PHYSICAL REVIEW
LETTERS

VOLUME 86 28 MAY 2001 NUMBER 22

Mathematical Analysis of Coupled Parallel Simulations

Michael R. Shirts and Vijay S. Pande
Department of Chemistry, Stanford University,
Stanford, California 94305-5080
(Received 10 October 2000)

A set of parallel replicas of a single simulation can be statistically coupled to closely approximate long trajectories. In many cases, this produces nearly linear speedup over a single simulation (M times faster with M simulations), rendering previously intractable problems within reach of large computer clusters. Interestingly, by varying the coupling of the parallel simulations, it is possible in some systems to obtain greater than linear speedup. The methods are generalizable to any search algorithm with long residence times in intermediate states.

DOI: 10.1103/PhysRevLett.86.4983 PACS numbers: 02.70.Ns, 82.20.Uv, 82.20.Wt, 87.15.Aa

Many simulations of interest to science become intractable for a single computer as the number of degrees of freedom increase. Thus, one of the most important problems of scientific computing is large-scale parallelization of algorithms. Traditional parallelization schemes require extremely fast communication between multiple processors and frequently do not scale to large numbers of processors. However, it is possible to statistically couple many simulations run in parallel to obtain a result that is equivalent to a longer simulation, effectively parallelizing this process [1]. In this way, rare events can be simulated in a much shorter time than by single processor simulations, or even than by running the same number of uncoupled simulations. This parallelization can utilize much slower communication times (minutes to hours, as opposed to microseconds) and thus makes many computationally intensive simulations accessible to clusters of PC's, instead of only large supercomputers, and in many cases may scale to hundreds or thousands of processors [2].

It is obvious that running multiple simulations will improve the rate at which the phase space of the simulation is explored. However, in general M independent simulations will not explore the simulation space in the same way as a single simulation that is M times faster. For example, the time-dependent behavior of the system under study may not be captured by this set of parallel simulations. The question is whether the M simulations can be coupled in some way in order to accurately reproduce properties of a single long simulation.

At least one case of such a coupling algorithm already exists, for multistate systems where the total process time is dominated by the waiting time of transitions between states. It was first applied by Voter [1] to small solid-state systems under the name of "parallel replica dynamics" and later by Baker *et al.* [3] to large-scale atomistic protein simulations as "ensemble dynamics." In this case, M simulations are run in parallel. When one of the M simulations makes a transition to another state, all of the M simulations are reset to the new state of the transitioning simulation, and the M simulations continue from there, repeating this process of finding transitions. We call this coupling scheme "transition coupling," as the independent simulations interact only when they undergo certain long time scale transitions. While it has currently been applied to molecular dynamics simulations of condensed phases, it works for any algorithm whose total time is dominated by waiting times for rare events of interest.

In this paper, we present a formalism for calculating the computational speedup (the increase in speed obtained using M processors versus a single processor, assuming one simulation per processor) of different coupling schemes for arbitrary probability distributions, and for interpreting the simulation data to predict rates. A physical argument has previously been presented demonstrating that on an energy landscape with exponential transitions beween states, transition coupling should give an exactly linear speedup of rates with number of simulations and that the distribution of frequencies of different reaction pathways should be preserved [1]. Here we present a more mathematical and generalizable derivation, as well as extensions to various types of more complicated processes. We find that tremendous speedups in simulating rare events can be obtained, rendering previously intractable problems within reach. We first present the algorithm in its most abstract form and then show how it can be applied to a specific type of computationally demanding problem, atomistic condensed matter simulations.

Parallel simulations with two discrete states.—We treat the system to be simulated as occupying one of N discrete states, with some time-dependent probability of instantaneous transition $P_{ij}(t)$ from state i to any one of the other states j at time t. No other physical assumptions are made about the states at this point.

We first deal with a transition between only two states, with M independent simulations. If there is a probability $P_{12}(t)$ of transition from state 1 to state 2 at time t, then the probability $P_{M,12}(t)$ of the first of M simulations making the transition at time t will be M times the probability of one simulation making a transition in that interval (since any of the M simulations could be first), multiplied by the probability of the $M - 1$ other simulations not yet having made a transition at time t, yielding

$$P_{M,12}(t) = MP_{12}(t)\left(1 - \int_0^t P_{12}(t')\,dt'\right)^{M-1}. \qquad (1)$$

We note that this can be expressed as

$$P_{M,12}(t) = -\frac{d}{dt}[1 - F_{12}(t)]^M, \qquad (2)$$

using $F(t) \equiv \int_0^t P(t')\,dt'$, the cumulative probability function.

There is one particularly interesting case. If the movement between states is a Poisson process (a process in which the rate of transition is both low and constant with time), then we will have an exponential probability $P_{12}(t) = k\exp(-kt)$ of passing from state 1 to state 2, with average time $\langle t_1\rangle = 1/k$, and median time (which is often a more relevant measure, since for most simulations we are often more interested in the typical first-passage time than in the tails of the distribution) of $t_{1/2} = \ln 2/k$. Then $F_{12}(t) = 1 - \exp(-kt)$, meaning

$$P_{M,12}(t) = -\frac{d}{dt}[1 - F_{12}(t)]^M = Mk\exp(-Mkt). \qquad (3)$$

This is identical to the probability distribution of a single simulation with the replacement $k \rightarrow Mk$. Since $\langle t\rangle = 1/k$ and $t_{1/2} = \ln 2/k$ are both functions of $1/k$, we obtain identically linear scaling in time with M independent (not coupled in any way) simulations.

However, if the probability distribution of transition is nonexponential, this linear increase in rates will not hold with M parallel simulations. The exact details will depend upon the type of simulation and can be easily computed using the formula above. Some simple cases are of interest, however. Given a monotonic probability distribution, if the probability density is less convex (i.e., has a shorter tail) than an exponential distribution, then $\langle t\rangle$ will be greater than $\frac{1}{M}\langle t\rangle$. If the probability distribution is more convex (with a longer tail than an exponential), then the average times of first passage will be less than $\frac{1}{M}\langle t\rangle$. In this latter case, the effective rate increases greater than linearly with number of simulations, a useful fact that will be elaborated on later in this Letter.

Parallel simulations with many states.—Many systems of interest will have some number of non-negligible intermediate states, and we now therefore model a system with N states. In the most general case, we can express the probability density of transition to the final state N of N states from inital state 1 in the form

$$P_{1N}(t) = \sum_{\substack{\text{all possible paths}\\ \text{from 1 to } N}} \int_0^{t_1}\int_0^{t_2}\cdots\int_0^{t_{n-2}} P_{12}(t_1)P_{23}(t_2 - t_1)\cdots P_{n-1n}(t - t_{n-2})\prod_{i=1}^{n-2} dt_i\,. \qquad (4)$$

In this case, the $P_{ij}(t)$'s represent probabilities from the ith and jth states along the selected path, each of which has a number of total states traversed $n \geq N$. For arbitrary probability distributions of transition $P_{ij}(t)$ there is little to be gained in examining this general case with this formalism. However, for any specific case, the probability transition of M uncoupled parallel replicas with arbitrary transition probabilities can be found by applying Eq. (2) to the probability distribution generated in Eq. (4). The case of transition coupled simulations can be obtained by first applying Eq. (2) to each probability $P_{ij}(t)$ and then chaining these probabilities as in Eq. (4).

There is one specific case that bears closer examination. If we assume an exponential distribution for each of the individual transition probabilities $P_{ij}(t) = k_{ij}\exp(-k_{ij}t)$, then this can be interpreted as a system of differential equations in the occupation of each state with constant coefficients. The matrix form of the differential equation is $\dot{\mathbf{x}} = K\mathbf{x}$, with the matrix K of transition rate constants k_{ij}, and can be solved to yield $\mathbf{x} = \sum_{i=1}^N \mathbf{c}_i\mathbf{e}_i\exp(\lambda_i t)$, where the $\mathbf{c}_i$ are constants dependent on boundary conditions, $\mathbf{e}_i$ and λ_i are the eigenvectors and eigenvalues of K, respectively, and $\mathbf{x}$ is the occupancy vector of the N states. In order to restrict ourselves to physical processes (population of all states non-negative; total population in all states equal to 1), we must add the requirements that the matrix K is Markovian, meaning that $k_{ij} \geq 0$ for $i \neq j$, and that $\sum_i^N k_{ij} = 0$. These conditions imply that there will be exactly one zero eigenvalue, and all other eigenvalues will have negative real parts, and be wholly real or existing as complex conjugates, and that $\sum_i c_i = 0$. If we stop the simulation in the final state (i.e., there is no back transition from the final state), then we must have $k_{Nj} = 0$ for all j. This also implies that $x_N(t) = F_{1N}(t)$, simplifying the math.

We are interested in the time-dependent probability of entry $P(t)$ into the state x_N given by $\frac{dF_{1N}(t)}{dt} = \frac{dx_N}{dt}$. The average time of transition to the Nth state $\langle t_{1N}\rangle$ is then

$$\int_0^\infty tP(t)\,dt = \int_0^\infty \sum_{i=1}^{N} c_i e_{iN} \lambda_i t \exp(\lambda_i t)\,dt = \sum_i \frac{c_i e_{iN}}{\lambda_i}, \tag{5}$$

where e_{iN} is the Nth component of the eigenvector $\mathbf{e}_i$.

Let us suppose that all the simulations are independent. For an arbitrary landscape in the exponential transition case, the probability of arriving in the final state will be of the form

$$P(t) = \frac{dx_N}{dt} = \sum_{i=1}^{N} c_i e_{iN} \lambda_i \exp(\lambda_i t), \tag{6}$$

where again all $\lambda < 0$, and the c_i are such that the probability is normalized. For notational simplicity, call $a_i = c_i e_{iN}$. Computing $F_{1M}(t)$ for this distribution, and applying Eq. (2) yields

$$P_M(t) = M\left[\sum_{i=1}^{N} a_i \lambda_i \exp(\lambda_i t)\right]\left[\sum_{i=1}^{N} a_i \exp(\lambda_i t)\right]^{M-1}. \tag{7}$$

We can see that we cannot make the simple replacement of $\lambda_i \rightarrow M\lambda_i$ to obtain a result similar to the $M = 1$ case with only an increased rate constant, as there are many cross terms, so in general, we will not have a simple speedup of M times with M parallel simulations if these simulations are not coupled.

For an arbitrary set of transition probabilities for uncoupled simulations, we can find a formula for the average time of first passage by using Eq. (2) and integrating by parts:

$$\langle t_M\rangle = \int_0^\infty tP_M(t)\,dt = -\int_0^\infty [1 - F_1(t)]^M\,dt. \tag{8}$$

For example, returning to the case of multiple exponentials, we obtain (noting that exactly one $\lambda_i = 0$)

$$\langle t_M\rangle = -\int_0^\infty \left(\sum_{i=1}^{N-1} a_i \exp(\lambda_i t)\right)^M dt. \tag{9}$$

For the general case, it is impossible to obtain an explicit formula for the median time $t_{\frac{1}{2},M}$ for general $P(t)$, although a simple implicit formula $F_1(t_{\frac{1}{2},M}) = 1 - 2^{-1/M}$ exists that can be solved numerically in any specific case.

Again in the case of exponential distributions, let us suppose transition coupling with the M simulations. In this context the transition coupling scheme consists of moving the $M - 1$ simulations that remain in state i after the first transition from i to j to the state j also. Since each transition from state i to j has an overall rate constant of M times the single simulation rate constant, then all entries in the rate matrix will simply be multiplied by M, as each corresponds to an individual transition. This results in the multiplication of the eigenvalues by M as well, which implies that $\langle t_M\rangle = \sum_i \frac{c_i e_{iN}}{M\lambda_i} = \frac{1}{M}\langle t_1\rangle$. The average time will still scale with $1/M$, as will the median time, since all time scales have changed by the same factor M. Using transition coupling, we can still therefore obtain an M times speedup using M simulations, and interpret the rate constant k as $M^{-1}k_{\text{eff}}$, the effective rate of our ensemble simulation.

Calculations with three states.—Although the above expressions are usually not transparent analytically, it allows us to find numerical solutions for a large number of systems. Here we present two that illustrate interesting scaling properties.

In the simplest case, we have an irreversible transition from state 1 to state 2 and an irreversible transition from state 2 to state 3. We can write this as a one-parameter system in r, where r is the ratio of the smaller rate to the larger rate. We see (Fig. 1) that we always scale less than M if the simulations are uncoupled. This lack of linearity can be significant at large numbers of simulations, even with low ratios of slow to fast rates. Using transition coupling, however, we get exactly linear speedup.

Processes with superlinear speedup.—It is also possible for uncoupled simulations to be faster than transition coupling. Imagine a system with three states. From an initial state, a simulation can go either to a trap or to the final state. Transitions between the trap and the final state are not allowed, and thus the trap is "off pathway." Traps are very common in complex systems, such as protein folding [4]. We can see dramatic increases in the speedup of $\langle t_M\rangle$ if we ignore the fast transitions between the trap and the initial state, particularly when the trap is deep (see Fig. 2).

The explanation for this superlinearity is that a typical path will spend a considerable amount of time flipping back

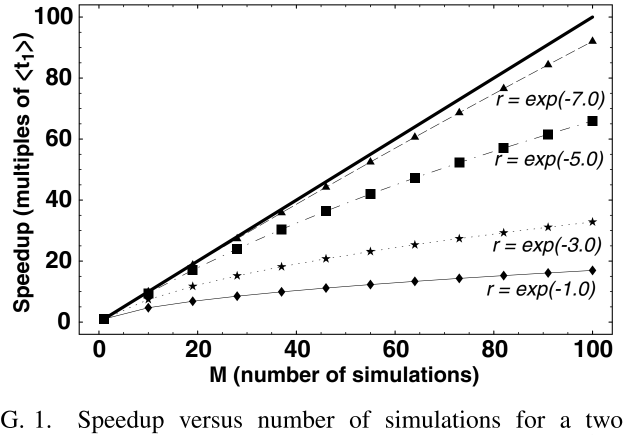

FIG. 1. Speedup versus number of simulations for a two-barrier system. Shown are plots for a range of $r = k_{12}/k_{23}$, where k_{12} is the rate constant of the slow barrier crossing, and k_{23} is the rate constant of the fast barrier crossing. If a system has two sequential transitions that must be crossed to arrive at the final state, then using uncoupled parallel simulations can lead to substantial deviation from full linear scaling (bold line).

and forth from the trap to the initial state. Using multiple simulations with transition coupling would increase the rate at which the system would flip back and forth, but the same percentage of time would still be spent in escaping from the trap. However, if the simulations are uncoupled, then not all simulations will be moved into the trap when a single simulation moves into the trap. These other simulations will thus have an increased chance of finding the "productive" exit, eliminating disproportionately long pathways to the final state. As mentioned, average times are therefore reduced dramatically (Fig. 2), but even the median times scale at a rate better than M, sometimes by as much as an order of magnitude. Note we do not have to be aware of what this off-pathway trap state looks like to benefit from the speedup, only that it exists.

The relationship between individual simulation time and overall time.—If we preserve the overall distribution of transition rates from one state to the others, as occurs with transition coupling with exponential probability distributions, then the distribution of the sum of all parallel simulation times will be equal to the distribution of single simulation times. If the individual transition probability is nonexponential, then the total parallelized simulation time will be different from single simulation time because the actual simulated process is now subtly different. If the system scales superlinearly, it is because the ensemble dynamics run is equivalent to an atypically fast trajectory. We are able to gain greater than 100% efficiency in superlinear scaling systems by avoiding slower transition paths. If the system scales sublinearly, then we are predominantly simulating paths that occur more slowly than a typical event. For example, if we have a small minimum which is not detected, each simulation will spend time in the associated minimum, whereas in the exactly linear case, the sum of time spent in the minimum by all simulations would be equivalent to the single simulation time. Thus, rates are exactly conserved with exponential transition processes and transition coupling, but are distorted in other cases.

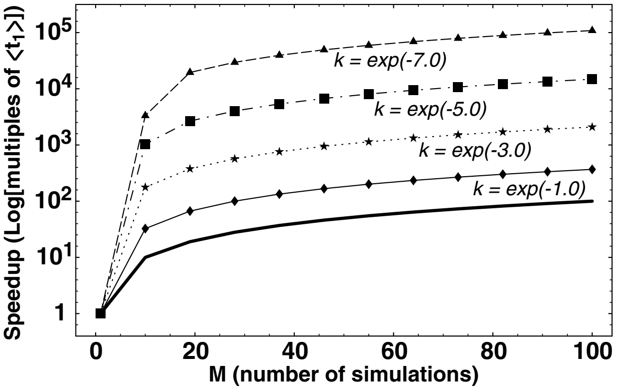

FIG. 2. Speedup in average time of first passage versus number of simulations for a three-state system with a trap. Plots are shown for different values of the escape rate k_{trap}, and therefore depth of the trap, with rate of transition from initial state to trap and to final state both equal to 1.0. An off-pathway trap with 1/20th the escape rate would lead to speedup in average time of arrival at the final state of about 2000 with only 100 simulations using uncoupled simulations. "Deeper" traps would lead to even greater than linear scaling for average time of first passage. Transition coupling gives a linear speedup (bold line).

Application to physical systems.—The above analysis is for instantaneous transition between clearly delineated states. How is this abstract transition model similar to physical systems of interest? We may also ask are the simple results for exponential distributions of transition times applicable for any physical systems?

In the case of deep, rugged free energy landscapes, these conditions are approximately met, and implementations of transition coupling give speedup which is linear within the margin of error [3]. Transitions correspond to crossings of barriers between local minima in the free energy. If barrier crossing times are negligible with respect to waiting times for barrier crossing, as we expect in the case of long time scale motions, we will have a near approximation to instantaneous transitions. If additionally, the decorrelation time within the minima is much less than the waiting time, we will have a Poisson process, thus implying exponential distribution of escape times from each individual minimum. Note that in order to capture the majority of the speedup, only the minima involved in rate determining steps must meet this criteria.

Unfortunately, it is not always feasible to successfully identify all transitions from free energy minima. The most promising approach currently appears to be using the energy variance of a simulation to identify changes in local free energy minima, for the same reason that large peaks in heat capacity indicate changes in state in a first-order transition [5], but it is not clear under which conditions this is valid.

There are physical situations which result in the breakdown of linearity in transition coupling. First, time is required for equilibration within a given free energy minimum. Second, there is a minimum barrier crossing time required for a simulation to cross each barrier. Simulations cannot be sped up to less than the sum of all barrier crossing times along the fastest path, so in all simulations, we must eventually reach a limit beyond which scaling fails. For example, a process that naturally takes 100 ns in nature will not necessarily be sped up to 100 fs by using one million simulations, as the minimum time necessary could easily be greater than 100 fs. This implies that finite barrier crossing times will tend to lead to sublinear scaling, so all ensemble dynamics simulations will eventually be limited by the barrier crossing time. However, there are strong reasons to believe that this fast time is significantly larger than the typical time in many condensed phase problems, perhaps on the order of 100 to 1000 times or more [6].

The characterizations described in this paper indicate that this method should be highly effective for hundreds or even thousands of computers. Consider again the simulation of protein folding dynamics. While the fastest proteins fold in 10 μs, a single CPU can simulate only 1 ns/day, thus requiring about 30 CPU years. With a 1000 processor cluster, and suitable parallel coupling, one can simulate 1 μs/day, rendering the problem tractable.

M. R. S. gratefully acknowledges support from the Fannie and John Hertz Foundation. The authors also acknowledge helpful comments from Scott Sheffield and Madhav Ranganathan.

[1] A. F. Voter, Phys. Rev. B **57**, 13 985 (1998).

[2] M. Shirts and V. S. Pande, Science **290**, 1903 (2000).

[3] I. Baker, J. Chapman, M. R. Shirts, S. Elmer, B. Nakatani, and V. S. Pande, J. Phys. Chem. (to be published).

[4] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, Proteins: Struct., Funct., Genet. **21**, 167 (1995).

[5] V. S. Pande and D. S. Rokhsar, Proc. Natl. Acad. Sci. U.S.A. **96**, 1273 (1999).

[6] D. Chandler, J. Chem. Phys. **68**, 2959 (1978).