View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Dagstuhl Research Online Publication Server

Very Large Cliques are Easy to Detect *
(Preliminary Version)

A. E. Andreev
LSI Logic Corporation, AE-187
1551 McCarthy Blvd.
CA 95305 Milpitas, USA
andreev@lsil.com

S. Jukna
Institute of Mathematics
Akademijos 4
LT-08663 Vilnius, Lithuania

May 25, 2006

Abstract

It is known that, for every constant $k \geq 3$, the presence of a k-clique (a complete subgraph on k vertices) in an n-vertex graph cannot be detected by a monotone boolean circuit using fewer than $\Omega((n/\log n)^k)$ gates. We show that, for every constant k, the presence of an $(n-k)$-clique in an n-vertex graph can be detected by a monotone circuit using only $O(n^2 \log n)$ gates. Moreover, if we allow unbounded fanin, then $O(\log n)$ gates are enough.

Keywords: Clique function, monotone circuits, perfect hashing, vertex cover, critical graphs

AMS subject classification: 05C35, 05C60, 68Q17, 68R10, 94C30

1 Introduction

We consider the well-known Clique function CLIQUE(n, s). This is a monotone boolean function on $\binom{n}{2}$ boolean variables representing the edges of an undirected graph G on n vertices, whose value is 1 iff G contains an s-clique. We are interested in proving good *upper bounds* on the size of monotone circuits with fanin-2 AND and OR gates computing CLIQUE(n, s).

The only non-trivial *upper* bound for CLIQUE(n, s) we are aware of is a non-monotone upper bound $O(n^{2.5\lceil s/3\rceil})$ given in [3, 10] (see also [1]). This bound is obtained by a reduction to boolean matrix multiplication. Until recently, no *monotone* circuits better than DNFs for this function were known.

A trivial depth-2 formula—its minimal DNF—has $\binom{n}{s} - 1$ fanin-2 OR gates. Can we reduce the number of gates by allowing larger depth? In particular, can this number be made polynomial in n for growing s?

*The research reported herein was initiated by the discussions during the Dagstuhl-Seminar "Complexity of Boolean Functions" (March 2006).

Dagstuhl Seminar Proceedings 06111
Complexity of Boolean Functions
http://drops.dagstuhl.de/opus/volltexte/2006/609

That it is impossible to save even one OR gate using so-called *multilinear* monotone circuits—where inputs to each AND gate are computed from disjoint sets of variables—was recently shown by Krieger in [8]: for any s, multilinear monotone circuits for $\mathrm{CLIQUE}(n, s)$ require $\binom{n}{s} - 1$ OR gates—just as many as the minimal DNF of this function! That substantial savings are impossible even in the class of *all* monotone circuits follows from well known lower bounds on the monotone circuit complexity of the clique function obtained by Razborov [12] and numerically improved by Alon and Boppana [1]: for every constant $s \geq 3$, the function $\mathrm{CLIQUE}(n, s)$ cannot be computed by a monotone circuit using fewer than $\Omega((n/\log n)^s)$ gates, and for growing s we need at least $2^{\Omega(\sqrt{s})}$ gates, as long as $s \leq (n/\log n)^{2/3}/4$.

By a padding argument, this implies that even $\mathrm{CLIQUE}(n, n-k)$ requires super-polynomial number of gates, as long as $k \leq n/2$ grows faster than $\log^3 n$. To see this, let m be the maximal number such that $m - s \leq k$ where $s = (m/\log m)^{2/3}/4$. Then $s = \Omega(k^{2/3})$ and $\mathrm{CLIQUE}(m, s)$ is a sub-function of (i.e. can be obtained by setting to 1 some variables in) $\mathrm{CLIQUE}(n, n-k)$: just consider only the n-vertex graphs containing a fixed clique on $n - m$ vertices connected to all the remaining vertices (the rest may be arbitrary). Thus, the function $\mathrm{CLIQUE}(m, s)$, and hence also the function $\mathrm{CLIQUE}(n, n-k)$, requires at least $2^{\Omega(\sqrt{s})} = 2^{\Omega\left(k^{1/3}\right)}$ gates, which is super-polynomial (in n) for $k = \omega(\log^3 n)$.

But what is the complexity of $\mathrm{CLIQUE}(n, n-k)$ when k is indeed small, say, constant—can then this function be computed by a monotone circuit using much fewer than $\binom{n}{k}$ OR gates?

For $k = 1$ this was recently answered affirmatively by Krieger in [8]: the function $\mathrm{CLIQUE}(n, n-1)$ can be computed by a monotone $\Pi\Sigma\Pi$-formula using only $O(\log n)$ OR gates. (Note that a DNF for this function has $n-1$ OR gates.) The argument of [8] uses the existence of particular error-correcting codes to encode $(n-1)$-cliques, and does not seem to work for $k > 1$.

In this paper we use another argument (based on perfect hashing) to obtain a more general result: a logarithmic number of OR gates is enough for *every* constant k, and a polynomial number of gates is enough also for growing k, as long as $k = O(\sqrt{\log n})$. Moreover, we can define the desired $\Pi\Sigma\Pi$-formulas *explicitely.*

2 Results

Theorem 2.1. *For every constant k, the function* $\mathrm{CLIQUE}(n, n-k)$ *can be computed by a monotone* $\Pi\Sigma\Pi$*-formula containing at most* $O(\log n)$ *OR gates.*

This theorem is a direct consequence of the following more general result which, for every constant k, allows us also to explicitely *construct* such a formula.

Recall that a vertex cover in a graph H is a set of its vertices containing at least one endpoint of each edge. The vertex cover number of H, denoted by $\tau(H)$, is the minimum size of such a set. A graph is *τ-critical* if removal of any its edges reduces the vertex cover number. For example, there are only two τ-critical non-isomorphic graphs H with $\tau(H)=2$, a triangle and a graph consisting of two disjoint edges. Erdős, Hajnal and Moon [4] prove that every τ-critical graph has at most $\binom{\tau(H)+1}{2}$ edges.

In what follows, let $\mathcal{G}(r,k)$ denote the set of all τ-critical graphs on $[r]=\{1,\ldots,r\}$ with $\tau(H)=k+1$.

Given a family F of functions $f:[n]\to[r]$, where $[n]=\{1,\ldots,n\}$, consider the following monotone $\Pi\Sigma\Pi$-formula

$$\Phi_F(X)=\bigwedge_{H\in\mathcal{G}(r,k)}\ \bigwedge_{f\in F}\ \bigvee_{\{a,b\}\in E(H)}\ \bigwedge_{e\in f^{-1}(a)\times f^{-1}(b)} x_e.$$

This formula rejects a given graph $G=([n],E)$ iff there exists a graph $H\in\mathcal{G}(r,k)$ and a function $f\in F$ such that for each edge $\{a,b\}$ of H there is at least one edge in the complement $\overline{G}$ of G between $f^{-1}(a)$ and $f^{-1}(b)$. The formula $\Phi_F(X)$ has at most

$$(|\mathcal{G}(r,k)|+|F|)\binom{k+2}{2}\le 2^{O(k^2\log(r/k))}+O(k^2|F|)$$

OR gates, which is linear in $|F|$ if both r and k are constants.

A family F of functions $f:[n]\to[r]$ is *s-perfect* ($n,r\ge s$) if for every subset $S\subseteq[n]$ of size $|S|=s$ there is an $f\in F$ such that $|f(S)|=|S|$. That is, for every s-element subset of $[n]$ at least one function in F is one-to-one when restricted to this subset. Such families are also known in the literature as (n,r,s)-perfect hash families.

Theorem 2.2. *Let F be a (n,r,s)-perfect hash family with $s=2(k+1)$. Then the formula $\Phi_F(X)$ computes* CLIQUE$(n,n-k)$.

Using a simple probabilistic argument, Mehlhorn [9] shows that (n,r,s)-perfect hash families F of size $|F|\le se^{s^2/r}\log n$ exist. Thus, taking $r=s$, we obtain that for every k, a desired monotone $\Pi\Sigma\Pi$-formula for CLIQUE$(n,n-k)$ with

$$\varphi(n,k)=2^{O(k^2)}+O(k^2e^{2k}\log n)$$

OR gates exists. This is $O(\log n)$ for every constant k, and polynomial for $k=O(\sqrt{\log n})$. Recall that already for $k=\omega(\log^3 n)$, any monotone circuit for CLIQUE$(n,n-k)$ requires a super-polynomial number of gates.

Remark 2.3. In the class of $\Pi\Sigma\Pi$-formulas, the upper bound $\varphi(n,k)$ cannot be substantially improved because, as shown by Radhakrishnan [11], every (not necessarily monotone) $\Pi\Sigma\Pi$-formula for the threshold function T^n_{n-k} with $k<(\log\log n)^2$ requires at least $2^{\Omega(\sqrt{k/\ln k})}\log n$ OR gates. This lower bound implies the same lower bound for CLIQUE$(n,n-k)$, because T^n_s is a monotone projection of CLIQUE(n,s): just assign all variables x_e of CLIQUE(n,s) with $i\in e$ the value of the i-th variable of T^n_s, that is, identify complete stars with their centers.

Using *explicit* (n, r, s)-perfect hash families we can obtain explicit formulas. For any constant s, (n, r, s)-perfect hash families F with $|F| = O(\log^s n)$ can be obtained by the following simple construction.

Let $M = \{m_{a,i}\}$ be an $n \times b$ matrix with $b = \lceil \log n \rceil$ whose rows are distinct 0-1 vectors of length b. Let $F = \{f_1, \ldots, f_b\}$ be the family of functions $f_i : [n] \to \{0,1\}$ determined by the columns of M; hence, $f_i(a) = m_{a,i}$. Let G be an arbitrary $(s+1)$-perfect family of functions $g : \{0,1\}^s \to [r]$. Bondy's theorem [2] says that the projections of any set of $s+1$ distict binary vectors on some set of s coordinates must all be distinct. Hence, for any set $a_1, \ldots, a_{s+1}$ of $s+1$ rows there exist s columns $f_{i_1}, \ldots, f_{i_s}$ such that all $s+1$ vectors $\vec{v}_j = (f_{i_1}(a_j), \ldots, f_{i_s}(a_j))$, $j = 1, \ldots, s+1$ are distinct. Since the family G is $(s+1)$-perfect, at least one function $g \in G$ will take different values on all these $s+1$ vectors. Hence, the function $h(x) = g(f_{i_1}(x), \ldots, f_{i_s}(x))$ takes different values on all $s+1$ points $a_1, \ldots, a_{s+1}$, as desired. Thus, taking the superposition of functions from G with s-tuples of functions from F, we obtain a family H of

$$|H| \leq \binom{|F|}{s} \cdot |G| = O(|G| \log^s n)$$

functions $h : [n] \to [r]$ which is $(s+1)$-perfect.

If s is constant then, for example, we can take $r = 2^s$ and let G consist of the single function $g(x) = \sum_{i=1}^{s} x_i 2^{i-1}$. Then $|H| = O(\log^s n)$. To make the range size r smaller one can use, for example, the fact, due to Fredman, Komlós and Szemerédi [5], that if p is a prime larger than m, then the functions $g_1, g_2, \ldots, g_{p-1}$ with $g_\alpha(x) = (\alpha x \bmod p) \bmod r$ form a family of perfect (m, r, s)-hash functions for every $r \geq s^2$. Using this fact, we can reduce the range size r till $r = s^2$ at the cost of increasing the size of the family G till $|G| = O(2^k)$.

Anyway, for constant k, Theorem 2.2 and our construction yields

Corollary 2.4. *For every constant k, there is an explicit monotone* $\Pi\Sigma\Pi$-*formula for* $\mathrm{CLIQUE}(n, n-k)$ *using only* $O(\log^{2k+2} n)$ *OR gates.*

Remark 2.5. For fixed values of r and s, infinite classes of (n, r, s)-perfect hash families F even with $|F| = O(\log n)$ were constructed by Wang and Xing in [13] using algebraic curves over finite fields. Using this (more involved) construction one can achieve the upper bounds stated in Theorem 2.1 by explicit monotone $\Pi\Sigma\Pi$-formulas.

Remark 2.6. Let k be an arbitrary constant. In the proof of Theorem 2.2 we construct a monotone $\Sigma\Pi\Sigma$-formula with $O(\log n)$ OR gates for the dual function of $\mathrm{CLIQUE}(n, n-k)$. (Recall that a dual of a boolean function $f(x_1, \ldots, x_n)$ is the function $f^*(x_1, \ldots, x_n) = \neg f(\neg x_1, \ldots, \neg x_n)$, where "$\neg$" denotes negation.) Moreover, this formula is *multilinear*, i.e. inputs to each its AND gate are computed from disjoint sets of variables. On the other hand, Krieger [8] shows that *every* monotone multilinear circuit for $\mathrm{CLIQUE}(n, n-k)$ requires at least $\binom{n}{k} - 1$ OR gates. This gives an example of a boolean function, whose dual requires much larger multilinear circuits than the function itself.

3 Proof of Theorem 2.2

Instead of the function CLIQUE$(n, n-k)$ it will be convenient to consider the dual function CLIQUE$^*(n, n-k)$. Note that this function accepts a given graph $G = ([n], E)$ iff G has no independent set with $n-k$ vertices, which is equivalent to $\tau(G) \geq k+1$. Hence, the graphs in $\mathcal{G}(n,k)$ are the smallest (with respect to the number of edges) graphs accepted by CLIQUE$^*(n, n-k)$. Recall that $\mathcal{G}(n,k)$ consists of all τ-critical graphs on $[r] = \{1, \ldots, r\}$ with $\tau(H) = k+1$. We will construct a monotone $\Sigma\Pi\Sigma$-formula for CLIQUE$^*(n, n-k)$. Replacing OR gates by AND gates (and vice versa) in this formula we obtain a monotone $\Pi\Sigma\Pi$-formula for CLIQUE$(n, n-k)$.

Important for our construction is that the number of non-isolated vertices in graphs $H \in \mathcal{G}(n,k)$ depends only on k, and not on n. This is a direct consequence of a result, due to Hajnal [6], that in a τ-critical graph without isolated vertices every independent set of size s has at least s neighbors. (For completeness, we include a short proof of this interesting result in the appendix.)

Claim 3.1. *Every graph in $\mathcal{G}(n,k)$ has at most $s = 2(k+1)$ non-isolated vertices.*

Proof. Let $G = (V, E)$ be a τ-critical graph without isolated vertices which cannot be covered by k vertices. Since G is minimal, it can be covered by some set S of $|S| = k+1$ vertices and by no smaller set. Hence the complement $T = V - S$ is an independent set. By Hajnal's theorem, the set T must have at least $|T|$ neighbors. Since all these neighbors must lie in S, the desired upper bound $|V| = |S| + |T| \leq 2|S| \leq 2(k+1)$ on the total number of vertices follows. $\square$

Let now F be an arbitrary s-perfect family of functions $f : [n] \to [r]$, and consider the following monotone $\Sigma\Pi\Sigma$-formula

$$\Phi^*(X) = \bigvee_{H \in \mathcal{G}(r,k)} \bigvee_{f \in F} K_{f,H}(X),$$

where

$$K_{f,H}(X) = \bigwedge_{\{a,b\} \in E(H)} \bigvee_{e \in f^{-1}(a) \times f^{-1}(b)} x_e.$$

To verify that this formula computes CLIQUE$^*(n, n-k)$, is enough to show that:

(i) $\tau(G) \geq k+1$ for every graph G accepted by $\Phi^*(X)$, and

(ii) $\Phi^*(X)$ accepts all graphs from $\mathcal{G}(n,k)$.

To show (i), suppose that $\Phi^*(X)$ accepts some graph G. Then this graph must be accepted by some sub-formula $K_{f,H}$ with $f \in F$ and $H \in \mathcal{G}(r,k)$. That is, for every edge $\{a,b\}$ in H there must be an edge in G joining some vertex $i \in f^{-1}(a)$ with some vertex $j \in f^{-1}(b)$. Hence, if a set S covers the edge $\{i,j\}$, then the set $f(S)$ must cover the edge $\{a,b\}$. Thus, if S is a minimal vertex cover in G, then $f(S)$ is a vertex cover in H, implying that $\tau(G) = |S| \geq |f(S)| \geq \tau(H) = k+1$.

To show (ii), take an arbitrary graph $G = ([n], E)$ in $\mathcal{G}(n,k)$. By Claim 3.1, G has at most s non-isolated vertices. By the definition of F, some function $f : [n] \to [r]$ must be one-to-one on these vertices. Consider the graph $H = ([r], E')$ with $\{a,b\} \in E'$ iff $\{f^{-1}(a), f^{-1}(b)\} \cap E \neq \emptyset$. Since $G \in \mathcal{G}(n,k)$ and f is one-to-one on all non-isolated vertices of G, the graph H belongs to $\mathcal{G}(r,k)$. Moreover, for every edge $\{a,b\}$ of H, the pair $e = \{i,j\}$ with $f(i) = a$ and $f(j) = b$ is an edge of G, implying that $x_e = 1$. This means that the sub-formula $K_{f,H}$ of $\Phi^*(X)$, and hence, the formula itself must accept G.

This completes the proof of Theorem 2.2. □

Acknowledgment

We are thankful to Matthias Krieger for interesting discussions.

References

[1] N. Alon, R. Boppana, The monotone circuit complexity of Boolean functions, Combinatorica, 7(1) (1987), 1–22.

[2] J. A. Bondy, Induced subsets, J. Combin. Theory (B), 12 (1972), 201–202.

[3] F. Chung, R. M. Karp, in: Open problems proposed at the NSF Conf. on Complexity Theory, Eugene, Oregon 1984, SIGACT News 16 (1984), 46.

[4] P. Erdős, A. Hajnal, J. Moon, Mathematical notes, Am. Math. Monthly, 71 (1964), 1107–1110.

[5] M. Fredman, J. Komlós, E. Szemerédi, Storing a sparse table with $O(1)$ worst-case access time, J. of the ACM, 31(3) (1984), 538–544.

[6] A. Hajnal, A theorem on k-saturated graphs, Canad. Math. J. 17 (1965) 720–724.

[7] L. Lovász, Combinatorial Problems and Exercises, North-Holland, 1979.

[8] M. Krieger, On the incompresability of monotone DNFs, in: Proc. 15th Int. Symp. on Fundamentals of Computation Theory (FCT'05), Liskiewicz, M., Reischuk, R. (Eds.), Lect. Notes in Comput. Sci., vol. 3623, Springer-Verlag (2005), pp. 32–43. Journal version to appear in: Theory of Computing Systems.

[9] K. Mehlhorn, On the program size of perfect and universal hash functions, in: Proc. 23rd Annual IEEE Symposium on Foundations of Computer Science (1982) 170–175.

[10] J. Nešetřil, S. Poljak, On the complexity of the subgraph problem, CMUC 26 (1985) 415–419.

[11] J. Radhakrishnan, ΣΠΣ threshold formulas, Combinatorica 14(3) (1994), 345–374.

[12] A. A. Razborov, Lower bounds on the monotone complexity of some Boolean functions, Soviet Math. Doklady 31 (1985) 354–357.

[13] H. Wang, C. Xang, Explicit constructions of perfect hash families from algebraic curves over finite fields, J. Comb. Theory (A) 93 (2001) 112–124.

4 Appendix

Theorem 4.1 (Hajnal [6])**.** *In a τ-critical graph without isolated vertices, every independent set of size s has at least s neighbors.*

Proof. (due to Lovász [7]) Let $G = (V, E)$ be a τ-critical graph without isolated vertices. Then G is also α-critical in that removal of any its edge increases its independence number $\alpha(G)$, i.e. the maximum size of an independent set in G. An independent set T is *maximal* if $|T| = \alpha(G)$.

Let us first show that every vertex belongs to at least one maximal independent set but not to all such sets. For this, take a vertex x and an edge $e = \{x, y\}$. Remove e from G. Since G is α-critical, the resulting graph has an independent set T of size $\alpha(G) + 1$. Since T was not independent in G, $x, y \in T$. Then $T - \{x\}$ is an independent set in G of size $|T - \{x\}| = \alpha(G)$, i.e. is a maximal independent set avoiding the vertex x, and $T - \{y\}$ is a maximal independent set containing x.

Hence, if X is an arbitrary independent set in G, then the intersection of X with *all* maximal independent sets in G is empty. It remains therefore to show that, if Y is an arbitrary independent set, and S is an intersection of Y with an arbitrary number of maximal independent sets, then

$$|N(Y)| - |N(S)| \geq |Y| - |S|,$$

where $N(Y)$ is the set of all neighbors of Y, i.e. the set of all vertices adjacent to at least one vertex in Y. Since an intersection of independent sets is an independent set, it is enough to prove the claim for the case when T is a maximal independent set and $S = Y \cap T$. Since clearly $N(S) \subseteq N(Y) - T$, we have

$$\begin{aligned} |N(Y)| - |N(S)| &\geq |N(Y) \cap T| = |T| - |S| - |T - Y - N(Y)| \\ &= \alpha(G) - |S| + |Y| - |(T \cup Y) - N(Y)| \geq |Y| - |S|, \end{aligned}$$

where the last inequality holds because the set $(T \cup Y) - N(Y)$ is independent. ∎