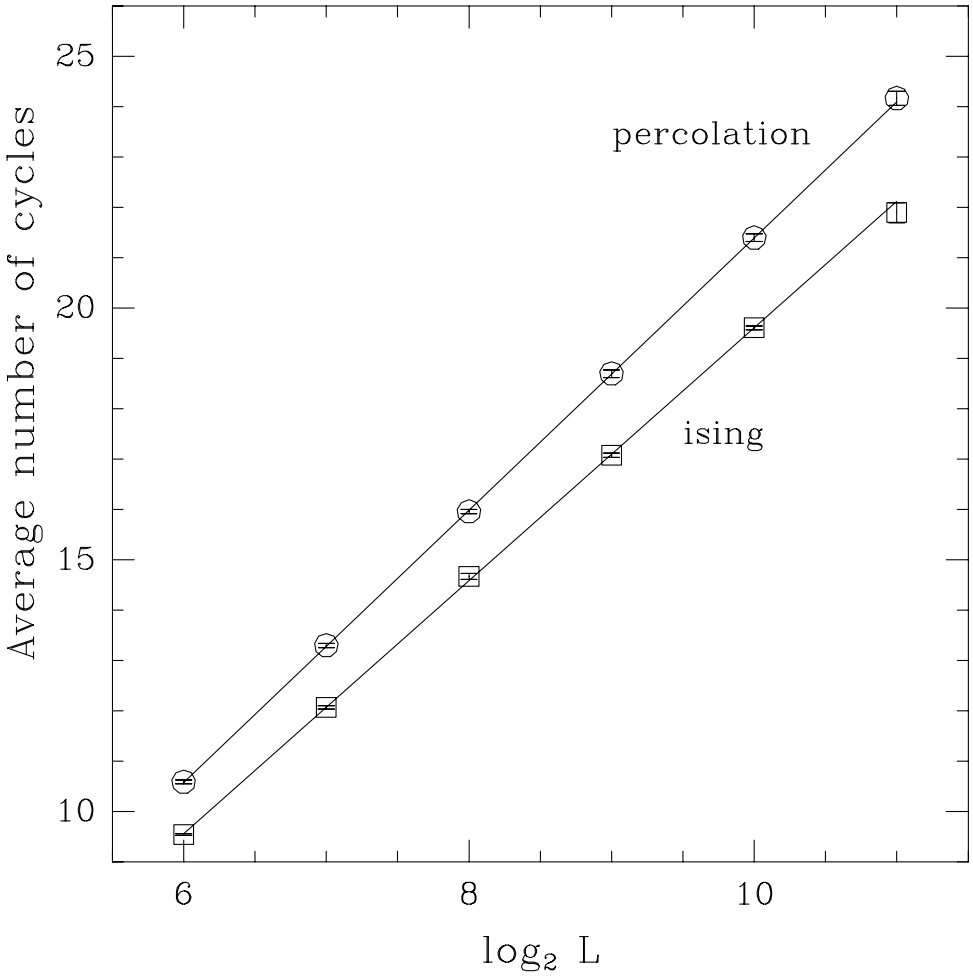

Figure 1. Number of multi-grid cycles needed to converge as a function of $\log_2 L$ for the case of bond percolation (circles) and the Ising model both at the critical point in two dimensions (squares). The lines are fits to the power law aL^z. The number of configurations ranges from 100 to 1000 for the smaller values of L.

[13] N. Avico *et al.*, *A 100 Gigaflops Parallel Computer,* Roma *La Sapienza* preprint 733 (Roma, Italy, April 1990).

[14] S. Aoki *et al.*, *Proposal for a Lattice Gauge Theory Teraflops Computer,* 1991 (unpublished).

[15] M. Fischler *et al.*, Nucl. Phys. **B** (Proc. Suppl.) **17**, 263 (

[16] Y. Iwasaki *et al.*, Nucl. Phys. **B** (Proc. Suppl.) **17**, 259 (

[17] A. D. Sokal, *Multi-Grid Monte Carlo for Lattice Field Theories,* lectures given at the Winter College on "Multilevel Techniques Physics", ICTP, Trieste, January 1991.

[18] X.-J. Li and A. D. Sokal, to be published.

[19] R. C. Brower, P. Tamayo and B. York, J. Stat. Phys. **63**,

[20] J. Apostolakis, P. Coddington and E. Marinari, in pre

[21] D. Hillis, *The Connection Machine*, (MIT Press, Cambridge, M 1985).

[22] U. Wolff, Nucl. Phys. **B** (Proc. Suppl.) **17**, 93 (1990).

[23] J.-S. Wang and R. H. Swendsen, Physica A **167**, 565 (1990

[24] R. Rusack, private communication.

[4] C. M. Fortuin and P. W. Kasteleyn, Physica **57**, 536 (197

[5] See for example B. M. McCoy and T. T. Wu, *The Two-Dimensiona Ising Model*, (Harvard University Press, Cambridge, Mass.

[6] R. G. Edwards and A. D. Sokal, Phys. Rev. D **38**, 2009 (19

[7] X.-J. Li and A. D. Sokal, Phys. Rev. Lett. **63**, 827 (198

[8] E. M. Reingold, J. Nievergelt and N. Deo, *Combinatorial Al Theory and Practice* (Prentice-Hall, Englewood Cliffs, N.J E. Horowitz and S. Sahni, *Fundamentals of Computer Algorithms*, (Computer Science Press, Rockville, Maryland, 1978)

[9] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, (Acad Press, New York, 1982).

[10] C. F. Baillie and P. D. Coddington, *Cluster Identification Algori Spin Models - Sequential and Parallel*, to be published in Concurren Practice and Experience.

[11] R. G. Edwards, X.-J. Li and A. D. Sokal, *Sequential and Ve Algorithms for Computing the Connected Components of an Undirected Graph*, in preparation.

[12] E. Marinari and C. Rovelli, in preparation.

algorithms which have been shown to reduce critical slo
different systems (for reviews of clus[21,22,23] gorithms, s
The algorithm could also be used in other applications
beling, such as image analysis [9]. As an example, the issue of compon
labeling is relevant to the next generation of high energ
since on-line reconstruction of traces will demand fas
This problem is being studied, for example, in relation
Superconducting Supercollider (SSC) detectors [24].

We believe that algorithms of this kind, together wit
with a simple but effective communication network, wil
in future studies of critical phenomena and of lattice
continuum limit.

Acknowledgements

This work was done using Connection Machines at the No
lel Architecture Center at Syracuse University, Sandia
Rice University, and the Advanced Computing Laborator
National Laboratory. Work supported in part by the Cent
Parallel Computation with NSF cooperative agreement N
and a grant from the IBM Corporation.

References

[1] A. D. Sokal, in *Computer Simulation Studies in Condensed Matter Physics: Recent Developments*, eds. D. P. Landau *et al.* (Springe
Verlag, Berlin-Heidelberg, 1988).

[2] A. D. Sokal, *How to Beat Critical Slowing Down - 1990 Update*, in Proc.
of Conference 'Lattice 90', Tallahassee, October 199
in Nucl. Phys. B (Proc. Suppl.).

[3] R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **58**, 86

ate long distance connections are set *on* since the sites
to belong to the same cluster, and they become fast long
nication channels (displaying the labels using color g
changes due to merging of large precursor sub-clusters
opened up).

We have implemented our code on the Connection Machin
typical massively parallel SIMD computer [2]. The mapping of the
hypercubic connections to the physical structure of th
some specific *power of* 2 communications, which are are exe
the speed of local communications. These are just the
local communications which we need for the efficient impl
algorithm.

In order to test the algorithm we have analyzed two ty
cally relevant cases: the bond percolation model (site
probability p), and the Ising model, both in two dimensi
behavior of the labeling algorithm is not relevant for t
we are really interested in is the *average* time to label p
configurations which occur in the simulation of these mo
greatest at the critical point, which for the square lattice is $p = \frac{1}{2}$ for bond
percolation model, and the Curie temperature for the Is
obtained our data by averaging over a large number of di
of the site connections, taken from configurations at th
two models, in order to get statistically significant re
obtain the scaling behavior of our algorithm.

In Fig. 1 we show the average number of iterations nee
as a function of L for large case of bond percolation (circle
Ising model (squares). The logarithmic slowing down is
not see any sign of power-law behavior (so $y = 0$), or of a
the logarithm. Since each iteration of the algorithm i
cycle of L steps (each step taking the same amount of time
computational complexity L^2. Hence each L step algorithm adds only
a $(\log L)^2$ term to the overall slowing down of a spin model clu

We have experimented with many variations and optimi
algorithm (see [20]). These can greatly increase the perform
algorithm, and make a substantial difference in a realis
large lattices.

Our algorithm is very general, and can be applied to a

performance degrades very fast with increasing volume, because clusters at a second order phase transition are cluster whose diameter is of order L, and for any local the minimum label has to diffuse across this large cluste

Our method is based on two main ideas. The first is to us approach in propagating cluster labels. Boolean conne 2^m, for $m = 1, \ldots, l-1$ (where the l) a, at ri c o u s ii l z t e i $L = t$ 2 h e x and the y direction (in the $d = 2$ case) by a logical AND of con $m-1$. For example, the distance 2 connection is between sit set *on* if sites x a n d s i t e s a n d $\hat{i} + 2\hat{x}$ are both connected. This is done in all directions of the lattice, for all levels algorithm works for a lattice of any dimensionality.

The second idea is that inter-site connections can b means that a connection (between two sites at a generic was originally *off* can be declared to be *on* if at any time du process the two sites are found to have the same label, i that they must belong to the same connected cluster. Cl between two sites at a distance M can in this way be *on* ev path between the two sites contains sites belonging to components: the existence of an *on* connection between t plies that a connected path joining the two sites exist improvement greatly reduces the number of iterations ne the final values of the labels.

Thus, during one multi-scale label updating cycle ea turn at each of its $2d$ neighbors at all levels m of the mul It will update its label when necessary and *update its connecti* the level $m-1$ connections, and also using connection i cycle of the algorithm sweeps all l connection levels, a solves the trivial case where all connections are *on*.

Clearly at the beginning of the dynamical procedure connections are all *off*, and due to the fractal structure will take many iterations before a significant number of tions are improved. It is thus very useful to tailor the levels as a function of the cycle number [20] useful at th beginning, while using longer distance connections is n end of the procedure. Here we just show results for the s the depth is constant. What eventually happens is that s

that the same lower bound for the dynamical critical exp to cluster schemes and to the mixed cluster]). The multi-grid main problem is including what we know about the large d of the system in the updating schemes.

In this note we propose a regular, completely synchr algorithm for cluster labeling. We show that it does not law computational slowing down (i.e. $y = 0$) in the two cas percolation and Ising models at their critical points. tive on a general SIMD machine with some very basic non-l In the following we will assume that the machine allows cation between sites which are at distance of 2^m in any direction of the physical lattice. These are the only non-local in order to build an algorithm which is not affected by p down. Note that these are the same connections that are n efficiently implement a Fast Fourier Transform algorithm would be provided, for example, by a machine with a hype We think that the results we present here strongly suppo kind of effective non-local communication for the next ge computers.

Our method has some similarities with the one propos Tamayo and York [19] that it is a SIMD, multi-grid style a however it is much simpler, and seems to have better scal will report in a separate paper [20] variations and improvement this algorithm, and the performance of these different m a more detailed comparison to the method of Brower *et al.*

The simplest local component labeling algorithm on a computer is based on label propagation [9, 10]. We start with a different label on each site, and with a list of first neighbor conn whether a given pair of sites is connected or not (the variables in the following: *off* means no connection is pr that there is a connection). Each site then looks to its right and below), and if it is connected to this neighb neighboring label, if it is less then its own label. Eve situation is reached which gives a correct labeling of label of a cluster being the minimum initial label of a cluster. This algorithm suffers from computational sl the cases of interest (spin and percolation models at t

which y is zero, the advantages of cluster update algori
local algorithms may be offset by the computational comp
the clusters.

This could be the case if our computer were parallel ra
tial, in which case the problem of component labeling i
cated] [. 10 Here the information regarding the connectivit
ical part of the lattice is only contained in a single p
information far away (on the lattice, and hence also on
take a long time. Our aim is to find a parallel labeling
computational slowing down (i.e. $y = 0$). Because of ou
which we will clarify in the following, we will conside
the parallel computer is a Single Instruction Multiple l

In order to handle critical slowing down, a parallel
some physical switching capabilities allowing fast and
tions between regions that are far away in the physical s
simulation. It does not seem possible, deep in the crit
pletely avoid all forms of non-local data transmission
patterns of switching networks require complex hardwa
cost and the potential failure rate of the system, we ar
algorithms which run on machines with simple communicat
have good asymptotic scaling behavior.

Our effort is thus also aimed towards a better comprehe
tures that will be required in the next generation of c
for lattice QCD. On machines with a speed of the order o
refs, [1, 4 3, 5 1] 6 for information on the different projects, s
are already at a very advanced stage), lattice QCD will
large correlation lengths, for both the gauge interact
It seems that the use of some non-local algorithms will b
essential: effective multi-grid methods will be probab
the quark propagators, and cluster algorithms could pl
in speeding up the pure gauge dynamics. Many of these ma
some kind of SIMD architecture, and the topology of the
will be very important. For a dedicated machine, keepi
simple as possible is a crucial goal, but on the other han
of supporting all the relevant algorithms.

The ideas which make cluster algorithms effective are
grid methods (see for example the Al]a, na Sno dk ta hel fe ac tu re not

Our understanding of critical phenomena, connected
of some typical length scale of a system, has benefited
scale numerical simulations. Large computer power and
probe the theories deep into the critical region, allo
understanding of renormalization group and scaling id
down is the typical drawback of a large scale simulatio
if the correlation length ξ of the system is becoming lar
will have a correlation time (the number of iterations
statistically independent, where the dynamic critical
exponent z, $z \geq 2$. [1]

The proposal of Swendsen and Wang [3] critical slowing do
exploits the Fortuin-Kasteleyn representation [4] of the Ising model [5]
The partition function can be rewritten as a sum over p
(or, according to the very illuminating suggestion of Fortuin and Kasteleyn, a
model). Using this equivalence we can implement non-lo
system by updating whole clusters of spins at a time. Alt
critical slowing down is not completely eliminated, the method is significantly
effective because the maximum cluster size diverges whe
length diverges, so we can make very large non-local cha
point.

The dominant computational aspect of this algorithmi
of the clusters. This is the classic problem of connect
ing, [6]: on a d-dimensional lattice, we have $V = L^d$ sites, with bonds between
neighboring sites which can be *on* (a connection) or *off* (no
our goal is to end up with the same label on all connected
labels for all disconnected clusters.

Let us define a computational exponent y for this prob
time to label the connected components of a cluster grows as L^{d+y} for large L,
Monte Carlo simulation, this cost is compounded by the n
needed to generate an independent configuration, which
critical point. On a serial machine, labeling algorithms
which are guaranteed to scale with V, or worst case $y = 0$ (see e.g.
refs. [7, 8, 9, 10, 11, 12] for a discussion of sequential labeling alg
means that the overall computational cost of a Monte Car
algorithm at the critical point is L^{d+z}. For trivial generalizations of local
algorithm, the overall cost will be L^{d+y} times the cost of
labeling. Hence if for some reason we cannot use a labe

SCCS-103
CRPC-TR91155

A MULTI-GRID CLUSTER LABELING SCHEME

John Apostolakis, Paul Coddington
and Enzo Marinari[a]

Physics Department,
Syracuse University,
Syracuse, N.Y. 13244, U.S.A.

June 6, 1991

Abstract

We introduce a simple multi-scale algorithm for connected component labeling on parallel computers, which we apply to the problem of labeling clusters in spin and percolation models. We show that it is only logarithmically slowed down in the critical limit of bond percolation and the Ising model. We also discuss, in light of the proposed Teraflop computers optimized for lattice gauge theories and other lattice problems, the minimum requirements for simple computer switchboard architectures for which one can efficiently implement multi-scale algorithms to fight critical slowing down.

[a]: Permanent Address: Dipartimento di Fisica, Universita *Vergata*, Via E. Carnevale, 00173 Roma, Italy.