The Annals of Probability
1992, Vol. 20, No. 3, 1109–1146

ROUGH LARGE DEVIATION ESTIMATES FOR SIMULATED ANNEALING: APPLICATION TO EXPONENTIAL SCHEDULES

BY OLIVIER CATONI

Ecole Normale Supérieure, Paris

Simulated annealing algorithms are time inhomogeneous controlled Markov chains used to search for the minima of energy functions defined on finite state spaces. The control parameters, the so-called cooling schedule, control the probability that the energy should increase during one step of the algorithm.

Most of the studies on simulated annealing have dealt with limit theorems, such as characterizing convergence conditions on the cooling schedule, or giving an equivalent of the law of the process for one fixed cooling schedule.

In this paper we derive *finite time* estimates. These estimates are *uniform in the cooling schedule and in the energy function*.

With new technical tools, we gain a new insight into the algorithm. We give a sharp upper bound for the probability that the energy is close to its minimum value. Hence we characterize the *optimal convergence rate*. This involves a new constant, the "*difficulty*" *of the energy landscape*.

We calculate two cooling schedules for which our bound is almost reached. In one case it is reached up to a multiplicative constant *for one energy function*. In the other case it is reached in the sense of logarithmic equivalence *uniformly in the energy function*. These two schedules are both *triangular*: There is one different schedule for each finite simulation time. For each fixed finite time the second schedule has the currently used but previously mathematically unjustified exponential form.

Finally, the title is "Rough large deviation estimates" because we have computed sharper ones (i.e., with sharp multiplicative constants) in two other papers.

Contents

Received April 1990; revised November 1990.
AMS 1980 *subject classifications*. Primary 60F10; secondary 60J10, 93E25.
Key words and phrases. Simulated annealing, large deviations, time inhomogeneous Markov chains, optimal cooling schedules, exponential cooling schedules.

Introduction. This is a finite time study of simulated annealing algorithms. It is based on large deviation estimates which are uniform in the cooling schedule and the energy function.

Simulated annealing algorithms have their origin in statistical mechanics. They simulate a system in contact with a heat bath of decreasing temperature. The system is described by the canonical ensemble $(Z^{-1}\exp(-U_i/T))_{i\in E}$.

The simulation is done by iterative random perturbations which leave the canonical distribution invariant. Temperature T is a decreasing function of time $n \in \mathbb{N}$. Hence from the mathematical point of view we have a time inhomogeneous controlled Markov chain. The control parameters are the temperatures $(T_n)_{n\in\mathbb{N}}$ and represent the action of the heat bath.

It is a computer simulation, hence the state space E is finite.

This method has been considered in computer science for its own sake as a global optimization technique [22, 12]. The reason is that the canonical distribution concentrates on the states of global minimum energy as temperature decreases to 0.

In fact convergence is a delicate issue as soon as the energy function U is not convex, since the system may get quenched in a state which is only a local minimum of the energy. This is what happens if the decrease of the temperature is too fast.

The fact of escaping from a local minimum of the system is a large time behaviour: At constant temperature, it will on the average take place after a time which is an exponential function of the temperature. Estimating the law of such events is part of large deviations theory.

The reference work on the subject is by Freidlin and Wentzell [11], who invented the method. They have studied the time homogeneous case for stochastic differential equations $dX_t = b(X_t)\,dt + \varepsilon\, dw_t$. They introduce a Markov chain on the attractors of the unperturbed system which covers the case of "simulated annealing at constant temperature."

1. The state of the art. Convergence of annealing algorithms has been an active subject of research during the past few years. The different approaches to the subject can be roughly catalogued in the following way:

Most authors are working with the assumption that temperature is slowly decreasing [6, 15, 16, 18, 24]. This is a legitimate assumption when one is thinking of simulated annealing from the point of view of statistical mechanics, because the canonical ensemble describes the system in equilibrium. Thus the simulation has a clear physical meaning only when temperature decreases slowly enough for the system to be always almost at equilibrium. On the contrary, from the point of view of optimization, there is no reason for restricting temperature to decrease slowly if it is possible to do better by going faster. The aim is no longer to be almost at equilibrium, but to concentrate around the minimum as efficiently as possible.

1. Some authors are even doing the technical job at almost constant temperature [18, 24]. They assume that for each temperature T they have some

process $(X_n^T)_{n \in \mathbb{N}}$ which satisfies for some positive constants C and D, independent of the temperature,

$$(1) \qquad Ce^{-U(i,j)/T} \leq P\big(X_n^T = j | X_{n-1}^T = i\big) \leq De^{-U(i,j)/T}, \qquad n \in \mathbb{N}.$$

(Tsitsiklis), or that for any positive ε,

$$(2) \qquad e^{-(U(i,j)+\varepsilon)/T} \leq P\big(X_n^T = j | X_{n-1}^T = i\big) \leq e^{-(U(i,j)-\varepsilon)/T}, \qquad n \in \mathbb{N},$$

for T small enough (Hwang and Sheu).

Then Hwang and Sheu study the case $T_n = c/\ln n$ using estimates on the constant case on appropriate time intervals. They show that for $c > d$, d being the famous "critical depth" of the energy landscape, the density of the law of X_n with regard to the canonical law at temperature T_n tends to a positive limit. By a clever analysis of intervals $[m, n]$ for which $\exp(-1/T_n) \geq \frac{1}{2}\exp(-1/T_m)$, Tsitsiklis shows that a necessary and sufficient condition of convergence is $\sum_n e^{-d/T_n} = +\infty$. This was first proved by Hájek.

2. Holley, Kusuoka and Stroock study the first nonzero eigenvalue of the Dirichlet form associated with the infinitesimal generator of the continuous time process. They use it to estimate the derivative of the density of the law of X_t towards the canonical distribution at temperature T_t. Their results are bounds on the L_p norm of this density for cooling schedules $T_t = c/\ln t$ with $c > d$.

3. Chiang and Chow use the forward Kolmogorov equation to incrementally establish estimates of $P(X_n = i)/P(X_n = j)$ for couples (i, j) in the same component of coarser and coarser subdivisions of the state space into "cycles." They make assumptions on the derivative of the cooling schedule that essentially lead to $T_n = c/\ln n$. They prove the same result as Hwang and Sheu.

Finally, Hájek tackles directly the problem of exit from a domain in the inhomogeneous case. Though he does not seem to know about Wentzell and Freidlin, he considers the law of the exit time and point from cycles, as studied by Wentzell and Freidlin in the homogeneous case. Cycles have the property (cf. [11]) that the law of the exit time concentrates around its mean in the sense that for any positive ε,

$$(3) \qquad \lim_{T \to 0} P\big(E(\tau)^{1-\varepsilon} \leq \tau \leq E(\tau)^{1+\varepsilon}\big) = 1.$$

Unfortunately Hájek's estimates are not sharp. He proves that if τ is the exit time from a cycle, then

$$(4) \qquad P(\tau \geq r) \geq \exp\left(-\Gamma\left(e^{-d/T_0} + \int_0^r e^{-d/T_s}\, ds\right)\right),$$

but, whereas some reversed inequality of the same type is expected, he can only prove that

$$(5) \qquad E\left(\int_0^\tau \exp(-d/T_s)\, ds\right) < +\infty$$

when

$$\int_0^{+\infty} \exp(-d/T_s)\, ds = +\infty. \tag{6}$$

There is a great gap between the two estimates, and this did not allow Hájek to go further than his necessary and sufficient condition for convergence (the one which was proved again by Tsitsiklis).

2. A finite time point of view on simulated annealing. Simulation means that not only the state space but also computer time is finite. The question then arises: What can we do best with some finite large computer time? It is different from asking what we can do best asymptotically. Indeed, in the first case, we can change the temperature values each time we are doing a new experiment with some new allowed computer time.

Another approach from the classical point of view is to ask: What can we do if we have no precise information about the energy function? For instance, one knows that the choice of a cooling schedule of the form $T_n = c/\ln n$ should depend on the value of some *critical depth* d of the energy landscape. Hence if we do not know d, which is in many situations very likely, but maybe only some bounds on d, we are in trouble.

To address these questions, we need estimates which are *uniform in the cooling schedule and the energy function*. What will remain fixed are the state space and the communications between states.

Our estimates will be of the type used by Wentzell and Freidlin; they will be on the law of the couple (exit time, exit point) from strict subdomains of the state space.

The knowledge of such distributions for any subdomains gives very precise information on the behaviour of the system.

Let A be a subdomain of the state space E. We can estimate the probability of being in A at time n by estimating successively the probability of having stayed in A from the beginning of time, the probability of having jumped once to $E - A$ and back, jumped twice, and so on.

Our measure of convergence will be

$$M(n) = \sup_{i \in E} P(U(X_n) \geq \alpha | X_0 = i), \tag{7}$$

the probability that the energy at time n is superior to α in the worst case. Time n is the time when we stop the computer simulation. We are studying the case of a deterministic stopping rule; we choose n in advance. When the convergence measure $M(n)$ is large, convergence is poor. Convergence here means that $U(X_n) < \alpha$—the energy at time n is inferior to some level α close to the minimum value of the energy.

We give a lower bound for $M(n)$: The quantity

$$\inf_{T_1, \dots, T_n} M(n), \tag{8}$$

cannot approach 0 faster than some power of $1/n$. Surprisingly, this power does not involve the critical depth, which is linked to the second eigenvalue of the transition matrix, but a new constant, which we baptized the *difficulty of the energy landscape*.

In order to show that our power of $1/n$ is sharp, we build cooling schedules for which it is reached. As could be expected, our cooling schedules will depend on the horizon n.

This could be expected because for one given schedule T_k defined for all times k, the convergence measure $M(n)$ can be suspected (Holley and Stroock [16]) to depend asymptotically only on the second eigenvalue of the transition matrix, that is, only on the critical depth of the energy landscape. But our bound on $M(n)$ depends on the difficulty of the energy landscape, which can take arbitrary values for one fixed value of the critical depth. Hence, in general, $M(n)$ cannot follow the same power of $1/n$ as our bound when the cooling schedule is independent of the simulation time n. (The materials for a rigorous proof are in [4].)

Thus our cooling schedules will have two parameters: the simulation time n and the current time k. The schedules $(T_k^n)_{1 \le k \le n}$ will be defined on a triangle of the parameter space, hence the names triangular cooling schedules and triangular algorithms. Rather than working directly with closed formulas, we will derive an upper bound for the convergence measure $M(n)$, calculated with some supplementary assumptions on the triangular cooling schedule $(T_k^n)_{1 \le k \le n}$.

These assumptions are a set of inequalities on the finite differences $1/T_k^n - 1/T_{k+1}^n$. There is one inequality for each state i in the state space E. The inequalities depend on the level α chosen to define convergence.

Once again our assumptions are oriented towards a finite time study of the process. We want to prove that $M(n) \le \varepsilon \operatorname{Card}(E)$, and we will assume the finite difference inequality linked with any given state $i \in E$ of energy $U_i \ge \alpha$ to hold only as long as $P(X_k = i) > \varepsilon$.

This can be contrasted with the usual approach (Chiang and Chow [6]) where one assumes one finite difference inequality to hold for all times.

When we decrease ε, the time during which each inequality must hold increases, but in all cases, when k grows, the number of inequalities decreases and T_k^n is allowed to decrease faster and faster. We choose n such that $P(X_k = i) \le \varepsilon$ for all $i \in E$ of energy $U_i \ge \alpha$. Hence at the end, when k approaches n, T_k^n is allowed to decrease arbitrarily fast and even to be set to 0. These kinds of schedule could not be handled in the usual asymptotic framework.

We give two closed formulas for T_k^n. The first one depends on the energy U, through a sequence of typical depths and energy values. For this choice of the cooling schedule, the measure of convergence satisfies

$$M(n) \le \frac{1}{n^p} \times \text{constant}, \tag{9}$$

with sharp constant p (for α small enough).

The second one does not depend on the en•rgy U. For each n, T_k^n is exponential in k. For this choice of the temperatures we have

$$\lim_{n \to +\infty} \frac{\ln M(n)}{\ln n} \le -p \tag{10}$$

and the convergence is uniform on the energy U when we assume some bounds on some characteristics of U.

Our result shows how the rate of exponential decrease of the temperature has to be tuned to the time of simulation. In practice, however, it is possible to do the contrary: Choose some rate and let the algorithm run until no moves are observed. Our result shows that if this is done, the link between the convergence measure M(stopping time) and the typical value of the stopping time will be optimal in the sense of logarithmic equivalents.

This is the first mathematical justification of the use of exponential schedules. Exponential schedules arise in a natural way when one tries to satisfy our set of finite difference inequalities uniformly in the energy U. We have not chosen to study them because they were widely used. It simply turned out that finite time oriented theoretical calculations rejoined common practice.

The paper will now follow the plan we have just sketched: description of the model, large deviations estimates, upper bound on the convergence rate [i.e., lower bound on $M(n)$], lower bound on the convergence rate for a restricted family of schedules and study of two closed formulas for those schedules.

3. Description of the model. The state space E will be a finite set, the energy U will be any function U: $E \mapsto \mathbb{R}_+$ such that $\min_{i \in E} U_i = 0$ (this normalization does not restrict the generality). The communication kernel q: $E \times E \mapsto [0, 1]$ will be a symmetric irreducible Markov kernel on $E \times E$:

$$\begin{aligned} &\sum_j q(i,j) = 1, \qquad i \in E, \\ &q(i,j) = q(j,i), \qquad \sup_{n \in \mathbb{N}} q^n(i,j) > 0, \qquad i,j \in E. \end{aligned} \tag{11}$$

Given an energy landscape (E, q, U) and a "temperature" $T \in \mathbb{R}_+$, we define the transition kernel at temperature T by

$$\begin{aligned} &p_T(i,j) = e^{-(U_j - U_i)^+/T} q(i,j), \qquad i,j \in E, i \ne j \\ &p_T(i,i) = 1 - \sum_{j \in E,\, j \ne i} e^{-(U_j - U_i)^+/T} q(i,j), \end{aligned} \tag{12}$$

where $x^+ = \sup\{x, 0\}$.

A simulated annealing algorithm on the energy landscape (E, U, q) will be given by $(E, q, U, T, \mathscr{L}_0, X)$, where $(T_n)_{n \in \mathbb{N}^*}$ is a nonincreasing sequence in $\mathbb{R}_+$, $\mathscr{L}_0$ is a probability distribution on E, $(X_n)_{n \in \mathbb{N}}$ is a Markov chain on E

with initial distribution $\mathscr{L}_0$ and transition kernel

$$P(X_n = j | X_{n-1} = i) = p_{T_n}(i, j). \tag{13}$$

On the state space E we will consider the partitions induced by the communications relations $\mathscr{R}_\lambda$ at level $\lambda \in \mathbb{R}$:

$$(i, j) \in \mathscr{R}_\lambda \Leftrightarrow \sup_{n \in \mathbb{N}} (q_\lambda)^n(i, j) > 0, \tag{14}$$

where

$$q_\lambda(i, j) = \begin{cases} q(i, j), & \text{if } U_i \vee U_j \leq \lambda, \\ 0, & \text{otherwise}. \end{cases} \tag{15}$$

We use the convention that $q_\lambda^0 = \mathrm{Id}$, so that $(i, i) \in \mathscr{R}_\lambda$, for any $i \in E$.

The set $\mathscr{C}(E, q, U)$ of cycles of (E, q, U) will be the set of the components of E for the relations $\mathscr{R}_\lambda$, $\lambda \in \mathbb{R}$:

$$\mathscr{C}(E, q, U) = \bigcup_{\lambda \in \mathbb{R}} E/\mathscr{R}_\lambda. \tag{16}$$

The boundary $B(C)$ of C will be

$$B(C) = \{j \in E - C | \exists\, i \in C, q(i, j) > 0\}. \tag{17}$$

The energy of C will be

$$U(C) = \min_{i \in C} U_i, \tag{18}$$

its depth

$$H(C) = \min_{j \in B(C)} \max_{i \in C} (U_j - U_i)^+, \tag{19}$$

its principal boundary

$$\tilde{B}(C) = \{i \in B(C) | U_i \leq U(C) + H(C)\}, \tag{20}$$

and its bottom

$$F(C) = \{i \in C | U_i = U(C)\}. \tag{21}$$

The inclusion relation on $\mathscr{C}(E, q, U)$ defines a tree, the leaves being the points of E and the root being E itself. Two cycles which cannot be compared have empty intersection.

Hence, for any subset A of E, the maximal elements of $\mathscr{P}(A) \cap \mathscr{C}(E, q, U)$ form a partition $\mathscr{M}(A)$ of A, which will be called the maximal partition of A.

The maximal partition of a cycle C is $\{C\}$. In this case it is also interesting to consider the "natural partition" $\mathscr{N}(C)$ of C formed by the maximal elements of $(\mathscr{P}(C) - \{C\}) \cap \mathscr{C}(E, q, U)$. It is easily seen that, for any $G, G' \in \mathscr{N}(C)$, $H(G) + U(G) = H(G') + U(G') = \lambda(C)$.

The difficulty of a cycle $C \in \mathscr{C}(E, q, U)$ is

$$\delta(C) = \frac{H(C)}{U(C)}. \tag{22}$$

The difficulty of an energy landscape (E, q, U) at level $\alpha > 0$ is

$$(23)\qquad D_\alpha(E,q,U) = \max\{\delta(C)|C \in \mathscr{C}(E,q,U), U(C) \geq \alpha\}.$$

REMARK. For α small enough,

$$(24)\qquad D_\alpha = \max\{\delta(C)|C \in \mathscr{M}(E - F(E))\}.$$

The depth of any subset A will be

$$(25)\qquad H(A) = \max\{H(C)|C \in \mathscr{M}(A)\}.$$

Another useful extension of the notion of depth is the following. Let $H_1, \dots, H_r$ be the decreasing sequence of the depths of the cycles of $\mathscr{P}(E - F(E))$. Let us put by convention $H(E) = H_0 = +\infty$ and

$$(26)\qquad F_k = \bigcup_{C \in \mathscr{C}(E,q,U)| H(C) \geq H_k} F(C).$$

We can see that $F_0 = F(E)$ and that E is the increasing sum of the sets F_k. We can see also that if $f \in F_k - F_{k-1}$ and if $f \in G_f \in \mathscr{M}(E - F_{k-1})$, then $H(G_f) = H_k$. We will put $H_f = H_k$. Hence H_f is defined for any $f \in E$.

We will often allow ourselves to choose one finite cooling schedule $(T_n^N)_{1 \leq n \leq N}$ for each value of the horizon $N \in \mathbb{N}^*$. The corresponding collection of finite annealing algorithms $(E, q, U, (T_n^N)_{1 \leq n \leq N}, \mathscr{L}_0, (X_n^N)_{1 \leq n \leq N})$ will be called a triangular annealing algorithm. We will also say that $(T_n^N)_{1 \leq n \leq N}$ is a triangular cooling schedule. [The finite Markov chains $(X_n^N)_{1 \leq n \leq N}$ are realized on distinct probability spaces; no stochastic relation is assumed between them.]

For any subset A of E, the exit time from A will be

$$(27)\qquad \tau(A, m) = \inf\{n > m | X_n \notin A\}.$$

With this definition we have always $\tau(A, m) \geq m + 1$.

In our study of the Markov chain X we will consider the following families of tensors with indices in state space E and in time space $\mathbb{Z}$. For any $A, B \subset E$ and any $i, j \in E$, $m, n \in \mathbb{Z}$, we define

$$(28)\qquad M(A,B)_{i,m}^{j,n} = \begin{cases} P(X_n = j, X_{n-1} \notin B, \tau(A,m) \geq n | X_m = i), \\ \qquad m < n, i \in E, j \in B, \\ 0, \qquad \text{otherwise.} \end{cases}$$

[The cooling schedule T, and therefore (X_n), are extended to $n \in \mathbb{Z}$ by $T_n = T_1$ for $n \leq 1$.] In the same way we define

$$(29)\qquad L(A,B)_{i,m}^{j,n} = \begin{cases} P(X_n = j, \tau(A,m) > n | X_m = i), & i \in E, j \in B, m \leq n, \\ 0, & \text{otherwise.} \end{cases}$$

For two tensors $(V_{i,m}^{j,n})_{i,j \in E,\, m,n \in \mathbb{Z}}$ and $(W_{i,m}^{j,n})_{i,j \in E,\, m,n \in \mathbb{Z}}$ we will form the usual inner tensor product

$$(30)\qquad [VW]_{i,m}^{j,n} = \sum_{h \in E,\, k \in \mathbb{Z}} V_{i,m}^{h,k} W_{h,k}^{j,n}.$$

4. Large deviation estimates. Our approach is inspired by Wentzell and Freidlin. We study the law of exit point and time from a domain. The technical advantage of the method is localization—this law is independent of what is outside the domain.

The case of a domain restricted to one point is trivial. Some domains behave almost like points—these are the cycles of E. Viewed from outside, a cycle behaves like a point at low temperature; the exit point and time calculated with the assumption that the starting point is within the cycle does not depend on this starting point. The typical behavior of the system is to visit each state in the cycle a large number of times before leaving it.

Before leaving a cycle, the system spends nonetheless the major part of its time in the states of lowest energy. Hence a *partial equilibrium* is reached. Our method is to single out one point f of lowest energy in the cycle C. We study the behavior of the system after its last visit to this point f. Hence we restrict the domain to $C - \{f\}$. We will consider that this point f is in contact with some "particles bath": At each time n some new particle X^n appears in state f which has a further destiny X_k^n. We suppose that this particle follows the simulated annealing dynamics within $C - \{f\}$ and is absorbed outside. We estimate the measure

$$E\left(\sum_{\substack{n \\ n \le k}} \delta_{X_k^n}\right) \tag{31}$$

on $C - \{f\}$. It turns out that this measure is finite, because there is absorption at least in f.

Then we replace our spring of particles at state f by an estimate of the probability of hitting state f at each time n. This gives an estimate of the probability of leaving a cycle which can be calculated from estimates on the strict subdomains of the cycle under study.

It leads to an induction argument on the size of the cycle.

In the course of this induction, we have to generalize our estimates to any kind of subdomain (we have to cover the case of $C - \{f\}$ which is not a cycle). The survival probability in a given domain is of the same order or less than the survival probability in the deepest subcycle of the domain.

The probabilities of survival are expressed by comparison with *survival kernels* Q_m^n satisfying

$$\sum_{k=m}^{n} Q_m^k = \left(1 - (1+b) \prod_{k=m+1}^{n} \left(1 - ae^{-H/T_k}\right)^+\right)^+. \tag{32}$$

These kernels can be thought of as *inhomogeneous exponential laws*.

Using loose bounds on the flow of communications through the boundary of domains, we derive estimates which are uniform in the energy function and temperatures.

Composition lemmas for some Markov kernels on the integers. The study of the Markov chains $(X_n)_{n \in \mathbb{N}}$ leads in a natural way to the consideration of some kernels on the integers depending on the cooling schedule T which is considered.

For any constants $H \geq 0$, $a > 0$ and $b > 0$, $\mathscr{D}^r(H, a, b)$ will be the family of kernels Q: $\mathbb{Z} \times \mathbb{Z} \to [0, 1]$ such that $Q_m^n = 0$ if $m > n$ (we will say that Q is an increasing kernel on the integers), and for any $m < n \in \mathbb{Z}$,

$$1 - (1 + b) \prod_{l=m+1}^{n-1} (1 - ae^{-H/T_l}) \leq \sum_{k=m}^{n-1} Q_m^k \leq 1, \tag{33}$$

which implies that

$$\sum_{k=n}^{+\infty} Q_m^k \leq (1 + b) \prod_{l=m+1}^{n-1} (1 - ae^{-H/T_l}). \tag{34}$$

(We make the convention that a product is null if it contains negative factors.) It implies also that Q is a Markov transition kernel when $\sum_{n \geq 1} e^{-H/T_n} = +\infty$.

In the same way, by $Q \in \mathscr{D}^l(H, a, b)$ we mean that Q is an increasing kernel on the integers such that

$$1 - (1 + b) \prod_{l=m+1}^{n-1} (1 - ae^{-H/T_l}) \leq \sum_{k=m+1}^{n} Q_k^n \leq 1, \qquad m < n, \tag{35}$$

which implies that

$$\sum_{k=-\infty}^{m} Q_k^n \leq (1 + b) \prod_{l=m+1}^{n-1} (1 - ae^{-H/T_l}). \tag{36}$$

For these definitions of classes, we have the following composition lemmas, which are satisfied for any cooling schedule.

LEMMA 4.1. *If Q, $R \in \mathscr{D}^r(H, a, b)$ [resp., $\in \mathscr{D}^l(H, a, b)$], then for any $\lambda \in [0, 1]$ the kernel $\lambda Q + (1 - \lambda)R$ is of class $\mathscr{D}^r(H, a, b)$ [resp., $\mathscr{D}^l(H, a, b)$].*

The proof is easy.

LEMMA 4.2. *If Q, $R \in \mathscr{D}^r(H, a, b)$ [resp., $\in \mathscr{D}^l(H, a, b)$], then the composed kernel QR is of class $\mathscr{D}^r(H, a', b')$ [resp., $\mathscr{D}^l(H, a', b')$] for some positive constants a' and b'.*

PROOF. We can assume, lowering its value if necessary, that $a \leq \frac{1}{2}$. We will give the proof for right classes. By integration by parts we have

$$\begin{aligned}\sum_{k=m}^{n-1} \{QR\}_m^k &\geq \sum_{k=m}^{n-1} Q_m^k \left(1 - (1 + b) \prod_{l=k+1}^{n-1} (1 - ae^{-H/T_l})\right)^+ \\ &\geq \sum_{k=m}^{n-2} S_m^k (1 + b) ae^{-H/T_{k+1}} \prod_{l=k+2}^{n-1} (1 - ae^{-H/T_l}),\end{aligned} \tag{37}$$

with

$$S_m^k = \sum_{l=m}^{k} Q_m^l \geq \left(1 - (1+b) \prod_{l=m+1}^{k} (1 - ae^{-H/T_l})\right). \tag{38}$$

Hence

$$\begin{aligned} \sum_{k=m}^{n-1} \{QR\}_m^k &\geq (1+b) \sum_{k=m+1}^{n-1} ae^{-H/T_k} \prod_{l=k+1}^{n-1} (1 - ae^{-H/T_l}) \\ &\quad - (1+b)^2 \sum_{k=m}^{n-2} \prod_{l=m+1}^{k} (1 - ae^{-H/T_l}) ae^{-H/T_{k+1}} \\ &\quad \times \prod_{l=k+2}^{n-1} (1 - ae^{-H/T_l}). \end{aligned} \tag{39}$$

The second term of the right member is larger than

$$-\frac{(1+b)^2}{1-a} \prod_{l=m+1}^{n-1} (1 - ae^{-H/T_l}) \left(\sum_{l=m+1}^{n-1} ae^{-H/T_l} \right). \tag{40}$$

Because for any collection of positive numbers x_l we have

$$\sum x_l \leq \prod (1 + x_l), \tag{41}$$

it is also larger than

$$\begin{aligned} &-\frac{2(1+b)^2}{1-a} \prod_{l=m+1}^{n-1} (1 - ae^{-H/T_l}) \left(1 + \frac{a}{2} e^{-H/T_l}\right) \\ &\geq -\frac{2(1+b)^2}{1-a} \prod_{l=m+1}^{n-1} \left(1 - \frac{a}{2} e^{-H/T_l}\right). \end{aligned} \tag{42}$$

As for the first term in the right member of equation (39), it is equal to

$$(1+b)\left(1 - \prod_{l=m+1}^{n-1} (1 - ae^{-H/T_l})\right); \tag{43}$$

hence

$$\sum_{k=m}^{n-1} \{QR\}_m^k \geq 1 - \left(1 + \frac{2(1+b)^2}{1-a}\right) \prod_{l=m+1}^{n-1} \left(1 - \frac{a}{2} e^{-H/T_l}\right). \tag{44}$$ □

The main theorems. The aim of this subsection is to prove the two following theorems.

THEOREM 4.3. *For any state space (E, q) there are positive constants K, K', a, b, c and d such that for any energy function U the following induction hypothesis $\mathscr{H}_n$ is true for any $n \in \mathbb{N}$:*

$\mathscr{H}_n(1)$ For any $C \in \mathscr{C}(E)$ s.t. $|C| \le n$, for any $i \in C$, for any $A \subset C$ such that $i \in A$, for any $j \in B(C)$, there is $Q \in \mathscr{D}^l(H(C-A), a, b)$ [resp., $Q \in \mathscr{D}^r(H(C-A), a, b)$] such that

$$M(C-A, E-C)_{i,m}^{j,n} \le Ke^{-(U_j-U_i)^+/T_{m+1}}Q_m^n, \tag{45}$$

and moreover we have

$$P\big(\tau(C,m) > n | X_m = i\big) \ge c \prod_{l=m+1}^{n} (1 - de^{-H(C)/T_l}). \tag{46}$$

$\mathscr{H}_n(2)$ For any $C \in \mathscr{C}(E)$ such that $|C| \le n$, for any $i \in C$, any $j \in B(C)$, there exists $Q \in \mathscr{D}^r(H(C), a, b)$ such that

$$M(C, E-C)_{i,m}^{j,n} \ge K'e^{-(U_j-U(C)-H(C))^+/T_n}Q_m^n. \tag{47}$$

As a corollary, for any $A \subset E$ such that $|A| \le n$, for any $i \in A$,

$$P\big(\tau(A,m) > n | X_m = i\big) \le (1+b) \prod_{l=m+1}^{n} (1 - ae^{-H(A)/T_l}). \tag{48}$$

$\mathscr{H}_n(3)$ For any cycle $C \in \mathscr{C}(E)$ such that $|C| \le n$, for any $A \subset C$, for any $i \in C$, and $j \in B(C)$, there is $Q \in \mathscr{D}^l(H(C-A), a, b)$ [resp., $Q \in \mathscr{D}^r(H(C-A), a, b)$] such that

$$M(C-A, E-C)_{i,m}^{j,n} \le Ke^{-(U_j-H(C)-U(C))^+/T_{m+1}}Q_m^n. \tag{49}$$

$\mathscr{H}_n(4)$ For any $A \subset E$ such that $|A| \le n$, for any $C \in \mathscr{M}(A)$, for any $i \in A$ there exists $Q \in \mathscr{D}^l(H(A), a, b)$ [resp., $Q \in \mathscr{D}^r(H(A), a, b)$] such that

$$M(A, C)_{i,m}^{E,n} \le KQ_m^n. \tag{50}$$

THEOREM 4.4. *For any state space with communications (E, q) there are positive constants a, b, K such that for any energy function U for any k, such that $0 \le k \le r$, for any $i \in F_k$ we have*

$$L(E-F_k, E-F_k)_{i,m}^{j,n} \le Ke^{-(U_j-U_i)^+/T_{m+1}}Q_m^n \tag{51}$$

with $Q \in \mathscr{D}^l(H_{k+1}, a, b)$.

The meaning of this last theorem is the following. We localize our study to $E - F_k$, replacing the states of F_k by absorbing states. Thus we eliminate the deepest states, those states which keep a long memory of the past. In $E - F_k$, the deepest cycles are of depth H_{k+1}; hence the memory of the past vanishes at speed $\prod_{l=m+1}^{n}(1 - ae^{-H_{k+1}/T_l})$.

Now consider that some point $i \in F_k$ is in contact with some particles bath, generating at each time m a new process $(X_n^m)_{n \ge m}$ with absorption on F_k (only particles jumping in $E - F_k$ at time $m+1$ survive, and they die when they return to F_k). What the theorem says is

$$(52)\qquad \sum_{l\le m} E(\mathbf{1}(X_l^n = j))e^{(U_j-U_i)^+/T_{l+1}} \le K \prod_{s=m+1}^{n-1} (1 - ae^{-H_{k+1}/T_s}).$$

It is best understood at constant temperature. It means then that the mean number of particles at site j is of order less than or equal to $e^{-(U_j-U_i)^+/T}$ and that the contribution of "old" particles decreases exponentially fast with their age. The exponential rate is given by H_{k+1}.

The right member of the inequality decreases with H_{k+1}. Hence when k grows, the estimate is sharper and sharper, but at the same time it is localized to a smaller and smaller set of states.

In fact the estimate is sharp for $j \in F_{k+1}$, because the survival time of a particle starting from j and absorbed on F_k is then exactly of order $\prod_l(1 - ae^{-H_{k+1}/T_l})$. For points $j \in E - F_{k+1}$ this may not be the case.

PROOF OF THEOREM 4.3. It is not hard to see that $\mathscr{H}_1$ is true. Let us carry on the induction step of the proof. Let us assume that $\mathscr{H}_{n-1}$ is true.

POINT $\mathscr{H}_n(1)$. Let $\{C_0, \dots, C_r\}$ be the natural partition of C. We can assume that $i \in C_0$. We have

$$(53)\qquad \begin{aligned} M(C-A, E-C)_i^j &= M(C_0 - A, E - C)_i^j \\ &\quad + \sum_{k=0}^{r} \{M(C_0 - A, C - (C_0 \cup A))M(C - A, C_k - A) \\ &\qquad \times M(C_k - A, E - C)\}_i^j. \end{aligned}$$

For any k, $C_k - A$ is a union of cycles of $\mathscr{M}(C - A)$:

$$(54)\qquad C_k - A = \bigcup_s G_{k,s},$$

and we have

$$(55)\qquad M(C - A, C_k - A) \le \sum_s M(C - A, G_{k,s}).$$

We can apply $\mathscr{H}_{n-1}(4)$ to $C - A$; hence we find a $Q \in \mathscr{D}^l(H(C - A), a, b)$ [resp., $Q \in \mathscr{D}^r(H(C - A), a, b)$] such that

$$(56)\qquad M(C - A, C_k - A) \le KQ.$$

From $\mathscr{H}_{n-1}(1)$ we see that there exists $Q \in \mathscr{D}^l(H(C_0 - A), a, b)$ [resp., $\in \mathscr{D}^r(H(C_0 - A), a, b)$] such that

$$(57)\qquad M(C_0 - A, C - (C_0 \cup A))_{i,m}^{E,n} \le e^{-(H(C_0) + U(C_0) - U_i)T_{m+1}} Q_m^n$$

and from $\mathscr{H}_{n-1}(3)$ we see that for any $h \in C_k - A$ there is $Q \in \mathscr{D}^l(H(C_k - A), a, b)$ [resp., $\in \mathscr{D}^r(H(C_k - A), a, b)$] such that

$$(58)\qquad M(C_k - A, E - C_k)_h^j \le KQ_m^n e^{-(U_j - H(C_k) - U(C_k))/T_{m+1}}.$$

By composition of these three estimates we get that there exist positive constants K, a, b such that

$$M(C-A, E-C)_{i,m}^{j,n} \leq Ke^{-(U_j-U_i)T_{m+1}}Q_m^n \tag{59}$$

with $Q \in \mathscr{D}^l(H(C-A), a, b)$ [resp., $\in \mathscr{D}^r(H(C-A), a, b)$].

We will now establish that there are positive constants c and d such that for any $i \in C$,

$$P\big(\tau(C, m) > n | X_m = i\big) \geq c \prod_{l=m+1}^{n} \left(1 - de^{-H(C)/T_l}\right), \tag{60}$$

with the convention that the product is equal to 0 if some of its factors are negative.

Let us remark that in this formula we cannot assume that $d < 1$; hence it may be meaningful only for small enough values of the temperature. To prove it we write, f being some point of $F(C)$,

$$\begin{aligned} 1 &- P(\tau(C, m) > n | X_m = i) \\ &\leq \sum_{l=m+1}^{n} M(C - \{f\}, E - C)_{i,m}^{E,l} \\ &\quad + \sum_{k=m+1}^{n-1} \sum_{l=k+1}^{n} P(X_k = f, \tau(C, m) > k | X_m = i) M(C - \{f\}, E - C)_{f,k}^{E,l} \\ &\leq 1 - \gamma + K \sum_{k=m+1}^{n} e^{-H(C)/T_k}, \end{aligned} \tag{61}$$

this last line being a consequence of $\mathscr{H}_n(1)$ and of $\mathscr{H}_{n-1}(2)$ applied to the cycles of $\mathscr{M}(C - \{f\})$, which proves that the probability to get out of $C - \{f\}$ in $\{f\}$ admits a positive lower bound of the form γ, and therefore proves that the probability to get out of C without visiting $\{f\}$ is bounded from above by $1 - \gamma$. (We can assume here that $\sum e^{-H(C)/T_k} = +\infty$, putting $T_k = T_n$ for $k \geq n$.)

Let us define

$$R(H, c, m) = \inf\left\{ n > m \,\Big|\, \sum_{l=m+1}^{n} e^{-H/T_l} \geq c \right\}. \tag{62}$$

Assume that $e^{-H(C)/T_m} \leq \gamma/(4K)$, let $u_0 = m$ and

$$u_{n+1} = R\left(H(C), \frac{\gamma}{4K}, u_n\right). \tag{63}$$

We have for $k \in \mathbb{N}$

$$P\big(\tau(C, u_k) > u_{k+1} | X_{u_k} = i\big) \geq \frac{\gamma}{2}. \tag{64}$$

Moreover, for any family (x_l) of positive numbers,

$$\prod (1 - x_l) \leq e^{-\Sigma x_l}; \tag{65}$$

hence, putting

$$K_2 = \sup\left(\frac{4K}{\gamma} \ln\left(\frac{2}{\gamma}\right), \frac{4K}{\gamma} \right), \tag{66}$$

we have

$$P\big(\tau(C, u_k) > u_{k+1} | X_{u_k} = i\big) \geq \prod_{l=u_k+1}^{u_{k+1}} \big(1 - K_2 e^{-H(C)/T_l}\big). \tag{67}$$

Hence for any $n \geq m$,

$$P\big(\tau(C, m) > n | X_m = i\big) \geq \frac{\gamma}{2} \prod_{l=m+1}^{n} \big(1 - K_2 e^{-H(C)/T_l}\big). \tag{68}$$

The case $e^{-H(C)/T_{m+1}} > \gamma/(4K)$ is trivial, with the convention that the product is equal to 0 when one of its factors is negative (which is then the case). This ends the proof of inequality (60).

POINT $\mathscr{H}_n(2)$. Let i and j be fixed. Let $C_i \in \mathscr{N}(C)$ be such that $i \in C_i$ and let $C_j \in \mathscr{N}(C)$ be such that $j \in \tilde{B}(C_j)$. Let $\lambda(C) = U(C_j) + H(C_j)$. There is a sequence $C_0, \ldots, C_r$ of cycles of $\mathscr{N}(C)$ such that $C_0 = C_i$, $C_r = C_j$ and $\tilde{B}(C_{k-1}) \cap C_k \neq \varnothing$, for $k = 1, \ldots, r$. From $\mathscr{H}_{n-1}(2)$ and the composition lemmas we deduce that there are $K_1, a, b, Q \in \mathscr{D}^r(\lambda(C) - U(C), a, b)$ such that

$$M(C, E - C)_{i,m}^{j,n} \geq K_1 e^{-(U_j - H(C_j) - U(C_j))/T_n} Q_m^n; \tag{69}$$

hence

$$\begin{aligned} &\sum_{l=m+1}^{n} e^{(U_j - U(C) - H(C))/T_l} M(C, E - C)_{i,m}^{j,l} \\ &\quad \geq K_1 e^{-(U(C) + H(C) - \lambda(C))/T_n} \\ &\qquad \times \left(1 - (1 + b) \prod_{l=m+1}^{n} \big(1 - a e^{-(\lambda(C) - U(C))/T_l}\big)\right). \end{aligned} \tag{70}$$

For any fixed m let us put $u_0 = m$ and

$$u_{k+1} = R\big(\lambda(C) - U(C), K_2, u_k\big), \tag{71}$$

with $K_2 = a^{-1} \ln(2(1 + b))$. We have, putting $K_3 = K_1/2$,

$$\begin{aligned} &\sum_{l=u_k+1}^{u_{k+1}} e^{(U_j - U(C) - H(C))/T_l} M(C, E - C)_{i,u_k}^{j,l} \\ &\quad \geq K_3 e^{-(U(C) + H(C) - \lambda(C))/T_{u_{k+1}}}. \end{aligned} \tag{72}$$

The case $|C| = 1$ is straightforward; hence we may assume that $|C| \geq 2$. In case $|C| \geq 2$, there is some constant $\chi < 1$, namely,

$$\chi = 1 - \inf\{q(i,j) | (i,j) \in E^2, i \neq j, q(i,j) > 0\}, \tag{73}$$

such that for any $i \in C$ and any m,

$$P(\tau(C,m) > m + 1 | X_m = i) \geq 1 - \chi. \tag{74}$$

Hence we have from $\mathscr{H}_n(1)$

$$P(\tau(C,m) > n | X_m = i) \geq c \prod_{l=m+1}^{n} \left(1 - (de^{-H(C)/T_l}) \wedge \chi\right). \tag{75}$$

For any $j \in B(C)$ we have (using the Markov property)

$$\begin{aligned} &\sum_{l=m+1}^{u_k} e^{(U_j - H(C) - U(C))/T_l} M(C, E - C)_{i,m}^{j,l} \\ &\quad \geq \sum_{s=1}^{k} K_3 e^{-(U(C)+H(C)-\lambda(C))/T_{u_s}} c \prod_{l=m+1}^{u_{s-1}} \left(1 - (de^{-H(C)/T_l}) \wedge \chi\right). \end{aligned} \tag{76}$$

For any $n \geq m$, let k be such that $u_k \leq n < u_{k+1}$. We have

$$\begin{aligned} &\sum_{l=m+1}^{n} e^{(U_j - H(C) - U(C))/T_l} M(C, E - C)_{i,m}^{j,l} \\ &\quad \geq \sum_{s=1}^{k} \frac{K_3}{K_2 + 1} \sum_{l=u_s+1}^{u_{s+1}} e^{-H(C)/T_l} c \prod_{l=m+1}^{u_{s-1}} \left(1 - (de^{-H(C)/T_l}) \wedge \chi\right) \\ &\quad \geq \frac{K_3}{K_2 + 1} \frac{c}{d} \prod_{s=m+1}^{u_1} \left(1 - (de^{-H(C)/T_s}) \wedge \chi\right) Q_m^n, \end{aligned} \tag{77}$$

with

$$Q_m^n = \begin{cases} \left((de^{-H(C)/T_n}) \wedge \chi\right) \prod_{l=u_1+1}^{n-1} \left(1 - (de^{-H(C)/T_l}) \wedge \chi\right), & n > u_1, \\ 0, & \text{otherwise}. \end{cases} \tag{78}$$

To end the proof it is enough to remark that

$$\begin{aligned} \prod_{l=m+1}^{u_1} \left(1 - (de^{-H(C)/T_l}) \wedge \chi\right) &\geq \exp\left(\frac{\ln(1-\chi)}{\chi} d \sum_{l=m+1}^{u_1} e^{-H(C)/T_l}\right) \\ &\geq \exp\left(\frac{\ln(1-\chi)}{\chi} d(K_2 + 1)\right) \end{aligned} \tag{79}$$

and that $Q \in \mathscr{D}^r(H(C), a_3, b_3)$ with $a_3 = d \wedge \chi$ and

$$b_3 = \exp\left(-\frac{\ln(1-\chi)}{\chi} d(K_2 + 1) \right) - 1.$$

This ends the proof of the second point.

Let us now come to its corollary. Let $A \subset E$ be such that $|A| \le n$.

From $\mathscr{H}_n(2)$ we deduce that there are positive constants a_1, $K_1 \le 1$ and b_1 such that

$$\begin{aligned} &P(\tau(A,m) > n | X_m = i) \\ &\quad \le 1 - K_1\left(1 - (1+b_1) \prod_{l=m+1}^{n} \left(1 - a_1 e^{-H(A)/T_l}\right)\right). \end{aligned} \tag{80}$$

Indeed for any $i \in A$, there is a finite sequence $C_0, \dots, C_r \in \mathscr{M}(A)$ such that $i \in C_0$, $\tilde{B}(C_{k-1}) \cap C_k \ne \varnothing$, $k = 1, \dots, r$ and $\tilde{C}_r \cap (E - A) \ne \varnothing$, and we can apply $\mathscr{H}_n(2)$ and the compositions lemmas to $[M(C_0, C_1)M(C_1, C_2) \cdots M(C_r, E - A)]_i^E$.

Let us put $u_0 = m$ and

$$u_{n+1} = R\left(H(A), \frac{\ln(2(1+b_1))}{a_1}, u_n\right). \tag{81}$$

We have

$$P\left(\tau(A, u_n) > u_{n+1} | X_{u_n} = i\right) \le 1 - \frac{K_1}{2} \le \prod_{l=u_n+1}^{u_{n+1}} \left(1 - ae^{-H(A)/T_l}\right), \tag{82}$$

with

$$a = \frac{K_1}{2}\left(1 + \frac{\ln(2(1+b_1))}{a_1}\right)^{-1}. \tag{83}$$

Now let $n \ge m$ be fixed. There is $k \in \mathbb{N}$ such that $u_k \le n < u_{k+1}$. We have

$$\begin{aligned} P(\tau(A,m) > n | X_m = i) &\le P(\tau(A,m) > u_k | X_m = i) \\ &\le \prod_{l=m+1}^{u_k} \left(1 - ae^{-H(A)/T_l}\right) \\ &\le (1+b) \prod_{l=m+1}^{n} \left(1 - ae^{-H(A)/T_l}\right), \end{aligned} \tag{84}$$

with $b = (1 - K_1/2)^{-1} - 1$. This ends the proof of the corollary.

Point $\mathscr{H}_n(3)$. Let $f \in F(C)$. We have

$$\begin{aligned} M(C-A, E-C)_{i,m}^{j,n} &= M(C - (A \cup \{f\}), E - C)_{i,m}^{j,n} \\ &\quad + \sum_{k=m+1}^{n-1} P(X_k = f, \tau(C-A, m) > k | X_m = i) \\ &\quad \times M(C - (A \cup \{f\}), E - C)_{f,k}^{n,j}. \end{aligned} \tag{85}$$

Moreover, letting (C_s) be the maximal partition of $C - (A \cup \{f\})$,

$$\begin{aligned} M(C - (A \cup \{f\}), E - C) \\ = \sum_s \{M(C - (A \cup \{f\}), C_s) M(C_s - A, E - C)\}. \end{aligned} \tag{86}$$

According to $\mathscr{H}_{n-1}(4)$,

$$M(C - (A \cup \{f\}), C_s)_{i,m}^{E,n} \le K Q_m^n, \tag{87}$$

with $Q \in \mathscr{D}^l(H(C - A), a, b)$ [resp., $\in \mathscr{D}^r(H(C - A), a, b)$], and according to $\mathscr{H}_{n-1}(3)$,

$$M(C_s - A, E - C)_{h,m}^{j,n} \le K e^{-(U_j - H(C_s) - U(C_s))^+ / T_{m+1}} Q_m^n, \tag{88}$$

with $Q \in \mathscr{D}^l(H(C_s - A), a, b)$ [resp., $\in \mathscr{D}^r(H(C_s - A), a, b)$]. Hence

$$M(C - (A \cup \{f\}), E - C)_{i,m}^{j,n} \le K e^{-(U_j - H(C) - U(C))^+ / T_{m+1}} Q_m^n, \tag{89}$$

with $Q \in \mathscr{D}^l(H(C - A), a, b)$ [resp., $\in \mathscr{D}^r(H(C - A), a, b)$].

Moreover, from $\mathscr{H}_n(1)$,

$$M(C - (A \cup \{f\}), E - C)_{f,m}^{j,n} \le K e^{-(U_j - U(C))^+ / T_{m+1}} Q_m^n, \tag{90}$$

with $Q \in \mathscr{D}^l(H(C - (A \cup \{f\})), a, b)$ [resp., $\in \mathscr{D}^r(H(C - (A \cup \{f\})), a, b)$], and from point $\mathscr{H}_n(2)$ we have

$$\begin{aligned} P(X_k = f, \tau(C - A, m) > k | X_m = i) \\ \le (1 + b) \prod_{l=m+1}^{k} (1 - a e^{-H(C-A)/T_l}), \end{aligned} \tag{91}$$

hence

$$\begin{aligned} \sum_{k=m+1}^{n-1} P(X_k = f, \tau(C - A, m) > k | X_m = i) \\ \times M(C - (A \cup \{f\}), KE - C)_{f,k}^{n,j} \\ \le K e^{-(U_j - U(C) - H(C))^+ / T_{m+1}} Q_m^n, \end{aligned} \tag{92}$$

with $Q \in \mathscr{D}^l(H(C - A), a, b)$ [resp., $\in \mathscr{D}^r(H(C - A), a, b)$].

POINT $\mathscr{H}_n(4)$. Consider some cycle $G \in \mathscr{M}(A)$ such that $H(G) = H(A)$, and some $g \in F(G)$. We have

$$\begin{aligned} M(A, C)_{i,m}^{E,n} = M(A - \{g\}, C)_{i,m}^{E,n} \\ + \sum_{k=m+1}^{n} P(\tau(A, m) > k, X_k = g | X_m = i) \\ \times M(A - \{g\}, C)_{g,k}^{E,n}. \end{aligned} \tag{93}$$

Moreover,

$$\begin{aligned} M(A - \{g\}, C)_{g,m}^{E,n} = \{M(G - \{g\}, E - G) M(A - \{g\}, C)\}_{g,m}^{E,n} \\ + M(G - \{g\}, C)_{g,m}^{E,n}. \end{aligned} \tag{94}$$

Applying $\mathscr{H}_n(1)$ to G and $\mathscr{H}_{n-1}(4)$ to $A - \{g\}$, we get that

$$M(A - \{g\}, C)_{g,m}^{E,n} \le K e^{-H(A)/T_{m+1}} Q_m^n, \tag{95}$$

with $Q \in \mathscr{D}^l(H(A), a, b)$ [resp., $\in \mathscr{D}^r(H(A), a, b)$]. We conclude by combining this equation with equation (93) and hypothesis $\mathscr{H}_{n-1}(4)$ applied to $A - \{g\}$ again.

Thus we have established that $\mathscr{H}_n$ is satisfied for any $n \le |E|$. □

PROOF OF THEOREM 4.4. For any $0 \le k \le r$ we have

$$L(E - F_k, E - F_k) = \sum_{l=k+1}^{r} L(E - F_k, F_l - F_{l-1}). \tag{96}$$

Conditioning by the last visit to $F_{k+1} - F_k$ when this set has been visited gives

$$\begin{aligned} L(E - F_k, F_l - F_{l-1}) &= L(E - F_k, F_{k+1} - F_k) L(E - F_{k+1}, F_l - F_{l-1}) \\ &\quad + L(E - F_{k+1}, F_l - F_{l-1}) \\ &= (L(E - F_k, F_{k+1} - F_k) + I) L(E - F_{k+1}, F_l - F_{l-1}), \end{aligned} \tag{97}$$

where $I_{i,m}^{j,n} = \delta_i^j \delta_m^n$, δ being Kronecker's symbol. Hence

$$\begin{aligned} L(E - F_k, F_l - F_{l-1}) &= \prod_{s=k+1}^{l-1} (L(E - F_{s-1}, F_s - F_{s-1}) + I) L(E - F_{l-1}, F_l - F_{l-1}). \end{aligned} \tag{98}$$

Thus

$$L(E - F_k, E - F_k) = \prod_{s=k+1}^{r} (L(E - F_{s-1}, F_s - F_{s-1}) + I) - I. \tag{99}$$

Hence it is enough to prove Theorem 4.4 with the kernel $L(E - F_k, E - F_k)$ replaced by the kernel $L(E - F_k, F_{k+1} - F_k)$.

For any $f \in F_k$, any $g \in F_{k+1} - F_k$, let G_g be the cycle of $\mathscr{M}(E - F_k)$ which contains g. We have

$$\begin{aligned} L(E - F_k, F_{k+1} - F_k)_{f,m}^{g,n} &\le L(E - F_k, F_{k+1} - F_k)_{f,m}^{G_g,n} \\ &\le \sum_{j \in G_g} \sum_{l=m+1}^{n} M(E - F_k, G_g)_{f,m}^{j,l} P(\tau(l, G_g) > n | X_l = j). \end{aligned} \tag{100}$$

Let C be the largest cycle containing f and not g. We have

$$\begin{aligned} M(E - F_k, G_g)_{f,m}^{E,n} &\le [M(C - F_k, E - C) M(E - F_k, G_g)]_{f,m}^{E,n} \\ &\le K e^{-(H(C) + U(C) - U_f)/T_{m+1}} Q_m^n, \end{aligned} \tag{101}$$

with $Q \in \mathscr{D}^r(H_{k+1}, a, b)$.

If $U_g < U_f$, then $g \notin G_f$; hence $H(C) + U(C) - U_f \ge H(G_f) \ge H_k \ge H_{k+1}$.

If $U_g \ge U_f$ we have nonetheless $H(C) + U(C) \ge H(G_g) + U_g$.

Hence in any case

$$M(E - F_k, G_g)_{f,m}^{E,n} \le K e^{-(H_{k+1} + (U_g - U_f)^+)/T_{m+1}} Q_m^n, \tag{102}$$

with $Q \in \mathscr{D}^r(H_{k+1}, a, b)$. Hence

$$L(E - F_k, F_{k+1} - F_k)_{f,l}^{g,n} \leq \sum_{m=l}^{n} e^{-(H_{k+1}+(U_g-U_f)^+)/T_{l+1}} Q_l^m (1 + b) \prod_{s=m+1}^{n} (1 - ae^{-H_{k+1}/T_s}), \tag{103}$$

with $Q \in \mathscr{D}^r(H_{k+1}, a, b)$. Let us use the convention that

$$\prod_{s=m+1}^{n} (1 - ae^{H_{k+1}/T_s}) = 1$$

when $m \geq n$. The function

$$m \mapsto \prod_{s=m+1}^{n} (1 - ae^{-H_{k+1}/T_s})$$

being nondecreasing in m (for fixed n), we have for l fixed:

$$\begin{aligned}
&\sum_{m=l}^{n} Q_l^m \prod_{s=m+1}^{n} (1 - ae^{-H_{k+1}/T_s}) \\
&\quad \leq \sum_{m=l}^{+\infty} Q_l^m \prod_{s=m+1}^{n} (1 - ae^{-H_{k+1}/T_s}) \\
&\quad \leq \sum_{m=l}^{+\infty} \left(\sum_{s=m}^{+\infty} Q_l^s - \sum_{s=m+1}^{+\infty} Q_l^s \right) \prod_{s=m+1}^{n} (1 - ae^{-H_{k+1}/T_s}) \\
&\quad \leq \sum_{m=l+1}^{+\infty} \sum_{s=m}^{+\infty} Q_l^s \left(\prod_{s=m+1}^{n} (1 - ae^{-H_{k+1}/T_s}) - \prod_{s=m}^{n} (1 - ae^{-H_{k+1}/T_s}) \right) \\
&\qquad + \sum_{s=l}^{+\infty} Q_l^s \prod_{s=l+1}^{n} (1 - ae^{-H_{k+1}/T_s}) \\
&\quad \leq \sum_{m=l+1}^{n} (1 + b) \prod_{s=l+1}^{m-1} (1 - ae^{-H_{k+1}/T_s}) ae^{-H_{k+1}/T_m} \prod_{s=m+1}^{n} (1 - ae^{H_{k+1}/T_s}) \\
&\qquad + \prod_{s=l+1}^{n} (1 - ae^{-H_{k+1}/T_s}) \\
&\quad \leq \frac{(1 + b)}{1 - a} \left(\sum_{s=m+1}^{n} ae^{-H_{k+1}/T_s} \right) \prod_{s=l+1}^{n} (1 - ae^{-H_{k+1}/T_s}) \\
&\qquad + \prod_{s=l+1}^{n} (1 - ae^{-H_{k+1}/T_s}) \\
&\quad \leq \frac{2(1 + b)}{1 - a} \prod_{s=l+1}^{n} \left(1 + \frac{a}{2} e^{-H_{k+1}/T_s}\right) (1 - ae^{-H_{k+1}/T_s}) \\
&\qquad + \prod_{s=l+1}^{n} (1 - ae^{-H_{k+1}/T_s}) \\
&\quad \leq \frac{2(1 + b)}{1 - a} \prod_{s=l+1}^{n} \left(1 - \frac{a}{2} e^{-H_{k+1}/T_s}\right) + \prod_{s=l+1}^{n} (1 - ae^{-H_{k+1}/T_s}).
\end{aligned} \tag{104}$$

Hence there are positive constants a' and K such that

$$(105)\qquad \begin{aligned} &L(E - F_k, F_{k+1} - F_k)^{g,n}_{f,l} \\ &\quad \le K e^{-(U_g - U_f)^+/T_{l+1}} a' e^{-H_{k+1}/T_{l+1}} \prod_{s=l+2}^{n} \left(1 - a' e^{-H_{k+1}/T_s}\right). \end{aligned} \qquad \square$$

5. An upper bound for the convergence rate. We give here a proof of an upper bound for the convergence rate based on Theorem 4.3. We had given a proof of a similar bound in our thesis [5], but it was longer and relied on sharp large deviation estimates which are themselves harder to prove than rough ones.

THEOREM 5.1. *For any annealing algorithm* $(E, q, U, T, \mathscr{L}_0, X)$ *we have*

$$(106)\qquad P(X_n = i) \ge \Big(\inf_{j \in E} \mathscr{L}_0(j)\Big) e^{-U_i/T_n}, \qquad i \in E.$$

PROOF. The proof is by induction on n. Assume that the theorem is proved for $n - 1$. We have for any $i \in E$,

$$(107)\qquad \begin{aligned} P(X_n = i) &= \sum_{f \in E} P(X_{n-1} = f) p_{T_n}(f, i) \\ &\ge \Big(\inf_{j \in E} \mathscr{L}_0(j)\Big) \sum_{f \in E} e^{-U_f/T_{n-1}} p_{T_n}(f, i) \\ &\ge \Big(\inf_{j \in E} \mathscr{L}_0(j)\Big) \sum_{f \in E} e^{-U_f/T_n} p_{T_n}(f, i) \\ &\ge e^{-U_i/T_n} \inf_{j \in E} \mathscr{L}_0(j). \end{aligned} \qquad \square$$

THEOREM 5.2. *For any state space* (E, q), *there are positive constants* a, b *such that for any energy function* U, *for any positive energy level* α *and for any cooling schedule* T *we have*

$$(108)\qquad \sup_{i \in E} P(U(X_n) \ge \alpha | X_0 = i) \ge b(an)^{-D_\alpha^{-1}(U)}, \qquad n > 0.$$

PROOF. Let $C \in \mathscr{C}(E, U)$ be such that $\delta(C) = D_\alpha(U)$ and $U(C) \ge \alpha$. Let

$$(109)\qquad \mathscr{L}_0(j) = \frac{1}{|E|}, \qquad j \in E.$$

We will show that

$$(110)\qquad \begin{aligned} b(an)^{-D_\alpha^{-1}(U)} \le P(U(X_n) \ge \alpha) &= \frac{1}{|E|} \sum_{i \in E} P(U(X_n) \ge \alpha | X_0 = i) \\ &\le \sup_{i \in E} P(U(X_n) \ge \alpha | X_0 = i). \end{aligned}$$

We have

$$P(X_n = C) = \sum_{i \in C} P(X_0 = i) P(\tau(C,0) > n | X_0 = i) \tag{111}$$
$$+ \sum_{k<n} \sum_{i \in C} \sum_{j \in B(C)} P(X_k = j) q(j,i) P(\tau(C,k) > n | X_k = i).$$

Here we will deal separately with the case $|C| = 1$.

If $|C| = 1$ we deduce from equation (111) that there exist positive constants c, d and K, with $d \leq 1$, such that

$$P(X_n \in C) \geq c \prod_{k=1}^{n} (1 - d e^{-H(C)/T_k}) \tag{112}$$
$$+ \sum_{k=1}^{n} K e^{-(H(C)+U(C))/T_k} \prod_{l=k+1}^{n} (1 - d e^{-H(C)/T_l}).$$

If $|C| \geq 2$, then putting $\lambda = 1 - \inf\{q(i,j) | i \neq j, \ q(i,j) > 0\}$ we have for any $i \in C$,

$$P(\tau(C,m) > m + 1 | X_m = i) \geq 1 - \lambda; \tag{113}$$

hence from $\mathscr{H}(1)$ there are positive constants c and d such that

$$P(\tau(C,m) > n | X_m = i) \geq c \prod_{l=m+1}^{n} (1 - (d e^{-H(C)/T_l}) \wedge \lambda). \tag{114}$$

Thus in this case

$$P(X_n \in C) \geq c \prod_{l=1}^{n} (1 - (d e^{-H(C)/T_l}) \wedge \lambda) \tag{115}$$
$$+ \sum_{k=1}^{n} K e^{-(H(C)+U(C))/T_k} \prod_{l=k+1}^{n} (1 - (d e^{-H(C)/T_l}) \wedge \lambda)$$

(but of course we cannot assume here that $d \leq 1$).

We will end the proof by computing a lower bound for the minimum of the right-hand side of equation (112) or (115) over all possible sequences $T_1, \dots, T_n$.

We will make the computations for equation (115) and leave the very similar case of equation (112) to the reader. Let us consider the sequence $R_0, \dots, R_n$, with $R_0 \leq c$ and

$$R_n = R_{n-1}(1 - (d e^{-H(C)/T_n}) \wedge \lambda) + K e^{-(H(C)+U(C))/T_n}. \tag{116}$$

We have from equation (115),

$$P(X_n \in C) \geq R_n. \tag{117}$$

It is a simple computation to see that if

$$\left(\frac{d}{K} \left(1 + \frac{U(C)}{H(C)} \right)^{-1} R_{n-1} \right)^{H(C)/U(C)} \leq \frac{\lambda}{d}, \tag{118}$$

then

$$\min_{T_n} R_n = R_{n-1} - \frac{d}{1+\delta}\left(\frac{d}{K(1+\delta^{-1})}\right)^{\delta} R_{n-1}^{(1+\delta)} \tag{119}$$

[and if

$$\left(\frac{d}{K}\left(1+\frac{U(C)}{H(C)}\right)^{-1} R_{n-1}\right)^{H(C)/U(C)} \geq \frac{\lambda}{d}, \tag{120}$$

then

$$\min_{T_n} R_n = R_{n-1}(1-\lambda) + K\left(\frac{\lambda}{d}\right)^{(1+\delta^{-1})}]. \tag{121}$$

In light of this, let us put

$$R_0 = c \wedge \frac{K}{d}\left(\frac{\lambda}{d}\right)^{\delta^{-1}}. \tag{122}$$

Then we deduce from equation (119) that

$$\min_{T_1,\dots,T_n} R_n = S_n \tag{123}$$

satisfies

$$S_0 = R_0,$$
$$S_n - S_{n-1} = -\frac{d}{1+\delta}\left(\frac{d}{K(1+\delta^{-1})}\right)^{\delta} S_{n-1}^{(1+\delta)}. \tag{124}$$

Hence $S_n < S_{n-1}$ and

$$\begin{aligned} S_n^{-\delta} - S_{n-1}^{-\delta} &\leq -\delta(S_n - S_{n-1}) S_n^{-(1+\delta)} \\ &\leq \delta \frac{d}{1+\delta}\left(\frac{d}{K(1+\delta^{-1})}\right)^{\delta}\left(\frac{S_{n-1}}{S_n}\right)^{1+\delta}, \end{aligned} \tag{125}$$

but

$$\begin{aligned} \frac{S_{n-1}}{S_n} &= \left(1 - \frac{d}{1+\delta}\left(\frac{dS_{n-1}}{K(1+\delta^{-1})}\right)^{\delta}\right)^{-1} \\ &\leq \left(1 - d\left(\frac{dS_0}{K}\right)^{\delta}\right)^{-1} \\ &\leq (1-\lambda)^{-1}; \end{aligned} \tag{126}$$

hence

$$S_n^{-\delta} - S_{n-1}^{-\delta} \leq K^{-\delta}\left(\frac{d}{1-\lambda}\right)^{1+\delta}, \tag{127}$$

thus

$$(128)\qquad \begin{aligned} S_n^{-\delta} &\le S_0^{-\delta} + nK^{-\delta}\left(\frac{d}{1-\lambda}\right)^{1+\delta} \\ &\le n\left(S_0^{-\delta} + K^{-\delta}\left(\frac{d}{1-\lambda}\right)^{1+\delta}\right), \qquad n \ge 1. \end{aligned}$$

Moreover,

$$(129)\qquad S_0^{-\delta} \le c^{-\delta} \vee \left(\frac{d}{K}\right)^{\delta}\frac{d}{\lambda} \le \left(\frac{d}{\lambda} \vee 1\right)\left(c^{-1} \vee \frac{d}{K}\right)^{\delta},$$

so

$$(130)\qquad S_n^{-\delta} \le n\left(\left(\frac{d}{\lambda} \vee 1\right)\left(c^{-1} \vee \frac{d}{K}\right)^{\delta} + \frac{d}{1-\lambda}\left(\frac{d}{K(1-\lambda)}\right)^{\delta}\right).$$

Thus, putting

$$(131)\qquad \begin{aligned} b &= \left(c^{-1} \vee \frac{d}{K(1-\lambda)}\right)^{-1}, \\ a &= 2\left(\frac{d}{\lambda} \vee \frac{d}{1-\lambda} \vee 1\right), \end{aligned}$$

we have

$$(132)\qquad S_n^{-\delta} \le b^{-\delta} a n,$$

or

$$(133)\qquad S_n \ge \frac{b}{(an)^{\delta^{-1}}}, \qquad n \ge 1. \qquad \square$$

6. A lower bound for the convergence rate. In this section we will prove a technical proposition, from which we will derive the theorems of the two following sections. This proposition has already been discussed in Section 2. Let us give some more explanations.

The aim is to get an upper bound for the convergence measure when convergence is measured at level α. For this purpose, we fix some positive ε and try to find conditions on the cooling schedule for which

$$(134)\qquad \sup_{i \in E} P(X_n = f \mid X_0 = i) \le K(\varepsilon \vee e^{-U_f/T_n}), \qquad n > 0,$$

for any state f of energy $U_f \ge \alpha$.

We use successively Theorem 4.4 for $F_0, F_1, \dots, F_{r-1}$. We start with a trivial estimate on F_0:

$$(135)\qquad P(X_n = f \mid X_0 = i) \le 1, \qquad f \in F_0.$$

Theorem 4.4 allows us to transform it into an estimate on $E - F_0$. We get that

$$\sup_{i \in E} P(X_n = f | X_0 = i) \le (1 + b) \prod_{l=1}^{n} (1 - a e^{-H_1/T_l}) + K \sum_{m=-\infty}^{n} e^{-U_f/T_{m+1}} Q_m^n, \tag{136}$$

with $Q \in \mathscr{D}^l(H_1, a, b)$. We have extended the summation to negative m's and the definition of T_n to $n \in \mathbb{Z}$ by putting $T_n = T_0$ for $n \le 0$.

The reason for this extension is that, as $m \mapsto e^{-U_f/T_{m+1}}$ is decreasing, integration by parts shows that Q_m^n can be replaced by

$$(1 + b) a e^{-H_1/T_{m+1}} \prod_{l=m+2}^{n-1} (1 - a e^{-H_1/T_l}), \tag{137}$$

hence by

$$\frac{(1 + b)}{1 - a} a e^{-H_1/T_{m+1}} \prod_{l=m+2}^{n} (1 - a e^{-H_1/T_l}). \tag{138}$$

Now we want to find some condition on the cooling schedule $(T_m)_{m \in \mathbb{N}}$ for which

$$\sum_{m=-\infty}^{n} e^{U_f(1/T_n - 1/T_{m+1})} a e^{-H_1/T_{m+1}} \prod_{l=m+2}^{n} (1 - a e^{H_1/T_l}) \tag{139}$$

stays bounded.

One way to achieve this is to arrange that

$$e^{U_f(1/T_n - 1/T_{m+1})} \le \prod_{l=m+2}^{n} \left(1 + \frac{a}{2} e^{-H_1/T_l}\right). \tag{140}$$

This leads us to ask for

$$U_f\left(\frac{1}{T_{l+1}} - \frac{1}{T_l}\right) \le \frac{a}{4} e^{-H_1/T_l}. \tag{141}$$

We have seen that Theorem 4.4 applied to F_0 is sharp only for $f \in F_1$. Hence we will apply it only for $f \in F_1$. Accordingly we impose the preceding finite difference inequality only for $f \in F_1$. For $f \in F_2$ it is sharper to use the estimates we already have for $f \in F_1$ and Theorem 4.4 applied to F_2. We proceed in the same way with $F_2, F_3, \dots, F_{r-1}$.

We thus get that

$$\sup_{i \in E} P(X_n = f | X_0 = i) \le K e^{-U_f/T_n} \tag{142}$$

as soon as

$$U_f\left(\frac{1}{T_{l+1}} - \frac{1}{T_l}\right) \le \frac{a}{2} e^{-H_f/T_l}, \qquad f \in E. \tag{143}$$

Let us recall that $H_f = H_k$ for $f \in F_k - F_{k-1}$ by definition.

Now we can make the remark that we get more than we want. This will allow us to weaken our assumptions in the perspective of convergence at level α and of a finite time study.

First let us remark that we really need our finite difference inequalities only as long as $e^{-H_f/T_n} > \varepsilon$, for states f of energy $U_f \geq \alpha$. For states f of energy $U_f < \alpha$ we want it to hold only as long as $e^{-\alpha/T_n} > \varepsilon$. We can handily express this by thresholding the finite difference at level

$$\frac{\ln(\varepsilon^{-1})}{U_f \vee \alpha}. \tag{144}$$

A last remark is that we can weaken our finite difference inequalities even more by thresholding the energy U at some level $\eta \geq \alpha$. This thresholding of the energy function will allow us to find cooling schedules which satisfy our weakened finite difference inequalities uniformly in the energy U.

All these remarks introduce the following precise formulation.

PROPOSITION 6.1. *For any state space (E, q), there are positive constants B and K such that for any constants $\eta \geq 0$, $0 < \varepsilon < 1$, $\alpha \geq 0$, any energy function U and any cooling schedule T such that for any $f \in E$,*

$$\begin{aligned}(U_f \wedge \eta)\left(\frac{1}{T_m} \wedge \frac{\ln(\varepsilon^{-1})}{(U_f \vee \alpha) \wedge \eta} - \frac{1}{T_{m-1}} \wedge \frac{\ln(\varepsilon^{-1})}{(U_f \vee \alpha) \wedge \eta}\right) \\ \leq Be^{-H_f/T_{m-1}}, \qquad m > 0,\end{aligned} \tag{145}$$

we have for any $f \in E$,

$$\begin{aligned}&\sup_{i \in E} P(X_n = f | X_0 = i) \\ &\quad \leq Ke^{\eta/T_0}\left(e^{-(U_f \wedge \eta)/T_n} \vee \varepsilon^{(U_f \wedge \eta)/((U_f \vee \alpha) \wedge \eta)}\right), \qquad n > 0.\end{aligned} \tag{146}$$

Consequently in this case

$$\sup_{i \in E} P(U(X_n) \geq \alpha | X_0 = i) \leq Ke^{\eta/T_0}\left(e^{-(\alpha \wedge \eta)/T_n} \vee \varepsilon\right). \tag{147}$$

PROOF. Let (E, q) be fixed. Let $B = a/4$, where a is the constant of Theorem 4.4. We will assume without loss of generality that $a \leq 1$. Let $\varepsilon, \eta, \alpha, U, T$ be as in Proposition 6.1.

We are going to prove by induction on k the following assertion.

$\mathscr{U}_k$: There is a positive constant K such that for any $f \in F_k$,

$$\sup_{i \in E} L(E, E)_{i,0}^{f,n} \leq Ke^{\eta/T_0}\left(e^{-(U_f \wedge \eta)/T_n} \vee \varepsilon^{(U_f \wedge \eta)/((U_f \vee \alpha) \wedge \eta)}\right). \tag{148}$$

Assertion $\mathscr{U}_0$ is trivial because $L(E, E)_{i,0}^{j,n} \le 1$. Let us assume that we have proved $\mathscr{U}_{k-1}$. We have for any $i \in E$, and $g \in F_k - F_{k-1}$,

$$L(E, F_k - F_{k-1})_{i,0}^{g,n} = L(E - F_{k-1}, F_k - F_{k-1})_{i,0}^{g,n} + \{L(E, F_{k-1}) L(E - F_{k-1}, F_k - F_{k-1})\}_{i,0}^{g,n}; \tag{149}$$

hence

$$\begin{aligned} L(E, E)_{i,0}^{g,n} &\le (1+b) \prod_{s=1}^{n} (1 - ae^{-H_k/T_s}) \\ &\quad + \sum_{f \in F_{k-1}} \sum_{m=-\infty}^{n} Ke^{\eta/T_0} (e^{-(U_f \wedge \eta)/T_m} \vee \varepsilon^{(U_f \wedge \eta)/((U_f \vee \alpha) \wedge \eta)}) e^{-(U_g - U_f)^+/T_m} \\ &\quad \times (1+b) ae^{-H_k/T_m} \prod_{s=m+1}^{n} (1 - ae^{-H_k/T_s}) \\ &\le (1+b) \prod_{s=1}^{n} (1 - ae^{-H_k/T_s}) \\ &\quad + \sum_{m=-\infty}^{n} K(1+b) \mathrm{Card}(F_{k-1}) e^{\eta/T_0} (e^{-(U_g \wedge \eta)/T_m} \vee \varepsilon^{(U_g \wedge \eta)/((U_g \vee \alpha) \wedge \eta)}) \\ &\quad \times ae^{-H_k/T_m} \prod_{s=m+1}^{n} (1 - ae^{-H_k/T_s}). \end{aligned} \tag{150}$$

The first inequality in (150) is a consequence of Theorem 4.4. It can be justified by integrating by parts and remarking that

$$m \mapsto (e^{-(U_f \wedge \eta)/T_m} \vee \varepsilon^{(U_f \wedge \eta)/((U_f \vee \alpha) \wedge \eta)}) e^{-(U_g - U_f)/T_m} \tag{151}$$

is nonincreasing. As for the second inequality,

$$e^{-(U_f \wedge \eta)/T_m} e^{-(U_g - U_f)^+/T_m} \le e^{-(U_g \wedge \eta)/T_m}, \tag{152}$$

and if $\varepsilon \ge e^{-((U_f \vee \alpha) \wedge \eta)/T_m}$, then if $U_g \le U_f$,

$$\varepsilon^{(U_f \wedge \eta)/((U_f \vee \alpha) \wedge \eta)} e^{-(U_g - U_f)^+/T_m} \le \varepsilon^{(U_g \wedge \eta)/((U_g \vee \alpha) \wedge \eta)}, \tag{153}$$

and if $U_g > U_f$,

$$\begin{aligned} \varepsilon^{(U_f \wedge \eta)/((U_f \vee \alpha) \wedge \eta)} e^{-(U_g - U_f)/T_m} &\le \varepsilon^{(U_f \wedge \eta)/((U_f \vee \alpha) \wedge \eta)} \varepsilon^{(U_g - U_f)/((U_f \vee \alpha) \wedge \eta)} \\ &\le \varepsilon^{(U_g \wedge \eta)/((U_f \vee \alpha) \wedge \eta)} \le \varepsilon^{(U_g \wedge \eta)/((U_g \vee \alpha) \wedge \eta)}; \end{aligned} \tag{154}$$

hence, in any case,

$$\begin{aligned} &(e^{-(U_f \wedge \eta)/T_m} \vee \varepsilon^{(U_f \wedge \eta)/((U_f \vee \alpha) \wedge \eta)}) e^{-(U_g - U_f)^+/T_m} \\ &\quad \le (e^{-(U_g \wedge \eta)/T_m} \vee \varepsilon^{(U_g \wedge \eta)/((U_g \vee \alpha) \wedge \eta)}). \end{aligned} \tag{155}$$

Let us find an upper bound for the ratio

$$\rho = \frac{e^{-(U_g \wedge \eta)/T_m} \vee \varepsilon^{(U_g \wedge \eta)/((U_g \vee \alpha) \wedge \eta)}}{e^{-(U_g \wedge \eta)/T_n} \vee \varepsilon^{(U_g \wedge \eta)/((U_g \vee \alpha) \wedge \eta)}}. \tag{156}$$

We have

$$\rho = \prod_{s=m+1}^{n} r_s, \tag{157}$$

where

$$r_s = \frac{e^{-(U_g \wedge \eta)/T_{s-1}} \vee \varepsilon^{(U_g \wedge \eta)/((U_g \vee \alpha) \wedge \eta)}}{e^{-(U_g \wedge \eta)/T_s} \vee \varepsilon^{(U_g \wedge \eta)/((U_g \vee \alpha) \wedge \eta)}}. \tag{158}$$

We have

$$\begin{aligned} r_s &= \exp\left((U_g \wedge \eta)\left(\frac{1}{T_s} \wedge \frac{\ln(\varepsilon^{-1})}{(U_g \vee \alpha) \wedge \eta} - \frac{1}{T_{s-1}} \wedge \frac{\ln(\varepsilon^{-1})}{(U_g \vee \alpha) \wedge \eta}\right)\right) \\ &\le \exp(Be^{-H_g/T_{s-1}}) \\ &\le 1 + \frac{a}{2} e^{-H_k/T_{s-1}}. \end{aligned} \tag{159}$$

Hence

$$\rho \le \prod_{s=m}^{n-1}\left(1 + \frac{a}{2} e^{-H_k/T_s}\right) \le \left(1 + \frac{a}{2}\right) \prod_{s=m+1}^{n}\left(1 + \frac{a}{2} e^{-H_k/T_s}\right). \tag{160}$$

In the same way,

$$\begin{aligned} &\frac{1}{e^{-(U_g \wedge \eta)/T_n} \vee \varepsilon^{(U_g \wedge \eta)/((U_g \vee \alpha) \wedge \eta)}} \\ &\quad \le \left(e^{(U_g \wedge \eta)/T_0} \wedge \varepsilon^{-(U_g \wedge \eta)/((U_g \vee \alpha) \wedge \eta)}\right)\left(1 + \frac{a}{2}\right) \prod_{s=1}^{n}\left(1 + \frac{a}{2} e^{-H_k/T_s}\right) \\ &\quad \le e^{\eta/T_0}\left(1 + \frac{a}{2}\right) \prod_{s=1}^{n}\left(1 + \frac{a}{2} e^{-H_k/T_s}\right). \end{aligned} \tag{161}$$

Coming back to equation (150) we have

$$\begin{aligned} \sup_{i \in E} L(E, E)_{i,0}^{g,n} &\le \left(e^{-(U_g \wedge \eta)/T_n} \vee \varepsilon^{(U_g \wedge \eta)/((U_g \vee \alpha) \wedge \eta)}\right) \\ &\quad \times \Bigg\{(1 + b)\left(1 + \frac{a}{2}\right) e^{\eta/T_0} \prod_{s=1}^{n}\left(1 - \frac{a}{2} e^{-H_k/T_s}\right) \\ &\quad\quad + \operatorname{Card}(F_{k-1}) K e^{\eta/T_0} 2\left(1 + \frac{a}{2}\right)(1 + b) \\ &\quad\quad \times \sum_{m=-\infty}^{n} \frac{a}{2} e^{-H_k/T_m} \prod_{s=m+1}^{n}\left(1 - \frac{a}{2} e^{-H_k/T_s}\right)\Bigg\} \\ &\le (2K \operatorname{Card}(F_{k-1}) + 1)\left(1 + \frac{a}{2}\right)(1 + b) e^{\eta/T_0} \\ &\quad \times \left(e^{-(U_g \wedge \eta)/T_n} \vee \varepsilon^{(U_g \wedge \eta)/((U_g \vee \alpha) \wedge \eta)}\right). \end{aligned} \tag{162}$$

This proves $\mathscr{U}_k$ from $\mathscr{U}_{k-1}$ and hence, by induction on k, Proposition 6.1. □

7. Almost optimal cooling schedules. We are going to show that in Theorem 5.2, for any given energy U, the constant $D_\lambda(U)$ is sharp for λ small enough.

More precisely, for any energy level $\lambda > 0$ let us put

$$\tilde{D}_\lambda = \sup_{f \in E - F(E)} \frac{H_f}{U_f \vee \lambda}; \tag{163}$$

the constant D_λ will be sharp when $D_\lambda = \tilde{D}_\lambda$, which is the case for U fixed and $\lambda > 0$ small enough.

We will say that the triangular cooling schedule $(T_n^N)_{1 \le n \le N}$ is almost λ-optimal if there is a constant K such that

$$\max_{i \in E} P\big(U(X_N^N) \ge \lambda | X_0 = i\big) \le \frac{K}{N^{D_\lambda^{-1}}}. \tag{164}$$

THEOREM 7.1. *For any energy landscape (E, q) any energy function U and any energy level $\lambda > 0$, there exist a triangular cooling schedule T_n^N and a positive constant K such that*

$$\sup_{i \in E} P\big(U(X_N^N) \ge \lambda | X_0 = i\big) \le K N^{-\tilde{D}_\lambda^{-1}}, \qquad N > 0. \tag{165}$$

Hence when $\tilde{D}_\lambda = D_\lambda$, T is almost λ-optimal.

Let us introduce the technical proof of this theorem with some remarks.

We are going to use Proposition 6.1 for one fixed energy function U. Once U is fixed we see that some of our finite difference inequalities are redundant.

For a given H_k, the most constraining inequality is obtained for the largest value of U_f, $f \in F_k - F_{k+1}$. On the other hand, for one given state $f \in E - F_0$, the steepest cooling schedule corresponding to the finite difference inequality indexed by f is

$$\frac{1}{T_n} = \frac{1}{H_f} \ln\left(\frac{H_f B n}{U_f} + e^{H_f / T_0} \right). \tag{166}$$

Hence for n large, the most constraining equations are obtained for the largest value of H_f, which is H_1.

Among the inequalities corresponding to states $f \in F_1 - F_0$, those with minimum energy U_f will have the longest life. Hence we will consider $\tilde{H}_1 = H_1$ and $\tilde{U}_1 = \min_{f \in F_1 - F_0} U_f$. We will apply Proposition 6.1 with threshold $\eta = \tilde{U}_1$. With this threshold the most constraining equation is

$$\tilde{U}_1 \left(\frac{1}{T_{n+1}} - \frac{1}{T_n} \right) \le B e^{-\tilde{H}_1 / T_n}. \tag{167}$$

The steepest solution to our leading finite difference inequality has the form

$$(168)\qquad \frac{1}{T_n} = \frac{1}{\tilde{H}_1}\ln\left(\frac{\tilde{H}_1 Bn}{\tilde{U}_1} + e^{-\tilde{H}_1/T_0}\right).$$

We find the usual rate $T_n \sim \tilde{H}_1/\ln n$ linked with the first eigenvalue of the transition kernel.

Let us remark that the introduction of the constant $(1/\tilde{H}_1)\ln(\tilde{H}_1 B/\tilde{U}_1)$ is crucial in order to obtain the right speed of convergence. This fact has been clarified by a more precise study in our thesis and in [4].

All this is good as long as $e^{-\tilde{U}_1/T_n} \geq \varepsilon$, that is, as long as time

$$(169)\qquad m_1 = \left\lfloor \frac{\tilde{U}_1}{B\tilde{H}_1}\left(\varepsilon^{-(\tilde{H}_1/\tilde{U}_1)} - e^{\tilde{H}_1/T_0}\right)\right\rfloor.$$

Afterwards our leading finite difference inequality disappears from the set of constraints needed in Proposition 6.1. Hence another one is going to take the leadership.

To see which one, we have to consider all the surviving inequalities. These are the inequalities indexed by the states f of energy $U_f < \tilde{U}_1$. Among them, the most constraining ones for large times are obtained for states f such that

$$(170)\qquad H_f = \max\left\{H_g,\, U_g < \tilde{U}_1\right\} = \tilde{H}_2.$$

Thus all surviving equations are dominated by

$$(171)\qquad \tilde{U}_1\left(\frac{1}{T_{n+1}} - \frac{1}{T_n}\right) \leq Be^{-\tilde{H}_2/T_n}.$$

The steepest solution to this equation has the form

$$(172)\qquad \frac{1}{T_n} = \frac{1}{\tilde{H}_2}\ln\left(\frac{\tilde{H}_2 B}{\tilde{U}_1}(n - m_1) + \varepsilon^{-(\tilde{H}_2/\tilde{U}_1)}\right).$$

We will use this formula as long as there remains in the set of equations some inequality with right member $Be^{-\tilde{H}_2/T_n}$, that is, up to time

$$(173)\qquad m_2 = m_1 + \left\lfloor \frac{\tilde{U}_1}{B\tilde{H}_2}\left(\varepsilon^{-(\tilde{H}_2/\tilde{U}_2)} - \varepsilon^{-(\tilde{H}_2/\tilde{U}_1)}\right)\right\rfloor,$$

where

$$(174)\qquad \tilde{U}_2 = \min\left\{U_g,\, H_g = \tilde{H}_2\right\}.$$

We carry on this process, introducing two sequences $\tilde{H}_k$ and $\tilde{U}_k$ as long as the set of equations in the hypothesis of Proposition 6.1 is nonempty, that is, as long as $\tilde{U} \geq \lambda$. The last value of $\tilde{H}$ may be 0. This case has to be treated separately, as will be clear in the following detailed proof.

PROOF OF THEOREM 7.1. Let (E, q, U) be fixed.

Let us define the sequence of depths $(\tilde{H}_k)_{k=1,\dots,r}$ and the sequence of energy values $(\tilde{U}_k)_{k=1,\dots,r}$ by

$$(175)\qquad \begin{aligned} \tilde{H}_1 &= \max\{H(C) | C \in \mathscr{M}(E - F(E))\}, \\ \tilde{U}_1 &= \min\{U(C) | C \in \mathscr{M}(E - F(E)),\, H(C) \geq \tilde{H}_1\}, \\ &\vdots \\ \tilde{H}_k &= \max\{H(C) | C \in \mathscr{M}(E - F(E)),\, U(C) < \tilde{U}_{k-1}\}, \\ \tilde{U}_k &= \min\{U(C) | C \in \mathscr{M}(E - F(E)),\, H(C) \geq \tilde{H}_k\}, \\ &\vdots \end{aligned}$$

where the definition carries on by induction as long as the involved sets are nonempty.

Let

$$(176)\qquad r_\lambda = r \wedge \inf\{k | \tilde{U}_k \leq \lambda\}.$$

For any fixed $\varepsilon \in]0, 1[$ we will build a finite schedule $T_n^{N_\varepsilon}$ such that

$$(177)\qquad \sup_{i \in E} P\left(U\left(X_{N_\varepsilon}^{N_\varepsilon}\right) \geq \lambda | X_0 = i\right) \leq K\varepsilon,$$

and we will show that for some constant K' we have $N_\varepsilon \leq K' \varepsilon^{-\tilde{D}_\lambda}$.

Let us choose some positive constant T_0.

Let B be as in Proposition 6.1. We define the sequence of times $(m_k)_{k=1,\dots,r_\lambda}$ by

$$(178)\qquad \begin{aligned} m_k = &\left\lceil \frac{\tilde{U}_1}{B\tilde{H}_1} \left(\varepsilon^{-\tilde{H}_1/\tilde{U}_1} - e^{\tilde{H}_1/T_0}\right) \right\rceil \\ &+ \sum_{s=2}^{k} \left\lceil \frac{\tilde{U}_{k-1}}{B\tilde{H}_k} \left(\varepsilon^{-(\tilde{H}_k/\tilde{U}_k)} - \varepsilon^{-(\tilde{H}_k/\tilde{U}_{k-1})}\right) \right\rceil, \qquad m < r_\lambda, \end{aligned}$$

where $\lceil x \rceil = \min\{n \in \mathbb{Z} | n \geq x\}$.

If $\tilde{H}_{r_\lambda} > 0$ we define m_{r_λ} by

$$(179)\qquad m_{r_\lambda} = m_{r_\lambda - 1} + \left\lceil \frac{\tilde{U}_{r_\lambda - 1}}{B\tilde{H}_{r_\lambda}} \left(\varepsilon^{-(\tilde{H}_{r_\lambda}/\lambda)} - \varepsilon^{-(\tilde{H}_{r_\lambda}/\tilde{U}_{r_\lambda - 1})}\right) \right\rceil.$$

If $\tilde{H}_{r_\lambda} = 0$, we put

$$(180)\qquad m_{r_\lambda} = m_{r_\lambda - 1} + \frac{\ln(A\varepsilon^{-1})}{\ln(\rho^{-1})},$$

where the constants A and ρ are such that the transition kernel p_0 at temperature $T = 0$ satisfies

$$\sup_{i \in E} p_0^n(i, F_{r_\lambda}) \le A\rho^n. \tag{181}$$

(Let us recall that $E - F_{r_\lambda}$ is in this case the set of the local and global minima of the energy landscape.)

Let us put $N_\varepsilon = m_{r_\lambda}$, and let us define $(T_n^{N_\varepsilon})_{0 \le n \le N_\varepsilon}$ by

$$\frac{1}{T_n} = \begin{cases} \dfrac{1}{\tilde{H}_1} \ln\left(\dfrac{B\tilde{H}_1}{\tilde{U}_1} n + e^{\tilde{H}_1/T_0} \right), & 0 < n < m_1, \\ \dfrac{1}{\tilde{H}_k} \ln\left(\dfrac{B\tilde{H}_k}{\tilde{U}_{k-1}} (n - m_{k-1}) + \varepsilon^{-(\tilde{H}_k/\tilde{U}_{k-1})} \right), & \\ & m_{k-1} \le n < m_k,\ k < r_\lambda, \end{cases} \tag{182}$$

and

(i) in the case $\tilde{H}_{r_\lambda} > 0$,

$$\frac{1}{T_n} = \frac{1}{\tilde{H}_{r_\lambda}} \ln\left(\frac{B\tilde{H}_{r_\lambda}}{\tilde{U}_{(r_\lambda - 1)}} (n - m_{(r_\lambda - 1)}) + \varepsilon^{-(\tilde{H}_{r_\lambda}/\tilde{U}_{(r_\lambda - 1)})} \right), \quad m_{(r_\lambda - 1)} \le n \le m_{r_\lambda}; \tag{183}$$

(ii) in the case $\tilde{H}_{r_\lambda} = 0$,

$$\frac{1}{T_{m_{(r_\lambda - 1)}}} = \frac{1}{\tilde{U}_{(r_\lambda - 1)}} \ln(\varepsilon^{-1}), \qquad T_n = 0, \quad m_{(r_\lambda - 1)} < n \le m_{r_\lambda}. \tag{184}$$

Let us assume for a while that $\tilde{H}_{r_\lambda} > 0$. With our choice of cooling schedule, the hypotheses of Proposition 6.1 are satisfied for $\eta = \tilde{U}_1 \vee \lambda$, and $\alpha = \lambda$; hence we have

$$\sup_{i \in E} P\big(U(X_{N_\varepsilon}) \ge \lambda | X_0 = i\big) \le K e^{(\tilde{U}_1 \vee \lambda)/T_0} \big(e^{-\lambda / T_{N_\varepsilon}^{N_\varepsilon}} \vee \varepsilon \big), \tag{185}$$

but we have chosen N_ε such that $e^{-\lambda/T_{N_\varepsilon}^{N_\varepsilon}} = \varepsilon$; hence

$$\sup_{i \in E} P\big(U(X_{N_\varepsilon}) \ge \lambda | X_0 = i\big) \le K e^{(\tilde{U}_1 \vee \lambda)/T_0} \varepsilon. \tag{186}$$

If $\tilde{H}_{r_\lambda} = 0$, we can apply Proposition 6.1 up to time $m_{(r_\lambda - 1)}$. We get that

$$\sup_{i \in E} P\Big(U\big(X_{m_{(r_\lambda - 1)}}\big) \ge \tilde{U}_{(r_\lambda - 1)} | X_0 = i\Big) \le K e^{(\tilde{U}_1 \vee \lambda)/T_0} \varepsilon; \tag{187}$$

hence

$$\sup_{i \in E} P\Big(U\big(X_{N_\varepsilon}\big) \ge \tilde{U}_{(r_\lambda - 1)} | X_0 = i\Big) \le K e^{(\tilde{U}_1 \vee \lambda)/T_0} \varepsilon. \tag{188}$$

Moreover,

$$\sup_{i \in E} P\left(\lambda \le U(X_{N_\varepsilon}) < \tilde{U}_{r_\lambda - 1} | X_0 = i\right) \le \varepsilon. \tag{189}$$

Indeed, as $H_{r_\lambda} = 0$, for any $f \in E$ such that $\lambda \le U_f < \tilde{U}_{r_\lambda - 1}$, we have $H_f = 0$. Hence

$$\sup_{i \in E} P\left(U(X_{N_\varepsilon}) \ge \lambda | X_0 = i\right) \le \varepsilon(K+1) e^{(\tilde{U}_1 \vee \lambda)/T_0}. \tag{190}$$

Finally we have, putting $\tilde{U}_0 = \tilde{U}_1$,

$$N_\varepsilon \le \begin{cases} r_\lambda \sup_{k=1,\dots,r_\lambda} \left(\dfrac{\tilde{U}_{(k-1)}}{B\tilde{H}_k} \right) \varepsilon^{-\tilde{D}_\lambda}, & \text{if } \tilde{H}_{r_\lambda} > 0, \\ (r_\lambda - 1) \sup_{k=1,\dots,r_\lambda - 1} \left(\dfrac{\tilde{U}_{(k-1)}}{B\tilde{H}_k} \right) \varepsilon^{-\tilde{D}_\lambda} + \dfrac{\ln(A\varepsilon^{-1})}{\ln(\lambda^{-1})}, & \text{if } \tilde{H}_{r_\lambda} = 0. \end{cases} \tag{191}$$

Hence (for U fixed) there is a positive constant K' such that

$$\varepsilon \le \frac{K'}{N_\varepsilon^{\tilde{D}_\lambda^{-1}}}. \tag{192}$$

□

8. Logarithmically almost optimal exponential cooling schedules. For a given energy landscape (E, q, U) and a given positive energy level λ, we will say that a triangular cooling schedule $(T_n^N)_{1 \le n \le N}$ is logarithmically almost λ-optimal if

$$\lim_{N \to +\infty} (\ln N)^{-1} \ln \max_{i \in E} P\left(U(X_N^N) \ge \lambda | X_0 = i\right) = -D_\lambda^{-1}. \tag{193}$$

In this section we want to use Proposition 6.1 uniformly in the energy function U.

We will look for cooling schedules $(T_n^N)_{1 \le n \le N}$ such that $n \mapsto T_n^N$ is *convex*. In this case, for any state f, the finite difference inequality

$$\begin{gathered} (U_f \wedge \eta)\left(\frac{1}{T_{n+1}} - \frac{1}{T_n} \right) \le B e^{-H_f/T_n}, \\ e^{-((U_f \vee \alpha) \wedge \eta)/T_n} \ge \varepsilon, \end{gathered} \tag{194}$$

has to be checked only at time

$$N_f = \sup\{n, e^{-((U_f \vee \alpha) \wedge \eta)/T_n} \le \varepsilon\}. \tag{195}$$

If the threshold η is large enough compared with the values of U, namely, if $\eta \ge H_1/\tilde{D}_\alpha$, then for any state f we have $H_f \le \tilde{D}_\alpha((U_f \vee \alpha) \wedge \eta)$; hence

$$e^{-H_f/T_{N_f}} \ge \varepsilon^{H_f/((U_f \vee \alpha) \wedge \eta)} \ge \varepsilon^{\tilde{D}_\alpha}. \tag{196}$$

Moreover, $(U_f \wedge \eta) \le T_{N_f} \ln(\varepsilon^{-1})$; hence the finite difference equation indexed by f in Proposition 6.1 can be strengthened into

$$T_{N_f}\left(\frac{1}{T_{N_f+1}} - \frac{1}{T_{N_f}}\right) \le \frac{Be^{\tilde{D}_\alpha}}{\ln(\varepsilon^{-1})}, \tag{197}$$

with the supplementary hypothesis that $n \mapsto 1/T_n$ is convex.

Thus we see that it is possible to find one triangular cooling schedule for which the hypotheses of Proposition 6.1 are satisfied uniformly for all energies U satisfying $H_1 \le \tilde{D}_\alpha \eta$. Any triangular schedule satisfying the finite difference inequality

$$T_n\left(\frac{1}{T_{n+1}} - \frac{1}{T_n}\right) \le \frac{B\varepsilon^{\tilde{D}_\alpha}}{\ln(\varepsilon^{-1})} \tag{198}$$

and the convexity assumption will do.

The steepest one is given by

$$\frac{1}{T_n} = A\exp\left(\frac{B\varepsilon^{\tilde{D}_\alpha}}{\ln(\varepsilon^{-1})} n\right). \tag{199}$$

These remarks introduce the following precise formulation ($\ln_k$ is the kth iterate of the logarithm function).

THEOREM 8.1. *For any energy landscape with communications (E, q) there exists a positive constant B such that for any positive constants A, η, there is a positive constant K such that for any positive constant $\delta \le \eta$, for any energy function U satisfying $(H(E - F(E)))/(\tilde{D}_\delta(E, U, q)) \le \eta$, for any initial distribution $\mathscr{L}_0$ and for any triangular cooling schedule (T_n^N) such that*

$$\begin{aligned} 1/T_n^N &= A\exp(n\xi), \\ \xi &\le B\frac{\varepsilon^{\tilde{D}_\delta}}{\ln \varepsilon^{-1}}, \\ N &\ge \frac{1}{\xi}\left[\ln_2(\varepsilon^{-1}) - \ln(\delta A)\right], \end{aligned} \tag{200}$$

the triangular annealing algorithm (E, U, q, T, X) satisfies

$$P\left(U\left(X_N^N\right) \ge \delta\right) \le K\varepsilon. \tag{201}$$

COROLLARY 8.2. *If we take*

$$\xi = (B \wedge 1)\frac{\varepsilon^{\tilde{D}_\delta}}{\ln(\varepsilon^{-1})}$$

and

$$N = \frac{1}{\xi}\left[\ln(\xi^{-1}) - \ln(\delta A)\right], \tag{202}$$

or

$$N = \frac{1}{\xi}\left[\ln_2(\xi^{-1}) + \ln(\tilde{D}_\delta^{-1}) - \ln(\delta A)\right] \tag{203}$$

or

$$\xi = \frac{1}{N}[\ln N - \ln(\delta A)] \tag{204}$$

or

$$\xi = \frac{1}{N}\left[\ln_2 N + \ln(\tilde{D}_\delta^{-1}) - \ln(\delta A)\right], \tag{205}$$

the hypotheses of Theorem 8.1 *are fulfilled for N (or* ξ^{-1} *or* ε^{-1}*) large enough, and* $1/T_n^N = A\exp(\xi n)$ *is logarithmically almost* δ*-optimal when* $D_\delta = \tilde{D}_\delta$.

REMARK. In this context, it is not unreasonable to use an adaptive stopping rule for the choice of N, since for fixed ξ the annealing algorithm corresponding to

$$\frac{1}{T_n} = A\exp(n\xi) \tag{206}$$

is not ergodic and $U(X_n)$ is almost surely constant for n large [the stopping rule can be that $U(X_n)$ has been constant for σ moves].

REMARK ON THE CHOICE OF A. The proof shows that it is wise to take A such that $A(H(E - F(E)))/\tilde{D}_\delta(E, U, q))$ is not too big, say

$$A = \frac{\tilde{D}_\delta(E, U, q)}{H(E - F(E))}. \tag{207}$$

PROOF OF THEOREM 8.1. We have for any $f \in E$,

$$\begin{aligned}
&(U_f \wedge \eta)\left(\frac{1}{T_m} \wedge \frac{\ln(\varepsilon^{-1})}{(U_f \vee \delta) \wedge \eta} - \frac{1}{T_{m-1}} \wedge \frac{\ln(\varepsilon^{-1})}{(U_f \vee \delta) \wedge \eta}\right)\\
&\quad\le (U_f \wedge \eta)\xi\left(\frac{1}{T_m} \wedge \frac{\ln(\varepsilon^{-1})}{(U_f \vee \delta) \wedge \eta}\right)\\
&\quad\le \xi\ln(\varepsilon^{-1}) \le Be^{\tilde{D}_\delta}\\
&\quad\le \varepsilon^{H_f/((U_f \vee \delta) \wedge \eta)} \le \exp\left(-H_f\left(\frac{\ln(\varepsilon^{-1})}{(U_f \vee \delta) \wedge \eta} \wedge \frac{1}{T_{m-1}}\right)\right).
\end{aligned} \tag{208}$$

Hence inequality (145) is satisfied with $\alpha = \delta$ when $1/T_{m-1} \le (\ln(\varepsilon^{-1}))/((U_f \vee \delta) \wedge \eta)$, hence for any m, because it is otherwise trivial. Thus we deduce from Proposition 6.1 that for some constant K,

$$\sup_{i \in E} P(U(X_n) \ge \delta | X_0 = i) \le Ke^{A\eta}(e^{-\delta/T_n} \vee \varepsilon); \tag{209}$$

hence for $N \geq \xi^{-1}(\ln_2(\varepsilon^{-1}) - \ln(\delta A))$ we have

$$\sup_{i \in E} P(U(X_N) \geq \delta | X_0 = i) \leq K e^{A\eta} \varepsilon. \tag{210}$$

The proof of the corollary involves only elementary verifications, stemming from the inequality

$$\ln(\xi^{-1}) \geq \ln(B^{-1}) + \tilde{D}_\delta \ln(\varepsilon^{-1}) + \ln_2(\varepsilon^{-1}). \tag{211}$$ □

Conclusion. This study had two aims: to perform the large deviations program directly in the time inhomogeneous case; and to apply it to get finite time results.

It is good news that the large deviations program can be performed with *uniform estimates* as well in the cooling schedule as in the energy function. Without uniform constants, we would not have been able to derive the optimal convergence rate and almost optimal schedules.

The finite time point of view is a natural description of practical computer experiments, because in practice the simulation time is always finite. Hence it is not artificial to assume that one repeats the same experiment with different simulation times and is allowed to change the whole cooling schedule at each time. What can be regarded as only a slight increase of freedom in the control of the process—allowing triangular cooling schedules—interestingly enough brings results qualitatively different from the asymptotic ones.

This new point of view has shed light on a previously unnoticed fundamental constant: *the difficulty of the energy landscape*, which governs the optimal finite time convergence rate.

Unexpectedly, this constant is not linked with the critical depth of the energy landscape. More generally in the finite time setting, when the system approaches the end of simulation time, the important eigenvalue of the transition matrix is no more the second one. This is one of the mathematical translations of the fact that simulated annealing is performing a *hierarchial search*. At the beginning, the important eigenvalue is linked with the deepest nonglobal maxima. Then the leadership is taken by states of decreasing depths.

In other terms, the probability of being in the deepest local minima is decreased at high temperature. When temperature becomes lower this probability becomes almost constant and the effect of carrying on the simulation is to decrease the probability of being in shallower local minima.

This hierarchical behaviour has a *practical* consequence: In order to achieve the optimal convergence rate, it is necessary to tune the whole cooling schedule to the simulation time, *even in the limit of large simulation times*. For instance, if you double the simulation time, it will be more efficient to increase both the time spent at high temperatures and the time spent at low temperatures than it would be to increase only the time spent at low temperatures.

This is clearly reflected in the construction of a piecewise logarithmic cooling schedule (Theorem 7.1). The number of logarithmic patches depends on the energy function but not on the simulation time. What depends on the simulation time is the length of the interval on which each formula is used.

Another practical result is that it is possible to have a cooling schedule which is *uniformly efficient for a whole range of energy functions*. This schedule is of exponential form. Hence the finite point of view brings the first justification for exponential schedules. But it brings even more—it establishes that *exponential schedules should be triangular* and gives the nature of the dependence of the constant to be put in the exponential on the simulation time.

Finally, it is hoped that this paper brings a methodology which could be used to answer other questions about simulated annealing. The large deviations results constitute a localization method. A systematic calculus is established. It tells how to compute an estimate of the probability of any large time behaviour which can be described as a succession of jumps from one cycle to another one, or more generally from one subdomain of the state space to another. It covers the study of simulated annealing with absorption (cf. [8]) and of simulated annealing with absorption coupled with a "particles bath."

Proposition 6.1 gives the efficiency of a cooling schedule in terms of a set of finite-time–finite-difference inequalities. We have exploited it in two extreme situations: when everything is known about the energy function and when nothing is known. In practical intermediate situations, it is possible to use it to derive cooling schedules adapted to the available information.

Let us point out eventually that we have followed another avenue in other papers ([3]–[5]); instead of aiming at uniformity in the energy function, we got asymptotically sharp multiplicative constants for one fixed energy function. Proving "sharp large deviation estimates" is technically much harder, but gives a precise theoretical understanding of what is going on at low temperatures.

Acknowledgment. The present work has been directed by Professor Robert Azencott, to whom thanks are due for his useful suggestions and encouragements.

REFERENCES

[1] Azencott, R. (1988). Simulated annealing. *Séminaire Bourbaki 40ième année, 1987–1988* **697**.

[2] Catoni, O. (1988). Grandes déviations et décroissance de la température dans les algorithmes de recruit. *C. R. Acad. Sci. Paris Sér. I* **307** 535–538.

[3] Catoni, O. (1991a). Sharp large deviations estimates for simulated annealing algorithms. *Ann. Inst. H. Poincaré* **27** 291–383.

[4] Catoni, O. (1991b). Applications of sharp large deviations estimates to optimal cooling schedules. *Ann. Inst. H. Poincaré* **27** 463–518.

[5] Catoni, O. (1990). Large deviations for annealing. Dissertation, Université Paris-Sud Orsay.

[6] Chiang, T. S. and Chow, Y. (1988). On the convergence rate of annealing processes. *SIAM J. Control Optim.* **26** 1455–1470.

[7] Chiang, T.-S. and Chow, Y. (1989). A limit theorem for a class of inhomogeneous Markov processes. *Ann. Probab.* **17** 1483–1502.

[8] CHIANG, T.-S. and CHOW, Y. (1990a). The asymptotic behaviour of simulated annealing processes with absorption. Unpublished manuscript.

[9] CHIANG, T.-S. and CHOW, Y. (1992). On occupation times of annealing processes. *Bull. Inst. Math. Acad. Sinica*. To appear.

[10] DOBRUSHIN, R. L. (1956). Central limit theorems for non-stationary Markov chains. I, II. *Theory Probab. Appl.* **1** 65–80, 329–383. (English translation.)

[11] FRIEDLIN, M. I. and WENTZELL, A. D. (1984). *Random Perturbations of Dynamical Systems*. Springer, New York.

[12] GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Trans. Pattern Analysis and Machine Intelligence* **6** 721–741.

[13] GIDAS, B. (1985). Non-stationary Markov chains and convergence of the annealing algorithms. *J. Statist. Phys.* **39** 73–131.

[14] HÁJEK, B. (1988). Cooling schedules for optimal annealing. *Math. Oper. Res.* **13** 311–329.

[15] HOLLEY, R. A., KUSUOKA, S. and STROOCK, D. W. (1989). Asymptotics of the spectral gap with applications to the theory of simulated annealing. *J. Funct. Anal.* **83** 333–347.

[16] HOLLEY, R. and STROOCK, D. (1988). Simulated annealing via Sobolev inequalities. *Comm. Math. Phys.* **115** 553–569.

[17] HWANG, C. R. and SHEU, S. J. (1986). Large time behaviours of perturbed diffusion Markov processes with applications. I, II and III. Unpublished manuscript.

[18] HWANG, C. R. and SHEU, S. J. (1988a). Singular perturbed Markov chains and exact behaviours of simulated annealing process. Unpublished manuscript.

[19] HWANG, C. R. and SHEU, S. J. (1988b). On the weak reversibility condition in simulated annealing. Unpublished manuscript.

[20] IOSIFESCU, M. and THEODORESCU, R. (1969). *Random Processes and Learning*. Springer, New York.

[21] ISAACSON, D. L. and MADSEN, R. W. (1973). Strongly ergodic behaviour for non-stationary Markov processes. *Ann. Probab.* **1** 329–335.

[22] KIRKPATRICK, S., GELATT, C. D. and VECCHI, M. P. (1983). Optimization by simulated annealing. *Science* **220** 621–680.

[23] TSITSIKLIS, J. N. (1988). A survey of large time asymptotics of simulated annealing algorithms. In *Stochastic Differential Systems, Stochastic Control Theory and Applications. IMA Vol. Math. Appl.* **10** (W. Fleming and P. L. Lions, eds.). Springer, New York.

[24] TSITSIKLIS, J. N. (1989). Markov chains with rare transitions and simulated annealing. *Math. Oper. Res.* **14** 70–90.

DIAM—INTELLIGENCE ARTIFICIELLE
ET MATHÉMATIQUES
LABORATOIRE DE MATHÉMATIQUES, UA 762 DU CNRS
DÉPARTMENT DE MATHÉMATIQUES ET D'INFORMATIQUE
ECOLE NORMALE SUPÉRIEURE
45, RUE D'ULM 75005 PARIS
FRANCE