb{I}_{\{k=2\}})}\mathcal{N}^i_{a_{k-1}/N}(\mathrm{d}t)-\sum_{i=1}^{X_k^N(t-)}\mathcal{N}^i_{b_k}(\mathrm{d}t)\\
&\quad -\sum_{i=1}^{X_1^N(t-)X_k^N(t-)}\mathcal{N}^i_{a_k/N}(\mathrm{d}t)+\sum_{i=1}^{X_{k+1}^N(t-)}\mathcal{N}^i_{b_{k+1}}(\mathrm{d}t),
\end{aligned}$$

and for all $k>N$, $X_k^N(t)\equiv 0$, with $f(t-)$ being the limit on the left of the function f at $t>0$.

In order to separate the drift part and the martingale part, it is convenient to introduce the corresponding martingales $(D^N(t))$. For $k\geq 1$, let

$$\begin{aligned}
D_{2k-1}^N(t) &= \int_0^t\left(\sum_{i=1}^{X_1^N(u-)(X_k^N(u-)-\mathbb{I}_{\{k=1\}})}\mathcal{N}^i_{a_k/N}(\mathrm{d}u)-\frac{1}{N}a_kX_1^N(u)(X_k^N(u)-\mathbb{I}_{\{k=1\}})\,\mathrm{d}u\right),\\
D_{2k}^N(t) &= \int_0^t\left(\sum_{i=1}^{X_{k+1}^N(u-)}\mathcal{N}^i_{b_{k+1}}(\mathrm{d}u)-b_{k+1}X_{k+1}^N(u)\,\mathrm{d}u\right).
\end{aligned}$$

Clearly, for every fixed $i \in \mathbb{N}^+$, $(D_i^N(t))$ is local martingale in $\mathbb{R}$ with previsible increasing process

$$\left(\langle D_i^N \rangle(t)\right) = \left(N \int_0^t s_i\left(\frac{X^N(u)}{N}\right) \mathrm{d}u - \mathbb{I}_{\{i=1\}} \int_0^t a_1 \frac{X_1^N(u)}{N} \mathrm{d}u\right), \tag{6.6}$$

where the operator $s = (s_i, i \geq 1)$ is defined by relation (6.2), additionally the cross-variation processes are null, i.e., for any $i \neq j$,

$$\left(\langle D_i^N, D_j^N \rangle(t)\right) = (0).$$

This is in fact one of the motivations to introduce the variables $(D_k^N(t))$ and the functionals $(\tau(\cdot))$ and $(s(\cdot))$.

A simple calculation shows that, for any N, the process $(X^N(t))$ satisfies the relation

$$X^N(t) = X^N(0) + N \int_0^t \tau\left(s\left(\frac{X^N(u)}{N}\right)\right) \mathrm{d}u + \tau\left(D^N(t)\right) + 2a_1 \int_0^t \left(\frac{X_1^N(u)}{N}\right) e_1 \, \mathrm{d}u, \tag{6.7}$$

where τ is defined by relation (6.3) and $e_1 = (1, 0, 0, \ldots)$. Therefore the fluctuation process $(W^N(t))$, see relation (6.4), satisfies

$$\begin{aligned} W^N(t) = W^N(0) &+ \frac{1}{\sqrt{N}} \tau\left(D^N(t)\right) + \frac{2a_1}{\sqrt{N}} \int_0^t \left(\frac{X_1^N(u)}{N}\right) e_1 \, \mathrm{d}u \\ &+ \frac{1}{2} \int_0^t \tau\left(\nabla s\left(\frac{X^N(u)}{N}\right) \cdot W^N(u) + \nabla s\left(c(u)\right) \cdot W^N(u)\right) \mathrm{d}u. \end{aligned} \tag{6.8}$$

If we let N go to infinity, then $(X^N(t)/N)$ converges to $c(t)$ and provides that, for all $i \geq 1$,

$$\lim_{N \to +\infty} \left(\langle D_i^N / \sqrt{N} \rangle(t)\right) = \int_0^t s_i(c(u)) \, \mathrm{d}u.$$

We can expect that the process $(D^N(t)/\sqrt{N})$ converges to a stochastic integral

$$\int_0^t \mathrm{Diag}\left(\sqrt{s(c(u))}\right) \cdot \mathrm{d}\beta(u),$$

with $(\beta(t))$ defined in relation (6.5) in the introduction. See page 339 in Ethier and Kurtz's [57] for example. Then, formally, the limiting process for $(W^N(t))$, provided that it exists, would be the solution of the SDE (6.5).

In order to give the well-posedness of the infinite dimensional process (6.5), we introduce the following notations.

Definition 6.1. *Let $\mathcal{X}_+$ be the phase space of the Becker-Döring model with initial density less than 1,*

$$\mathcal{X}_+ := \left\{ c = (c_k, k \in \mathbb{N}^+) \,\middle|\, \forall k,\ c_k \geq 0;\ \sum_{k \geq 1} k c_k \leq 1 \right\}.$$

For $c(0) \in \mathcal{X}_+$, the solution of the Becker-Döring equation (BD) is a continuous function taking values in $\mathcal{X}_+$.

Definition 6.2. *One assumes that $w=(w_n)$ is a fixed non-decreasing sequence of positive real numbers such that*
— $\|1/w\|_{l_1}:=\sum_{n\geq 1} 1/w_n<\infty$;
— $\lim_{n\to\infty} w_n^{1/n}\leq 1$;
— *there exists a constant γ_0 such that for all $n\in\mathbb{N}^+$, $w_{2n}\leq\gamma_0 w_n$.*
One denotes by $L_2(w)$ the associated L_2-space,

$$L_2(w):=\left\{z\in\mathbb{R}^{\mathbb{N}^+}\,\middle|\,\sum_{n=1}^{\infty} w_n z_n^2<\infty\right\},$$

its inner product is defined by, for z, $z'\in L_2(w)$

$$\langle z,z'\rangle_{L_2(w)}=\sum_{n=1}^{\infty} w_n z_n z_n',$$

and its associated norm is $\|z\|_{L_2(w)}=\sqrt{\langle z,z\rangle_{L_2(w)}}$.

An orthonormal basis $(\vec{h}_n)_{n=1}^{\infty}$ of $L_2(w)$ is defined as, for $n\geq 1$,

$$\vec{h}_n=(0,\dots,1/\sqrt{w_n},0,\dots). \tag{6.9}$$

As it will be seen in Section 6.3, $L_2(w)$ is a convenient space to ensure a boundedness property of the linear mapping $\tau\circ\nabla s(\cdot)$, which is essential for the study of fluctuation process (6.5).

We now turn to the conditions on the rates and the initial state of the Becker-Döring equations.

Assumptions 6.1.

(a) The kinetic rates (a_k), (b_k) are positive and bounded, i.e.,

$$\Lambda:=\max_{k\geq 1}\{a_k,b_k\}<\infty.$$

(b) There exists a positive increasing sequence $r\in\mathbb{R}_+^{\mathbb{N}^+}$, such that
— *$r_k\geq w_k$, for all $k\geq 1$;*
— *there exists a fixed constant γ_r, such that $r_{2k}\leq\gamma_r r_k$, for all $k\geq 1$;*
— *(w_k/r_k) is a decreasing sequence converging to 0,*
where (w_n) is the sequence introduced in Definition 6.2.
The initial state of the process $(X_k^N(0))$ is assumed to satisfy

$$\sup_N \mathbb{E}\left(\sum_{k\geq 1}\frac{r_k X_k^N(0)}{N}\right)<\infty.$$

(c) The initial state of the first order process, $c(0)=(c_k(0))\in\mathcal{X}_+$, is such that

$$\sum_{k\geq 1} w_k c_k(0)<\infty \text{ and } \lim_{N\to\infty}\mathbb{E}\left(\sum_{k\geq 1}\left|\frac{X_k^N(0)}{N}-c_k(0)\right|\right)=0$$

hold.

(d) *The initial state of the centered process, the random variables* $W^N(0)$, $N\geq 1$, *defined by relation* (6.4) *satisfies the relation*

$$C_0 := \sup_N \mathbb{E}\left(\|W^N(0)\|^2_{L_2(w)}\right) < +\infty,$$

and the sequence $(W^N(0))$ *converges in probability to an* $L_2(w)$*-valued random variable* $W(0)$.

For a detailed discussion of Assumption (a), see Section 6.3.1 below. Assumption (b) gives conditions on the moments of the initial state of the process. For example, by taking $r_k = k^\beta$ with $\beta > 1$, one can study the central limit problem in the Hilbert space $L_2(w)$ with weights $w_k = k^\alpha$ for some $\alpha \in (1, \beta)$. Note that this condition is more, but not much more, demanding than the conservation of mass relation

$$\sum_{k\geq 1} k X_k^N(t)/N \equiv 1,$$

which is always satisfied. This assumption gives therefore a quite large class of initial distributions for which a central limit theorem holds.

Some Notations. For any $T>0$, let $\mathcal{D}_T := D([0,T], L_2(w))$ be the space of càdlàg functions on $[0,T]$ taking values in $L_2(w)$. Since $L_2(w)$ is a separable and complete space, there exists a metric on $\mathcal{D}_T$, such that $\mathcal{D}_T$ is a separable and complete space. See the chapter 3 in Billingsley [19] for details.

Let $\mathcal{L}(L_2(w))$ be the set of linear operators on $L_2(w)$ and the $\|\cdot\|_{\mathcal{L}(L_2(w))}$ be the associated norm for linear operators, *i.e.* for any $f \in \mathcal{L}(L_2(w))$,

$$\|f\|_{\mathcal{L}(L_2(w))} = \sup_{\substack{z\in L_2(w)\\ \|z\|_{L_2(w)}\leq 1}} \|f(z)\|_{L_2(w)}.$$

For any $f, g \in \mathcal{L}(L_2(w))$, $f \circ g$ denotes the composition. We call $A \in \mathcal{L}(L_2(w))$ to be in a trace class if there exist a constant C, such that for all $z \in L_2(w)$, $\|Az\|_{L_2(w)} \leq C\|z\|_{L_2(w)}$ and it has a finite trace, i.e.,

$$\mathrm{Tr}A := \sum_{n\geq 1} \langle Ah_n, h_n\rangle_{L_2(w)} < \infty.$$

6.3 SDE in Hilbert space

In this section, we are going study the existence and uniqueness of process

$$W(t) = W(0) + \int_0^t \tau\Big(\nabla s(c(u)) \cdot W(u)\Big)\,\mathrm{d}u + \int_0^t \tau\Big(\mathrm{Diag}\left(\sqrt{s(c(u))}\right) \cdot \mathrm{d}\beta(u)\Big), \quad (6.10)$$

when Assumptions (a) and (c) hold.

Clearly, the process $(W(t))$ has the form of a stochastic differential equation in the infinite dimensional space $\mathbb{R}^{\mathbb{N}^+}$. It has a stochastic integral

$$\mathcal{M}(t):=\int_0^t \tau\left(\mathrm{Diag}\left(\sqrt{s(c(u))}\right)\cdot \mathrm{d}\beta(u)\right)$$

with respect to $\beta(t)=\sum_{n\geq 1}\beta_n(t)\vec{e}_n$, which is a cylindrical Wiener process (c.f. Yor [169]) in the Hilbert space $L^2(\mathbb{R}^{\mathbb{N}^+})$. Here $(c(t))$ is the unique solution of Becker-Döring equation (6.1) with initial state $c(0)$. However, the mapping τ is unbounded in $L^2(\mathbb{R}^{\mathbb{N}^+})$. Therefore, we consider the SDE (6.10) in the Hilbert space $L_2(w)$ (Definition 6.2). In the following, we are going to show that

1. for any $c\in\mathcal{X}_+$, the linear mapping $\tau\circ\nabla s(c)\ :\ L_2(w)\to L_2(w)$ is bounded (and therefore Lipschitz);
2. the stochastic process $(\mathcal{M}(t))$ is well-defined in $L_2(w)$ and is a martingale.

Finally, by using the results in Section 7.1 of Da Prato and Zabczyk [39], we give the proof of the existence and uniqueness of the solutions of the equation (6.10).

Lemma 6.1. *τ is a continuous linear mapping from $L_2(w)$ to $L_2(w)$: there exists a finite constant $\gamma_\tau(w)$, such that for any $z\in L_2(w)$,*

$$\|\tau(z)\|_{L_2(w)}\leq\gamma_\tau(w)\|z\|_{L_2(w)}.$$

Proof. One only needs to verify that τ is bounded. For any $z\in L_2(w)$,

$$\|\tau(z)\|^2_{L_2(w)}=R_1+R_2,$$

where

$$R_1=w_1\left(-\sum_{k\geq 1}(1+\mathbb{I}_{\{k=1\}})z_{2k-1}+\sum_{k\geq 2}(1+\mathbb{I}_{\{k=2\}})z_{2k-2}\right)^2,$$
$$R_2=\sum_{k\geq 2}w_k\left(z_{2k-3}-z_{2k-2}-z_{2k-1}+z_{2k}\right)^2.$$

By using Cauchy-Schwarz inequality, we have

$$R_1\leq 4w_1\sum_{k\geq 1}\frac{1}{w_k}\|z\|^2_{L_2(w)}=4w_1\left\|\frac{1}{w}\right\|_{l_1}\|z\|^2_{L_2(w)},$$

and by using the increasing property of weights w

$$\begin{aligned}R_2&\leq 4\sum_{k\geq 2}w_k\left(z^2_{2k-3}+z^2_{2k-2}+z^2_{2k-1}+z^2_{2k}\right)\\&\leq 4\sum_{k\geq 2}\left(w_{2k-3}z^2_{2k-3}+w_{2k-2}z^2_{2k-2}+w_{2k-1}z^2_{2k-1}+w_{2k}z^2_{2k}\right)+w_2z_1^2\\&\leq(8+w_2/w_1)\|z\|^2_{L_2(w)}.\end{aligned}$$

In conclusion, one has $\|\tau(z)\|_{L_2(w)}\leq\gamma_\tau(w)\|z\|_{L_2(w)}$, for

$$\gamma_\tau(w)=\left(4w_1\left\|\frac{1}{w}\right\|_{l_1}+8+\frac{w_2}{w_1}\right)^{1/2}.$$

□

Lemma 6.2. *Under Assumption (a), for any $c \in \mathcal{X}_+$ that satisfies $\sum_{k\geq 1} w_k c_k < \infty$, there exists a finite constant $\gamma(c,w)$, such that for any $x \in L_2(w)$,*

$$\|\nabla s(c)\cdot x\|_{L_2(w)} \leq \gamma(c,w)\|x\|_{L_2(w)}.$$

Moreover, under Assumption (a) and (c), if $(c(t))$ is the solution of the Becker-Döring equation (6.1) with initial state $c(0)$, then for any finite time T, one has

$$\sup_{u\leq T} \gamma(c(u),w) < +\infty.$$

Proof.

$$\begin{aligned}\|\nabla s(c)\cdot x\|^2_{L_2(w)} &= \sum_{k\geq 1} w_{2k-1} a_k^2 (x_1 c_k + c_1 x_k)^2 + \sum_{k\geq 2} w_{2k-2} b_k^2 x_k^2 \\ &\leq 2\gamma_0\Lambda^2 \left(\sum_{k\geq 1} w_k c_k^2\right) x_1^2 + 2\gamma_0\Lambda^2 \sum_{k\geq 1} w_k x_k^2 + \gamma_0\Lambda^2 \sum_{k\geq 2} w_k x_k^2 \leq \gamma(c,w)\|x\|^2_{L_2(w)},\end{aligned}$$

where for any $c \in \mathcal{X}_+$

$$\gamma(c,w) = \gamma_0\Lambda^2 \left(3 + 2\sum_{k\geq 1} w_k c_k\right).$$

The Theorem 2.2 in Ball et al [16] gives that $\sup_{u\leq T} w_k c_k(u) < \infty$ under Assumption (a) and (c). □

6.3.1 Remark on the assumptions on the coefficients (a_i) and (b_i)

We recall the main results for the existence and uniqueness of a solution of the deterministic Becker-Döring ODEs.

1. From Theorem 2.2 in Ball et al. [15], existence and uniqueness hold under the condition

$$a_i(g_{i+1} - g_i) = O(g_i) \qquad \text{and} \qquad \sum_{i=1}^{+\infty} g_i c_i(0) < +\infty,$$

   for a positive increasing sequence g satisfying $\min_i (g_{i+1} - g_i) \geq \delta > 0$. For example, when $g_i = i^2$, *i.e.*, $a_i = O(i)$ and the second moment of the initial state is bounded, the Becker-Döring equation is well-posed.

2. Conditions from Theorem 2.1 in Laurençot and Mischler [103],
   There exists some constant $K > 0$ such that

$$a_i - a_{i-1} < K \text{ and } b_i - b_{i-1} < K,$$

   for any $i \in \mathbb{N}$.

It should be noted that Laurençot and Mischler [103] have a stronger conditions on the coefficients but, contrary to Ball et al. [15], not any on the initial state.

The known conditions necessary to get the convergence of the first order of the stochastic Becker-Döring model to the solution of the deterministic Becker-Döring ODEs are more demanding, they are given by Theorem 2 in Jeon [83], they are

$$\lim_{i\to\infty} \frac{a_i}{i} = 0 \qquad \text{and} \qquad \lim_{i\to\infty} b_i = 0.$$

Note that law of large number in this chapter requires growth rates (a_i) to be sublinear and break rates (b_i) to be vanishing.

For second order convergence, additional conditions seem however to be necessary to establish the well-posedness of the limiting fluctuation process defined by SDE (6.5). First of all, for any $c \in \mathcal{X}_+$, the linear operator $\tau \circ \nabla s(c)$ does seem to have monotonicity or symmetry properties that could give an alternative construction of the solution of SDE (6.5). Hence, boundedness properties of the linear operator $\tau \circ \nabla s(c)$ have to be used to get existence and uniqueness results of the solution of SDE (6.5).

For any $k \geq 2$, it is easy to check that the kth coordinate of the vector $\tau \circ \nabla s(c) \cdot h_k$ is $-(b_k + a_k c_1)/\sqrt{w_k}$, in particular the relation

$$\|\tau \circ \nabla s(c)\|_{\mathcal{L}(L_2(w))} \geq \|\tau \circ \nabla s(c) \cdot h_k\|_{L_2(w)} \geq (b_k + a_k c_1)^2 \qquad \text{for all } k,$$

holds. Hence the operator $\tau \circ \nabla s(c)$ is unbounded in any weighted L_2 space if a boundedness property for the sequences (a_i) and (b_i) does not hold. By using similar arguments, we observe that the operator $\tau \circ \nabla s(c)$ is unbounded in the state spaces l_1 and l_∞ as well if this property does not hold.

Proposition 6.1. *If Assumption (a) and (c) hold, then the process $(\mathcal{M}(t))$ is a well-defined continuous, square-integrable, martingale.*

Proof. By definition of the stochastic integral with respect to a cylindrical Wiener process (Chapter 4 in Da Prato and Zabczyk [39]), it is sufficient to verify that,

$$\int_0^T \sum_{n \geq 1} \|\tau \circ \mathrm{Diag}\left(\sqrt{s(c(u))}\right) \cdot e_n\|^2_{L_2(w)} \, \mathrm{d}u < \infty.$$

By using Lemma 6.1,

$$\|\tau \circ \mathrm{Diag}\left(\sqrt{s(c(u))}\right) \cdot e_n\|^2_{L_2(w)} \leq \gamma_\tau(w)^2 \|\mathrm{Diag}\left(\sqrt{s(c(u))}\right) \cdot e_n\|^2_{L_2(w)} = \gamma_\tau(w)^2 w_n s_n(c(u)).$$

Therefore,

$$\begin{aligned}
&\int_0^T \sum_{n \geq 1} \|\tau \circ \mathrm{Diag}\left(\sqrt{s(c(u))}\right) \cdot e_n\|^2_{L_2(w)} \, \mathrm{d}u \\
&\quad \leq \gamma_\tau(w)^2 \int_0^T \sum_{k \geq 1} (w_{2k-1} a_k c_1(u) c_k(u) + w_{2k} b_{k+1} c_{k+1}(u)) \, \mathrm{d}u \\
&\quad \leq \gamma_0 \gamma_\tau(w)^2 \Lambda T \sup_{s \leq T} \sum_{k \geq 1} w_k c_k(s) < \infty.
\end{aligned}$$

The last inequality is valid under Assumption (a) and (c). See Ball et al. [16] for details. □

Theorem 6.1. *For any measurable $L_2(w)$-valued random variable $W(0)$, the SDE (6.10) has a unique strong solution in $L_2(w)$. If, in addition, for $p \geq 1$, $\mathbb{E}\|W(0)\|^{2p}_{L_2(w)} < \infty$, then*

$$\mathbb{E}\left(\sup_{0 \leq t \leq T} \|W(s)\|^{2p}_{L_2(w)}\right) < \infty.$$

Proof. In Proposition 6.1, we proved that the martingale part $\mathcal{M}(t)$ of the SDE (6.10) is well-defined and continuous. By using Lemma 6.1 and Lemma 6.2, the drift part is Lipschitz and linear on any $x \in C([0,T], L_2(w))$:

$$\|\tau\left(\nabla s(c(u)) \cdot x(u)\right)\|_{L_2(w)} \leq \gamma_\tau(w) \sup_{0\leq u\leq T} \gamma(c(u), w) \sup_{0\leq u\leq T} \|x(u)\|_{L_2(w)}.$$

Therefore, by using the results in Section 7.1 of Da Prato and Zabczyk [39], we can obtain the strongly existence, uniqueness, continuous and bounded results of SDE (6.10). □

In Ball et al. [15] it is shown that, under the assumptions

$$z_s^{-1} := \limsup_{i\to\infty} R_i^{1/i} < \infty \qquad \text{and} \quad \sup_{0\leq z<z_s} \sum_{i\geq 1} iR_i z^i > 1, \tag{6.11}$$

where, for $k\geq 1$, $R_k = \prod_{i=2}^{i}(a_{i-1}/b_i)$, then the equation

$$\sum_{k\geq 1} kR_k z^k = 1$$

has a unique solution $z=\widetilde{c}_1$. Moreover if $\widetilde{c}_k = R_k(\widetilde{c}_1)^i$, then $(\widetilde{c}_i)_{i\geq 1}$ is the unique fixed point of Becker-Döring equations (BD).

Proposition 6.2 (Fluctuations at equilibrium)**.** *Under condition* (6.11)*, at equilibrium, the strong solution of SDE* (6.10) *can be represented as*

$$\widetilde{W}(t) = \mathcal{T}(t)\widetilde{W}(0) + \int_0^t \mathcal{T}(t-s)\tau\Big(\mathrm{d}\mathcal{B}(s)\Big), \tag{6.12}$$

where $(\mathcal{T}(t))$ *is the semi-group associated with linear operator* $\tau\circ\nabla s(\widetilde{c})$*,*

$$\mathcal{T}(t) := e^{\tau\circ\nabla s(\widetilde{c})t},$$

and $(\mathcal{B}(t))$ *is a* $\widetilde{Q}$*-Wiener process in* $L_2(w)$ *where*

$$\widetilde{Q} = \mathrm{Diag}\left(w_n s_n(\widetilde{c}), n\geq 1\right).$$

Moreover, the stochastic convolution part

$$\widetilde{W}_{sc}(t) := \int_0^t \mathcal{T}(t-s)\tau\,\mathrm{d}\mathcal{B}(s)$$

is Gaussian, continuous in mean square, has a predictable version and

$$\left(\left\langle \widetilde{W}_{sc}(t)\right\rangle\right) = \left(\int_0^t \mathcal{T}(r)\tau\widetilde{Q}\left(\mathcal{T}(r)\tau\right)^*\mathrm{d}r\right).$$

Proof. At equilibrium, the SDE (6.10) becomes

$$\widetilde{W}(t) = \widetilde{W}(0) + \int_0^t \tau\left(\nabla s(\widetilde{c})\cdot\widetilde{W}(u)\right)\mathrm{d}u + \int_0^t \tau\left(\mathrm{Diag}\left(\sqrt{s(\widetilde{c})}\right)\cdot\mathrm{d}\beta(u)\right). \tag{6.13}$$

It is a linear equation with additive noise. The noise part can be expressed by $\tau(\mathcal{B}(t))$ where

$$\mathcal{B}(t):=\int_0^t \mathrm{Diag}\left(\sqrt{s(\widetilde{c})}\right)\cdot \mathrm{d}\beta(u)=\sum_{n\geq 1}\sqrt{w_n s_n(\widetilde{c})}\beta_n(t)h_n.$$

Recall that (h_n) defined by (6.9) is an orthonormal basis in $L_2(w)$. By definition of z_s and the assumption $\lim_{n\to\infty} w_n^{1/n}=1$, if $0\leq z<z_s$, then

$$\sum_{n\geq 1} w_n R_n z^n<\infty$$

holds. Let

$$\widetilde{Q}=\mathrm{Diag}\left(w_n s_n(\widetilde{c}), n\geq 1\right),$$

then it is trace class, i.e.,

$$\mathrm{Tr}\widetilde{Q}=\sum_{n\geq 1} w_n s_n(\widetilde{c})\leq 2\gamma_0\Lambda\sum_{n\geq 1} w_n\widetilde{c}_n<\infty.$$

Therefore, $\mathcal{B}(t)$ is a $\widetilde{Q}$-Wiener process in $L_2(w)$ (c.f. Section 4.1 in [39]).

Again, by using Lemma 6.1 and Lemma 6.2, we can see that the linear operator $\tau\circ\nabla s(\widetilde{c})$ is bounded, and therefore, the associated semi-group $\mathcal{T}$ is uniformly continuous and satisfies that for any $z\in L_2(w)$,

$$\|\mathcal{T}(t)z\|_{L_2(w)}\leq e^{\gamma_\tau(w)\gamma(\widetilde{c},w)t}\|z\|_{L_2(w)}.$$

Therefore, on any finite time interval $[0,T]$,

$$\int_0^t \mathrm{Tr}\left(\mathcal{T}(r)\tau\widetilde{Q}\tau^*\mathcal{T}^*(r)\right)\mathrm{d}r=\int_0^t \|\mathcal{T}(r)\tau\widetilde{Q}^{1/2}h_n\|^2_{L_2(w)}\,\mathrm{d}r$$
$$\leq\gamma_\tau(w)^2 Te^{2\gamma_\tau(w)\gamma(\widetilde{c},w)T}\mathrm{Tr}\widetilde{Q}<\infty.$$

By applying the Theorem 5.4 in Da Prato and Zabczyk [39], the process (6.12) is the unique weak solution of SDE (6.13). By Theorem 5.2 in [39], the stochastic convolution part $(\widetilde{W}_{sc}(t))$ is Gaussian, continuous in mean square, has a predictable version and for all $t\in[0,T]$,

$$\left\langle\widetilde{W}_{sc}(t)\right\rangle=\int_0^t \mathcal{T}(r)\tau\widetilde{Q}\tau^*\mathcal{T}^*(r)\,\mathrm{d}r.$$

To show that the process defined by relation (6.12) is also a strong solution, we use Theorem 5.29 in [39]. It is sufficient to check

$$\sum_{n\geq 1}\|\tau\nabla s(\widetilde{c})\tau\widetilde{Q}^{1/2}h_n\|^2_{L_2(w)}<\infty.$$

Then, stochastic integral $\tau\circ\nabla s(\widetilde{c})\cdot\widetilde{W}_{sc}(t)$ is a well defined continuous square-integrable martingale on $[0,T]$. The process $\tau\circ\nabla s(\widetilde{c})\cdot\widetilde{W}(t)$ has square integrable trajectories. Therefore, $(\widetilde{W}(t))$ is a strong solution. The proposition is proved. □

6.4 Convergence of the Fluctuation Processes

Recall that the fluctuation processes $(W^N(t))$ satisfies relation (6.8),

$$W^N(t)=W^N(0)+\frac{1}{\sqrt{N}}\tau\left(D^N(t)\right)+\frac{2a_1}{\sqrt{N}}\int_0^t\left(\frac{X_1^N(u)}{N}\right)e_1\,\mathrm{d}u$$
$$+\frac{1}{2}\int_0^t\tau\left(\nabla s\left(\frac{X^N(u)}{N}\right)\cdot W^N(u)+\nabla s\left(c(u)\right)\cdot W^N(u)\right)\mathrm{d}u.$$

The goal of this section is to prove that, when N is going to infinity, the process $(W^N(t))$ is converging in distribution to $(W(t))$, the solution of the SDE (6.10). We will first prove some technical lemmas and the convergence of scaled process $(X^N(t)/N)$. Then, we will prove the tightness and convergence of the local martingales $(D^N(t))$ and, with the help of these results, we will get the tightness of $(W^N(t))$ and identify the limit.

Lemma 6.3. *Under Assumptions (a) and (b) then, for any $T>0$, one has*

$$\zeta_T:=\sup_N\sup_{t\le T}\mathbb{E}\left(\sum_{k\ge1}\frac{r_kX_k^N(t)}{N}\right)<+\infty, \tag{6.14}$$

$$\kappa_T:=\sup_N\mathbb{E}\left(\sup_{t\le T}\sum_{k\ge1}\frac{\left|\mathcal{M}_k^N(t)\right|}{\sqrt{N}}\right)<+\infty, \tag{6.15}$$

where $\mathcal{M}^N(t)=\tau(D^N(t))$.

Proof. By using the SDE (6.7), we get that, for any N,

$$\mathbb{E}\left(\sum_{k\ge1}\frac{r_kX_k^N(t)}{N}\right)=\mathbb{E}\left(\sum_{k\ge1}\frac{r_kX_k^N(0)}{N}\right)$$
$$+\mathbb{E}\int_0^t r_2\left(a_1\frac{X_1^N(u)(X_1^N(u)-1)}{N^2}-b_2\frac{X_2^N(u)}{N}\right)\mathrm{d}u$$
$$+r_1\mathbb{E}\int_0^t\left(-\sum_{i\ge1}(1+\mathbb{I}_{\{i=1\}})a_i\frac{X_1^N(u)(X_i^N(u)-\mathbb{I}_{\{i=1\}})}{N^2}+\sum_{i\ge2}(1+\mathbb{I}_{\{i=2\}})b_i\frac{X_i^N(u)}{N}\right)\mathrm{d}u$$
$$+\mathbb{E}\int_0^t\left(\frac{X_1^N(u)}{N}\sum_{i\ge2}(r_{i+1}-r_i)a_i\frac{X_i^N(u)}{N}+\sum_{i\ge3}(r_{i-1}-r_i)b_i\frac{X_i^N(u)}{N}\right)\mathrm{d}u$$
$$\le\mathbb{E}\left(\sum_{k\ge1}\frac{r_kX_k^N(0)}{N}\right)+(r_2+r_1)\,\Lambda t+\gamma_r\Lambda\mathbb{E}\int_0^t\sum_{i\ge1}r_i\frac{X_i^N(u)}{N}\,\mathrm{d}u,$$

where Λ is defined in Assumption (a). We apply Gronwall's inequality, then, for any $t\le T$,

$$\mathbb{E}\left(\sum_{k\ge1}\frac{r_kX_k^N(t)}{N}\right)\le e^{\gamma_r\Lambda T}\left(\mathbb{E}\left(\sum_{k\ge1}\frac{r_kX_k^N(0)}{N}\right)+(r_2+r_1)\Lambda T\right).$$

Therefore, there exists a constant ζ_T, such that the relation (6.14) holds.

For the other inequality, we first see

$$\begin{aligned}
&\mathbb{E}\left(\sup_{t\le T}\sum_{k\ge1}\frac{\left|\mathcal{M}_k^N(t)\right|}{\sqrt{N}}\right)\\
&\le\sum_{k\ge1}\frac{1}{w_k}+\mathbb{E}\left(\sup_{t\le T}\sum_{k\ge1}\frac{\left|\mathcal{M}_k^N(t)\right|}{\sqrt{N}}\mathbb{I}_{\left\{\frac{\left|\mathcal{M}_k^N(t)\right|}{\sqrt{N}}\ge\frac{1}{w_k}\right\}}\right)\\
&\le2+\mathbb{E}\left(\sup_{t\le T}\sum_{k\ge1}\frac{w_k\mathcal{M}_k^N(t)^2}{N}\right)\le2+\mathbb{E}\left(\sum_{k\ge1}\frac{w_k\langle\mathcal{M}_k^N(T)\rangle}{N}\right).
\end{aligned}$$

By using the expression of the increasing process (6.6) and the definition of τ, we get

$$\begin{aligned}
\sum_{k\ge2}\frac{w_k}{N}\left\langle\mathcal{M}_k^N(T)\right\rangle\le&\Lambda\int_0^T\sum_{k\ge1}(w_{k+1}\mathbb{I}_{\{k\ge2\}}+w_k)\frac{X_k^N(u)}{N}\,\mathrm{d}u\\
&+\Lambda\int_0^T\sum_{k\ge2}(w_k+w_{k-1}\mathbb{I}_{\{k\ge3\}})\frac{X_k^N(u)}{N}\,\mathrm{d}u.
\end{aligned}$$

It implies

$$\begin{aligned}
\sup_N\mathbb{E}\left(\sum_{k\ge2}\frac{w_k}{N}\left\langle\mathcal{M}_k^N(T)\right\rangle\right)\le&w_1\Lambda T+(3+\gamma_0)\Lambda T\sup_N\sup_{t\le T}\mathbb{E}\sum_{k\ge2}\frac{w_k}{N}X_k^N(t)\\
\le&w_1\Lambda T+(3+\gamma_0)\Lambda T\zeta_T.
\end{aligned}$$

For $k=1$,

$$\left\langle\frac{\mathcal{M}_1^N(T)}{N}\right\rangle\le2\int_0^T\sum_{i\ge1}\left(a_i\frac{X_1^N(u)X_i^N(u)}{N^2}+b_{i+1}\frac{X_{i+1}^N(u)}{N}\right)\mathrm{d}u\le4\Lambda T.$$

Therefore, there exists a positive constant κ_T, such that

$$\sup_N\mathbb{E}\left(\sup_{t\le T}\sum_{k\ge1}\frac{\left|\mathcal{M}_k^N(t)\right|}{\sqrt{N}}\right)\le\kappa_T$$

holds.

□

Now we prove the law of large numbers under Assumptions (a), (b) and (c) in the L_1-norm. This result will be used in the proof of Theorem 6.3. We should note that we could not directly apply the Theorem 2 of Jeon [83] since it requires the conditions $\lim_{i\to\infty}a_i/i=0$ and $\lim_{i\to\infty}b_i=0$.

Theorem 6.2. (Law of large numbers) *Under Assumptions* (a), (b) *and* (c), *for any* $T>0$, *one has*

$$\lim_{N\to\infty}\mathbb{E}\left(\sup_{t\le T}\sum_{k\ge1}\left|\frac{X_k^N(t)}{N}-c_k(t)\right|\right)=0,$$

where $(c(t))$ *is the unique solution of Becker-Döring equation* (BD) *with initial state* $c(0)\in\mathcal{X}_+$.

Proof. From SDE (6.7), we have that

$$\frac{X^N(t)}{N}-c(t)=\frac{X^N(0)}{N}-c(0)+\int_0^t \tau\left(s\left(\frac{X^N(u)}{N}\right)-s\left(c(u)\right)\right)\mathrm{d}u \\ +\frac{2a_1}{N}\int_0^t\left(\frac{X_1^N(u)}{N}\right)e_1\,\mathrm{d}u+\frac{\mathcal{M}^N(t)}{N}.$$

With a simple calculation, we get the relation

$$\sum_{k\geq 1}\left|\frac{X_k^N(t)}{N}-c_k(t)\right|\leq\sum_{k\geq 1}\left|\frac{X_k^N(0)}{N}-c_k(0)\right|+\frac{2\Lambda t}{N} \\ +\int_0^t 8\Lambda\sum_{k\geq 1}\left|\frac{X_k^N(u)}{N}-c_k(u)\right|\mathrm{d}u+\sum_{k\geq 1}\left|\frac{\mathcal{M}_k^N(t)}{N}\right|,$$

and, by Gronwall's inequality and relation (6.15), we have

$$\lim_{N\to\infty}\mathbb{E}\sup_{t\leq T}\sum_{k\geq 1}\left|\frac{X_k^N(t)}{N}-c_k(t)\right|=0.$$

□

Proposition 6.3. *Under Assumptions (a) and (b), the sequence of local martingales*

$$\left(\frac{1}{\sqrt{N}}D^N(t)\right)=\left(\frac{1}{\sqrt{N}}D_i^N(t)\right)_{i\geq 1}$$

is tight for the convergence in distribution in $\mathcal{D}_T$.

Proof. Recall that for all $i>N$, $(D_i^N(t))\equiv 0$ and $(X_i^N(t))\equiv 0$. By using Jensen's and Doob's inequalities, for any fixed T and N, we get

$$\mathbb{E}\left(\sup_{t\leq T}\left\|\frac{1}{\sqrt{N}}D^N(t)\right\|_{L_2(w)}\right)\leq\sqrt{\mathbb{E}\left(\sup_{t\leq T}\left\|\frac{1}{\sqrt{N}}D^N(t)\right\|_{L_2(w)}^2\right)} \\ =\sqrt{\mathbb{E}\left(\sup_{t\leq T}\sum_{i=1}^N w_i\frac{1}{N}D_i^N(t)^2\right)}\leq\sqrt{\sum_{i=1}^N w_i\frac{1}{N}\mathbb{E}\left(\langle D_i^N(T)\rangle\right)}.$$

From the expression of the increasing processes (6.6), we have

$$\frac{1}{N}\sum_{k\geq 1}w_{2k-1}\mathbb{E}\left(\langle D_{2k-1}^N(T)\rangle\right)\leq\gamma_0\Lambda\int_0^T\mathbb{E}\left(\sum_{k\geq 1}\frac{w_kX_k^N(u)}{N}\right)\mathrm{d}u,$$

and

$$\frac{1}{N}\sum_{k\geq 2}w_{2k-2}\mathbb{E}\left(\langle D_{2k-2}^N(T)\rangle\right)\leq\gamma_0\Lambda\int_0^T\mathbb{E}\left(\sum_{k\geq 2}\frac{w_kX_k^N(u)}{N}\right)\mathrm{d}u.$$

Therefore, by using inequality (6.14), one gets

$$\sup_N \mathbb{E} \sup_{t\le T} \left\| \frac{1}{\sqrt{N}} D^N(t) \right\|_{L_2(w)} \le \sqrt{2\gamma_0 \Lambda T \sup_N \sup_{t\le T} \mathbb{E} \left(\sum_{k\ge 1} \frac{w_k X_k^N(t)}{N} \right)} < \infty,$$

it gives that

$$\lim_{a\to\infty} \limsup_{N\to\infty} \mathbb{P} \left(\sup_{t\le T} \left\| \frac{1}{\sqrt{N}} D^N(t) \right\|_{L_2(w)} \ge a \right)$$
$$\le \lim_{a\to\infty} \frac{1}{a} \sup_N \mathbb{E} \left(\sup_{t\le T} \left\| \frac{1}{\sqrt{N}} D^N(t) \right\|_{L_2(w)} \right) = 0.$$

Since the Skorohod distance is weaker than the uniform distance in $\mathcal{D}_T$, we estimate the modulus of continuity with uniform distance to prove the tightness in the Skorohod space. For any $\varepsilon>0$, $\delta>0$ and any $L\in\mathbb{N}$, we have

$$\mathbb{P} \left(\sup_{t,s\le T, |t-s|<\delta} \left\| \frac{1}{\sqrt{N}} D^N(t) - \frac{1}{\sqrt{N}} D^N(s) \right\|_{L_2(w)} \ge \varepsilon \right) \le \mathbb{P} \left(\sup_{t\le T} \sum_{i>L} r_i \frac{1}{N} D_i^N(t)^2 \ge \frac{\varepsilon^2 r_L}{4 w_L} \right)$$
$$+ \mathbb{P} \left(\sup_{t,s\le T, |t-s|<\delta} \sum_{i=1}^{L} w_i \left(\frac{1}{\sqrt{N}} D_i^N(t) - \frac{1}{\sqrt{N}} D_i^N(s) \right)^2 \ge \frac{\varepsilon^2}{2} \right).$$

By using the inequality (6.14),

$$\mathbb{P} \left(\sup_{t\le T} \sum_{i>L} r_i \frac{1}{N} D_i^N(t)^2 \ge \frac{\varepsilon^2 r_L}{4 w_L} \right) \le \frac{4 w_L}{\varepsilon^2 r_L} \mathbb{E} \left(\sum_{i>L} r_i \frac{1}{N} \langle D_i^N(T) \rangle \right)$$
$$\le \frac{8 \gamma_r w_L \Lambda}{\varepsilon^2 r_L} T \sup_N \sup_{t\le T} \mathbb{E} \left(\sum_{i\ge 1} \frac{r_i X_i^N(t)}{N} \right) \le \frac{8 \gamma_r w_L \Lambda}{\varepsilon^2 r_L} T \zeta_T.$$

By using the Assumption (b), $\lim_{k\to\infty} w_k / r_k = 0$, for any constant $\eta>0$, there exist a constant L, such that

$$\mathbb{P} \left(\sup_{t\le T} \sum_{i\ge L} r_i \frac{1}{N} D_i^N(t)^2 \ge \frac{\varepsilon^2 r_L}{4 w_L} \right) \le \frac{\eta}{2}.$$

The processes $(D_i^N(t)/\sqrt{N})$ $i=1,\ldots,L$, live in finite dimensional space and each of them has an increasing process that is uniformly continuous almost surely. Therefore, for N large enough, we have

$$\mathbb{P} \left(\sup_{t,s\le T, |t-s|<\delta} \sum_{i=1}^{L} w_i \left(\frac{1}{\sqrt{N}} D_i^N(t) - \frac{1}{\sqrt{N}} D_i^N(s) \right)^2 \ge \frac{\varepsilon^2}{2} \right) \le \frac{\eta}{2},$$

consequently, by using Theorem 13.2 in Billingsley [19], the sequence of processes $(D^N(t)/\sqrt{N})$ is tight in $\mathcal{D}_T$.

□

Proposition 6.4. *For the convergence in distribution of random process, uniformly on compact sets,*

$$\lim_{N\to+\infty}\left(\frac{1}{\sqrt{N}}D^N(t)\right)=(D(t)):=\left(\int_0^t \mathrm{Diag}\left(\sqrt{s(c(u))}\right)\cdot \mathrm{d}\beta(u)\right).$$

Proof. From the previous proposition, we have that, for $T>0$, the sequence $(D^N(t)/\sqrt{N})$ is tight in $D([0,T],L_2(w))$. Let $D'(t)$ be a possible limit. For any $d\in\mathbb{N}^*$, let $\mathcal{P}_d$ be the projection from $\mathbb{R}^{\mathbb{N}^+}$ to $\mathbb{R}^d$, i.e., for any $z\in\mathbb{R}^{\mathbb{N}^+}$, $\mathcal{P}_d(z)=(z_1,\dots,z_d)$. It is easy to check that $(\mathcal{P}_d(D'(t)))$ is a limit of a subsequence of $(\mathcal{P}_d(D^N(t)/\sqrt{N}))$ for the weak convergence in probability in the L_2-norm. By using Theorem 1.4 page 339 of Ethier and Kurtz [57], we know that for any $d\in\mathbb{N}^*$, the equality $(\mathcal{P}_d(D'(t)))=(\mathcal{P}_d(D(t)))$ holds in distribution. Hence, by Kolmogorov's theorem, we have the equality in distribution $(D'(t))=(D(t))$. □

We can now state our main result.

Theorem 6.3 (Functional Central Limit Theorem)**.** *Under Assumptions* (a)–(d)*, then the fluctuation process* $(W^N(t))$ *defined by relation* (6.4) *converges in distribution to the* $L_2(w)$*-valued process* $(W(t))$*, the unique strong solution of the SDE* (6.10)*.*

Proof. From relation (6.8), process $(W^N(t))$ satisfies

$$W^N(t)=W^N(0)+\frac{1}{\sqrt{N}}\tau\left(D^N(t)\right)+\int_0^t \tau\circ\nabla s\left(c(u)\right)\cdot W^N(u)\,\mathrm{d}u$$
$$+\frac{2a_1}{\sqrt{N}}\int_0^t\left(\frac{X_1^N(u)}{N}\right)e_1\,\mathrm{d}u+\int_0^t\tau\left(\Delta^N(u)\right)\mathrm{d}u$$

where

$$\Delta^N(u)=\frac{1}{2}\left(\nabla s\left(\frac{X^N(u)}{N}\right)+\nabla s\left(c(u)\right)\right)\cdot W^N(u)-\nabla s\left(c(u)\right)\cdot W^N(u).$$

By direct calculations, we have that for $k\geq 1$

$$\Delta^N_{2k-1}(u)=\left(1+\mathbb{I}_{\{k=1\}}\right)a_k\left(\frac{X_1^N(u)}{N}-c_1(u)\right)W_k^N(u),$$
$$\Delta^N_{2k}(u)=0,$$

and then

$$\|\Delta^N(u)\|_{L_2(w)}\leq 2\sqrt{\gamma_0}\Lambda\left|\frac{X_1^N(u)}{N}-c_1(u)\right|\|W^N(u)\|_{L_2(w)}\leq 2\sqrt{\gamma_0}\Lambda\|W^N(u)\|_{L_2(w)}.$$

Let $\Gamma=\gamma_\tau(w)(\sup_{u\leq T}\gamma(c(u),w)+2\sqrt{\gamma_0}\Lambda)$, by using Lemma 6.1 and Lemma 6.2, for all $t\leq T$, one has

$$\|W^N(t)\|_{L_2(w)}\leq\|W^N(0)\|_{L_2(w)}+\gamma_\tau(w)\frac{1}{\sqrt{N}}\|D^N(t)\|_{L_2(w)}+\frac{2\sqrt{w_1}\Lambda t}{\sqrt{N}}$$
$$+\int_0^t\Gamma\|W^N(u)\|_{L_2(w)}\,\mathrm{d}u.$$

Thanks to Gronwall's lemma, for any $t \leq T$, we have

$$\|W^N(t)\|_{L_2(w)} \leq e^{\Gamma T}\left(\|W^N(0)\|_{L_2(w)} + \gamma_\tau(w)\frac{1}{\sqrt{N}}\|D^N(t)\|_{L_2(w)} + \frac{2\sqrt{w_1}\Lambda T}{\sqrt{N}}\right),$$

which implies the relation

$$\begin{aligned}&\sup_N \mathbb{E}\sup_{t\leq T}\|W^N(t)\|^2_{L_2(w)}\\ &\quad\leq 3e^{2\Gamma T}\left(\sup_N \mathbb{E}\|W^N(0)\|^2_{L_2(w)} + \gamma_\tau(w)^2\sup_N \mathbb{E}\sup_{t\leq T}\frac{1}{N}\|D^N(t)\|^2_{L_2(w)} + \frac{4w_1\Lambda^2T^2}{N}\right).\end{aligned}$$

In the proof of Proposition 6.3, we have shown that

$$\sup_N \mathbb{E}\sup_{t\leq T}\frac{1}{N}\|D^N(t)\|^2_{L_2(w)} < \infty,$$

therefore, if

$$\sup_N \mathbb{E}\|W^N(0)\|^2_{L_2(w)} < C_0,$$

then there exists a finite constant C_T such that

$$\sup_N \mathbb{E}\sup_{t\leq T}\|W^N(t)\|^2_{L_2(w)} < C_T.$$

For the tightness of $(W^N(t))$, for any $\varepsilon > 0$, $\eta > 0$, by using Lemma 6.1 and Lemma 6.2, we get that

$$\begin{aligned}\mathbb{P}\left(\sup_{t,s\leq T,|t-s|<\delta}\|W^N(t)-W^N(s)\|_{L_2(w)} \geq \varepsilon\right) &\leq \mathbb{P}\left(\Gamma\delta\sup_{u\leq T}\|W^N(u)\|_{L_2(w)} \geq \frac{\varepsilon}{2}\right)\\ &\quad + \mathbb{P}\left(\sup_{t,s\leq T,|t-s|<\delta}\left\|\frac{D^N(t)}{\sqrt{N}} - \frac{D^N(s)}{\sqrt{N}}\right\|_{L_2(w)} \geq \frac{\varepsilon}{4\gamma_\tau(w)}\right) + \mathbb{I}_{\left\{\frac{2\sqrt{w_1}\Lambda\delta}{\sqrt{N}} > \frac{\varepsilon}{4}\right\}}\end{aligned}$$

holds. Choose $\delta_1 > 0$ such that

$$\delta_1 < \sqrt{\frac{\eta}{2C_T}}\frac{\varepsilon}{2\Gamma},$$

then, for $\delta \in (0, \delta_1)$ and $N \geq 1$,

$$\mathbb{P}\left(\Gamma\delta\sup_{u\leq T}\|W^N(u)\|_{L_2(w)} \geq \frac{\varepsilon}{2}\right) \leq \left(\frac{2\Gamma\delta}{\varepsilon}\right)^2 \mathbb{E}\sup_{u\leq T}\|W^N(u)\|^2_{L_2(w)} < \frac{\eta}{2}.$$

According to the proof of Proposition 6.3, there exist $\delta_2 > 0$ and N_0, such that, for $\delta \in (0, \delta_2)$ and $N > N_0$, the relation

$$\mathbb{P}\left(\sup_{t,s\leq T,|t-s|<\delta}\left\|\frac{D^N(t)}{\sqrt{N}} - \frac{D^N(s)}{\sqrt{N}}\right\|_{L_2(w)} \geq \frac{\varepsilon}{4\gamma_\tau(w)}\right) \leq \frac{\eta}{2}$$

holds. In conclusion, for any $\delta<\delta_1\wedge\delta_2\wedge(\varepsilon/(8\Lambda\sqrt{w_1}))$ and $N>N_0$,

$$\mathbb{P}\left(\sup_{t,s\leq T,|t-s|<\delta}\|W^N(t)-W^N(s)\|_{L_2(w)}\geq\varepsilon\right)\leq\eta.$$

It is then easy to check that

$$\lim_{a\to\infty}\limsup_{N\to\infty}\mathbb{P}\left(\sup_{t\leq T}\|W^N(t)\|_{L_2(w)}\geq a\right)=0.$$

Therefore, the process $(W^N(t))$ is tight in $D([0,T],L_2(w))$. To identify the limit, note that

$$\begin{aligned}\mathbb{E}\sup_{t\leq T}\left\|\int_0^t\tau\left(\Delta^N(u)\right)\mathrm{d}u\right\|_{L_2(w)}&\leq\gamma_\tau(w)\mathbb{E}\int_0^T\|\Delta^N(u)\|_{L_2(w)}\,\mathrm{d}u\\&\leq 2\sqrt{\gamma_0}\gamma_\tau(w)T\Lambda\mathbb{E}\left(\sup_{u\leq T}\left|\frac{X_1^N(u)}{N}-c_1(u)\right|\|W^N(u)\|_{L_2(w)}\right)\\&\leq 2\sqrt{\gamma_0}\gamma_\tau(w)T\Lambda\left(2\mathbb{E}\sup_{u\leq T}\left|\frac{X_1^N(u)}{N}-c_1(u)\right|\right)^{1/2}\left(\mathbb{E}\sup_{u\leq T}\|W^N(u)\|^2_{L_2(w)}\right)^{1/2}.\end{aligned}$$

By using Theorem 6.2, this term is vanishing as N goes to infinity. From Proposition 6.4, one conclude that any limit of $(W^N(t))$ satisfies SDE (6.10). The theorem is proved. □

References

[1] Aghajani, R., Robert, P., and Sun, W. A large scale analysis of unreliable stochastic networks. *Ann. Appl. Probab. 28*, 2 (04 2018), 851–887.

[2] Alanyali, M., and Dashouk, M. On power-of-choice in downlink transmission scheduling. In *2008 Information Theory and Applications Workshop* (Jan 2008), pp. 12–17.

[3] Aldous, D. *Probability approximations via the Poisson clumping heuristic*. Springer-Verlag, New York, 1989.

[4] Aldous, D., and Diaconis, P. Strong uniform times and finite random walks. *Advances in Applied Mathematics 8* (1987), 69–97.

[5] Aldous, D. J. Deterministic and stochastic models for coalescence (aggregation and coagulation): A review of the mean-field theory for probabilists. *Bernoulli 5*, 1 (1999), 3–48.

[6] Alfonsi, A., Corbetta, J., and Jourdain, B. Evolution of the wasserstein distance between the marginals of two Markov processes. arXiv:1606.02994, 2016.

[7] Alistarh, D., Kopinsky, J., Li, J., and Nadiradze, G. The power of choice in priority scheduling. In *Proceedings of the ACM Symposium on Principles of Distributed Computing* (New York, NY, USA, 2017), PODC '17, ACM, pp. 283–292.

[8] Anderson, D. F., and Kurtz, T. G. Continuous time Markov chain models for chemical reaction networks. In *Design and Analysis of Biomolecular Circuits*, H. Koeppl, G. Setti, M. di Bernardo, and D. Densmore, Eds. Springer New York, 2011, pp. 3–42.

[9] Anderson, R. F., and Orey, S. Small random perturbation of dynamical systems with reflecting boundary. *Nagoya Math. J. 60* (1976), 189–216.

[10] Andreis, L., Pra, P. D., and Fischer, M. Mckean-vlasov limit for interacting systems with simultaneous jumps. Preprint, arXiv:1704.01052.

[11] Andrews, J. M., and Roberts, C. J. A Lumry-Eyring nucleated polymerization model of protein aggregation kinetics: 1. Aggregation with pre-equilibrated unfolding. *J Phys Chem B 111*, 27 (Jul 2007), 7897–7913.

[12] Arnold, V. I. *Dynamical Systems III*. "Encyclopaedia of Mathematical Sciences book series". Springer Berlin Heidelberg, 1988.

[13] Athreya, K. B., and Ney, P. E. *Branching processes*. Springer-Verlag, New York, 1972. Die Grundlehren der mathematischen Wissenschaften, Band 196.

[14] Azar, Y., Broder, A. Z., Karlin, A. R., and Upfal, E. Balanced allocations. *SIAM journal on computing 29*, 1 (1999), 180–200.

[15] Ball, J. M., Carr, J., and Penrose, O. The Becker-Döring cluster equations: Basic properties and asymptotic behaviour of solutions. *Comm. Math. Phys. 104*, 4 (Dec 1986), 657–692.

[16] Becker, R., and Döring, W. Kinetische behandlung der keimbildung in übersättigten dämpfen. *Annalen der Physik 416*, 8 (1935), 719–752.

[17] Bertoin, J. *Random fragmentation and coagulation processes*, vol. 102 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006.

[18] Bertoin, J. Markovian growth-fragmentation processes. *Bernoulli 23*, 2 (05 2017), 1082–1101.

[19] Billingsley, P. Convergence of probability measures. *John Wiley & Sons, INC* (2009), 1–287.

[20] Bingham, N. H. Fluctuation theory for the Ehrenfest urn. *Advances in Applied Probability 23*, 3 (1991), 598–611.

[21] Borthakur, D. HDFS architecture guide. *HADOOP APACHE PROJECT http://hadoop.apache.org/* (2008).

[22] Bozaykut, P., Ozer, N. K., and Karademir, B. Regulation of protein turnover by heat shock proteins. *Free Radic. Biol. Med. 77* (Dec 2014), 195–209.

[23] Bramson, M., Lu, Y., and Prabhakar, B. Decay of tails at equilibrium for fifo join the shortest queue networks. *The Annals of Applied Probability 23*, 5 (10 2013), 1841–1878.

[24] Braun, W., and Hepp, K. The vlasov dynamics and its fluctuations in the $1/n$ limit of interacting classical particles. *Comm. Math. Phys. 56*, 2 (1977), 101–113.

[25] Brémaud, P. *Point processes and queues: Martingale dynamics*. Springer-Verlag, New York, 1981. Springer Series in Statistics.

[26] Brown, T. A martingale approach to the Poisson convergence of simple point processes. *The Annals of Probability 6*, 4 (1978), 615–628.

[27] Budhiraja, A., and Friedlander, E. Diffusion approximations for load balancing mechanisms in cloud storage systems. arXiv preprint arXiv:1706.09914, 2017.

[28] Budhiraja, A., Mukherjee, D., and Wu, R. Supermarket model on graphs. Arxiv *https://arxiv.org/abs/1801.02979*.

[29] Calvez, V., Doumic, M., and Gabriel, P. Self-similarity in a general aggregation–fragmentation problem. application to fitness analysis. *Journal de Mathématiques Pures et Appliquées 98*, 1 (2012), 1–7.

[30] Caputo, P., Dai Pra, P., and Posta, G. Convex entropy decay via the Bochner-Bakry-Emery approach. *Annales de l'institut Henri Poincaré 45*, 3 (08 2009), 734–753.

[31] Carrillo, J. A., McCann, R. J., and Villani, C. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Revista Matemática Iberoamericana 19*, 3 (2003), 971–1018.

[32] Cattiaux, P., Guillin, A., and Malrieu, F. Probabilistic approach for granular media equations in the non-uniformly convex case. *Probability Theory and Related Fields 140*, 1-2 (2008), 19–40.

[33] Chang, C. C., and Yau, H.-T. Fluctuations of one-dimensional Ginzburg-Landau models in nonequilibrium. *Comm. Math. Phys. 145*, 2 (1992), 209–234.

[34] Chang, F., Dean, J., Ghemawat, S., Hsieh, W. C., Wallach, D. A., Burrows, M., Chandra, T., Fikes, A., and Gruber, R. E. Bigtable: A distributed storage system for structured data. In *Proceedings of the 7th USENIX Symposium on Operating Systems Design and Implementation - Volume 7* (Berkeley, CA, USA, 2006), OSDI'06, USENIX Association, pp. 15–15.

[35] Chen, H., and Mandelbaum, A. Discrete flow networks: bottleneck analysis and fluid approximations. *Mathematics of Operation Research 16*, 2 (May 1991), 408–446.

[36] Chun, B.-G., Dabek, F., Haeberlen, A., Sit, E., Weatherspoon, H., Kaashoek, M. F., Kubiatowicz, J., and Morris, R. Efficient replica maintenance for distributed storage systems. In *Proceedings of the 3rd Conference on Networked Systems Design & Implementation - Volume 3* (Berkeley, CA, USA, 2006), NSDI'06, USENIX Association, pp. 4–4.

[37] Comets, F. Nucleation for a long range magnetic model. *Ann. Inst. H. Poincaré Probab. Statist 23*, 2 (1987), 135–178.

[38] Comets, F. *Directed polymers in random environments*, vol. 2175 of *Lecture Notes in Mathematics*. Springer, Cham, 2017. Lecture notes from the 46th Probability Summer School held in Saint-Flour, 2016.

[39] Da Prato, G., and Zabczyk, J. *Stochastic Equations in Infinite Dimensions*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1992.

[40] Dabek, F., Li, J., Sit, E., Robertson, J., Kaashoek, F. F., and Morris, R. Designing a DHT for low latency and high throughput. In *the 1st Symposium on Networked Systems Design and Implementation* (San Francisco, CA, USA, March 2004).

[41] Davis, J. K., and Sindi, S. S. A study in nucleated polymerization models of protein aggregation. *Applied Mathematics Letters 40* (2015), 97–101.

[42] Dawson, D. A. Measure-valued Markov processes. In *École d'Été de Probabilités de Saint-Flour XXI—1991*, vol. 1541 of *Lecture Notes in Math.* Springer, Berlin, 1993, pp. 1–260.

[43] DeCandia, G., Hastorun, D., Jampani, M., Kakulapati, G., Lakshman, A., Pilchin, A., Sivasubramanian, S., Vosshall, P., and Vogels, W. Dynamo: Amazon's highly available key-value store. In *Proceedings of Twenty-first ACM SIGOPS Symposium on Operating Systems Principles* (New York, NY, USA, 2007), SOSP'07, ACM, pp. 205–220.

[44] den Hollander, F. *Random polymers*, vol. 1974 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2009. Lectures from the 37th Probability Summer School held in Saint-Flour, 2007.

[45] Deschamps, J., Hingant, E., and Yvinec, R. Boundary value for a nonlinear transport equation emerging from a stochastic coagulation-fragmentation type model. ArXiv e-prints, Dec. 2014.

[46] Desvillettes, L., and Villani, C. On the trend to global equilibrium in spatially inhomogeneous entropy-dissipating systems: the linear Fokker-Planck equation. *Communications on Pure and Applied Mathematics 54*, 1 (2001), 1–42.

[47] Diaconis, P. *Group representations in probability and statistics.* Institute of Mathematical Statistics, Hayward, 1988.

[48] Diaconis, P., Graham, R. L., and Morrison, J. A. Asymptotic analysis of a random walk on a hypercube with many dimensions. *Random Structures Algorithms 1*, 1 (1990), 51–72.

[49] Dobson, C. The generic nature of protein folding and misfolding. In *Protein Misfolding, Aggregation, and Conformational Diseases*, V. Uversky and A. Fink, Eds., vol. 4 of *Protein Reviews*. Springer US, 2006, pp. 21–41.

[50] Dobson, C. M. Protein folding and misfolding. *Nature 426* (2003), 884–890.

[51] Doumic, M., Eugène, S., and Robert, P. Asymptotics of stochastic protein assembly models. *SIAM Journal on Applied Mathematics 76*, 6 (Nov. 2016), 2333–2352.

[52] Durrett, R., Granovsky, B. L., and Gueron, S. The equilibrium behavior of reversible coagulation-fragmentation processes. *Journal of Theoretical Probability 12*, 2 (Apr 1999), 447–474.

[53] Eden, K., Morris, R., Gillam, J., MacPhee, C. E., and Allen, R. J. Competition between primary nucleation and autocatalysis in amyloid fibril self-assembly. *Biophysical Journal 108*, 3 (2015), 632–643.

[54] Ehrenfest, P. Uber zwei bekannte einwande gegen das boltzmannsche h theorem. *Phys. Z 8* (1907), 311–314.

[55] El Karoui, N., and Chaleyat-Maurel, M. *Temps locaux*, vol. 52-53. Société Mathématique de France, 1978, ch. Un problème de réflexion et ses applications au temps local et aux équations différentielles stochastiques sur $\mathbb{R}$, pp. 117–144. Exposés du Séminaire J. Azéma-M. Yor, Paris, 1976–1977.

[56] Ercolani, N., Jansen, S., and Ueltschi, D. Random partitions in statistical mechanics. *Electronic Journal of Probability 19* (2014), 37 pp.

[57] Ethier, S. N., and Kurtz, T. G. *Markov processes: Characterization and convergence*. John Wiley & Sons Inc., New York, 1986.

[58] Eugène, S., Xue, W.-F., Robert, P., and Doumic, M. Insights into the variability of nucleated amyloid polymerization by a minimalistic model of stochastic protein assembly. *Journal of Chemical Physics 144*, 17 (2016), 175101.

[59] Evans, M. R., and Hanney, T. Nonequilibrium statistical mechanics of the zero-range process and related models. *Journal of Physics A: Mathematical and General 38*, 19 (2005), R195.

[60] Fallat, S. M., and Johnson, C. R. *Totally nonnegative matrices*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2011.

[61] Feuillet, M., and Robert, P. A scaling analysis of a transient stochastic network. *Advances in Applied Probability 46*, 2 (June 2014), 516–535.

[62] Firestone, M., Levie, R. D., and Rangarajan, S. On one-dimensional nucleation and growth of "living" polymers I. homogeneous nucleation. *Journal of Theoretical Biology 104*, 4 (1983), 535 – 552.

[63] Freidlin, M., and Wentzell, A. *Random Perturbations of Dynamical Systems*. No. v. 260 in Grundlehren der mathematischen Wissenschaften. Springer, 1998.

[64] Fristedt, B. The structure of random partitions of large integers. *Transactions of the American Mathematical Society 337*, 2 (1993), 703–735.

[65] Ganguly, A., Ramanan, K., Robert, P., and Sun, W. A Large-Scale network with moving servers. In *Proc. of Sigmetrics workshop MAMA* (June 2017).

[66] Ghemawat, S., Gobioff, H., and Leung, S.-T. The Google file system. In *the 9th symposium on Operating systems principles* (New York, NY, USA, October 2003), pp. 29–43.

[67] Gillam, J., and MacPhee, C. Modelling amyloid fibril formation kinetics: mechanisms of nucleation and growth. *Journal of Physics: Condensed Matter 25*, 37 (2013), 373101.

[68] Godrèche, C. *From Urn Models to Zero-Range Processes: Statics and Dynamics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, pp. 261–294.

[69] Godrèche, C., and Luck, J.-M. Nonequilibrium dynamics of urn models. *Journal of Physics: Condensed Matter 14*, 7 (2002), 1601–1615.

[70] Graham, C. Chaoticity on path space for a queueing network with selection of the shortest queue among several. *Journal of Applied Probability 37*, 1 (2000), 198–211.

[71] Graham, C. Functional central limit theorems for a large network in which customers join the shortest of several queues. *Probability Theory and Related Fields 131*, 1 (Jan 2005), 97–120.

[72] Guo, M. Z., Papanicolaou, G. C., and Varadhan, S. R. S. Nonlinear diffusion limit for a system with nearest neighbor interactions. *Comm. Math. Phys. 118*, 1 (1988), 31–59.

[73] Harrison, J., and Reiman, M. Reflected Brownian motion on an orthant. *Annals of Probability 9*, 2 (1981), 302–308.

[74] Heneghan, A., Wilson, P., and Haymet, A. Heterogeneous nucleation of supercooled water, and the effect of an added catalyst. *Proceedings of the National Academy of Sciences 99*, 15 (2002), 9631–9634.

[75] Hingant, E., and Yvinec, R. Deterministic and Stochastic Becker-Döring equations: Past and Recent Mathematical Developments. arXiv preprint arXiv: 1609.00697, 2016.

[76] Horn, R. A., and Johnson, C. R. *Matrix analysis*. Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original.

[77] Hunt, P., and Kurtz, T. Large loss networks. *Stochastic Processes and their Applications 53*, 2 (1994), 363 – 378.

[78] Jabin, P.-E., and Niethammer, B. On the rate of convergence to equilibrium in the becker–döring equations. *Journal of Differential Equations 191*, 2 (2003), 518–543.

[79] Jacobsen, M. *Point process theory and applications*. Probability and its Applications. Birkhäuser Boston, Inc., Boston, MA, 2006.

[80] Jacod, J. *Calcul stochastique et problèmes de martingales*, vol. 714 of *Lecture Notes in Mathematics*. Springer, Berlin, 1979.

[81] Jacod, J., and Shiryaev, A. N. *Limit theorems for stochastic processes*, vol. 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1987.

[82] Jelasity, M., Montresor, A., Jesi, G. P., and Voulgaris, S. The Peersim simulator. http://peersim.sourceforge.net/.

[83] Jeon, I. Existence of gelling solutions for coagulation-fragmentation equations. *Comm. Math. Phys. 194*, 3 (Jun 1998), 541–567.

[84] Joulin, A. A new Poisson-type deviation inequality for Markov jump processes with positive Wasserstein curvature. *Bernoulli. Official Journal of the Bernoulli Society for Mathematical Statistics and Probability 15*, 2 (2009), 532–549.

[85] Kac, M. Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 3: Contributions to Astronomy and Physics* (Berkeley, Calif., 1956), University of California Press, pp. 171–197.

[86] Karlin, S., and McGregor, J. Ehrenfest urn models. *Journal of Applied Probability 2* (1965), 352–376.

[87] Karthik, A., Mukhopadhyay, A., and Mazumdar, R. R. Choosing among heterogeneous server clouds. *Queueing Systems 85*, 1 (Feb 2017), 1–29.

[88] Kasahara, Y., and Watanabe, S. Limit theorems for point processes and their functionals. *Journal of the Mathematical Society of Japan 38*, 3 (1986), 543–574.

[89] Kashchiev, D. *Nucleation*. Butterworth-Heinemann, 2000.

[90] Kashchiev, D. Protein polymerization into fibrils from the viewpoint of nucleation theory. *Biophys. J. 109*, 10 (Nov 2015), 2126–2136.

[91] Kashchiev, D. Modeling the effect of monomer conformational change on the early stage of protein self-assembly into fibrils. *The Journal of Physical Chemistry B 121*, 1 (2017), 35–46.

[92] Keilson, J. *Markov chains Models-Rarity and exponentiality*, vol. 28 of *Applied Mathematical sciences*. Springer Verlag, New York, 1979.

[93] Kelly, F. P. Loss networks. *Ann. Appl. Probab. 1*, 3 (08 1991), 319–378.

[94] Khasminskii, R. *Stochastic stability of differential equations*. Mechanics–analysis. Sijthoff & Noordhoff, 1980.

[95] Khasminskii, R., and Krylov, N. On averaging principle for diffusion processes with null-recurrent fast component. *Stochastic Processes and their applications 93*, 2 (2001), 229–240.

[96] Kingman, J. F. C. *Poisson processes*. Oxford studies in probability, 1993.

[97] Kolchin, V. F., Sevastyanov, B. A., and Chistyakov, V. P. *Random allocations*. Winston, 1978.

[98] Kurtz, T. Averaging for martingale problems and stochastic approximation. In *Applied Stochastic Analysis, US-French Workshop* (1992), vol. 177 of *Lecture notes in Control and Information sciences*, Springer Verlag, pp. 186–209.

[99] Kurtz, T. G. The relationship between stochastic and deterministic models for chemical reactions. *The Journal of Chemical Physics 57*, 7 (1972), 2976–2978.

[100] Kurtz, T. G. Strong approximation theorems for density dependent markov chains. *Stochastic Processes and their Applications 6*, 3 (1978), 223 – 240.

[101] Lakshman, A., and Malik, P. Cassandra: A decentralized structured storage system. *SIGOPS Oper. Syst. Rev. 44*, 2 (Apr. 2010), 35–40.

[102] Last, G., and Brandt, A. *Marked point processes on the real line*. Probability and its Applications (New York). Springer-Verlag, New York, 1995.

[103] Laurençot, P., and Mischler, S. From the becker–döring to the lifshitz–slyozov–wagner equations. *Journal of Statistical Physics 106*, 5 (Mar 2002), 957–991.

[104] Legtchenko, S., Monnet, S., Sens, P., and Muller, G. RelaxDHT: A churn-resilient replication strategy for peer-to-peer distributed hash-tables. *ACM Trans. Auton. Adapt. Syst. 7*, 2 (July 2012), 28:1–28:18.

[105] Luçon, E., and Stannat, W. Mean field limit for disordered diffusions with singular interactions. *The Annals of Applied Probability 24*, 5 (2014), 1946–1993.

[106] Luczak, M. J., and McDiarmid, C. Asymptotic distributions and chaos for the supermarket model. *Electronic Journal of Probability 12* (2007), no. 3, 75–99.

[107] Maguluri, S. T., Srikant, R., and Ying, L. Stochastic models of load balancing and scheduling in cloud computing clusters. In *2012 Proceedings IEEE INFOCOM* (March 2012), pp. 702–710.

[108] Markowich, P. A., and Villani, C. On the trend to equilibrium for the Fokker-Planck equation: an interplay between physics and functional analysis. *Matemática Contemporânea 19* (2000), 1–29. VI Workshop on Partial Differential Equations, Part II (Rio de Janeiro, 1999).

[109] McKean, H. P. Propagation of chaos for a class of non-linear parabolic equations. In *Stochastic Differential Equations (Lecture Series in Differential Equations, Session 7, Catholic Univ., 1967)*. Air Force Office Sci. Res., Arlington, Va., 1967, pp. 41–57.

[110] McManus, J. J., Charbonneau, P., Zaccarelli, E., and Asherie, N. The physics of protein self-assembly. preprint, Feb. 2016.

[111] Meyn, S., and Tweedie, R. *Markov chains and stochastic stability*. Communications and control engineering series. Springer, 1993.

[112] Mitzenmacher, M., Richa, A. W., and Sitaraman, R. The power of two random choices: A survey of techniques and results. In *in Handbook of Randomized Computing* (2000), pp. 255–312.

[113] Mohamed, H., and Robert, P. A probabilistic analysis of some tree algorithms. *Annals of Applied Probability 15*, 4 (Nov. 2005), 2445–2471.

[114] Montresor, A., and Jelasity, M. PeerSim: A scalable P2P simulator. In *Proc. of the 9th Int. Conference on Peer-to-Peer (P2P'09)* (Seattle, WA, Sept. 2009), pp. 99–100.

[115] Morris, A. M., Watzky, M. A., and Finke, R. G. Protein aggregation kinetics, mechanism, and curve-fitting: a review of the literature. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics 1794*, 3 (2009), 375–397.

[116] Müller, P. E. *Limiting Properties of a Continuous Local Mean-Field Interacting Spin System*. PhD thesis, Rheinischen Friedrich-Wilhelms-Universität Bonn, 2016.

[117] Murray, J. D. *Mathematical biology. I*, third ed., vol. 17 of *Interdisciplinary Applied Mathematics*. Springer-Verlag, New York, 2002. An introduction.

[118] Niethammer, B. Macroscopic limits of the Becker-Döring equations. *Commun. Math. Sci. 2*, 1 (05 2004), 85–92.

[119] Norris, J. R. Smoluchowski's coagulation equation: uniqueness, nonuniqueness and a hydrodynamic limit for the stochastic coalescent. *Ann. Appl. Probab. 9*, 1 (02 1999), 78–109.

[120] Nummelin, E. *General irreducible Markov chains and nonnegative operators*. Cambridge University Press, Cambridge, 1984.

[121] Ollivier, Y. Ricci curvature of Markov chains on metric spaces. *Journal of Functional Analysis 256*, 3 (2009), 810–864.

[122] Oosawa, F., and Asakura, S. *Thermodynamics of the polymerization of protein*. Academic Press, 1975.

[123] Ow, S.-Y., and Dunstan, D. E. A brief overview of amyloids and Alzheimer's disease. *Protein Science 23*, 10 (2014), 1315–1331.

[124] Papanicolaou, G. C., Stroock, D., and Varadhan, S. R. S. Martingale approach to some limit theorems. In *Papers from the Duke Turbulence Conference (Duke Univ., Durham, N.C., 1976), Paper No. 6*. Duke Univ., Durham, N.C., 1977, pp. ii+120 pp. Duke Univ. Math. Ser., Vol. III.

[125] Penrose, O. Metastable states for the Becker-Döring cluster equations. *Comm. Math. Phys. 124*, 4 (1989), 515–541.

[126] Penrose, O. Nucleation and droplet growth as a stochastic process. In *Analysis and stochastics of growth processes and interface models*. Oxford Univ. Press, Oxford, 2008, pp. 265–277.

[127] Philo, J. S., and Arakawa, T. Mechanisms of protein aggregation. *Curr Pharm Biotechnol 10*, 4 (Jun 2009), 348–351.

[128] Picconi, F., Baynat, B., and Sens, P. An analytical estimation of durability in DHTs. In *Distributed Computing and Internet Technology* (2007), T. Janowski and H. Mohanty, Eds., vol. 4882 of *Lecture Notes in Computer Science*, Springer, pp. 184–196.

[129] Pigolotti, S., Lizana, L., Otzen, D., and Sneppen, K. Quality control system response to stochastic growth of amyloid fibrils. *FEBS Letters 587*, 9 (2013), 1405–1410.

[130] Pinheiro, E., Weber, W.-D., and Barroso, L. A. Failure trends in a large disk drive population. In *5th USENIX Conference on File and Storage Technologies (FAST'07)* (2007), pp. 17–29. http://research.google.com/archive/disk_failures.pdf.

[131] Pitman, J. Exchangeable and partially exchangeable random partitions. *Probability Theory and Related Fields 102*, 2 (Jun 1995), 145–158.

[132] Prigent, S., Ballesta, A., Charles, F., Lenuzza, N., Gabriel, P., Tine, L. M., Rezaei, H., and Doumic, M. An efficient kinetic model for assemblies of amyloid fibrils and its application to polyglutamine aggregation. *Plos One 7*, 11 (2012). doi:10.1371/journal.pone.0043273.

[133] Pujo-Menjouet, L. *Étude de modèles mathématiques issus de la biologie du cycle cellulaire et de la dynamique des protéines*. Habilitation à diriger des recherches, Université Claude Bernard Lyon 1 - Institut Camille Jordan, Dec. 2016.

[134] Ramabhadran, S., and Pasquale, J. Analysis of long-running replicated systems. In *Proceedings IEEE INFOCOM 2006. 25TH IEEE International Conference on Computer Communications* (Barcelona, Spain, Apr. 2006), IEEE, pp. 1–9.

[135] Ramanan, K. Reflected diffusions defined via the extended skorokhod map. *Electronic Journal of Probability 11* (2006), 934–992.

[136] Ranjbar, M., and Rezakhanlou, F. Equilibrium fluctuations for a model of coagulating-fragmenting planar Brownian particles. *Comm. Math. Phys. 296*, 3 (Jun 2010), 769–826.

[137] Rhea, S., Godfrey, B., Karp, B., Kubiatowicz, J., Ratnasamy, S., Shenker, S., Stoica, I., and Yu, H. OpenDHT: a public DHT service and its uses. In *Proceedings of SIGCOMM* (2005), pp. 73–84.

[138] Robert, P. *Stochastic Networks and Queues*. Stochastic Modelling and Applied Probability Series. Springer-Verlag, New York, 2003.

[139] Robert, P., and Sun, W. On the asymptotic distribution of nucleation times of polymerization processes. Preprint, ArXiv: 1712.08347, Dec. 2017.

[140] Rodrigues, R., and Blake, C. When multi-hop peer-to-peer lookup matters. In *IPTPS'04: Proceedings of the 3rd International Workshop on Peer-to-Peer Systems* (San Diego, CA, USA, February 2004), pp. 112–122.

[141] Rogers, L. C. G., and Williams, D. *Diffusions, Markov processes, and martingales. Vol. 1: Foundations*, second ed. John Wiley & Sons Ltd., Chichester, 1994.

[142] Rogers, L. C. G., and Williams, D. *Diffusions, Markov processes, and martingales. Vol. 2*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000. Itô calculus, Reprint of the second (1994) edition.

[143] Ross, C. A., and Poirier, M. A. Protein aggregation and neurodegenerative disease. *Nat. Med. 10* (Jul 2004), S10–17.

[144] Rowstron, A. I. T., and Druschel, P. Storage management and caching in PAST, a large-scale, persistent peer-to-peer storage utility. In *the 8th ACM symposium on Operating Systems Principles* (December 2001), pp. 188–201.

[145] Rudin, W. *Real and complex analysis*, third ed. McGraw-Hill Book Co., New York, 1987.

[146] Shiga, T., and Tanaka, H. Central limit theorem for a system of markovian particles with mean field interactions. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 69*, 3 (1985), 439–459.

[147] Simon, V., Monnet, S., Feuillet, M., Robert, P., and Sens, P. Scattering and placing data replicas to enhance long-term durability. In *The 14th IEEE International Symposium on Network Computing and Applications (IEEE NCA15)* (Cambridge, United States, Sept. 2015), p. 6.

[148] Sipe, J. D., and Cohen, A. S. Review: history of the amyloid fibril. *J. Struct. Biol. 130*, 2-3 (Jun 2000), 88–98.

[149] Skorokhod, A. Stochastic equations for diffusion processes in a bounded region. *Theory Probab. Appl. 7* (1962), 3–23.

[150] Smith, H. L., Li, W., and Cheetham, M. E. Molecular chaperones and neuronal proteostasis. *Semin. Cell Dev. Biol. 40* (Apr 2015), 142–152.

[151] Spohn, H. Equilibrium fluctuations for interacting brownian particles. *Comm. Math. Phys. 103*, 1 (1986), 1–33.

[152] Spohn, H. *Large scale dynamics of interacting particles*. Springer Science & Business Media, 2012.

[153] Sun, W. A functional central limit theorem for the becker–döring model. *Journal of Statistical Physics 171*, 1 (Apr 2018), 145–165.

[154] Sun, W., Feuillet, M., and Robert, P. Analysis of large unreliable stochastic networks. *Annals of Applied Probability 26*, 5 (2016), 2959–3000.

[155] Sun, W., and Robert, P. Analysis of Large Urn Models with Local Mean-Field Interactions. Preprint, ArXiv: 1802.05064, Feb. 2018.

[156] Sun, W., Simon, V., Monnet, S., Robert, P., and Sens, P. Analysis of a stochastic model of replication in large distributed storage systems: A mean-field approach. In *ACM-Sigmetrics* (Urbana-Champaign, IL USA, June 2017), ACM.

[157] Szavits-Nossan, J., Eden, K., Morris, R. J., MacPhee, C. E., Evans, M. R., and Allen, R. J. Inherent variability in the kinetics of autocatalytic protein self-assembly. *Physical Review Letters 113* (Aug 2014), 098101.

[158] Sznitman, A. Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX — 1989* (1991), vol. 1464 of *Lecture Notes in Maths*, Springer-Verlag, pp. 167–243.

[159] Sznitman, A. S. Équations de type de boltzmann, spatialement homogènes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 66*, 4 (Sep 1984), 559–592.

[160] Sznitman, A.-S. Nonlinear reflecting diffusion process, and the propagation of chaos and fluctuations associated. *Journal of Functional Analysis 56*, 3 (1984), 311 – 336.

[161] Taylor, L., and Williams, R. Existence and uniqueness of semimartingale reflecting Brownian motions in an orthant. *Probability Theory and Related Fields 96* (1993), 283–317.

[162] Thai, M.-N. Birth and death process in mean field type interaction. arXiv:1510.03238, 2015.

[163] von Smoluchowski, M. Versuch einer mathematischen theorie der koagulationskinetik kolloider lösungen. *Zeitschrift fuer physikalische Chemie 92* (2010), 129 – 168.

[164] Voter, W., and Erickson, H. The kinetics of microtubule assembly. evidence for a two-stage nucleation mechanism. *The Journal of Biological Chemistry 259* (1984), 10430–10438.

[165] Vvedenskaya, N. D., Dobrushin, R. L., and Karpelevich, F. I. A queueing system with a choice of the shorter of two queues—an asymptotic approach. *Problemy Peredachi Informatsii 32*, 1 (1996), 20–34.

[166] Wegner, A., and Engel, J. Kinetics of the cooperative association of actin to actin filament. *Biophysical Chemistry 3*, 3 (1975), 215–225.

[167] Wu, C., and Shea, J. E. Coarse-grained models for protein aggregation. *Curr. Opin. Struct. Biol. 21*, 2 (Apr 2011), 209–220.

[168] Xue, W.-F., Homans, S. W., and Radford, S. E. Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly. *PNAS 105* (2008), 8926–8931.

[169] Yor, M. Existence et unicité de diffusions à valeurs dans un espace de Hilbert. *Annales de l'I.H.P. Probabilités et statistiques 10*, 1 (1974), 55–88.

[170] Yvinec, R., Bernard, S., Hingant, E., and Pujo-Menjouet, L. First passage times in homogeneous nucleation: Dependence on the total number of particles. *The Journal of Chemical Physics 144*, 3 (2016), 034106.

[171] Yvinec, R., D'Orsogna, M. R., and Chou, T. First passage times in homogeneous nucleation and self-assembly. *The Journal of Chemical Physics 137*, 24 (2012), 244107.

[172] Zhang, J., and Muthukumar, M. Simulations of nucleation and elongation of amyloid fibrils. *The Journal of Chemical Physics 130*, 3 (Jan 2009), 035102.

[173] Zhu, M. Equilibrium fluctuations for one-dimensional Ginzburg-Landau lattice model. *Nagoya Math. J. 117* (1990), 63–92.