Distributed Computing manuscript No.
(will be inserted by the editor)

Coupling and Self-Stabilization

Laurent Fribourg, Stéphane Messika and Claudine Picaronny

LSV, CNRS & ENS de Cachan,
61 av. du Prés. Wilson
94235 Cachan cedex, France
e-mail: {fribourg,messika,picaro}@lsv.ens-cachan.fr

The date of receipt and acceptance will be inserted by the editor

Summary. A randomized self-stabilizing algorithm $\mathcal{A}$ is an algorithm that, whatever the initial configuration is, reaches a set $\mathcal{L}$ of *legal configurations* in finite time with probability 1. The proof of convergence towards $\mathcal{L}$ is generally done by exhibiting a potential function φ, which measures the "vertical" distance of any configuration to $\mathcal{L}$, such that φ decreases with non-null probability at each step of $\mathcal{A}$. We propose here a method, based on the notion of coupling, which makes use of a "horizontal" distance δ between any pair of configurations, such that δ decreases in expectation at each step of $\mathcal{A}$. In contrast with classical methods, our coupling method does not require the knowledge of $\mathcal{L}$. In addition to the proof of convergence, the method allows us to assess the convergence rate according to two different measures. Proofs produced by the method are often simpler or give better upper bounds than their classical counterparts, as examplified here on Herman's mutual exclusion and Iterated Prisoner's Dilemma algorithms in the case of cyclic graphs.

1 Introduction

The notion of self-stabilization was introduced in computer science by Dijkstra [4]. A distributed algorithm $\mathcal{A}$ is self-stabilizing if, whatever the initial configuration it starts from, it reaches within a finite time a set $\mathcal{L}$ of "legal" configurations, i.e, configurations satisfying a desired property. Moreover, $\mathcal{L}$ is closed under $\mathcal{A}$: once $\mathcal{A}$ has reached $\mathcal{L}$, it never leaves it. Self-stabilizing systems have notably received much attention because they propose an elegant way of solving the problem of fault-tolerance [19]. Randomization is often employed in self-stabilization to break the symmetry in anonymous systems (see [5]). With randomized self-stabilizing algorithms, the convergence towards $\mathcal{L}$ is guaranteed with probability 1.

We show here that we can use the notion of *coupling*, as used in the field of Applied Probability, to prove the self-stabilization property and at the same time, the rate of convergence to the set of legal configurations. Coupling is a method used for analyzing the rate of convergence to equilibrium in Markov chain Monte Carlo experiments (see, e.g., [26]). The coupling time is the time that two faithful copies of a stochastic process coalesce together. Coupling time is generally used as an upper bound of the "mixing time" of a Markov chain $\mathcal{A}$, i.e., the time for the chain to be ε-close to its stationary distribution. We show here that self-stabilization of $\mathcal{A}$ follows from the *finiteness* of coupling time. The coupling time will be also used for deriving an upper bound on two measures of the rate of convergence of $\mathcal{A}$: the expected time of reaching $\mathcal{L}$ ("hitting time") and the time after which $\mathcal{L}$ has been reached with high probability ("ε-absorption time").

Comparison with related work.

Classically, self-stabilization of a randomized algorithm $\mathcal{A}$, seen as a Markov chain $(X_t)_{t=0}^{\infty}$, is shown by finding an integer-valued potential function φ on the set Ω of configurations that decreases with non-null probability until φ reaches a minimum value, say 0. In this case, assuming $\varphi(X_t) = 0 \Rightarrow X_t \in \mathcal{L}$, one is guaranteed that $\mathcal{L}$ has been reached. The expected time of hitting is calculated independently (see, e.g., [5,14]). There is another classical method for both showing self-stabilization w.r.t. $\mathcal{L}$ and analyzing the rate of convergence, as examplified in [9], which consists in finding an integer-valued potential function φ on Ω such that:

$\varphi(X_t) = 0 \Rightarrow X_t \in \mathcal{L}$, and

$E[\varphi(X_{t+1})] \leq \beta\varphi(X_t)$ for some β $(0 \leq \beta < 1)$.

This function φ can be seen as a "vertical" distance that separates X from $\mathcal{L}$ (See Figure 1).

Our new method basically consists in finding a coupling $(X_t, Y_t)_{t=1}^{\infty}$ (given arbitrary initial configurations $X_0 = x$ and $Y_0 = y$) where (X_t) and (Y_t) are faithful copies of $\mathcal{A}$, and a "horizontal" distance δ on $\Omega \times \Omega$ such that:

$E[\delta(X_{t+1}, Y_{t+1})] \leq \beta\delta(X_t, Y_t)$ for some β $(0 \leq \beta < 1)$.

For appropriate initial values y of Y_t, one can guarantee that, whatever the initial value x of X_t is, X_t has reached $\mathcal{L}$ when δ is null. Suppose indeed that y belongs to $\mathcal{L}$.

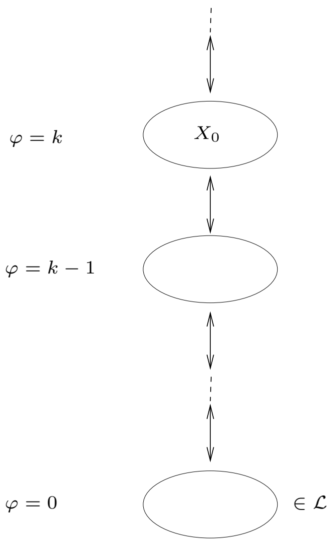

Fig. 1. Illustration of the classical method

Then $Y_t \in \mathcal{L}$ for all t (since $\mathcal{L}$ is closed), and $X_t \in \mathcal{L}$ when $\delta(X_t, Y_t) = 0$ (since then $X_t = Y_t$). See Figure 2.

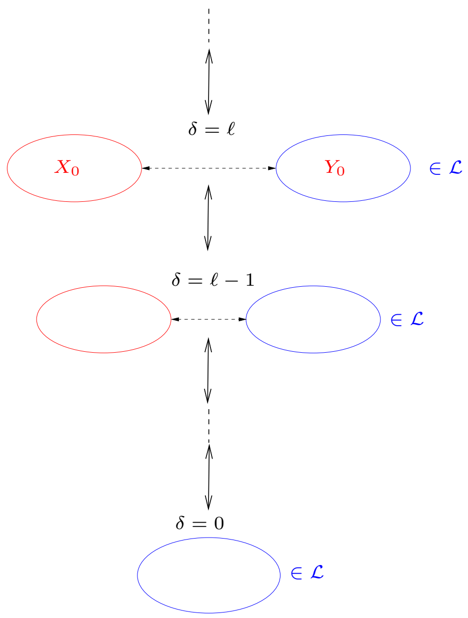

Fig. 2. Illustration of the coupling method

As explained below, the advantages of the coupling method are the following:

- it provides us not only with a proof of self-stabilization, but also with an upper bound for the hitting time and the ε-absorption time,
- it does not rely on the knowledge of $\mathcal{L}$,
- the discovery of δ and the evaluation of β can be greatly simplified by using various optimizations of coupling, such as path coupling (see [3]),
- on Herman's mutual exclusion and Iterated Prisoner's Dilemma algorithms, proofs produced by our method are simpler or give better upper bounds than their classical counterparts.

The method is however limited, because we have to assume that the scheduling of the randomized actions is fixed (e.g., synchronous), and the set of legal configurations $\mathcal{L}$ strongly connected. We will indicate how to relax the latter assumption on $\mathcal{L}$ at the end of the paper.

Plan of the paper. After some preliminaries on randomized distributed algorithms (Section 2), we define the concepts of self-stabilization and rate of convergence (Section 3). We then relate the notion of coupling to that of self-stabilization (Section 4), thus yielding a new method for proving self-stabilization (Section 5). The method is refined via the technique of path coupling in Section 6. We indicate how to relax a basic assumption on the legal configurations in Section 7, and conclude in Section 8.

2 Randomized Distributed Algorithms As Markov Chains

In a distributed system, the topology of the network of machines is generally given under the form of a graph $G = (V, E)$, where the set $V = \{1, \cdots, N\}$ of vertices corresponds to the locations of the machines. There is an edge between two vertices when the corresponding machines can communicate together. All the machines are here identical finite state machines. The space of states is Q. A *configuration* x of the network is the N-tuple of all the states of the machines. The set of configurations Q^N is denoted Ω. Given a configuration x of Ω, the state of the i-th machine is written $x(i)$. The communication between machines is done here through the reading of neighbors' states. Randomized distributed algorithms are characterized by a *scheduler* (or *adversary*), i.e., a mechanism which selects, at each step, a nonempty subset of machines, and a set of *actions* which applies simultaneously at each selected machine. In this paper, we suppose that the scheduler is fixed and memoryless (called "oblivious" in [25]): at each step, it selects a subset of machines depending on the current configuration only. For example, we will consider the case of a *synchronous scheduler* (resp. *randomized central scheduler*) which selects, at each step, all the machines (resp. a single machine randomly chosen). Once a machine is selected, its state (as well as possibly, the state of some of its neighbors) is changed by the action that applies. For a given memoryless scheduler, the randomized distributed algorithm can be seen as a Markov chain $\mathcal{A}$ of the form $(X_t)_{t=0}^{\infty}$ where X_t is a random variable taking its values on Ω (see, e.g., [8]): given configurations x and y, the probability at step t to go from x to y is a constant, $Pr(X_{t+1} = y | X_t = x)$ that depends on x and y only, not on t.

Example 1. We consider Herman's mutual exclusion algorithm [14]. The topology is a cyclic graph (ring) of N vertices, and the scheduler synchronous. The set of states is $Q = \{0, 1\}$, and the number of machines N is odd. At each step, the state of every machine $x(i)$ $(1 \leq i \leq N)$ is changed into $x'(i)$ as follows:

- if $x(i) \neq x(i-1)$ then $x'(i) = \neg x(i)$,
- if $x(i) = x(i-1)$ then $x'(i) = \begin{cases} 0 & \text{with probability } 1/2, \\ 1 & \text{with probability } 1/2. \end{cases}$

(When $i = 1$, $(i-1)$ stands for N. As usual, $\neg 0$ stands for 1, and $\neg 1$ for 0.)

Example 2. We consider the problem of the Iterated Prisoner's Dilemma, as modeled in [9]. The topology is a cyclic graph (ring) of N vertices, and the scheduler randomized central. The set of states is $Q = \{-, +\}$. At each step, a vertex i $(1 \leq i \leq N)$ is chosen uniformly at random, and the values $x(i)$ and $x(i+1)$ are changed into $x'(i)$ and $x'(i+1)$ respectively as follows:

- if $x(i) = x(i+1)$, then $x'(i) = x'(i+1) = +$,
- if $x(i) \neq x(i+1)$, then $x'(i) = x'(i+1) = -$.

(When $i = N$, $(i+1)$ stands here for 1.)

3 Self-Stabilization

Let us consider a Markov chain $\mathcal{A} = (X_t)_{t=0}^{\infty}$ on a finite set Ω. Two configurations x and y are in the same equivalence class if they are "inter-connected", i.e., if there exist t, u such that $Pr(X_t = y | X_0 = x) > 0$ and $Pr(X_u = x | X_0 = y) > 0$. Given two classes C and C', $C' \ll C$ means that $Pr(X_t = y | X_0 = x) > 0$ for some $x \in C, y \in C'$ and $t > 0$. The minimal classes for $\ll$ are called *ergodic sets*. More precisely:

Definition 1. *Let $\mathcal{M} \subset 2^{\Omega}$ be a set of configurations. $\mathcal{M}$ is an* ergodic set *if*
1. $\mathcal{M}$ is strongly connected, i.e.:
$\forall x, y \in \mathcal{M}: \quad Pr(X_t = y | X_0 = x) > 0$ *for some t, and*
2. $\mathcal{M}$ is closed, i.e.:
$(x \in \mathcal{M} \ \wedge \ Pr(X_1 = y | X_0 = x) > 0) \ \Rightarrow \ y \in \mathcal{M}$.

Every finite Markov chain has always at least one ergodic set (since finite partial ordering $\ll$ must have at least one minimal element). Moreover, two distinct ergodic sets are disjoint (since they are both strongly connected).

In the non-probabilistic context, we say that, given a set $\mathcal{L}$ of legal configurations, $\mathcal{A}$ is self-stabilizing w.r.t. $\mathcal{L}$ if, starting from any initial configuration, the system is guaranteed to reach a configuration of $\mathcal{L}$ within a finite number of transitions (see, e.g., [24]). For example, in mutual exclusion problems, a legal configuration is a configuration with a single token, which expresses the fact that only one machine can enjoy the resource. In the probabilistic context of Markov chains, the convergence property has to be guaranteed with probability 1. Formally:

Definition 2. *Given a closed set $\mathcal{L}$ of configurations, $\mathcal{A}$ is* self-stabilizing w.r.t. $\mathcal{L}$, *if $\mathcal{A}$ converges towards $\mathcal{L}$ (with probability 1), whatever the initial configuration is, i.e:*
$\forall x \in \Omega \ Pr(X_t \in \mathcal{L} | X_0 = x) \rightarrow 1$ *when $t \rightarrow \infty$.*.

Note that by Markov's theorem (see, e.g., [17]) $\mathcal{A}$ is self-stabilizing iff
$\forall x \ \exists t \ Pr(X_t \in \mathcal{L} | X_0 = x) > 0. \qquad (\Diamond).$
Given a Markov chain $\mathcal{A} = (X_t)_{t=0}^{\infty}$ and a closed set $\mathcal{L}$ of configurations, we are interested in methods for proving the self-stabilization property of $\mathcal{A}$ w.r.t. $\mathcal{L}$. We are also interested in evaluating the rate of convergence of $\mathcal{A}$ to $\mathcal{L}$. We will use two different measures of convergence: the "expected hitting time" and the "ε-absorption time".

The expected hitting time is the standard rate of convergence used in the self-stabilization community (see, e.g., [5], p. 118). It is the expected time for $\mathcal{A}$ to reach $\mathcal{L}$, starting from the "worst" configuration, i.e.:

Definition 3. *Given a Markov chain $\mathcal{A}$ and a set $\mathcal{L}$ of configurations, the* expected hitting time of $\mathcal{L}$ *(or more simply the* hitting time*) is:*
$$\mathbf{H}_{\mathcal{L}} = max_{x \in \Omega} \ E[H_{x\mathcal{L}}],$$
where $E[.]$ denotes expectation and
$$H_{x\mathcal{L}} = min\{t : \ X_t \in \mathcal{L} \mid X_0 = x\}.$$

We will also use a different rate of convergence, called here "ε-absorption time", that gives the time after which $\mathcal{L}$ has been reached with high probability.

Definition 4. *Given a Markov chain $\mathcal{A}$ and a closed set $\mathcal{L}$, the* time of ε-absorption by $\mathcal{L}$ *(or simply the* ε-absorption time*) is:*
$$\mathbf{\Theta}_{\mathcal{L}}(\varepsilon) = max_{x \in \Omega} \ \Theta_{x\mathcal{L}}(\varepsilon),$$
where $\Theta_{x\mathcal{L}}(\varepsilon) = min\{t : \ Pr(X_t \in \mathcal{L}) \geq 1 - \varepsilon \mid X_0 = x\}$.

So $\mathbf{\Theta}_{\mathcal{L}}(\varepsilon)$ is the minimal number of steps in which $\mathcal{A}$ reaches $\mathcal{L}$ with probability at least $1 - \varepsilon$. This notion is, for example, used in [9], for measuring the rate of convergence of the Iterated Prisoner's Dilemma. The notion is closed to the notion of "mixing time", that measures the number of steps after which the chain is ε-close of the "stationary distribution" of $\mathcal{A}$. [1] An upper bound on the mixing time is often computed by finding the "coupling time" (see, e.g., [23,26]), that is defined in Section 4.

Remark. Various notions of convergence rates are compared together in [1] and [20], but these studies concern only the case of "irreducible" chains where $\mathcal{L}$ and Ω coincide (all the configurations are legal and inter-connected) while, here, we are concerned with "reducible" chains where $\mathcal{L}$ is a strict subset of Ω.

Let us recall that a finite Markov chain always converges towards the set of "recurrent" configurations, that is the union of the ergodic sets (see, e.g., [17]). We have:

Proposition 1. *A Markov chain $\mathcal{A}$ on a finite set Ω is self-stabilizing w.r.t. the union of ergodic sets. Moreover: If $\mathcal{A}$ is self-stabilizing w.r.t. a subset $\mathcal{L}$ of Ω and $\mathcal{L}$ is ergodic, then $\mathcal{L}$ is the unique ergodic set.*

We mainly focus in this paper on Markov chains with a unique ergodic set. They correspond to the notion of "self-stabilizing" algorithms, as originally defined by Dijkstra in the deterministic framework [4], where all the

[1] Actually, the two notions coincide when the set $\mathcal{L}$ is reduced to a single configuration, as in the example of Iterated Prisoner's Dilemma.

legal configurations are inter-connected. It is generally easy to check that a given set $\mathcal{L}$ of configurations is ergodic for $\mathcal{A}$, as illustrated in Examples 3 and 4. What is difficult is to show the *uniqueness* of the ergodic set, i.e., the absence of any other ergodic set, besides $\mathcal{L}$: for example, for a mutual exclusion algorithm, the absence of any subset of "looping" configurations with two tokens.

Example 3. Consider Herman's algorithm in the case where N is odd. In a configuration, a "token" at position i ($1 \le i \le N$) corresponds to the presence of two contiguous states of the same value (00 or 11) at position $i-1$ and i. Since N is odd, any configuration contains always at least one token. It is easy to see that such a set is ergodic:
The set $\mathcal{L}$ is the set of the configurations with a single token. There are $2N$ such configurations: they are of the form $x_i = 01\cdots 010\mathbf{0}10\cdots 101$ where token 00 is at position i, or $x'_i = 10\cdots 101\mathbf{1}01\cdots 010$ where token 11 is at position i, for all $1 \le i \le N$. (Letters in bold indicate that they are subjet to randomized transitions.) Let us show that $\mathcal{L}$ is ergodic, i.e. closed and strongly connected.

Applying a transition to an arbitrary element of $\mathcal{L}$, say x_i, leads to the 'dual' element $x'_i = 10\cdots 101\mathbf{1}01\cdots 010$ with probability 1/2, where token 11 is at the same position, or to $x_{i+1} = 10\cdots 1010\mathbf{0}1\cdots 010$ with probability 1/2, where token 00 is at position one more right. This shows that $\mathcal{L}$ is closed (since x'_i and x_{i+1} belong to $\mathcal{L}$). Moreover, this shows that one can go from x_i to x'_i and x_{i+1} in one step; two elements of $\mathcal{L}$ are thus connected together within at most N steps. Hence $\mathcal{L}$ is strongly connected.

Example 4. In the Iterated Prisoner's Dilemma, the set $\mathcal{L}$ of legal configurations is the singleton made of the configuration $x^* = (+)^N$. Obviously, any action transforms x^* to itself. Hence, $\{x^*\}$ is trivially an ergodic set.

In the following, we assume that we are given a Markov chain $\mathcal{A}$ and an ergodic set $\mathcal{L}$, and we focus on the problem of proving the self-stabilization property of $\mathcal{A}$ w.r.t. $\mathcal{L}$. (The assumption of ergodicity for $\mathcal{L}$ will be relaxed in Section 7.) The following property will be useful.

Proposition 2. *Given a closed set $\mathcal{L}$, if $\mathbf{H}_\mathcal{L}$ is finite, then $\mathcal{A}$ is self-stabilizing w.r.t. $\mathcal{L}$.*

Proof. By contraposition. Suppose that $\mathcal{A}$ is not self-stabilizing. Then, from ($\diamondsuit$), we know that there exists $x \in \Omega$ such that

$$Pr(X_t \in \mathcal{L} | X_0 = x) = 0 \text{ for all } t \ge 0.$$

So $H_{x\mathcal{L}}$ takes always an infinite value. Therefore $E[H_{x\mathcal{L}}]$ is infinite, and so is $\mathbf{H}_\mathcal{L}$. □

In case $\mathcal{L}$ is not only closed, but ergodic, we have, using Prop. 1:

Proposition 3. *Given an ergodic set $\mathcal{L}$, if $\mathbf{H}_\mathcal{L}$ is finite, then $\mathcal{A}$ is self-stabilizing w.r.t. $\mathcal{L}$, and $\mathcal{L}$ is the unique ergodic set.*

4 Coupling

The method of "coupling" is an elementary probabilistic method for measuring the "agreement" time between the components of a stochastic process (see, e.g., [26,23]).

Definition 5. *A* coupling *for $\mathcal{A}$ is a Markov chain on $\Omega \times \Omega$ defining a stochastic process $(X_t, Y_t)_{t=1}^{\infty}$ with the properties*

1. *Each of the processes (X_t) and (Y_t) is a faithful copy of $\mathcal{A}$ (given initial configurations $X_0 = x$ and $Y_0 = y$).*
2. *If $X_t = Y_t$, then $X_{t+1} = Y_{t+1}$.*

Condition 1 ensures that each process, viewed in isolation, is just simulating $\mathcal{A}$ – yet the coupling may update X_t and Y_t simultaneously so that they will tend to move closer together, according to some notion of distance. Once the pair of configurations agree, condition 2 guarantees they agree from that time forward.

Definition 6. *Given a coupling (X_t, Y_t), the* (expected) coupling time *is:*

$$\mathbf{T} = max_{x \in \Omega, y \in \Omega}\ E[T_{x,y}],$$
where $T_{x,y} = min\{t : X_t = Y_t \mid X_0 = x, Y_0 = y\}$.

The coupling time is often computed as un upper bound on the mixing time, in order to show the property of "rapid mixing" for $\mathcal{A}$ (i.e, the fact that the mixing time is bounded above by a polynomial in N and $\ln(\frac{1}{\varepsilon})$). We show hereafter that the coupling time gives also an upper bound on the hitting time.

Theorem 1. *Given a Markov chain $\mathcal{A}$ and an ergodic set $\mathcal{L}$, if there exists a coupling of finite expected time $\mathbf{T}$, then:*

1. *The hitting time $\mathbf{H}_\mathcal{L}$ satisfies:* $\mathbf{H}_\mathcal{L} \le \mathbf{T}$.
2. *$\mathcal{L}$ is the unique ergodic set, and $\mathcal{A}$ is self-stabilizing w.r.t. $\mathcal{L}$.*

Proof. Let us suppose that there exists a coupling of finite expected $\mathbf{T}$, and let us show statements 1 and 2.

1. Recall that: $H_{x\mathcal{L}} = min\{t : X_t \in \mathcal{L} \mid X^0 = x\}$, and $T_{xy} = min\{t : X_t = Y_t \mid X^0 = x, Y^0 = y\}$. Suppose now that $y \in \mathcal{L}$. Then $Y_t \in \mathcal{L}$ since $\mathcal{L}$ is closed. Hence: $H_{x\mathcal{L}} \le T_{xy}$ for all $x \in \Omega, y \in \mathcal{L}$. And by taking the expectations, then the maxima of the two sides: $\mathbf{H}_\mathcal{L} \le \mathbf{T}$.
2. Uniqueness of $\mathcal{L}$ and self-stabilization of $\mathcal{A}$ follow from the finiteness of $\mathbf{H}_\mathcal{L}$ (statement 1) by Proposition 3.

□

5 Two Sufficient Criteria of Self-Stabilization

By Theorem 1, finding an upper bound on the time of coupling T allows us at once to prove the self-stabilization and to obtain an upper bound on the hitting time. Following classical results on mixing time (see e.g. [10]), we give hereafter two sufficient conditions for bounding the coupling time. In each case, this provides us additionally with an upper bound not only for the hitting time, but also for the ε-absorption time.

Theorem 2. *Given a Markov chain $\mathcal{A}$ and an ergodic set $\mathcal{L}$, suppose there exist a coupling (X_t, Y_t) and a function δ on $\Omega \times \Omega$ which takes values in $\{0, 1, \cdots, B\}$ such that:*

- *$\delta(X_t, Y_t) = 0$ iff $X_t = Y_t$, and*
- *there exists $\beta < 1$ such that, for all (X_t, Y_t):*
  $$E[\delta(X_{t+1}, Y_{t+1})] \leq \beta\delta(X_t, Y_t). \quad (1)$$

Then $\mathcal{L}$ is the unique ergodic set and $\mathcal{A}$ is self-stabilizing w.r.t. $\mathcal{L}$. Furthermore:

1. *The hitting time satisfies: $\mathbf{H}_{\mathcal{L}} \leq \frac{B}{1-\beta}$.*
2. *The ε-absorption time satisfies: $\mathbf{\Theta}_{\mathcal{L}}(\varepsilon) \leq \frac{\ln(B/\varepsilon)}{1-\beta}$.*

The proof of Theorem 2 relies on the following proposition:

Proposition 4. *Suppose that $D = (D_t)_{t=0}^{\infty}$ is a nonnegative stochastic process on $\{0, 1, \cdots, B\}$ such that $E[D_{t+1}] \leq \beta D_t$ (with $0 < \beta < 1$). Then if τ is the first time that $D_t = 0$, we have: $E[\tau] \leq B/(1-\beta)$.*

Proof. The process $Z(t) = (B - D_t) - (1-\beta)min(t, \tau)$ is a submartingale since $E[Z(t+1)] - Z(t) = D_t - E[D_{t+1}] - (1-\beta) \geq (1-\beta)(D_t - 1) \geq 0$. Moreover, τ is a stopping time for Z, and the differences $Z(t+1) - Z(t)$ are bounded. The Optional Stopping theorem for submartingales (see e.g., [27]) then applies, which yields: $E[Z_\tau] \geq Z_0$, i.e: $B - (1-\beta)E[\tau] \geq 0$. Hence: $E[\tau] \leq \frac{B}{1-\beta}$. □

Proof of Theorem 2. Let us consider an integer-valued function δ satisfying the assumptions of Theorem 2, and let us show statements 1 and 2. (The facts that $\mathcal{L}$ is the unique ergodic set, and $\mathcal{A}$ is self-stabilizing follow from statement 1, by Proposition 3.)

1. Consider two elements $x, y \in \Omega$, and the coupling (X_t, Y_t) starting from $(X_0, Y_0) = (x, y)$. Let D_t be the process defined by $D_t = \delta(X_t, Y_t)$ for $t \geq 0$. Since $\delta(X_t, Y_t) = 0$ iff $X_t = Y_t$, the quantity $T_{x,y}$ is the time required for D_t to reach 0. Consider the coupling (X_t, Y_t) which starts from $(X_0, Y_0) = (x, y)$. Therefore by Proposition 4, we have, for all $x, y \in \Omega$, $E[T_{x,y}] \leq B/(1-\beta)$. Now, from Theorem 1, we infer: $\mathbf{H}_{\mathcal{L}} \leq max_{x,y}\ E[T_{x,y}] \leq B/(1-\beta)$.
2. Since $E[\delta(X_{t+1}, Y_{t+1})] \leq \beta\delta(X_t, Y_t)$, we have $E[\delta(X_t, Y_t)] \leq \beta^t\delta(X_0, Y_0) \leq \beta^t B$. But, by Markov's inequality ($Pr(X \geq a) \leq E[X]/a$):
   $$Pr(\delta(X_t, Y_t) \geq 1) \leq E[\delta(X_t, Y_t)].$$
   Hence, for all $X_0, Y_0 \in \Omega$ and all $t > 0$:
   $$Pr(X_t \neq Y_t) = Pr(\delta(X_t, Y_t) > 0) = Pr(\delta(X_t, Y_t) \geq 1) \leq E[\delta(X_t, Y_t)] \leq \beta^t B.$$
   Therefore, for all $X_0, Y_0 \in \Omega$ and all $t > 0$:
   $$Pr(X_t = Y_t) \geq 1 - \beta^t B.$$
   Suppose that $Y_0 \in \mathcal{L}$. Then $Y_t \in \mathcal{L}$ (because $\mathcal{L}$ closed), and $X_t = Y_t$ implies $X_t \in \mathcal{L}$. So, for all $X_0 \in \Omega$ and all $t > 0$: $Pr(X_t \in \mathcal{L}) \geq 1 - \beta^t B$. It follows that $Pr(X_t \in \mathcal{L}) \geq 1 - \varepsilon$, as soon as $\beta^t B \leq \varepsilon$, i.e., $t \geq \frac{\ln(B/\varepsilon)}{\ln(1/\beta)}$. Hence $Pr(X_t \in \mathcal{L}) \geq 1 - \varepsilon$, as soon as $t \geq \frac{\ln(B/\varepsilon)}{1-\beta}$ (because $1 - \beta \leq \ln(\frac{1}{\beta})$). □

A similar theorem exists even when $\beta = 1$, i.e.: $E[\delta(X_{t+1}, Y_{t+1})] \leq \delta(X_t, Y_t)$, provided that the probability of $\delta(X_{t+1}, Y_{t+1}) \neq \delta(X_t, Y_t)$ can be bounded below.

Theorem 3. *Given a Markov chain $\mathcal{A}$ and an ergodic set $\mathcal{L}$, suppose there exist a coupling (X_t, Y_t) and a function δ on $\Omega \times \Omega$ which takes values in $\{0, 1, \cdots, B\}$ such that:*

- *$\delta(X_t, Y_t) = 0$ iff $X_t = Y_t$, and*
- *there exists $\alpha > 0$ such that, for all (X_t, Y_t) with $X_t \neq Y_t$:*
  $$E[\delta(X_{t+1}, Y_{t+1})] \leq \delta(X_t, Y_t) \wedge Pr(\delta(X_{t+1}, Y_{t+1}) \neq \delta(X_t, Y_t)) \geq \alpha. \quad (2)$$

Then $\mathcal{L}$ is the unique ergodic set and $\mathcal{A}$ is self-stabilizing w.r.t. $\mathcal{L}$. Furthermore:

1. *The hitting time satisfies: $\mathbf{H}_{\mathcal{L}} \leq B^2/\alpha$.*
2. *The ε-absorption time satisfies:*
   $$\mathbf{\Theta}_{\mathcal{L}}(\varepsilon) \leq \lceil e\frac{B^2}{\alpha}\rceil \lceil \ln(\tfrac{1}{\varepsilon})\rceil.$$

The proof of Theorem 3 is analogous to that of Theorem 2, but relies on the following proposition (whose proof follows that given in [10]; cf [21]):

Proposition 5. *Suppose that $D = (D_t)_{t=0}^{\infty}$ is a nonnegative stochastic process on $\{0, 1, \cdots, B\}$ such that $E[D_{t+1}] \leq D_t$. Furthermore suppose that $Pr(D_{t+1} \neq D_t) \geq \alpha$ (with $\alpha > 0$) when $D_t > 0$. Then if τ is the first time that $D_t = 0$, we have: $E[\tau] \leq B^2/\alpha$.*

Proof. The process $Z(t) = (B - D_t)^2 - \alpha t$ is a submartingale since $E[(D_{t+1} - D_t)^2] \geq \alpha$. (We have: $E[(D_{t+1} - D_t)^2] \geq Pr((D_{t+1} - D_t)^2 \geq 1) = Pr(D_{t+1} \neq D_t) \geq \alpha$.) Moreover, τ is a stopping time for Z, and the differences $Z(t+1) - Z(t)$ are bounded. The Optional Stopping theorem for submartingales then applies: $E[Z_\tau] = B^2 - \alpha E[\tau] \geq Z_0 = (B - D_0)^2$. Hence:
$E[\tau] \leq \frac{1}{\alpha}(B^2 - (B - D_0)^2) \leq \frac{B^2}{\alpha}$. □

Proof of Theorem 3. Let us consider an integer-valued function δ satisfying the assumptions of Theorem 3, and let us show statements 1 and 2. (The facts that $\mathcal{L}$ is the unique ergodic set, and $\mathcal{A}$ is self-stabilizing follow from statement 1, by Proposition 3.)

1. Consider two elements x, y of Ω and a coupling (X_t, Y_t) of initial element $(X_0, Y_0) = (x, y)$. Let $D_t = \delta(X_t, Y_t)$ for $t \geq 0$. Since $\delta(X_t, Y_t) = 0$ iff $X_t = Y_t$, the quantity $T_{x,y}$ is the time required for D_t to reach 0. Therefore by Proposition 5, we have, for all $x, y \in \Omega$, $E[T_{x,y}] \leq B^2/\alpha$. Now, from Theorem 1, we infer: $\mathbf{H}_{\mathcal{L}} \leq max_{x,y}\ E[T_{x,y}] \leq B^2/\alpha$.
2. Let $D_t = \delta(X_t, Y_t)$. It is easy to check that D_t satisfies the conditions of Prop. 5. Recall that $T_{x,y}$ is the first time that $X_t = Y_t$, hence $D_t = 0$, when $X_0 = x$ and $Y_0 = y$. It follows by Prop. 5:
   $E[T_{x,y}] \leq B^2/\alpha$. Let $T = \lceil eB^2/\alpha \rceil$, then by Markov's inequality, we have the probability that $T_{x,y} > T$ is at most e^{-1}. If we run s independent trials of length T then the probability that X_t and Y_t are not coupled by the end of the sT is at most e^{-s}. Therefore, for $t > T\lceil \ln(\varepsilon^{-1})\rceil$, the probability of $X_t = Y_t$ is at least

$1-\varepsilon$. Suppose that $Y_0 \in \mathcal{L}$. Then $Y_t \in \mathcal{L}$ (because $\mathcal{L}$ closed), and $X_t = Y_t$ implies $X_t \in \mathcal{L}$. So, for all $X_0 \in \Omega$ and all $t > T\lceil \ln(\varepsilon^{-1})\rceil$: $Pr(X_t \in \mathcal{L}) \geq 1-\varepsilon$. □

Therefore finding a coupling (X_t, Y_t) and a function δ such that (1) (resp. (2)) holds allows us to prove that $\mathcal{A}$ is self-stabilizing, and gives us an upper bound on two different rates of convergence.

6 Refinement of Coupling

6.1 Path Coupling

As pointed out in [23], it is often cumbersome to measure the expected change in distance between two arbitrary configurations. The method of *path coupling*, introduced by Bubley and Dyer [3], simplifies the approach by showing that only pairs of configurations that are "close" need to be considered. Path coupling involves defining a coupling (X_t, Y_t) by considering a *path*, or sequence $X_t = Z_0, Z_1, \cdots, Z_r = Y_t$ between X_t and Y_t where the Z_i satisfy certain conditions. The following version of the path coupling method is convenient:

Lemma 1. (Dyer and Greenhill [10]) *Let δ be a metric*[2] *defined on $\Omega \times \Omega$ which takes value in $\{0, \cdots, B\}$. Let U be a subset of $\Omega \times \Omega$ s.t.:*

For all $(X_t, Y_t) \in \Omega \times \Omega$, there exists a path $X_t = Z_0, Z_1, \cdots, Z_r = Y_t$ between X_t and Y_t such that $(Z_i, Z_{i+1}) \in U$ for $0 \leq i < r$ and $\sum_{i=0}^{r-1} \delta(Z_i, Z_{i+1}) = \delta(X_t, Y_t)$. (3)

Suppose there exist a coupling $(X, Y) \mapsto (X', Y')$ for the Markov chain $\mathcal{A}$ on all pairs $(X, Y) \in U$, and a constant $\beta \leq 1$ such that, for all $(X, Y) \in U$:

$E[\delta(X', Y')] \leq \beta\delta(X, Y)$. (4)

Then this coupling can be extended to a coupling for $\mathcal{A}$ on $\Omega \times \Omega$, which satisfies (4) for all $(X, Y) \in \Omega \times \Omega$.

Two configurations X and Y are said to be *adjacent* if $(X, Y) \in U$. The advantage of this lemma is that it allows us to check the crucial property (4) on the set U of adjacent pairs only, rather than on the entire space $\Omega \times \Omega$. Lemma 1 combined with Theorem 2 (resp. Theorem 3) allows us to enhance our coupling method for proving self-stabilization.

6.2 Application to Herman

Let us come back to Herman's algorithm (see Example 1).

Theorem 4. *For Herman's algorithm and N odd, there exist a subset U of $\Omega \times \Omega$, a metric δ on $\Omega \times \Omega$ taking value in $\{0, \cdots, N\}$ and satisfying condition (3), and a coupling such that:*

- $\forall (X_t, Y_t) \in U \quad E[\delta(X_{t+1}, Y_{t+1})] \leq \delta(X_t, Y_t)$, *and*
- $\forall (X_t, Y_t) \in \Omega \times \Omega$ *(with $X_t \neq Y_t$)* :
  $Pr(\delta(X_{t+1}, Y_{t+1}) \neq \delta(X_t, Y_t)) \geq 1/2$.

[2] I.e., a function such that: $\delta(X, Y) = 0$ iff $X = Y$, and $\delta(X, Z) \leq \delta(X, Y) + \delta(Y, Z)$, for all $X, Y, Z \in \Omega$.

Proof.

- *Subset U and metric δ.* We define δ as the Hamming distance: $\delta(X_t, Y_t)$ is the number of positions at which X_t and Y_t differ. The pair (X_t, Y_t) belongs to U iff $\delta(X_t, Y_t) = 1$. It is immediate to check condition (3) of Lemma 1.
- *Coupling.* The coupling is defined in order to force X_t and Y_t to do the same probabilistic choice, when they both have to perform a random action. In other words, for all i ($1 \leq i \leq N$):
  If $X_t(i) = X_t(i-1)$ and $Y_t(i) = Y_t(i-1)$ then
  $$X_{t+1}(i) = Y_{t+1}(i) = \begin{cases} 0 & \text{with probability } 1/2, \\ 1 & \text{with probability } 1/2. \end{cases}$$
- *Proof of $E[\delta(X_{t+1}, Y_{t+1})] = \delta(X_t, Y_t)$ on U.* Consider a pair $(X_t, Y_t) \in U$, and let ℓ be the position of disagreement between X_t and Y_t. In order to fix the ideas consider the following vector
  $$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} \nu_1 \ \nu_2 \ \cdots \ \nu_{\ell-2} \ 0 \ \mathbf{0} \ 0 \ \nu_{\ell+2} \ \cdots \ \nu_N \\ \nu_1 \ \nu_2 \ \cdots \ \nu_{\ell-2} \ 0 \ \mathbf{1} \ 0 \ \nu_{\ell+2} \ \cdots \ \nu_N \end{pmatrix}$$
  where all the ν_i are in $\{0, 1\}$, the figures in bold font correspond to positions ℓ. (The other cases are similar.) After one step, the state of all the machines at position $1, \cdots, N$ are updated. We have:
  $$\begin{pmatrix} X_{t+1} \\ Y_{t+1} \end{pmatrix} = \begin{pmatrix} \nu'_1 \ \nu'_2 \ \cdots \ \nu'_{\ell-2} \ \nu'_{\ell-1} \ ? \ ? \ \nu'_{\ell+2} \ \cdots \ \nu'_N \\ \nu'_1 \ \nu'_2 \ \cdots \ \nu'_{\ell-2} \ \nu'_{\ell-1} \ \mathbf{0} \ 1 \ \nu'_{\ell+2} \ \cdots \ \nu'_N \end{pmatrix}$$
  where '?' means "0 with prob. 1/2 and 1 with prob. 1/2". Note that, for $1 \leq i \leq \ell-1$ and $\ell+2 \leq i \leq N$, $X_{t+1}(i) = Y_{t+1}(i) = \nu'_i$ thanks to our coupling. So X_{t+1} and Y_{t+1} coincide everywhere except, perhaps, at positions ℓ or $\ell+1$. We have:
  $$\delta(X_{t+1}, Y_{t+1}) = \begin{cases} 0 & \text{with probability } 1/4, \\ 1 & \text{with probability } 1/2, \\ 2 & \text{with probability } 1/4. \end{cases}$$
  Hence $E[\delta(X_{t+1}, Y_{t+1})] = \delta(X_t, Y_t)$, for all $(X_t, Y_t) \in U$.
- *Proof of $Pr(\delta(X_{t+1}, Y_{t+1}) \neq \delta(X_t, Y_t)) \geq 1/2$.* Let us denote by q the number of disagreeing tokens (a disagreeing token is a position i such that $X_t(i-1) = X_t(i) \neq Y_t(i-1) = Y_t(i)$) and by p the number of zones of contiguous disagreeing positions. Let us identify the three sources of possible evolution of the set of disagreeing positions after a step:
  1. Thanks to the coupling, each disagreeing token $X_t(i-1) = X_t(i) \neq Y_t(i-1) = Y_t(i)$ evolves in a new agreeing position $X_{t+1}(i) = Y_{t+1}(i)$ with probability 1.
  2. Each first position in a disagreeing zone, say i, such that $X_t(i-1) = Y_t(i-1)$ and $X_t(i) \neq Y_t(i)$ can evolve in an agreeing position with probability 1/2. We denote by r the number of such i ($0 \leq r \leq p$).

3. Each first position in agreeing zone, say i, such that $X_t(i-1) \neq Y_t(i-1)$ and $X_t(i) = Y_t(i)$ can evolve in a disagreeeing position with probability $1/2$. We denote by s the number of such i ($0 \leq r \leq p$).

Cases 2 and 3 are depicted on Figure 3.

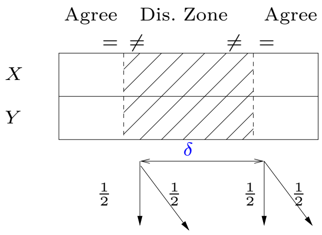

Fig. 3. Evolution of a disagreeing zone

We have: $\delta(X_{t+1}, Y_{t+1}) = \delta(X_t, Y_t) - q - r + s$. Therefore the event $\delta(X_{t+1}, Y_{t+1}) = \delta(X_t, Y_t)$ corresponds to all the cases where $q + r = s$. If $q > p$, such an event can never occur (probability 0). Otherwise, its probability is
$\frac{1}{4^p}\sum_{r=0}^{p-q}\binom{p}{r}\binom{p}{q+r} = \frac{1}{4^p}\sum_{r=0}^{p-q}\binom{p}{r}\binom{p}{p-q-r} \leq$
$\frac{1}{4^p}\binom{2p}{p-q} \leq \frac{1}{2^{2p}}\binom{2p}{p}$ (by Vandermonde's convolution [13])
$\leq \frac{1}{2}$ (by induction on p).
Hence $Pr(\delta(X_{t+1}, Y_{t+1}) \neq \delta(X_t, Y_t)) \geq \frac{1}{2}$.

□

Since $\delta(X_t, Y_t)$ takes its values in $\{0, 1, \cdots, N\}$, it then follows from Theorem 3, Lemma 1 and Theorem 4:

Corollary 1. *For N odd, Herman's algorithm is self-stabilizing w.r.t. the set $\mathcal{L}$ of configurations with a single token. Furthermore:*

1. *The hitting time satisfies:* $\mathbf{H}_{\mathcal{L}} \leq 2N^2$.
2. *The ε-absorption time satisfies:* $\mathbf{\Theta}_{\mathcal{L}}(\varepsilon) \leq 2eN^2\lceil\ln(\frac{1}{\varepsilon})\rceil$.

Note that the metric δ on $\Omega \times \Omega$ found here (Hamming distance) is much simpler than the function φ on Ω used by Herman, which involves the number of tokens of a configuration x together with the minimal distance between two tokens of x. Our method gives also directly an upper bound for the hitting time with no need for a separate analysis as done in Herman's work [14]. Besides, it gives a quadratic bound for the ε-absorption time (not considered by Herman).

The method can be applied in the same manner to several other self-stabilizing algorithms on cyclic graphs (e.g., mutual exclusion Flatebo-Datta's algorithm [11] with central randomized scheduler, Mayer-Ostrovsky-Yung's binary clock algorithm with synchronous scheduler [22]).

6.3 Application to Iterated Prisoner's Dilemma

Let us now consider Iterated Prisoner's Dilemma algorithm (Example 2). Recall that, in this case, the set $\mathcal{L}$ made of the unique configuration x^*, with $x^*(i) = +$ for all $1 \leq i \leq N$, is ergodic. Let us show that the algorithm is self-stabilizing.

Theorem 5. *For the Prisoner's Dilemma algorithm, there exist a subset U of $\Omega \times \Omega$, a metric δ on $\Omega \times \Omega$ taking value in $\{0, \cdots, 11N\}$ and satisfying condition (3), and a coupling such that, for all $(X_t, Y_t) \in U$:*
$$E[\delta(X_{t+1}, Y_{t+1})] \leq (1 - \tfrac{2}{29N})\delta(X_t, Y_t).$$

Proof.

- *Subset U.* Let $U = \cup_{1\leq k\leq 5} U_k$ where U_k ($1 \leq k \leq 5$) is the set of pairs (X, Y) such that X and Y coincide everywhere except on k contiguous positions where they differ. For any pair $(X, Y) \in U_k$ ($1 \leq k \leq 5$), let $\delta(X, Y) = a_k$ where a_k is a positive constant that will be determined later. By convention: $a_0 = 0$. The exact way of extending δ from U to the entire space $\Omega \times \Omega$ will be also explained later on.
- *Coupling.* The coupling $(X, Y) \mapsto (X', Y')$ is such that, at each step, the position chosen uniformly at random coincides for X and Y. (So, at each step, the state of the machine of the selected position, say j, and the state of the $j+1$-th machine are updated simultaneously in X and Y.)
- *Proof of $E[\delta(X', Y')] \leq \beta\delta(X, Y)$ with $\beta < 1$.* Let us show that, for appropriate values of a_k ($1 \leq k \leq 5$), and appropriate definition of $f(6)$, δ satisfies $E[\delta(X', Y')] \leq \beta\delta(X, Y)$ (with $\beta < 1$) for all $(X, Y) \in U$. Consider a pair $(X, Y) \in U$ with k contiguous disagreeing positions. Let i be the first disagreeing position. The vector
$$\begin{pmatrix} X \\ Y \end{pmatrix}$$
is of the form
$$\begin{pmatrix} \gamma_1 & \cdots & \gamma_{i-2} & \gamma_{i-1} & \gamma_i & \cdots & \gamma_{i+k-1} & \gamma_{i+k} & \cdots & \gamma_N \\ \gamma_1 & \cdots & \gamma_{i-2} & \gamma_{i-1} & \neg\gamma_i & \cdots & \neg\gamma_{i+k-1} & \gamma_{i+k} & \cdots & \gamma_N \end{pmatrix}$$
with $\gamma_\ell \in \{-, +\}$ for all $1 \leq \ell \leq N$. Suppose that the selected position j is such that $1 \leq j \leq i-2$ or $i+k \leq j \leq N$. Then $X(j) = Y(j)$ and $X(j+1) = Y(j+1)$, so $X'(j) = Y'(j)$ and $X'(j+1) = Y'(j+1)$, and the disagreement zone is not modified. Suppose now that the selected position j is equal to $i-1$. Then, after one step, we have:
$$\begin{pmatrix} X' \\ Y' \end{pmatrix}$$
is of the form
$$\begin{pmatrix} \gamma_1 & \cdots & \gamma_{i-2} & \gamma'_{i-1} & \gamma'_i & \gamma_{i+1} & \cdots & \gamma_{i+k-1} & \gamma_{i+k} & \cdots & \gamma_N \\ \gamma_1 & \cdots & \gamma_{i-2} & \neg\gamma'_{i-1} & \neg\gamma'_i & \neg\gamma_{i+1} & \cdots & \neg\gamma_{i+k-1} & \gamma_{i+k} & \cdots & \gamma_N \end{pmatrix}$$
where $\gamma'_{i-1} = \gamma'_i = +$ if $\gamma_{i-1} = \gamma_i$, and $\gamma'_{i-1} = \gamma'_i = -$ otherwise. This means that the disagreement zone has progressed one position at the left. A symmetrical case exists for $j = i+k-1$. We say that j is an "outer rim position". All the other possible cases for j are studied below.

Consider $(X, Y) \in U$. Let $[i, i+k-1]$ be the interval of contiguous disagreeing positions between X and Y (with $1 \leq k \leq 5$). The random choice of the selected machine j modifies the zone of disagreement iff j corresponds to:

- *Outer rim position:* This means that $j = i - 1$ or $j = i + k - 1$. There are two outer rim positions for every $1 \leq k \leq 5$. Choosing an outer rim position extends the disagreement zone by one. This happens with probability $2/N$, and contributes to modify $E[\delta]$ by: $\frac{2}{N}(a_5 - a_4)$ for $k = 4$, and $\frac{2}{N}(f(6) - a_5)$ for $k = 5$. [3]
- *Inner rim position:* This means that $j = i$ or $j = i+k-2$. There are no inner rim position if $k = 1$, one inner rim position if $k = 2$, and two inner rim positions if $k = 3, 4, 5$. Choosing an inner rim position decreases the disagreement zone by two. This happens with probability $1/N$ (resp. $2/N$) when $k = 2$ (resp. $k = 3, 4, 5$). It contributes to modify $E[\delta]$ by $\frac{1}{N}(a_0 - a_2) = -\frac{1}{N}a_2$ when $k = 2$, and by $\frac{2}{N}(a_{k-2} - a_k)$ when $k = 3, 4, 5$.
- *Internal position:* This means that $j = i + 1$ or $j = i + k - 3$. There are no internal position if $k = 1, 2$ or 3, one internal position if $k = 4$, and two internal positions if $k = 5$. For $k = 4$, choosing an internal position ($j = i + 1$) transforms the disagreement zone into two separated disagreement zones of length 1. This happens with probability $1/N$, and contributes to modify $E[\delta]$ by: $\frac{1}{N}(2a_1 - a_4)$. For $k = 5$, choosing an internal position ($j = i+1$ or $j = i+2$) transforms the disagreement zone into two separated disagreement zones of length 1 and 2. This happens with probability $2/N$, and contributes to modify $E[\delta]$ by: $\frac{2}{N}(a_1 + a_2 - a_5)$.

Accordingly, we have the following cases:

1. *Case k=1.* Then:
   $E[\delta(X', Y')] - \delta(X, Y) = \frac{2}{N}(a_2 - a_1)$.
   Hence $E[\delta(X', Y')] = \beta_1 \delta(X, Y)$
   with $\beta_1 = 1 - \frac{2}{N}\frac{a_1 - a_2}{a_1}$
   (using the fact that $\delta(X, Y)$ is equal here to a_1).
2. *Case k=2.* Then:
   $E[\delta(X', Y')] - \delta(X, Y)$
   $= \frac{1}{N}(2(a_3 - a_2) + (a_0 - a_2) = \frac{1}{N}(2a_3 - 3a_2)$.
   Hence $E[\delta(X', Y')] = \beta_2 \delta(X, Y)$
   with $\beta_2 = 1 - \frac{1}{N}\frac{3a_2 - 2a_3}{a_2}$
   (using the fact that $\delta(X, Y)$ is equal here to a_2).
3. *Case k=3.* Then:
   $E[\delta(X', Y')] - \delta(X, Y)$
   $= \frac{2}{N}((a_4 - a_3) + (a_1 - a_3))$
   $= \frac{2}{N}(a_4 - 2a_3 + a_1)$.
   Hence $E[\delta(X', Y')] = \beta_3 \delta(X, Y)$
   with $\beta_3 = 1 - \frac{2}{N}\frac{2a_3 - a_4 - a_1}{a_3}$
   (using the fact that $\delta(X, Y)$ is equal here to a_3).
4. *Case k=4.* Then:
   $E[\delta(X', Y')] - \delta(X, Y)$
   $= \frac{1}{N}(2(a_5 - a_4) + 2(a_2 - a_4) + (2a_1 - a_4))$
   $= \frac{1}{N}(2a_5 - 5a_4 + 2a_2 + 2a_1)$.
   Hence $E[\delta(X', Y')] = \beta_4 \delta(X, Y)$
   with $\beta_4 = 1 - \frac{1}{N}\frac{5a_4 - 2a_5 - 2a_2 - 2a_1}{a_4}$
   (using the fact that $\delta(X, Y)$ is equal here to a_4).
5. *Case k=5.* Then:
   $E[\delta(X', Y')] - \delta(X, Y)$
   $= \frac{2}{N}((f(6) - a_5) + (a_3 - a_5) + (a_1 + a_2 - a_5))$
   $= \frac{2}{N}(-3a_5 + f(6) + a_3 + a_2 + a_1)$.
   Hence $E[\delta(X', Y')] = \beta_5 \delta(X, Y)$
   with $\beta_5 = 1 - \frac{2}{N}\frac{3a_5 - f(6) - a_3 - a_2 - a_1}{a_5}$
   (using the fact that $\delta(X, Y)$ is equal here to a_5).

We have now to find $a_1, \cdots, a_5$ and a definition of f such that, for all $1 \leq k \leq 5$, β_k satisfies $0 < \beta_k < 1$. For any $p \geq 1$, let q and r be the quotient and remainder of p divided by 5; then:

$f(p) = qa_5 + a_r$ if $r \neq 1$,
$f(p) = (q-1)a_5 + a_4 + a_2$ if $p \neq 1 \wedge r = 1$.
$f(p) = a_1$ if $p = 1$.

So we have: $f(6) = a_4 + a_2$. [4] The coefficient β_5 of the equation $E[\delta(X', Y')] = \beta_5 \delta(X, Y)$ becomes
$\beta_5 = 1 - \frac{2}{N}\frac{3a_5 - a_4 - a_3 - 2a_2 - a_1}{a_5}$
A possible solution for solving $\{\beta_k < 1\}_{1 \leq k \leq 5}$, is now: $a_1 = 21, a_2 = 20, a_3 = 29, a_4 = 36, a_5 = 43$.
It follows $\beta \leq 1 - \frac{2}{29N}$, hence $\frac{1}{1-\beta} \leq \frac{29}{2}N$.

- *Function δ.* Let us now define a function δ on the whole space $\Omega \times \Omega$. The rough idea is to define $\delta(X, Y)$ as $\sum_{p=1}^{n} f(|W_p|)$, where $W_1, ..., W_n$ are maximal zones of contiguous positions where X and Y disagree. However, we have to correct this definition by taking into account the special case of two (or more) consecutive disagreement zones of length 1, separated only by one position. therefore we have to distinguish between maximal disagreement zones of length ≥ 2, and the other ones. Formally, a *disagreement zone of length $p \geq 2$* is a maximal sequence of indices $i, i+1, ..., i+p-1$ such that X and Y disagree on $i, i+1, ..., i+p-1$. An *alternating sequence of length $2p+1$ (with $p \geq 0$)* is a maximal sequence of indices $(i, i+1, ..., i+2p)$ such that X and Y disagree on $i, i+2, i+4, ..., i+2p$, and agree on $i+1, i+3, ..., i+2p-1$. Note that the length of an alternating sequence is an odd number. Given X and Y, let $\mathcal{A}$ be the set of alternating sequences, and $\mathcal{B}$ the set of disagreement zones of length ≥ 2. We define, for all $(X, Y) \in \Omega \times \Omega$:
  $\delta(X, Y) = \sum_{W \in \mathcal{B}} f(|W|) + \sum_{V \in \mathcal{A}} g(|V|)$,
  where g is defined on odd numbers as follows:
  $g(2q+1) = (q+1)a_2$ if q odd,

[3] This assumes that, for any pair of configurations X and Y that coincide everywhere except on 6 contiguous positions, $\delta(X, Y) = f(6)$; this assumption will be checked a posteriori when the formal definition of δ is given.

[4] In [12], we defined f in a simpler but erroneous manner by: $f(p) = qa_5 + a_r$, including the case where $r = 1$. With such a definition, $f(6) = a_5 + a_1$, and a possible solution for solving $\{\beta_k < 1\}_{1 \leq k \leq 5}$, is: $a_1 = 21, a_2 = 20, a_3 = 29, a_4 = 36, a_5 = 48$. The problem comes then from the fact: $f(6) = a_5 + a_1 > a_4 + a_2 = f(4) + f(2)$, which prevents δ to be a metric by violation of the triangular inequality.

$g(2q+1) = qa_2 + a_1$ if q even.
For example, for

$$\begin{pmatrix} 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0 \\ 0\,1\,0\,1\,0\,1\,1\,1\,0\,1\,0 \end{pmatrix}$$

there is an alternating sequence of length 3 (from position 2 to 4), a disagreement zone of length 3 (from position 6 to 8), and an alternating sequence of length 1 (10-th position), which gives: $\delta = g(3) + f(3) + g(1) = 2a_2 + a_3 + a_1$. Let us observe that, for any pair of configurations X and Y that coincide everywhere except on 6 contiguous positions, $\delta(X,Y) = f(6) = a_4 + a_2$, as assumed before.

- *Proof that δ satisfies* (3). Let us show condition (3): for all $x, y \in \Omega\times\Omega$, there is a path $x = z_0, z_1, \cdots, z_r = y$ with $(z_j, z_{j+1}) \in U$ (for all $1 \leq j < r$) and $\sum_{j=0}^{r-1} \delta(z_j, z_{j+1}) = \delta(x,y)$, where, for all $0 \leq j < r$, z_{j+1} is obtained from z_j by complementing exactly k contiguous states, for some $1 \leq k \leq 5$. Given $x, y \in \Omega$, let $W_1, ..., W_n$ (resp. $V_1, ..., V_m$) be the disagreement zones of length ≥ 2 (resp. alternating sequences) between x and y. The first step of the process consists in selecting a disagreement zone, say W_1, of z_0 if such a zone exists ($n \geq 1$), or an alternating sequence, say V_1, otherwise. Then z_1 is produced by complementing the states of z_0 corresponding to the k leftmost positions of W_1 (resp. V_1) for some $1 \leq k \leq 5$. More precisely, in case where W_1 is selected, z_1 is constructed by complementing the states of the 5 (resp. 4) leftmost positions of W_1 when $|W_1| \geq 7$ (resp. $|W_1| = 6$), or complementing all the states of W_1 otherwise (i.e., when $2 \leq |W_1| \leq 5$). If V_1 is selected, then z_1 is constructed from z_0 by complementing the states of the two leftmost positions of V_1 if $|V_1| \geq 3$, or complementing the state of the unique position of V_1 otherwise ($|V_1| = 1$). It is easy to see that, in all the cases, $\delta(z_1, y) = \delta(z_0, y) - \delta(z_0, z_1)$. The process is then applied to z_1, and so on iteratively until a step r for which $\delta(z_r, y) = 0$, i.e $z_r = y$. We have: $\sum_{j=0}^{r-1} \delta(z_j, z_{j+1}) = \delta(x,y)$. Let us observe that, given x and y, the choice of the disagreement zone (resp. alternating sequence) at each step modifies only the order in which the states are complemented from x to y, but does not affect the length r of the constructed path .
- *Proof that δ is a metric.* One can now show that δ is a metric, i.e.:
  1. $\forall x, y \in \Omega : \ \delta(x,y) = 0$ iff $x = y$.
  2. $\forall x, y, z \in \Omega : \ \delta(x,y) \leq \delta(x,z) + \delta(z,y)$.

  The first item holds because all the coefficients a_i are positive. The proof of the second item is done by induction on the length r of the path $z = z_0, z_1, z_2, ..., z_r = y$ satisfying (3) linking z to y (whose construction has been explained above). Let us show the induction step: $\delta(x,z) + \delta(z,y)$
  $= \delta(x,z) + (\delta(z,z_1) + \delta(z_1,y))$
  $= (\delta(x,z) + \delta(z,z_1)) + \delta(z_1,y)$
  $\geq \delta(x,z_1) + \delta(z_1,y)$ (by base case)
  $\geq \delta(x,y)$ (by induction hypothesis).
  It remains to show the base case, i.e.: $\delta(x,z) + \delta(z,z_1) \geq \delta(x,z_1)$, where z_1 has been obtained from z by complementing a zone, say D, of k contiguous states (for some $1 \leq k \leq 5$). Note that we have $\delta(z,z_1) = a_k$. Let $\ell, ..., \ell+k-1$ be the positions of D. Let $W_1, .., W_n$ (resp. $V_1, ..., V_m$) be the disagreement zones (resp. alternating sequences) of z with x. The proof is a tedious but simple case analysis according to the various possible positions of $\ell, ..., \ell+k-1$ with respect to $W_1, ..., W_n, V_1, ..., V_m$.
  The easy case is when D and the W_is, V_js do not interact together: no position of D coincides with or is next to any position of W_i or V_j. In this case, $\delta(x,z_1) = \delta(x,z) + \delta(z,z_1)$.
  A more difficult case is when D is next to the right of a disagreement zone (of length ≥ 2), say W_1: the rightmost position of W_1 is $\ell - 1$ (but no other position of D coincides with or is next to any position of W_i or V_j). Then, the disagreement zones and alternating sequences of z_1 w.r.t. x are the same as those of z, except that W_1 is replaced by the concatenation W_1D of W_1 and D. The inequality $\delta(x,z_1) \leq \delta(x,z) + \delta(z,z_1)$ then reduces to $f(|W_1D|) = f(|W_1| + k) \leq f(|W_1|) + a_k$. The latter inequality holds because, for all $s \geq 2$ and all $1 \leq k \leq 5$, we have: $f(s+k) \leq f(s) + a_k$.
  Another typical case is when D overlaps with W_1: the rightmost position of W_1 is $\ell + b - 1$ with $b \geq 1$ (but no other position of D coincides with or is next to any position of W_i or V_j). There are b positions in common to W_1 and D, s positions proper to W_1 on the left, and t positions proper to D on the right. We have: $|W_1| = s + b$ and $|D| = b + t$, where s, b and t are numbers (with $b + t \leq 5$). Let us assume that s and t are positive. The inequality $\delta(x,z_1) \leq \delta(x,z) + \delta(z,z_1)$ then reduces to $f(s) + f(t) \leq f(s+b) + f(t+b)$ in the case where s, t or $b \geq 2$, and to $g(3) \leq f(2) + f(2)$ in case $s = t = b = 1$. Both inequalities are easily checked (the second inequality is simply $2a_2 \leq 2a_2$).
  The analysis of all the other cases where D interacts with one (or more) disagreement zone or alternating sequence of z is similar.

Finally, let us note that the maximal value B of δ on $\Omega \times \Omega$ is at most $a_1\lceil\frac{N}{2}\rceil \leq 11N$. As a recapitulation, there exists a metric δ such that $E[\delta(X', Y')] \leq \beta\delta(X,Y)$, for $\beta \leq 1 - \frac{2}{29N}$. Furthermore, for these values of a_k, the maximal value B of δ on $\Omega \times \Omega$ is such that $B \leq 11N$. □

Therefore, from Theorem 2, Lemma 1 and Theorem 5, it follows:

Corollary 2. *Iterated Prisoner's Dilemma algorithm is self-stabilizing w.r.t. the set $\mathcal{L} = \{(+)^N\}$. Futhermore:*
1. *The hitting time satisfies:* $\mathbf{H}_{\mathcal{L}} \leq \frac{319}{2}N^2$.
2. *The ε-absorption time satisfies:*
   $\mathbf{\Theta}_{\mathcal{L}}(\varepsilon) \leq \frac{29}{2}N\ln(\frac{11N}{\varepsilon})$.

Thus the quasi-linear bound on the ε-time of absorption is obtained, as found in [9]. We retrieve also the quadratic bound on the hitting time found empirically in [18]. The proof presented here bears some resemblance with the proof by Dyer et al. in [9]: A function δ has been found here on $\Omega \times \Omega$ which satisfies $E\delta(X', Y') \leq \beta\delta(X, Y)$ (with $\beta < 1$), while they found a function φ on Ω satisfying $E\varphi(X') \leq \beta'\varphi(X)$ (with $\beta' < 1$) and $\varphi(\mathcal{L}) = 0$. Note that their function φ is simpler than δ (φ mainly involves isolated singletons ($-$) and doublets ($--$)). However, we obtain here a better ε-absorption time (linear factor $\beta = 18 < \beta' = 49/2$).

7 Extension

Up to now, we have focused on methods which apply when $\mathcal{A}$ has a unique ergodic set. We now give an extension of Theorem 3 which may apply when $\mathcal{A}$ has more than one ergodic set. (Theorem 2 extends similarly.)

Theorem 6. *Given a Markov chain $\mathcal{A}$ and two disjoint closed sets $\mathcal{L}_0$ and $\mathcal{L}_1$ ($\mathcal{L} = \mathcal{L}_0 \uplus \mathcal{L}_1$), suppose there exists a coupling (X_t, Y_t) and a function δ on $\Omega \times \Omega$ which takes values in $\{0, 1, \cdots, B\}$ such that:*

- $\delta(X_t, Y_t) = 0$ *iff* $X_t = Y_t$,
- $\delta(X_t, Y_t) = B \Rightarrow \delta(X_{t+1}, Y_{t+1}) = B$,
- $(\delta(X_t, Y_t) = B \wedge Y_t \in \mathcal{L}_0) \Rightarrow X_t \in \mathcal{L}_1$,
- *there exists $\alpha > 0$ such that, for all (X_t, Y_t) with $0 < \delta(X_t, Y_t) < B$:*
  $E[\delta(X_{t+1}, Y_{t+1})] \leq \delta(X_t, Y_t)$
  $\wedge \quad Pr(\delta(X_{t+1}, Y_{t+1}) \neq \delta(X_t, Y_t)) \geq \alpha$.

Then $\mathcal{A}$ is self-stabilizing w.r.t. $\mathcal{L}$. Futhermore:

1. *The expected hitting time satisfies:* $\mathbf{H}_{\mathcal{L}} \leq B^2/\alpha$.
2. *The ε-absorption time satisfies:*
   $$\mathbf{\Theta}_{\mathcal{L}}(\varepsilon) \leq \lceil e\tfrac{B^2}{\alpha}\rceil\lceil\ln(\tfrac{1}{\varepsilon})\rceil.$$

The proof of Theorem 6 is analogous to that of Theorem 3, but relies on the following proposition :

Proposition 6. *Suppose that $D = (D_t)_{t=0}^{\infty}$ is a nonnegative stochastic process on $\{0, 1, \cdots, B\}$ such that $E[D_{t+1}] \leq D_t$. Furthermore suppose that $Pr(D_{t+1} \neq D_t) \geq \alpha$ (with $\alpha > 0$) when $0 < D_t < B$. Then if τ is the first time that $D_t = 0 \vee D_t = B$, we have: $E[\tau] \leq B^2/\alpha$.*

Proof. Let D'_t be defined by: $D'_t = \begin{cases} D_t & \text{if } D_t \neq B, \\ 0 & \text{if } D_t = B. \end{cases}$
Let τ' be the first time at which D'_t reaches 0. We have $\tau' = \tau$ because $D'_t = 0$ iff $D_t = 0 \vee D_t = B$. On the other hand, it is easy to see that D'_t satisfies conditions of Prop. 5. Hence, we have $E[\tau] = E[\tau'] \leq B^2/\alpha$. ☐

Proof of Theorem 6. Let us consider an integer-valued function δ satisfying the assumptions of Theorem 6, and let us show statements 1 and 2. (The fact that $\mathcal{A}$ is self-stabilizing follows from statement 1, according to Proposition 2.)

1. Consider two elements x, y of Ω and a coupling (X_t, Y_t) of initial element $(X_0, Y_0) = (x, y)$ such that $y \in \mathcal{L}_0$. Let $D_t = \delta(X_t, Y_t)$ for $t \geq 0$. Since $D_t = 0$ iff $X_t = Y_t$ and $(D_t = B \wedge Y_t \in \mathcal{L}_0) \Rightarrow X_t \in \mathcal{L}_1$, the expected time $E[\tau]$ required for D_t to reach 0 or B, is an upper bound of the hitting time $\mathbf{H}_{\mathcal{L}}$. Since, by Proposition 6, we have: for all $x, y \in \Omega$, $E[\tau] \leq B^2/\alpha$, we have: $\mathbf{H}_{\mathcal{L}} \leq B^2/\alpha$.
2. Let $D_t = \delta(X_t, Y_t)$. It is easy to check that D_t satisfies the conditions of Prop. 6. Let $T'_{x,y}$ be the first time where $D_t = 0 \vee D_t = B$, starting with $D_0 = \delta(x, y)$ (i.e., $X_0 = x, Y_0 = y$). It follows by Prop. 6: $E[T'_{x,y}] \leq B^2/\alpha$. Let $T = \lceil eB^2/\alpha \rceil$, then by Markov's inequality, we have the probability that $T'_{x,y} > T$ is at most e^{-1}. If we run s independent trials of length T then the probability that $D_t \neq 0 \wedge D_t \neq B$ for $t \geq sT$ is at most e^{-s}. Therefore, for $t > T\lceil\ln(\varepsilon^{-1})\rceil$, the probability that $D_t = 0 \vee D_t = B$ is at least $1 - \varepsilon$. Suppose that $Y_0 \in \mathcal{L}_0$. Then $Y_t \in \mathcal{L}_0$ (because $\mathcal{L}_0$ is closed). Hence $D_t = 0$ implies $X_t \in \mathcal{L}_0$, while $D_t = B$ implies $X_t \in \mathcal{L}_1$ (since δ satisfies the assumptions of Theorem 6). So $X_t \in \mathcal{L}$ as soon as $D_t = 0 \vee D_t = B$. Therefore, for all $X_0 \in \Omega$ and all $t \geq T\lceil\ln(\varepsilon^{-1})\rceil$: $Pr(X_t \in \mathcal{L}) \geq 1 - \varepsilon$. ☐

As in Section 6, it suffices, thanks to Lemma 1, to check the condition $E[\delta(X_{t+1}, Y_{t+1})] \leq \delta(X_t, Y_t)$ of Theorem 6 on *adjacent* pairs (X_t, Y_t) only.

Example 5. Consider an undirected *regular* graph $G = (V, E)$, with N vertices and degree Δ, and a palette of colors $C = \{0, \cdots, q-1\}$. We suppose furthermore in this example that $q = 2$. A 0-coloring (resp. 1-coloring) is a function from V to $\{0\}$ (resp. $\{1\}$). Let $\mathcal{L}_0$ (resp. $\mathcal{L}_1$) be the 0-coloring (resp. 1-coloring) of G. Starting from any coloring of G, consider the following algorithm $\mathcal{A}$ (with a randomized central scheduler):

- Pick a vertex $v \in V$ and a neighbor w of v uniformly at random;

- recolor v with the color of w.

Let $\delta(X_t, Y_t)$ be the number of vertices at which X_t, Y_t differ, It is easy to show that, for (adjacent) colorings (X_t, Y_t):

$$E[\delta(X_{t+1}, Y_{t+1})] = \delta(X_t, Y_t). \qquad (5)$$

Furthermore, for any pair of colorings (X_t, Y_t) such that $0 < \delta(X_t, Y_t) < N$:

$$Pr(\delta(X_{t+1}, Y_{t+1}) \neq \delta(X_t, Y_t)) \geq \frac{1}{\Delta N} \qquad (6)$$

One infers from (5) and (6), using Theorem 6, that, starting from an arbitrary configuration of Ω, $\mathcal{A}$ is self-stabilizing w.r.t. $\mathcal{L}_0 \uplus \mathcal{L}_1$, and that the hitting time is cubic: $\mathbf{H}_{\mathcal{L}} \leq \Delta N^3$.

Theorem 6 can be used to treat the case where $\mathcal{L}$ is made of more than two ergodic sets, by using the idea of aggregating configurations together or "lumping" (see, e.g., [8]). For example, suppose that in Example 5, there are $q > 2$ colors. In this case $\mathcal{L}$ is composed of q ergodic sets, which correspond to the monochromatic colorings. One can reason as in Example 5, but with with $\mathcal{L}_0$ and $\mathcal{L}'_1$ (instead of $\mathcal{L}_0$ and $\mathcal{L}_1$), where $\mathcal{L}'_1$ is the set of all the colorings without color 0. Applying Theorem 6, one thus infers that the hitting time is cubic: Starting from an arbitrary coloring, the process reaches $\mathcal{L}_0$ or $\mathcal{L}'_1$, in at most ΔN^3 steps in average. So, any coloring losts at least one color after at most ΔN^3 steps in average. The hitting time for reaching a monochromatic color is therefore at most $(q-1)\Delta N^3$.

8 Conclusion

We have shown that the method of coupling, which is classically used to evaluate the rate of convergence to equilibrium of Monte Carlo Markov chains, can be used to prove self-stabilization of distributed algorithms in an original manner. It allows us also to analyze the rate of convergence of these algorithms according to two different measures. The method has been enhanced by using the refinement of coupling, called "path coupling". This suggests to explore applications of the method using other refinements of coupling, such as Huber's bounding chain method [15]. The basic method requires the set $\mathcal{L}$ of legal configurations to be strongly connected, but we indicated how to relax this assumption. We also believe that the method can be extended to the case where the scheduler is not fixed but arbitrary (and possibly "malicious"), such as in the case of Israeli-Jalfon's mutual exclusion [16] or randomized consensus protocols (e.g., [2]), using, for example, the technique of scheduler-luck games (see [6]).
Finally, let us indicate that the idea of coupling has been independently used in [7] for analyzing the convergence time of various protocols where processors communicate by messages.

Acknowledgements. We are grateful to Alistair Sinclair for helpful discussions on an earlier draft of this paper.

References

1. A. Aldous and J. Fill. *Reversible Markov Chains and Random Walks on Graph.* draft at http:/www.stat.Berkeley.EDU/users/aldous/book.html, To appear.
2. J. Aspnes and M. Herlihy. Fast randomized consensus using shared memory. *Journal of Algorithms*, 11(3):441–461, 1990.
3. R. Bubley and M. Dyer. Path coupling: A technique for proving rapid mixing in Markov chains. In *Proc. of the 38th Annual IEEE Symposium on Foundations of Computer Science (FOCS'97)*, pages 223–231, 1997.
4. E.W. Dijkstra. Self-stabilizing systems in spite of distributed control. *Communications of the ACM*, 17(11):643–644, Nov. 1974.
5. S. Dolev. *Self-Stabilization.* The MIT Press, 2000.
6. S. Dolev, A. Israeli, and S. Moran. Analyzing expected time by scheduler luck games. *IEEE transactions on Software Engineering*, 21(5):429–439, May 1995.
7. P. Duchon, N. Hanusse, and S. Tixeuil. Optimal randomized self-stabilizing mutual exclusion on synchronous rings. In *Proc. 18th Int. Conf. on Distributed Computing (DISC 2004), LNCS 3274*, pages 216–229. Springer, 2004.
8. M. Duflot, L. Fribourg, and C. Picaronny. Randomized distributed algorithms as Markov chains. In *Proc. 15th Int. Conf. on Distributed Computing (DISC 2001), LNCS 2180*, pages 240–254. Springer, 2001.
9. M. Dyer, L.A. Goldberg, C. Greenhill, G. Istrate, and M. Jerrum. Convergence of the Iterated Prisoner's Dilemma Game. *Combinatorics, Probability and Computing*, 11(2), 2002.
10. M.E. Dyer and C. Greenhill. A more rapidly mixing Markov chain for graph colorings. *Random Structures and Algorithms*, 13:285–317, 1998.
11. M. Flatebo and A.K. Datta. Two-state self-stabilizing algorithms for token rings. *IEEE Transactions on Software Engineering*, 20(6):500–504, June 1994.
12. L. Fribourg, S. Messika, and C. Picaronny. Coupling and Self-stabilization. In *Proc. 18th Int. Conf. on Distributed Computing (DISC 2004), LNCS 3274*, pages 201–215. Springer, 2004.
13. R.L. Graham, Knuth D.E., and Patashnik O. *Concrete Mathematics.* Addison-Wesley, 1994.
14. T. Herman. Probabilistic self-stabilization. *IPL*, 35(2):63–67, June 1990.
15. M. Huber. Exact sampling and approximate counting techniques. In *Proc. of the 30th Annual ACM Symp. on Theory of Computing (STOC'98)*, pages 31–40, 1998.
16. A. Israeli and M. Jalfon. Token management schemes and random walks yield self-stabilizing mutual exclusion. In *Proc. of the 9th Annual ACM Symposium on Principles of Distributed Computing (PODC'90)*, pages 119–131, 1990.
17. J.G. Kemeny and J.L. Snell. *Finite Markov Chains.* D. van Nostrand Co., 1969.
18. J.E. Kittock. Emergent conventions and the structure of multi-agent systems. In L. Nadel and D. Stein, editors, *Proc. of the 1993 Complex systems summer school.* Santa Fe Institute Studies in the Sciences of Complexity, Addison-Wesley, 1995.
19. L. Lamport. Solved problems, unsolved problems and non-problems in concurrency. In *Proc. of the 3rd ACM Symp. on Principles of Distributed Computing (PODC'84)*, pages 1–11, 1984.
20. L. Lovász and P. Winkler. Mixing Times. *Microsurveys in Discrete Probability*, pages 85–134, 1998.
21. M. Luby, D. Randall, and A. Sinclair. Markov chain algorithms for planar lattice structures (extended abstract). In *Proc. of the 36th Annual IEEE Symp. on Foundations of Computer Science (FOCS'95)*, pages 150–159, 1995.
22. A. Mayer, R. Ostrovsky, and M. Yung. Self-Stabilizing Algorithms for Synchronous Unidirectional Rings. In *Proc. of the 7th ACM-SIAM Symp. on Discrete Algorithms (SODA-96)*, 1996.
23. D. Randall. Mixing. In *Proc. of the 44th Annual IEEE Symp. on Foundations of Computer Science (FOCS'03)*, 2003.
24. M. Schneider. Self-stabilization. *ACM Computing Surveys*, 25:45–67, 1993.
25. R. Segala. *Modeling and Verification of Randomized Distributed Real-Time Systems.* PhD thesis, Massachusetts Institue of Technology, Jun. 1995.
26. A. Sinclair. Convergence rates for Monte Carlo experiments. In *Numerical Methods for Polymeric Systems*, pages 1–18. IMA Volumes in Mathematics & Its application, 1997.
27. D. Williams. *Probability with Martingales.* Cambridge University Press, 1991.