LABORATORY FOR
COMPUTER SCIENCE

MASSACHUSETTS
INSTITUTE OF
TECHNOLOGY

MIT/LCS/TM-336

# LOWER BOUNDS FOR RECOGNIZING SMALL CLIQUES ON CRCW PRAM'S

Paul Beame

August 1987

545 TECHNOLOGY SQUARE, CAMBRIDGE, MASSACHUSETTS 02139

# Lower Bounds for Recognizing Small Cliques on CRCW PRAM's

Paul Beame*†

Laboratory for Computer Science
Massachusetts Institute of Technology
545 Technology Square
Cambridge, MA 02139

August 1987

* Address starting fall 1987: Computer Science Department, FR-35, University of Washington, Seattle, WA 98195

† Supported by NSF grant PYI-25800

**Abstract**

We show that any CRCW PRAM which recognizes $k$-cliques in $n$-node graphs in time $T$ requires at least $n^{\Omega(k/T^2)}$ processors independent of its memory size. As a corollary we obtain essentially the same trade-off for unbounded fan-in circuits. We also demonstrate a similar but weaker trade-off for the memory size of CRCW PRAM's solving this problem independent of the number of processors. These bounds also answer an open question posed in [Ly1], i.e. they show that constant-depth circuits for recognizing $k$-cliques in $n$-node graphs require size $n^{\Theta(k)}$.

## 1. Introduction

There has been much recent success in proving lower bounds for problems in models of computation which permit operations on an unbounded number of items at unit cost. The first success in this area, namely producing super-polynomial lower bounds for constant-depth circuits with unbounded $\vee$ and $\wedge$ to compute parity, by Furst, Saxe, and Sipser [FSS], were quickly followed by stronger lower bounds for such circuits, by Ajtai independently [Aj1] and by Babai [Ba]. Also, lower bounds which produce a constant depth hierarchy of polynomial-size unbounded fan-in circuits were shown by Sipser in [Si]. With the exponential lower bounds for such circuits given by Yao ([Ya]) and subsequently improved to essentially optimal bounds by Hastad in [Ha1] and [Ha2], it has become clear that techniques for dealing with these circuits are quite powerful.

The lower bounds for constant-depth unbounded fan-in circuits actually produce lower bound trade-offs between depth and circuit size. Beame and Hastad ([Be1], [Be2], and [BH]) have extended these lower bound trade-offs to the much more powerful priority concurrent-read concurrent-write parallel random access machine (CRCW PRAM). This CRCW PRAM model has been an important and popular model for the design of parallel algorithms.

In another direction, Razborov [Ra] and Smolensky [Sm] extending and simplifying Razborov's work have shown strong lower bounds for majority and other symmetric functions on circuits which have unbounded fan-in modulo $p$ gates in addition to unbounded $\wedge$ and $\vee$ gates. Their techniques are quite different from those used for the other results in that they use approximation by small degree polynomials as opposed to restrictions.

One property which is shared by all of the Boolean functions for which the above lower bounds apply is that any representation of them in conjunctive or disjunctive normal form (CNF or DNF) requires long clauses, i.e. clauses whose length is polynomially related to the input size. In fact, whereas threshold functions with $n^\epsilon$ thresholds are hard, any threshold function with a $\log^{O(1)} n$ threshold can computed in constant depth and polynomial size ([AB]). Since proving bounds on the length of clauses in CNF and DNF is in some way fundamental part of most of the above proofs, at first glance one might be concerned it would be impossible to extend the lower bounds to functions that can be respresented with short clauses.

Results by Ajtai [Aj2] and Lynch [Ly1] show that this is not so. In [Aj2], Ajtai proved a weak super-polynomial lower bound for deciding if two nodes in a graph are reachable by a path of length $\log n$. Lynch's bound is a much stronger one. He showed that

if $k \leq \log n$ then any unbounded fan-in $\wedge$, $\vee$ $\neg$ circuit which finds $k$-cliques in a directed $n$-node graph in depth $d$ requires size $n^{\Omega(\sqrt{k/d^3})}$. This function can be represented in DNF with clauses of size $k^2 \leq \log^2 n$.

The main result of this paper improves Lynch's size lower bound and extends it to CRCW PRAM's. That is, we show that any CRCW PRAM which finds $k$-cliques in $n$-node graphs in time $T$ requires at least $n^{\Omega(k/T^2)}$ processors independent of the memory size. A similar but weaker trade-off is shown for the memory size of CRCW PRAM's solving this problem independent of the number of processors. The first bound implies essentially the same trade-off for unbounded fan-in circuits and answers an open question posed in [Ly1]. That is, it shows that constant-depth circuits for finding $k$-cliques in $n$-node graphs require size $n^{\Theta(k)}$.

While Lynch uses techniques that have a similar flavor to those in [Aj] and [Ba], our techniques extend those in [Ha1], [Ha2], and [BH]. We prove our bounds for inputs which are undirected graphs but it immediately follows that they hold for inputs representing directed graphs.

## 2. Definitions and Preliminaries

We begin with the definitions of the priority form of idealized CRCW PRAM, of processor and memory cell partitions, and of degrees in the same manner as in [Be1], [Be2], and [BH]. The input to the problems we consider will be undirected graphs so we will follow the usual convention of defining our parameters in terms of the number of nodes, n, and let $m = \binom{n}{2}$ denote the number of input variables.

**Definition:** A *CRCW PRAM* is a shared memory machine with processors $P_1, \ldots, P_{p(n)}$ which communicate through memory cells $C_1, \ldots, C_{c(n)}$. The input is initially stored in the first $m$ cells of memory, $C_1, \ldots, C_m$. Initially all cells other than the input cells contain the value 0. The output of the machine is the value in the cell $C_1$ at time $T(n)$.

Before each step $t$, processor $P_i$ is in state $q_i^t$. At time step $t$, depending on $q_i^t$, processor $P_i$ reads some cell $C_j$ of shared memory, then, depending on the contents, $(C_j)$, and $q_i^t$, assumes a new state $q_i^{t+1}$ and depending on this state, writes a value $v = v(q_i^{t+1})$ into some cell.

When several processors are attempting to write into a single cell at the same time step the one that succeeds will be the lowest numbered processor.

**Definition:** Let $M$ be a CRCW PRAM. For any processor $P_i$ the *processor partition*, $P(M, i, t)$, of the input set at time step $t$ is defined so that two inputs are in the same equivalence class of $P(M, i, t)$ if and only if they lead to the same state of processor $P_i$ at the end of time step $t$.

For any cell $C_j$ the *cell partition*, $C(M, j, t)$, of the input set at time $t$ is defined so that two inputs are in the same equivalence class of $C(M, j, t)$ if and only if they lead to the same contents of cell $C_j$ at the end of time step $t$.

**Definition:** Let $f$ be a Boolean function defined on a set $I \subseteq \{0,1\}^m$ A Boolean formula $F$ *represents* $f$ on $I$ if the inputs $x \in I$ satisfy $F$ exactly when $f(x) = 1$. Let the *maximum clause length* of a DNF formula $F$ be the maximum number of literals in any clause of

$F$. The (*Boolean*) *degree* of $f$ on $I$, $\delta(f)$, is the smallest maximum clause length of all disjunctive normal form (DNF) formulas representing $f$ on $I$. We extend this definition to sets of functions $\mathcal{F}$ by letting $\delta(\mathcal{F}) = \max_{f \in \mathcal{F}} \delta(f)$.

**Definition:** Let $A$ be a partition of a set $I \subseteq \{0,1\}^m$. Define the *degree* of $A$, $\delta(A)$, to be $\delta(\mathcal{F}_A)$ on $I$ where $\mathcal{F}_A$ is the set of characteristic functions of the equivalence classes of $A$ in $I$.

In this paper we will need a measure related to the degree defined above. We can extend the notion of degree by changing clause 'length' to any other monotone property of clauses. For the class of inputs we will be interested in, namely undirected graphs, we will be interpreting the $m$ inputs as the edges of a graph on $n$ nodes in a canonical way. In this case, a useful monotone property of clauses will be the number of nodes which are endpoints of edges appearing in the clause. We will write this *node degree* as $\delta_\nu$. For technical reasons we also will need to define a modified node degree in which we ignore some specified set $V \subseteq \{1,\ldots,n\}$ of the nodes, i.e. the monotone property is the number of nodes other than those in $V$ which are endpoints of edges appearing in the clause. We write the resulting degree measure as $\delta_\nu^V$. We will use $[V]^2$ for the set of input variables which have both endpoints in $V$.

**Definition:** A *restriction* $\pi$ *on* $K \subseteq \{1,\ldots,m\}$ is a function $\pi : K \to \{0,1,*\}$ where:

$$\pi(i) = \begin{cases} 1 & \text{means } x_i \text{ is set to 1} \\ 0 & \text{means } x_i \text{ is set to 0} \\ * & \text{means } x_i \text{ is unset} \end{cases}$$

We define the results of applying a restriction $\pi$ to a partition, $A\lceil_\pi$, a function, $f\lceil_\pi$, and a Boolean formula, $F\lceil_\pi$, in the natural way. If $\sigma$ and $\tau$ are restrictions then $\sigma\tau$ is a restriction which is the result of applying $\sigma$ first and then applying $\tau$. For any $K \subseteq \{1,\ldots,m\}$ define $Proj\{K\}$ to be the set of restrictions which assign 0 or 1 exactly to the inputs in $K$.

In several places we will need the following simple observation which parallels that contained in Lemma 3.1 of [BH] and extends that lemma to the more complicated definition of $\delta_\nu^V$.

**Lemma 2.1:** *Let $A$ be a partition of a set $I \subseteq \{0,1\}^m$. For every $Y \subseteq \{1,\ldots,n\}$ there exists a restriction $\sigma \in Proj\{[V \cup Y]^2 \setminus [V]^2\}$ such that $\delta_\nu^V(A) \leq |Y| + \delta_\nu^{V \cup Y}(A\lceil_\sigma)$.*

**Proof:** For each $\sigma \in Proj\{[V \cup Y]^2 \setminus [V]^2\}$ let $\mathcal{F}_\sigma$ be a set of DNF formulas which represent the characteristic functions of the equivalence classes in $A\lceil_\sigma$ and which have maximum number of nodes other than $V \cup Y$ appearing in each clause being at most $\delta_\nu^{V \cup Y}(A\lceil_\sigma)$. To each clause in $\mathcal{F}\lceil_\sigma$ append the clause $C_\sigma$ which is true on exactly those inputs in $\{0,1\}^m$ which agree with $\sigma$ to obtain a set of formulas $\tilde{\mathcal{F}}_\sigma$. By construction, the number of nodes other than those in $V$ which appear in any clause is at most $|Y| + \delta_\nu^{V \cup Y}(A\lceil_\sigma)$. Each class in $A$ can now be represented by a DNF formula which is the disjunction of formulas in various $\tilde{\mathcal{F}}_\sigma$. By definition of $\delta_\nu^V$ we have

$$\delta_\nu^V(A) \leq \max_{\sigma \in Proj\{[V \cup Y]^2 \setminus [V]^2\}} \delta_\nu^{V \cup Y}(A\lceil_\sigma) + |Y|.$$

The lemma follows immediately. □

We now include the following definitions and lemmas which are shown in detail in [Be2] and [BH].

**Definition:** We say that an input $x \in \{0,1\}^m$ *satisfies* a Boolean function $F : \{0,1\}^m \rightarrow \{0,1\}$ if $F(x) = 1$. We say that $x$ *falsifies* $F$ if $F(x) = 0$.

**Definition:** A *graded set of Boolean functions* is a set $\mathcal{G}$ of Boolean functions such that each $F \in \mathcal{G}$ has an associated positive integer *grade*, $\gamma(F)$ (or has grade $= \infty$) and no two functions of a given grade are simultaneously satisfiable.

**Definition:** For any graded set of Boolean functions, $\mathcal{G}$, *the partition determined by* $\mathcal{G}$, $\langle \mathcal{G} \rangle$, on $\{0,1\}^m$ is the partition such that $x, y \in \{0,1\}^m$ are in the same equivalence class if and only if:

(a) $x$ and $y$ both satisfy some function $F \in \mathcal{G}$, and $x$ and $y$ both falsify all $F' \in \mathcal{G}$ with $\gamma(F') < \gamma(F)$.

or (b) $x$ and $y$ both falsify all functions $F \in \mathcal{G}$.

**Lemma 2.2:** *Let $\mathcal{G}$ be a graded set of Boolean functions. If $\pi$ is a restriction then $\langle \mathcal{G} \rangle \lceil_\pi$ is the same partition as $\langle \mathcal{G} \lceil_\pi \rangle$ on $\{0,1\}^m \lceil_\pi$.*

As in [Be2] and [BH], we note that the above definitions can be carried over easily for Boolean formulas which *represent* the Boolean functions in the obvious way. Observe that if $\mathcal{F}$ represents $\mathcal{G}$ on $\{0,1\}^m \lceil_\pi$ then $\langle \mathcal{F} \rangle \lceil_\pi = \langle \mathcal{G} \rangle \lceil_\pi$. Also, the notion of degree applies to graded sets of Boolean functions simply using the natural definition of degree for sets of functions. It is easy to see that a graded set of Boolean functions $\mathcal{G}$ can be represented on $\{0,1\}^m \lceil_\pi$ by a graded set of DNF formulas $\mathcal{F}$, each with maximum clause length bounded by $\delta(\mathcal{G} \lceil_\pi)$.

**Definition:** Let $M$ be a CRCW PRAM. Define $\mathcal{G}(M, j, t)$ to be the graded set of Boolean functions as follows:

(i) For each positive integer $i$, the functions of grade $i$ in $\mathcal{G}(M, j, t)$ are the characteristic functions of those equivalence classes in $P(M, i, t)$ on which $P_i$ writes into cell $C_j$ during time step $t$.

(ii) The functions of grade $\infty$ in $\mathcal{G}(M, j, t)$ are all the characteristic functions of the equivalence classes in $C(M, j, t-1)$.

**Lemma 2.3:** *Let $M$ be a CRCW PRAM. $\langle \mathcal{G}(M, j, t) \rangle$ is a refinement of $C(M, j, t)$ on $\{0,1\}^m$.*

We follow essentially the same program for showing lower bounds on CRCW PRAM computations as in [Be2] and [BH]. That is, we show that after certain restrictions (which set more inputs as time progresses) are applied to the inputs, the processor and cell partitions have only small degree relative to the degree required to solve the problems. In using restrictions to obtain our lower bounds we must maintain a balance between the amount of degree reduction that a restriction achieves and the related simplification of the function required to be computed.

## 3. Lower Bounds for Clique

**Definition:** Let $Clique_k^n$ be the function which takes as input an undirected $n$ node graph and is equal to 1 if and only if the graph contains a clique on $k$ nodes.

**Theorem 3.1:** *If $M$ is a CRCW PRAM which computes the $Clique_k^n$ function for $k \leq \log n$ in time $T = T(n)$ then for sufficiently large $k$*

*(a) the total hardware $h(n) = p(n) + c(n)$ must be at least $n^{k/(89T^2)}$*

*(b) the number of processors $p(n)$ must be at least $n^{k/(89T^2)}$ even if the number of memory cells is infinite, and*

*(c) the number of memory cells $c(n)$ must be at least $n^{k/(43T^3)}$ even if the number of processors is infinite.*

In order to prove the existence of restrictions that satisfy these properties we need an appropriate probability space from which to choose restrictions. The distribution we use is essentially that introduced by Lynch [Ly] to prove his bounds for the clique problem on unbounded fan-in circuits.

**Definition:** Let $L \subseteq \{1, \ldots, n\}$. Define $R_{p,q}^L$ to be a probability space of restrictions on $[L]^2$ where for a random $\rho$ chosen from $R_{p,q}^L$, a set $S \subseteq L$ is chosen at random such that independently for each $v \in L$, $\Pr[\, v \in S \,] = p$ and $\Pr[\, v \notin S \,] = 1 - p$ and further that
1. For each $[u, v] \in [S]^2$, $\rho([u, v]) = *$
2. Independently for each $[u, v] \notin [S]^2$, $\Pr[\, \rho([u, v]) = 1 \,] = q$ and $\Pr[\, \rho([u, v]) = 0 \,] = 1 - q$.
We say that $\rho(V) = *$ if and only if $V \subseteq S$.

The outline above is now carried out by proving two lemmas. The first tells us that many nodes remain unset and the second tells us that the node degrees of the partitions do not increase.

**Lemma 3.1** *Let $q = n^{-8/k}$, $p' \leq 1/2$, and $L_0 = \{1 \cdots n\}$. If $\pi$ is chosen at random from $R_{p',q}^{L_0}$ and $p'n \geq k$ for $k$ sufficiently large then*

$$\Pr[\, \delta_\nu(Clique_k^n \lceil_\pi) \leq k/2 \,] \leq 3/4.$$

**Proof:** The distribution of the size of the set $S$ in the definition of $R_{p',q}^{L_0}$ is a binomial distribution with expected value $p'n$. Observe that, as in [BH], this random variable achieves it mean with probability at least $1/3$ for $p'n$ sufficiently large. Therefore, with probability $\geq 1/3$, $\pi$ leaves a clique of unset input variables on $p'n \geq k$ nodes.

Consider also the probability that, on the edge variables that $\pi$ sets to 1, $\pi$ produces a clique of size $\geq k/2$. This probability is easily bounded by

$$\binom{n}{k/2} q^{\binom{k/2}{2}} < n^{k/2} q^{\binom{k/2}{2}} = (nq^{(k/2-1)/2})^{k/2} = (n \cdot n^{-\frac{8}{k}(k-2)/4})^{k/2} \leq n^{-k/4} \leq 1/12$$

for $k$ sufficiently large.

Thus with probability at least $1/4$ $\pi$ leaves all the variables on a set of at least $k$ nodes unset and does not turn on the edges of any clique on $k/2$ nodes. Then, in order to force $Clique_k^n\lceil_\pi$ to be 1, more than $k/2$ of the remaining nodes must have all edges between them set to 1 and it is easily possible to force a clique by setting all the remaining edges to 1 so in this case the node-degree must be more than $k/2$. Thus the node-degree of $Clique_k^n\lceil_\pi$ is at most $k/2$ with probability at most $3/4$. □

**Lemma 3.2:** *Let $M$ be a CRCW PRAM just prior to a read or write operation, all of whose processor and cell partitions have node-degree at most $r \geq 1$ with variables from $\{x_{[u,v]}\}_{u,v\in L}$. Let $A$ be either an existing processor or cell partition of $M$ or a new cell partition resulting from a concurrent write of $M$. Choose $\rho$ at random from $R_{p,q}^L$ where $p, q \leq 1/2$. For $s > 0$ we have*

$$\Pr[\,\delta_\nu(A\lceil_\rho) \geq s\,] < [3p(2/q)^{s+r/2}r]^s.$$

Using Lemmas 2.2 and 2.3 we can obtain Lemma 3.2 from the following lemma by letting $F \equiv 0$ and $V = \phi$. The statement of this lemma is more complicated than that of the two similar lemmas in [Be2] and [BH] because of the exact way that, for the restrictions chosen at random from $R_{p,q}^L$, the existence of an unset variable can increase the likelihood that other variables are unset.

**Lemma 3.3:** *Let $\mathcal{G}$ be a graded set of DNF formulas on inputs $\{x_{[u,v]}\}_{[u,v]\in[L]^2\setminus[V]^2}$ with maximum number of nodes referred to in any clause bounded by $r \geq 1$ where $V \subseteq L \subseteq \{1,\ldots,n\}$. Let $\rho$ be a random restriction chosen from $R_{p,q}^L$. Let $F$ be an arbitrary function on $\{0,1\}^m$. Then, if $\langle \mathcal{G}\lceil_\rho\rangle$ is the partition determined by $\mathcal{G}\lceil_\rho$, for any $s \geq 0$ such that $s + |V| \leq w$ we have*

$$\Pr[\,\delta_\nu^V(\langle \mathcal{G}\lceil_\rho\rangle) \geq s \mid F\lceil_\rho = 0 \wedge \rho(V) = *\,] \leq \beta^s$$

*where $\beta > 0$ satisfies*

$$[\beta^{-1}(2/\min\{q, 1-q\})^{w+r/2}p/(1-p) + 1]^r = 2.$$

**Proof:** We first note that we only need to consider finite graded sets of formulas (i.e. $|\mathcal{G}|$ is finite). This follows since there are only a finite number of different input strings and so only a finite number of ways in which some formula in $\mathcal{G}$ can be satisfied and all smaller ones falsified. Also, it is trivial to see that the lemma holds for $s = 0$ or $\beta \geq 1$ so we can assume that $s > 0$ and $\beta < 1$.

The rest of the proof proceeds by induction on the total number of clauses in the formulas in $\mathcal{G}$. The intuitive idea is that as we work along the clauses one by one: if $\rho$ falsifies a clause, then we are left with essentially the same problem as before; if $\rho$ does not then, given the fact that it does not, it is much more likely that $\rho$ satisfies the clause and ensures that the remaining partition has only one class than that $\rho$ leaves any input in the clause unset.

In this proof for readability we will write $\delta^V\langle\mathcal{G}\rangle$ instead of $\delta_\nu^V(\langle\mathcal{G}\rangle)$.

BASE CASE: There are no clauses in the formulas in $\mathcal{G}$. In this case the formulas are all identically 0 and so all inputs are equivalent with respect to $\mathcal{G}$. Thus the partition determined by $\mathcal{G}\lceil_\rho$ consists of a single class so $\delta^V\langle\mathcal{G}\lceil_\rho\rangle = 0$ and the lemma holds for $\mathcal{G}$.

INDUCTION STEP: Assume that the lemma holds for all graded sets of formulas $\mathcal{G}'$ with fewer clauses than the formulas of $\mathcal{G}$. Let $F_1$ be a formula in $\mathcal{G}$ which has lowest grade among those formulas in $\mathcal{G}$ which are not identically 0; let $C_1$ be a clause of $F_1$. We can analyse the probability by considering separately the cases in which $\rho$ does or does not force clause $C_1$ to be 0. The failure probability, the probability that $\delta^V\langle\mathcal{G}\lceil_\rho\rangle \geq s$, is an average of the failure probabilities in these two cases. Thus

$$\begin{aligned}&\Pr[\,\delta^V\langle\mathcal{G}\lceil_\rho\rangle \geq s \mid F\lceil_\rho = 0 \wedge \rho(V) = *\,] \leq \\ &\quad \max(\,\Pr[\,\delta^V\langle\mathcal{G}\lceil_\rho\rangle \geq s \mid F\lceil_\rho = 0 \wedge C_1\lceil_\rho = 0 \wedge \rho(V) = *\,],\\ &\qquad \Pr[\,\delta^V\langle\mathcal{G}\lceil_\rho\rangle \geq s \mid F\lceil_\rho = 0 \wedge C_1\lceil_\rho \neq 0 \wedge \rho(V) = *\,]\,).\end{aligned}$$

The first term in the maximum is $\Pr[\,\delta^V\langle\mathcal{G}\lceil_\rho\rangle \geq s \mid (F \vee C_1)\lceil_\rho = 0 \wedge \rho(V) = *\,]$. Let $\tilde{F}_1$ be $F_1$ with clause $C_1$ removed; thus $F_1 = C_1 \vee \tilde{F}_1$ and $\tilde{F}_1 \neq F_1$. Let $\tilde{\mathcal{G}}$ be the same as $\mathcal{G}$ with formula $F_1$ replaced by $\tilde{F}_1$. In this case $C_1\lceil_\rho = 0$ so $F_1\lceil_\rho = \tilde{F}_1\lceil_\rho$ and thus $\langle\mathcal{G}\lceil_\rho\rangle = \langle\tilde{\mathcal{G}}\lceil_\rho\rangle$. In other words, when $C_1\lceil_\rho = 0$, the lemma requires a bound on $\Pr[\,\delta^V\langle\tilde{\mathcal{G}}\lceil_\rho\rangle \geq s \mid (F \vee C_1)\lceil_\rho = 0 \wedge \rho(V) = *\,]$. Since $\tilde{\mathcal{G}}$ has one fewer clause than $\mathcal{G}$ does, the inductive hypothesis implies that this probability is at most $\beta^s$.

The estimation of the second term in the maximum is more difficult. Let $T \subseteq L$ be the set of nodes appearing in clause $C_1$ and let $E \subseteq [T]^2$ be the set of edge variables appearing in $C_1$. By hypothesis $|T| \leq r$. Let $\rho_E$ be the restriction of $\rho$ to the edge variables in $E$. The condition that $C_1\lceil_\rho \neq 0$ is equivalent to the condition that $C_1\lceil_{\rho_E} \neq 0$. Let $Y$ be the subset of the nodes in $T \setminus V$ which are endpoints of edge variables to which $\rho_E$ assigns $*$; we denote the event that $Y$ is this subset by $*^V(\rho_E) = Y$. Then

$$\begin{aligned}&\Pr[\,\delta^V\langle\mathcal{G}\lceil_\rho\rangle \geq s \mid F\lceil_\rho = 0 \wedge C_1\lceil_{\rho_E} \neq 0 \wedge \rho(V) = *\,]\\ &= \sum_{Y \subseteq T\setminus V} \Pr[\,\delta^V\langle\mathcal{G}\lceil_\rho\rangle \geq s \wedge *^V(\rho_E) = Y \mid F\lceil_\rho = 0 \wedge C_1\lceil_{\rho_E} \neq 0 \wedge \rho(V) = *\,]. \qquad (1)\end{aligned}$$

Consider the case in which $Y = \phi$. Then $\rho_E$ sets every variable in $E \setminus [V]^2$ and since $\mathcal{G}$ has no variables from $[V]^2$, the value of $C_1$ is forced by $\rho_E$. But since we already know that $C_1\lceil_{\rho_E} \neq 0$ we must have $C_1\lceil_{\rho_E} = 1$. In this case every input satisfies $F_1\lceil_\rho$ and since $F_1$ has lowest grade we know that all inputs are equivalent with respect to the $\langle\mathcal{G}\lceil_\rho\rangle$. It follows that $\delta^V\langle\mathcal{G}\lceil_\rho\rangle = 0$ so the term corresponding to $Y = \phi$ has probability 0. The sum in (1) then becomes

$$\begin{aligned}&\Pr[\,\delta^V\langle\mathcal{G}\lceil_\rho\rangle \geq s \mid F\lceil_\rho = 0 \wedge C_1\lceil_{\rho_E} \neq 0 \wedge \rho(V) = *\,]\\ &= \sum_{Y \subseteq T\setminus V, Y \neq \phi} \Pr[\,\delta^V\langle\mathcal{G}\lceil_\rho\rangle \geq s \wedge *^V(\rho_E) = Y \mid F\lceil_\rho = 0 \wedge C_1\lceil_{\rho_E} \neq 0 \wedge \rho(V) = *\,]\\ &= \sum_{Y \subseteq T\setminus V, Y \neq \phi} \Pr[\,\delta^V\langle\mathcal{G}\lceil_\rho\rangle \geq s \mid F\lceil_\rho = 0 \wedge\ C_1\lceil_{\rho_E} \neq 0 \wedge \rho(V) = * \wedge *^V(\rho_E) = Y\,] \\ &\qquad \times \Pr[\,*^V(\rho_E) = Y \mid F\lceil_\rho = 0 \wedge C_1\lceil_{\rho_E} \neq 0 \wedge \rho(V) = *\,] \qquad (2)\end{aligned}$$

by simple conditional probability.

We tackle the latter term in each of these products first. If we let $\rho_E(Y) = *$ denote the event that every variable in $E \cap [Y]^2$ is unset by $\rho_E$ then elementary probability yields

$$\begin{aligned} &\Pr[\, *^V(\rho_E) = Y \mid F\lceil_\rho = 0 \wedge C_1\lceil_{\rho_E} \neq 0 \wedge \rho(V) = * \,] \\ \leq\ &\Pr[\, \rho_E(Y) = * \mid F\lceil_\rho = 0 \wedge C_1\lceil_{\rho_E} \neq 0 \wedge \rho(V) = * \,]\,. \end{aligned}$$

CLAIM:

$$\Pr[\, \rho_E(Y) = * \mid F\lceil_\rho = 0 \wedge C_1\lceil_{\rho_E} \neq 0 \wedge \rho(V) = * \,] \leq \frac{[p/(1-p)]^{|Y|}}{\min\{q, 1-q\}^{|Y|\cdot|V| + \binom{|Y|}{2}}}.$$

To see this consider any $\rho$ which satisfies $\rho_E(Y) = * \wedge F\lceil_\rho = 0 \wedge C_1\lceil_{\rho_E} \neq 0 \wedge \rho(V) = *$. By definition, the unset edge variables in $\rho$ must be from $[Y \cup V]^2 \setminus [V]^2$. We define a new restriction $\bar{\rho}$ which is obtained from $\rho$ by setting the variables in $[Y \cup V]^2 \setminus [V]^2$ that are in $E$ to 0 or 1 in the unique way that does not immediately force clause $C_1$ to 0. $\bar{\rho}$ still forces $F$ to 0 and still satisfies $\bar{\rho}(V) = *$. Thus $\bar{\rho}$ satisfies the last three conditions in the probability in question but not the first.

In changing $\rho$ to $\bar{\rho}$, the set $S$ of starred nodes in the definition of $R^L_{p,q}$ has had the nodes in $Y$ removed from it making $\bar{\rho}$ more likely than $\rho$ by a probability factor of $[(1-p)/p]^{|Y|}$. However, in the other aspect of the change to $\bar{\rho}$, some variables in $[Y \cup V]^2 \setminus [V]^2$ have had their values forced to 0 or 1. For each variable, the probability that it is set as required is as least $\min\{q, 1-q\}$. There are at most $|Y| \cdot |V| + \binom{|Y|}{2}$ of them and their probabilities are independent so this requirement decreases the likelihood of $\bar{\rho}$ by a factor of $\min\{q, 1-q\}^{|Y|\cdot|V| + \binom{|Y|}{2}}$. Thus

$$\frac{\Pr[\, \rho \,]}{\Pr[\, \bar{\rho} \,]} \leq \frac{[p/(1-p)]^{|Y|}}{\min\{q, 1-q\}^{|Y|\cdot|V| + \binom{|Y|}{2}}}.$$

Finally, we see that the operation which takes $\rho$ to $\bar{\rho}$ is uniquely invertible given $Y$; namely, take all variables in $E$ which have both endpoints in $Y \cup V$ and make them unset. The conditional probability we wish to estimate is by definition the quotient of the probability that a restriction satisfies all four conditions divided by the probability that it satisfies the last three. Thus probability in question is at most the above bound on the probability ratio of $\rho$ and $\bar{\rho}$ and the claim follows.

Now we look at the first term in each product in (2). The condition that $C_1\lceil_{\rho_E} \neq 0 \wedge *^V(\rho_E) = Y \wedge \rho(V) = *$ exactly specifies $\rho_E = \rho\lceil_E$ since it means that every variable in $E \setminus [V \cup Y]^2$ is set to 0 or 1 in the way which does not force the value of $C_1$ to 0 and that every variable in $V \cup Y$ is set to $*$. We let $F'$ be $F \vee G$ where $G\lceil_\rho = 0$ if and only if $\rho$ sets the variables in $E \setminus [V \cup Y]^2$ in the unique way that does not force clause $C_1$ to 0. Thus

$$\begin{aligned} &\Pr[\, \delta^V\langle \mathcal{G}\lceil_\rho \rangle \geq s \mid F\lceil_\rho = 0 \wedge C_1\lceil_{\rho_E} \neq 0 \wedge *^V(\rho_E) = Y \wedge \rho(V) = * \,] \\ =\ &\Pr[\, \delta^V\langle \mathcal{G}\lceil_\rho \rangle \geq s \mid F'\lceil_\rho = 0 \wedge *^V(\rho_E) = Y \wedge \rho(V) = * \,]\,. \end{aligned}$$

Now, the condition $*^V(\rho_E) = Y$ means that the variables in $[Y]^2$ are unset by $\rho$ and that the variables in $E \setminus [Y \cup V]^2$ are all set by $\rho$. The latter part of this condition is implied by the condition $F'\lceil_\rho = 0$. Thus we do not change the events by rewriting the probability as

$$\Pr[\, \delta^V\langle \mathcal{G}\lceil_\rho\rangle \geq s \mid F'\lceil_\rho = 0 \wedge \rho(V \cup Y) = * \,] \,.$$

If $|Y| \leq s$ then, by Lemma 2.1,

$$\begin{aligned}
&\Pr[\, \delta^V\langle \mathcal{G}\lceil_\rho\rangle \geq s \mid F'\lceil_\rho = 0 \wedge \rho(V \cup Y) = * \,] \\
&\leq \Pr[\, \exists \sigma \in Proj\{[Y \cup V]^2 \setminus [V]^2\},\ \delta^{V\cup Y}\langle(\mathcal{G}\lceil_\sigma)\lceil_\rho\rangle \ \geq\ s - |Y| \mid F'\lceil_\rho = 0 \wedge \rho(V \cup Y) = * \,] \\
&\leq \sum_{\sigma \in Proj\{[Y\cup V]^2 \setminus [V]^2\}} \Pr[\, \delta^{V\cup Y}\langle(\mathcal{G}\lceil_\sigma)\lceil_\rho\rangle \geq s - |Y| \mid F'\lceil_\rho = 0 \wedge \rho(V \cup Y) = * \,] \,. \qquad (3)
\end{aligned}$$

Because of the fact that $\sigma$ sets all inputs in $[Y \cup V]^2 \setminus [V]^2$, $E$ does not contain any variables in $[V]^2$, and $F'\lceil_\rho = 0$ we know that $\sigma\rho$ sets all the inputs in $E$ and thus forces the value of $C_1$. If $C_1\lceil_{\sigma\rho} = 1$ then all inputs in $\langle(\mathcal{G}\lceil_\sigma)\lceil_\rho\rangle$ are equivalent and thus $\delta^{V\cup Y}\langle(\mathcal{G}\lceil_\sigma)\lceil_\rho\rangle = 0 \leq s - |Y|$. Otherwise $C_1\lceil_{\sigma\rho} = 0$ and then $\langle(\mathcal{G}\lceil_\sigma)\lceil_\rho\rangle = \langle(\tilde{\mathcal{G}}\lceil_\sigma)\lceil_\rho\rangle$ since $\tilde{F}_1\lceil_{\sigma\rho'} = F_1\lceil_{\sigma\rho'}$. Thus the sum in (3) is equivalent to

$$\sum_{\sigma \in Proj\{[V\cup Y]^2 \setminus [V]^2\}} \Pr[\, \delta^{V\cup Y}\langle(\tilde{\mathcal{G}}\lceil_\sigma)\lceil_\rho\rangle \geq s - |Y| \mid F'\lceil_\rho = 0 \wedge \rho(V \cup Y) = * \,] \,.$$

Because $\tilde{\mathcal{G}}\lceil_\sigma$ has strictly fewer clauses than $\mathcal{G}$, has no inputs in $[V \cup Y]^2$, and since $s - |Y| + |V \cup Y| = s + |V| \leq w$ we can apply the inductive hypothesis to bound the probabilities in each term in this sum by $\beta^{s-|Y|}$. For each $Y$ the number of terms in the above sum is at most $|Proj\{[V \cup Y]^2 \setminus [V]^2\}| = 2^{|Y|\cdot|V| + \binom{|Y|}{2}}$ so we obtain a total bound of $2^{|Y|\cdot|V| + \binom{|Y|}{2}}\beta^{s-|Y|}$.

If $|Y| > s$ then we simply make the pessimistic assumption of failure, i.e. that the degree of the resulting partition is too large. Since $\beta < 1$ and $s - |Y| < 0$ we certainly have $1 < \beta^{s-|Y|}$. Thus

$$\Pr[\, \delta^V\langle \mathcal{G}\lceil_\rho\rangle \geq s \mid F\lceil_\rho = 0 \wedge C_1\lceil_{\rho_E} \neq 0 \wedge *^V(\rho_E) = Y \,]$$

is at most $2^{|Y|\cdot|V| + \binom{|Y|}{2}}\beta^{s-|Y|}$.

Finally, substituting these bounds in (2) we obtain a total failure probability of at most

$$\begin{aligned}
&\sum_{Y \subseteq T\setminus V, Y \neq \phi} \frac{[p/(1-p)]^{|Y|}}{\min\{q, 1-q\}^{|Y|\cdot|V| + \binom{|Y|}{2}}}\, 2^{|Y|\cdot|V| + \binom{|Y|}{2}}\, \beta^{s-|Y|} \\
&\qquad \leq \sum_{Y \subseteq T, Y \neq \phi} \Big[\Big(\frac{2}{\min\{q, 1-q\}}\Big)^{|V| + |Y|/2} p/(1-p)\Big]^{|Y|} \beta^{s-|Y|} \\
&\qquad \leq \sum_{Y \subseteq T, Y \neq \phi} \Big[\Big(\frac{2}{\min\{q, 1-q\}}\Big)^{w + r/2} p/(1-p)\Big]^{|Y|} \beta^{s-|Y|}
\end{aligned}$$

$$
\begin{aligned}
&= \beta^s \sum_{i=1}^{|T|} \binom{|T|}{i} \left[\frac{(2/\min\{q,1-q\})^{w+r/2}p}{\beta(1-p)}\right]^i \\
&= \beta^s \left[(\beta^{-1}\ (2/\min\{q,1-q\})^{w+r/2}p/(1-p)+1)^{|T|}-1\right] \\
&\leq \beta^s \left[(\beta^{-1}\ (2/\min\{q,1-q\})^{w+r/2}p/(1-p)+1)^{r}-1\right] \\
&= \beta^s
\end{aligned}
$$

using the definition of $\beta$. Thus the lemma holds for $\mathcal{G}$ and by induction we have proved the lemma. □

The following composition lemma is essential in allowing us to use Lemmas 3.1 and 3.2 in tandem.

**Lemma 3.4** *Let $L \subseteq \{1,\ldots,n\}$ and $0 \leq p_1, p_2, q \leq 1$. Choose $\sigma$ at random from $R^L_{p_1,q}$ then choose $\tau$ at random from $R^{L_\sigma}_{p_2,q}$ where $L_\sigma$ is the set of nodes in $L$ that are unset by $\sigma$. The distribution of $\sigma\tau$ is exactly the same as the distribution of a $\pi$ chosen at random from $R^L_{p_1p_2,q}$.*

**Proof:** In the case of either distribution, once it is decided to set an edge variable, its probability of being set to 0 is $1-q$ and to 1 is $q$, independent of all other edge variables. Thus, we merely have to show that the distributions of the sets of nodes which are chosen to be unset by $\sigma\tau$ and by $\pi$ are identical. It is easy to see that for each node in $L$, for either distribution, the probability that it is chosen to be unset is $p_1p_2$ independent of all the other nodes. The lemma follows. □

**Proof of Theorem 3.1:** Let $L_0$ be the set of all nodes, $\{1,\ldots,n\}$. The basic method of the proof will be to choose random restrictions from $R^{L_0}_{p',q}$ for appropriate choices of $q$ and $p'$ so that after $t$ steps the node-degrees of the processor and cell partitions will be too small to have computed $Clique^n_k$. In order to do this we will in fact keep $q$ fixed and let $p'$ decrease as $p^t$ for appropriately chosen $p$. This will amount to revealing a random graph with edge probability $q$ step by step with the portion of the graph still unknown being all edges on a set of $p^tn$ nodes after time $t$.

**Part (a):** CLAIM: Let $a = \log_n h(n) = \log h(n)/\log n$, let $s = \sqrt{2ak/11}$, let $p = n^{-4\sqrt{11a/2k}}$, and let $q = n^{-8/k}$. For $t \geq 0$ and a random $\pi_t$ chosen from $R^{L_0}_{p^t,q}$ with probability at least $1 - t/n$

$$
\begin{aligned}
&\max_i \delta_\nu(P(M,i,t)\lceil_{\pi_t}) \leq s, \\
\text{and}\quad &\max_j \delta_\nu(C(M,j,t)\lceil_{\pi_t}) \leq s.
\end{aligned}
$$

First we see how this claim implies the desired result. Observe that if $a \geq k/2$ or $T \geq \log n$ then we are done. Otherwise, assume that $k \leq np^T$. Consider a random $\pi_T$ chosen from $R^{L_0}_{p^T,q}$. By Lemma 3.1, with probability at least $1/4$, $Clique^n_k\lceil_{\pi_T}$ has node-degree at least $k/2$. However, by the claim, with probability at least $1 - \frac{\log n}{n}$

$$\delta_\nu(C(M,1,T)\lceil_{\pi_T}) \leq s < k/3.$$

Because the two failure probabilities sum to strictly less than 1 we can choose $\pi_T$ to be a restriction satisfying both these properties, contradicting the fact that $M$ computes $Clique_k^n$ in $T$ steps. Therefore the assumption is false and $p^T \leq (k/n) \leq n^{-\sqrt{88/89}}$ for $n$ sufficiently large. Thus $T\sqrt{a/k} \geq 1/\sqrt{89}$ or $a \geq k/(89T^2)$ which is as required for part (a).

We now show the claim by induction on $t$:

BASE CASE: At time 0 the processor partitions all consist of a single class with resulting degree of 0 and for each cell $C_j$, $C(M,j,0)$ is a partition which depends on at most one input bit so $\delta_\nu(C(M,j,0)) \leq 1 < s$. Thus $\pi_0$ is good with probability 1 as required by the claim.

INDUCTION STEP: Let $t \geq 0$. Assume the claim holds for $t$. By Lemma 3.4, a random $\pi_{t+1}$ chosen from $R^{L_0}_{p^{t+1},q}$ has the same probability distribution as $\pi_t\rho$ where $\pi_t$ is chosen at random from $R^{L_0}_{p^t,q}$ and then $\rho$ is chosen at random from $R^{L_t}_{p,q}$ where $L_t$ is the subset of nodes which are starred by $\pi_t$. Now by the induction hypothesis, with probability at least $1 - t/n$, $\pi_t$ satisfies

$$\delta_\nu(P(M,j,t)\lceil_{\pi_t}) \leq s$$
$$\delta_\nu(C(M,j,t)\lceil_{\pi_t}) \leq s.$$

We now assume that $\pi_t$ satisfies this condition and we will show that $\pi_t\rho$ will keep the degrees of the processor and cell partitions small with probability at least $1 - 1/n$. This will imply that $\pi_{t+1}$ is good with probability at least $1-(t+1)/n$ as required by the claim for $t+1$.

During the $(t+1)$-st step of the machine, each processor first reads some cell based on its current state and based on the value read it changes to a new state. Thus, for each $i$, the cell $C_j$ which processor $P_i$ reads depends only on the equivalence class in $P(M,i,t)$ containing the input. Also, this equivalence class and the equivalence class in $C(M,j,t)$ containing the input determines the new state of the processor. Therefore each equivalence class in $P(M,i,t+1)$ is an intersection of an equivalence class in $P(M,i,t)$ and one in $C(M,j,t)$ for some $j$. Then

$$\begin{aligned}\delta_\nu(P(M,i,t+1)\lceil_{\pi_t}) &\leq \delta_\nu(P(M,i,t)\lceil_{\pi_t}) + \max_j \delta_\nu(C(M,j,t)\lceil_{\pi_t}) \\ &\leq s + s = 2s. \qquad (a.1)\end{aligned}$$

Now, after $\pi_t$ is applied, all partitions have variables only from $L_t$. Therefore by Lemma 3.2, if we choose a $\rho$ at random from $R^{L_t}_{p,q}$ we have

$$\begin{aligned}\Pr[\,\delta_\nu((C(M,j,t+1)\lceil_{\pi_t})\lceil_\rho) \geq s\,] &\leq (6ps(2/q)^{2s})^s \\ &< (pn^{8s/k}2^{3s})^s \\ &= p^s n^{8s^2/k+3s^2/\log n} \\ &\leq n^{-4s\sqrt{11a/2k}} n^{11s^2/k} = n^{-4a+2a} = n^{-2a} \leq 1/nh(n)\end{aligned}$$

since $k \leq \log n$ and $h(n) \geq n$. Therefore we see that for a $\rho$ chosen at random from $R^{L_t}_{p,q}$,

$$\Pr[\,\max_j \delta_\nu(C(M,j,t+1)\lceil_{\pi_t\rho}) \geq s\,] < c(n)/nh(n). \qquad (a.2)$$

For each processor $P_i$ we already know that $\delta_\nu(P(M,i,t+1)\lceil_{\pi_t}) \leq 2s$. Since $P(M,i,t+1)\lceil_{\pi_t}$ depends only on the inputs in $L_t$, by Lemma 3.2 we have

$$\Pr[\ \delta_\nu((P(M,i,t+1)\lceil_{\pi_t})\lceil_\rho) \geq s\ ] < 1/nh(n)$$

as was the case for the cell partitions. Taking the maximum over all processors,

$$\Pr[\ \max_i \delta_\nu(P(M,i,t+1)\lceil_{\pi_t\rho}) \geq s\ ] < p(n)/nh(n). \qquad (a.3)$$

Therefore, putting $(a.2)$ and $(a.3)$ together, we see that with probability at least $1-1/n$, $\max_i \delta_\nu(P(M,i,t+1)\lceil_{\pi_t\rho}) \leq s$ and $\max_j \delta_\nu(C(M,j,t+1)\lceil_{\pi_t\rho}) \leq s$. By induction the claim for part (a) is proved. □

**Part (b):** CLAIM: Let $a = \log_n p(n) = \log p(n)/\log n$, let $s = \sqrt{2ak/11}$, let $p = n^{-4\sqrt{11a/2k}}$, and let $q = n^{-8/k}$. For $t \geq 1$ and a random $\pi_t$ chosen from $R^{L_0}_{p^t,q}$ with probability at least $1 - t/n$

$$\max_i \delta_\nu(P(M,i,t)\lceil_{\pi_t}) \leq s,$$
$$\text{and} \quad \max_j \delta_\nu(C(M,j,t)\lceil_{\pi_t}) \leq s.$$

The statement of part (b) follows from the claim exactly as in part (a).

We now show the claim by induction on $t$:

BASE CASE: This is identical to the base case in part (a).

INDUCTION STEP: Let $t \geq 0$. Assume the claim holds for $t$. By Lemma 3.4, a random $\pi_{t+1}$ chosen from $R^{L_0}_{p^{t+1},q}$ has the same probability distribution as $\pi_t\rho$ where $\pi_t$ is chosen at random from $R^{L_0}_{p^t,q}$ and then $\rho$ is chosen at random from $R^{L_t}_{p,q}$ where $L_t$ is the subset of nodes which are starred by $\pi_t$. Now by the induction hypothesis with probability at least $1 - t/n$, $\pi_t$ satisfies

$$\delta_\nu(P(M,j,t)\lceil_{\pi_t}) \leq s$$
$$\delta_\nu(C(M,j,t)\lceil_{\pi_t}) \leq s.$$

We now assume that $\pi_t$ satisfies this condition and we will show that $\pi_t\rho$ will keep the degrees of the processor and cell partitions small with probability at least $1 - 1/n$. This will imply that $\pi_{t+1}$ is good with probability at least $1-(t+1)/n$ as required by the claim for $t+1$.

Since the actual number of cells has no effect on the degrees of the partitions resulting from reads and state transitions, as in part (a):

$$\delta_\nu(P(M,i,t+1)\lceil_{\pi_t}) \leq \delta_\nu(P(M,i,t)\lceil_{\pi_t}) + \max_j \delta_\nu(C(M,j,t)\lceil_{\pi_t})$$
$$\leq s + s = 2s. \qquad (b.1)$$

As in part (a), we will show that the probability that a $\rho$ chosen at random from $R^{L_t}_{p,q}$ fails to have the correct properties is strictly less than $1 - 1/n$. The added complication

is that we do not have an *a priori* bound on the number of memory cells for which $\rho$ has to keep $\delta_\nu(C(M,j,t+1)\lceil_{\pi_t\rho}) \leq s$. The reason why this does not hurt us is that, by the inductive hypothesis, any memory cell $C_j$ which is not written into on any input in $\{0,1\}^n\lceil_{\pi_t}$ already satisfies $\delta_\nu(C(M,j,t+1)\lceil_{\pi_t}) \leq s$.

For each memory cell $C_j$ which is written into by some processor on an input in $\{0,1\}^n\lceil_{\pi_t}$, using the same reasoning as in part (a), we have

$$\Pr[\,\delta_\nu(C(M,j,t+1)\lceil_{\pi_t\rho}) \geq s\,] \leq (6ps(2/q)^{2s})^s. \qquad (b.2)$$

Also as in part (a), for each processor $P_i$,

$$\Pr[\,\delta_\nu(P(M,i,t+1)\lceil_{\pi_t\rho}) \geq s\,] \leq (6ps(2/q)^{2s})^s. \qquad (b.3)$$

Equation $(b.1)$ implies that, for inputs in $\{0,1\}^n\lceil_{\pi_t}$, the classes in the new state partition of each processor have characteristic functions represented by DNF formulas with maximum clause length bounded by $2s$. Since a DNF clause of length $\leq 2s$ is satisfied by a fraction of at least $1/2^{2s}$ of the possible inputs, each class in the partition $P(M,i,t+1)\lceil_{\pi_t}$ consists of a fraction of at least $1/2^{2s}$ of the possible inputs. This means that, for inputs in $\{0,1\}^n\lceil_{\pi_t}$, each processor can only be in one of $2^{2s}$ states and therefore can write into at most $2^{2s}$ different cells. Therefore the total number of cells for which $\rho$ must work is at most $2^{2s}p(n)$.

The argument above means that $(b.2)$ must be applied in at most $2^{2s}p(n)$ places and $(b.3)$ must be applied in $p(n)$ places. Thus the total probability that either $\max_i \delta_\nu(P(M,i,t+1)\lceil_{\pi_t\rho}) \geq s$ or $\max_j \delta_\nu(C(M,j,t+1)\lceil_{\pi_t\rho}) \geq s$ is bounded by

$$\begin{aligned}
(2^{2s}+1)p(n)(6ps(2/q)^{2s})^s &\leq 5^s p(n)(6ps(2/q)^{2s})^s \\
&= p(n)(30ps(2/q)^{2s})^s = p(n)(pn^{8s/k}2^{3s})^s \\
&< p(n)p^s n^{8s^2/k+3s^2/\log n} \leq p(n)n^{-4s\sqrt{11a/2k}}n^{11s^2/k} \\
&= p(n)n^{-4a+2a} = p(n)n^{-2a} \leq 1/n
\end{aligned}$$

since $p(n) \geq n$. Thus the total failure probability is strictly less than $1/n$ and the claim follows for $t+1$. By induction the claim for part (b) is proved. □

**Part (c):** CLAIM: Let $a = \log_n c(n) = \log c(n)/\log n$, let $s = (a^2k)^{1/3}$, let $p = n^{-\frac{7}{2}(a/k)^{1/3}}$, and let $q = n^{-8/k}$. For $0 \leq t \leq \frac{3}{10}(k/a)^{1/3} - 2$ and a random $\pi_t$ chosen from $R^{L_0}_{p^t,q}$ with probability at least $1 - t/n$

$$\max_i \delta_\nu(P(M,i,t)\lceil_{\pi_t}) \leq st,$$
$$\text{and} \quad \max_j \delta_\nu(C(M,j,t)\lceil_{\pi_t}) \leq s.$$

First we see how this claim implies the desired result. Observe that if $a \geq k/6$ or $T \geq \frac{3}{10}(k/a)^{1/3} - 2$ then we are done. Otherwise, assume that $k \leq np^T$. Consider a

random $\pi_T$ chosen from $R^{L_0}_{p^T,q}$. By Lemma 3.1, with probability at least $1/4$, $Clique^n_k\lceil_{\pi_T}$ has node-degree at least $k/2$. However, by the claim, with probability at least $1 - \frac{\log n}{n}$

$$\delta_\nu(C(M,1,T)\lceil_{\pi_T}) \leq s < k/3.$$

Because the two failure probabilities sum to strictly less than 1 we can choose $\pi_T$ to be a restriction satisfying both these properties, contradicting the fact that $M$ computes $Clique^n_k$ in $T$ steps. Therefore the assumption is false and $p^T \leq (k/n) \leq n^{-(343/344)^{1/3}}$ for $n$ sufficiently large. Thus $\frac{7}{2}T(a/k)^{1/3} \geq (343/344)^{1/3}$ and so $343T^3a/8k \geq 343/344$ from which we obtain $a \geq k/(43T)$ which is equivalent to the statement for part (c).

We now show the claim by induction on $t$:

BASE CASE: The base case follows for similar reasons to parts (a) and (b) except that we must note that in fact the initial processor partitions have degree 0 which is now strictly necessary.

INDUCTION STEP: Let $0 \leq t \leq \frac{3}{10}(a/k)^{1/3} - 3$. Assume the claim holds for $t$. By Lemma 3.4, a random $\pi_{t+1}$ chosen from $R^{L_0}_{p^{t+1},q}$ has the same probability distribution as $\pi_t\rho$ where $\pi_t$ is chosen at random from $R^{L_0}_{p^t,q}$ and then $\rho$ is chosen at random from $R^{L_t}_{p,q}$ where $L_t$ is the subset of nodes which are starred by $\pi_t$. Now by the induction hypothesis with probability at least $1 - t/n$, $\pi_t$ satisfies

$$\delta_\nu(P(M,j,t)\lceil_{\pi_t}) \leq st$$
$$\delta_\nu(C(M,j,t)\lceil_{\pi_t}) \leq s.$$

We now assume that $\pi_t$ satisfies this condition and we will show that $\pi_t\rho$ will keep the degrees of the processor and cell partitions small with probability at least $1 - 1/n$. This will imply that $\pi_{t+1}$ is good with probability at least $1-(t+1)/n$ as required by the claim for $t+1$.

By the same reasoning as that leading to equation $(a.1)$ it is clear that the new processor partitions resulting from reads and state transitions satisfy:

$$\begin{aligned}\delta_\nu(P(M,i,t+1)\lceil_{\pi_t}) &\leq \delta_\nu(P(M,i,t)\lceil_{\pi_t}) + \max_j \delta_\nu(C(M,j,t)\lceil_{\pi_t}) \\ &\leq st + s = s(t+1). \qquad (c.1)\end{aligned}$$

Thus, even before $\rho$ is applied, the processor partitions satisfy the conditions required.

For each memory cell $C_j$, since the new processor partitions have degree at most $s(t+1)$ by $(c.1)$ and since the old cell partitions have degree at most $s$, using the same reasoning as in the previous two cases, we have

$$\begin{aligned}\Pr[\,\delta_\nu(C(M,j,t+1)\lceil_{\pi_t\rho}) \geq s\,] &< [3ps(t+1)(2/q)^{s+s(t+1)/2}]^s = [3ps(t+1)(2/q)^{s(t+3)/2}]^s \\ &< [p2^{s(t+3)}q^{-s(t+3)/2}]^s \leq [p2^{s(t+3)}n^{4s(t+3)/k}]^s \\ &\leq [pn^{4s(t+3)/k+s(t+3)/\log n}]^s \\ &\leq [pn^{5s(t+3)/k}]^s \qquad \text{since } k \leq \log n \\ &= n^{-\frac{7}{2}s(a/k)^{1/3}}n^{5s^2(t+3)/k} \qquad (c.2)\end{aligned}$$

Because $t+3 \leq \frac{3}{10}(k/a)^{1/3}$, $5s^2(t+3)/k \leq \frac{3}{2}a$ and since $s(a/k)^{1/3} = a$ the probability in (c.2) is at most $n^{-2a} = \frac{1}{c(n)^2} \leq \frac{1}{nc(n)}$ since $c(n) \geq n$. There are $c(n)$ cells, so the total probability that $\max_j \delta_\nu(C(M,j,t+1)\lceil_{\pi_t\rho}) \geq s$ is at most $1/n$. Thus the total failure probability is strictly less than $1/n$ and the claim follows for $t+1$. By induction the claim for part (c) is proved. □

**Corollary 3.1:** *Any CRCW PRAM $M$ which computes the $Clique_k^n$ function for $k \leq \log n$ in time $T=T(n)$ then*

*(a) if the the number of processors $p(n) = n^{O(1)}$ then $T(n) \geq \Omega(\sqrt{k})$ even if the number of memory cells is infinite,*

*and (b) if the number of memory cells $c(n) = n^{O(1)}$ then $T(n) \geq \Omega(k^{1/3})$ even if the number of processors is infinite.*

**Proof:** From Theorem 3.1 part (b), for $n$ sufficiently large we have $p(n) \geq n^{k/(89T^2)}$. Since $p(n) = n^{O(1)}$ there is a constant $c_1$ such that $c_1 \geq k/(89T^2)$ and so $T \geq \sqrt{c_1/89} \cdot \sqrt{k}$ as required for part (a).

From Theorem 3.1 part (c), for $n$ sufficiently large we have $c(n) \geq n^{k/(43T^3)}$. Since $c(n) = n^{O(1)}$ there is a constant $c_2$ such that $c_2 \geq k/(43T^3)$ and so $T > (c_2/43)^{1/3} \cdot k^{1/3}$ as required for part (b). □

There is an obvious constant time algorithm to compute $Clique_k^n$ using $n^{O(k)}$ processors and memory cells to check for each of the $n^{O(k)}$ cliques on $k$ nodes. The following corollary shows that this algorithm achieves an asymptotically optimal exponent.

**Corollary 3.2:** *Any CRCW PRAM which computes the $Clique_k^n$ for $k \leq \log n$ in constant time $O(1)$ requires both the number of processors and the number of memory cells to be $n^{\Omega(k)}$.*

**Proof:** Substitute $T = O(1)$ into Theorem 3.1 parts (b) and (c). □

Using a standard simulation of unbounded fan-in circuits by CRCW PRAM's we obtain lower bounds for unbounded fan-in circuits as well.

**Corollary 3.3:** *Any unbounded fan-in circuit of depth $d$ computing the $Clique_k^n$ function of $n$ inputs where $k \leq \log n$ and $n$ is sufficiently large requires size*

$$n^{k/(81d^2)}.$$

*In particular, unbounded fan-in circuits of constant depth require size $n^{\Omega(k)}$ to compute $Clique_k^n$.*

□

## 4. Lower Bounds for other Graph Problems

Amongst graph problems of the form 'is graph G a subgraph of the input graph?' which includes the $Clique_k^n$ function described in section 3, it appears that, under certain conditions, similar lower bounds will follow for many of them using Lemma 3.2. The conditions seem to be based on the concept of the probability threshold of a graph property as described in [ES] and [Bo]. If we consider a random $n$ node graph with fixed edge probability, the probability threshold of a graph property is the value of the probability $q'$ around which the property changes from being almost certainly not true of the random graph to being almost certainly true of the graph.

The lemma that one would need corresponding to Lemma 3.1 would seem to require that $q'$ be an upper bound on the value of the probability $q$ to be used in the restrictions from $R_{p,q}^L$. In order to be useful, Lemmas 3.2 and 3.3 depend on $q$ not being too close to 0. Many interesting graph properties have thresholds which are too small for these lemmas to say anything interesting, but Erdös and Renyi have shown a number of other problems which have probability thresholds in the range for which the methods of the previous section should work (see [ES] and [Bo] for more details).

By reductions from the parity problem, lower bounds can be proved for many of the subgraph properties that have probability thresholds which are too small for the techniques above. However, a problem for which neither technique works is the problem of the existence of a path of length $\log n$ in a graph. Ajtai [Aj2] has shown a non-polynomial lower bound for constant-depth circuits computing this problem but the bound is a very weak one - for polynomial-size circuits it can produce no better than an $\Omega(\log^* n)$ depth lower bound. It would be interesting to obtain a significantly better lower bound for this problem. Recently Lynch [Ly2] has claimed such a bound.


## Acknowledgements

Thanks to Al Borodin and Steve Cook who suggested the problem and to Johan Hastad and Alan Woods for helpful discussions about this work.


## References


[Aj1] Ajtai, M. $\Sigma_1^1$-*Formulae on Finite Structures*, Annals of Pure and Applied Logic, vol. 24, 1983, pp. 1-48.

[Aj2] Ajtai, M. *First-order Definability on Finite Structures* I.B.M. Research Report RJ 4705 (50173), 1985.

[AB] Ajtai, M., and Ben-Or, M. *A Theorem on Probabilistic Constant Depth Computations*, Proc. 16th. ACM-STOC, 1984, pp. 471-474.

[Ba] Babai, L. private communication, 1984.

[Be1] Beame, P.W. *Limits on the Power of Concurrent-Write Parallel Machines*, Proc. 18th ACM-STOC 1986, pp. 169-176.

[Be2] Beame, P.W. *Lower Bounds in Parallel Machine Computation*, Ph.D. Thesis, University of Toronto TR 198/87, 1986.

[BH] Beame, P.W., and Hastad, J. *Optimal Bounds for Decision Problems on the CRCW PRAM*, Proc. 19th. ACM-STOC 1987, pp. 83-93.

[Bo] Bollobás, B. *Random Graphs* Academic Press, 1985.

[ES] Erdös, P., and Spencer, J. *Probabilistic Methods in Combinatorics* Academic Press, 1974.

[FSS] Furst, M., Saxe, J.B., and Sipser, M. *Parity, Circuits, and the Polynomial Time Hierarchy*, Mathematical Systems Theory, vol. 17, no. 1, 1984, pp. 13-28.

[Ha1] Hastad, J. *Almost Optimal Lower Bounds for Small Depth Circuits*, Proc. 18th ACM-STOC 1986, pp. 6-20.

[Ha2] Hastad, J. *Computational Limitations for Small Depth Circuits*, Ph.D. Thesis, M.I.T., 1986, also published under the same title by MIT Press.

[Ly1] Lynch, J. *A Depth-Size Tradeoff for Boolean Circuits with Unbounded Fan-in* Structure in Complexity Theory Proceedings, Lecture Notes in Computer Science 223, Springer Verlag, 1986, pp. 234-248.

[Ly2] Lynch, J. private communication, August 1987.

[Ra] Razborov, A.A. *Lower Bounds for the Size of Circuits of Bounded Depth with basis AND, XOR* (in Russian) Mat. Zametki, 1986.

[Si] Sipser, M. *Borel Sets and Circuit Complexity*, Proc. 15th ACM-STOC 1983, pp. 61-69.

[Sm] Smolensky, R. *Algebraic Methods in the Theory of Lower Bounds for Boolean Circuit Complexity*, Proc. 19th. ACM-STOC, 1987, pp. 77-82.

[SV] Stockmeyer, L.J., and Vishkin, U. *Simulation of Parallel Random Access Machines by Circuits*, SIAM J. Computing, vol 13(2), 1984, pp. 404-422.

[Ya] Yao, A.C. *Separating the Polynomial-Time Hierarchy by Oracles: Part I*, Proc. 26th. IEEE-FOCS 1985, pp. 1-10.