

# Very large cliques are easy to detect

A. E. Andreev\*      S. Jukna†

August 1, 2007

## 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 much fewer than  $n^k$  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 a logarithmic number of fanin-2 OR gates; the total number of gates does not exceed  $O(n^2 \log n)$ . Moreover, if we allow unbounded fanin, then a logarithmic number of gates is enough.

## 1 Introduction

Detecting the presence of a clique with a given number of vertices in graphs is one of the most prominent **NP**-complete problems. The corresponding to this problem boolean function  $\text{CLIQUE}(n, s)$  is a monotone boolean function of  $\binom{n}{2}$  boolean variables  $x_e$  representing the edges of an undirected graph  $G$  on  $n$  vertices, where  $x_e = 1$  iff the edge  $e$  is present in  $G$ . The value of the function is 1 iff  $G$  contains an  $s$ -clique, i.e. a complete subgraph on  $s$  vertices. Any proof that this function (say, for  $s = n/2$ ) cannot be computed using a polynomial number of AND, OR and NOT gates would imply the inequality  $\mathbf{P} \neq \mathbf{NP}$ . In the case of *monotone* circuits (containing only AND and OR gates but no NOT gates) this was proved by Razborov in [10]. But the question on whether the same also holds for non-monotone circuits remains widely open.

In this paper we are interested in proving good *upper bounds* on the size of monotone circuits with fanin-2 AND and OR gates computing  $\text{CLIQUE}(n, s)$ . The only non-trivial *upper* bound for  $\text{CLIQUE}(n, s)$  we are aware of is a non-monotone upper bound  $O(n^{2.5 \lceil s/3 \rceil})$  given in [4, 9] (see also [1] or Theorem 12.3 in [12]). This bound is obtained by a reduction to boolean matrix multiplication which can be done by a circuit of size  $o(n^{2.5})$  [3].

In this paper we are interested in *monotone* circuits for  $\text{CLIQUE}(n, s)$ . The simplest such circuit is a monotone DNF (disjunctive normal form): for each of the  $\binom{n}{s}$  potential  $s$ -cliques  $S$  we take an AND  $\bigwedge_{e \in \binom{S}{2}} x_e$ , testing whether  $S$  forms a clique in a given graph, and take an OR of these tests over all  $s$ -element subsets  $S$ . The resulting formula 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  $s$ ?

---

\*LSI Logic Corporation, AE-187, 1551 McCarthy Blvd., CA 95305 Milpitas, USA. Email: andreev@lsil.com

†Institute of Mathematics and Informatics, Akademijos 4, LT-08663 Vilnius, Lithuania

‡Research supported by the DFG grant SCHN 503/4-1.

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 [7]: for any  $s$ , any multilinear monotone circuit computing  $\text{CLIQUE}(n, s)$  requires  $\binom{n}{s} - 1$  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 [10] and numerically improved by Alon and Boppana [1]: for every constant  $s \geq 3$ , the function  $\text{CLIQUE}(n, s)$  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 simple padding argument (see Appendix A), this implies that even detecting cliques of size  $n - k$  requires super-polynomial number of gates, as long as  $k \leq n/2$  grows faster than  $\log^3 n$ . But what is the complexity of  $\text{CLIQUE}(n, n - k)$  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 [7]: the function  $\text{CLIQUE}(n, n - 1)$  can be computed by a monotone circuit using only  $O(\log n)$  OR gates. The corresponding circuit is a  $\Pi\Sigma\Pi$ -circuit: the first (next to the input variables) level consists of AND gates followed by a level of OR gates and a one more level of AND gates. Note that a DNF for this function requires  $n - 1$  OR gates.

The argument of [7] 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 also enough for growing  $k$ , as long as  $k = O(\sqrt{\log n})$ . Note that for  $k = \Omega(\sqrt{\log n})$  any DNF for  $\text{CLIQUE}(n, n - k)$  requires a super-polynomial number  $\binom{n}{k} - 1 = n^{\Omega(\sqrt{\log n})}$  of gates. Thus, increasing the depth of a circuit just by 1 leads to exponential savings.

Instead of constructing a circuit with few OR gates for the Clique function it will be convenient to construct a circuit with few AND gates for the dual function. Recall that a *dual* of a boolean function  $f(x_1, \dots, x_n)$  is the boolean function  $f^*(x_1, \dots, x_n) = \neg f(\neg x_1, \dots, \neg x_n)$ , where “ $\neg$ ” denotes negation. If  $f$  is monotone, then its dual  $f^*$  is also monotone.

In particular, the dual of  $\text{CLIQUE}(n, n - k)$  accepts a given graph  $G$  on  $n$  vertices iff  $G$  has no independent set with  $n - k$  vertices, which is equivalent to  $\tau(G) \geq k + 1$ , where  $\tau(G)$  is the vertex cover number of  $G$ . Recall that a vertex cover in a graph  $G$  is a set of its vertices containing at least one endpoint of each edge;  $\tau(G)$  is the minimum size of such a set. Hence, the dual of  $\text{CLIQUE}(n, n - k)$  is a monotone boolean function  $\text{VC}(n, k)$  of  $\binom{n}{2}$  boolean variables representing the edges of an undirected graph  $G$  on  $n$  vertices, whose value is 1 iff  $G$  does not have a vertex cover of cardinality  $k$ . We will construct a monotone  $\Sigma\Pi\Sigma$ -circuit for  $\text{VC}(n, k)$ . Replacing OR gates by AND gates (and vice versa) in this circuit yields a monotone  $\Pi\Sigma\Pi$ -circuit for  $\text{CLIQUE}(n, n - k)$ .

Our main result is the following:

**Theorem 1.** *For every  $k$ , the function  $\text{VC}(n, k)$  can be computed by a monotone  $\Sigma\Pi\Sigma$ -circuit containing at most  $N = O(k^3 e^{2k} \log n) + 2^{O(k^2)}$  fanin-2 AND gates. The total number of fanin-2 gates does not exceed  $n^2 N$ .*

Hence, for every fixed  $k$ , the function  $\text{CLIQUE}(n, n - k)$  can be computed by a monotone  $\Pi\Sigma\Pi$ -circuit containing at most  $O(\log n)$  fanin-2 OR gates; the total number of gates does

not exceed  $O(n^2 \log n)$ . This means that for every constant  $k$ , the total number of fanin-2 gates is almost linear in the number  $\binom{n}{2} = \Theta(n^2)$  of input variables. Moreover, the number of gates is polynomial, as long as  $k = O(\sqrt{\log n})$ . Recall that already for  $k = \omega(\log^3 n)$ , any monotone circuit for  $\text{CLIQUE}(n, n - k)$  requires a super-polynomial number of gates.

## 2 The construction

Theorem 1 is a direct consequence of the following more general result which, for every constant  $k$ , allows us also to explicitly *construct* the desired circuit.

We consider graphs on vertex-set  $[n] = \{1, \dots, n\}$ . We have a set  $X$  of  $\binom{n}{2}$  boolean variables  $x_e$  corresponding to edges. Each graph  $G = ([n], E)$  is specified by setting the values 0 and 1 to these variables:  $E = \{e : x_e = 1\}$ . The function  $\text{VC}(n, k)$  accepts  $G$  iff  $\tau(G) \geq k + 1$ .

A graph is  $\tau$ -critical if removal of any of its edges reduces the vertex cover number. For example, there are only two  $\tau$ -critical non-isomorphic graphs  $H$  with  $\tau(H) = 2$ , a triangle and a graph consisting of two disjoint edges.

Let  $\mathcal{G}(n, k)$  denote the set of all  $\tau$ -critical graphs on  $[n] = \{1, \dots, n\}$  with  $\tau(H) = k + 1$ . Observe that graphs in  $\mathcal{G}(n, k)$  are the smallest (with respect to the number of edges) graphs accepted by  $\text{VC}(n, k)$ .

Given a family  $F$  of functions  $f : [n] \rightarrow [r]$ , consider the following monotone  $\Sigma\Pi\Sigma$ -circuit

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

where

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

This circuit accepts 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  $G$  between  $f^{-1}(a)$  and  $f^{-1}(b)$ .

A family  $F$  of functions  $f : [n] \rightarrow [r]$  is  $s$ -perfect 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.

Using a simple probabilistic argument, Mehlhorn [8] shows that  $(n, r, s)$ -perfect hash families  $F$  of size  $|F| \leq \lceil se^{s^2/r} \log n \rceil$  exist for all  $2 \leq s \leq r \leq n$ . In particular,  $|F| = O(\log n)$  for every constant  $r$ .

**Theorem 2.** *Let  $F$  be an  $(n, r, s)$ -perfect hash family with  $s = 2(k + 1)$  and  $r \geq s$ . Then the circuit  $\Phi_F(X)$  computes  $\text{VC}(n, k)$ . Moreover, the number of fanin-2 AND gates in  $\Phi_F(X)$  does not exceed  $O(k^2 |F|) + (r/k)^{O(k^2)}$ .*

*Proof.* Recall that all graphs  $G$  in  $\mathcal{G}(n, k)$  have one and the same set  $[n] = \{1, \dots, n\}$  of vertices. Hence, some of the vertices may be *isolated*, i.e. may be incident with none of the edges of  $G$ . Important for our construction is that the number of non-isolated vertices in these graphs depends only on  $k$ , and not on  $n$ . This is a direct consequence of a result, due to Hajnal [6], that in a  $\tau$ -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 Appendix B.)

**Claim 1.** *Every  $\tau$ -critical graph  $G$  has at most  $2\tau(G)$  non-isolated vertices.*

*Proof.* Let  $G = (V, E)$  be an arbitrary  $\tau$ -critical graph, and let  $U \subseteq V$  be the set of non-isolated vertices of  $G$ . The induced subgraph  $G' = (U, E)$  has no isolated vertices and is still  $\tau$ -critical with  $\tau(G') = \tau(G)$ . Let  $S \subseteq U$  be an arbitrary vertex cover of  $G'$  with  $|S| = \tau(G)$ . The complement  $T = U - S$  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  $|U| = |S| + |T| \leq 2|S| \leq 2\tau(G)$  on the total number of non-isolated vertices of  $G$  follows.  $\square$

Let now  $F$  be an arbitrary  $s$ -perfect family of functions  $f : [n] \rightarrow [r]$ . We have to verify that then the circuit  $\Phi_F(X)$  computes  $\text{VC}(n, k)$ . Since the circuit is monotone, it is enough to show that:

- (a)  $\tau(G) \geq k + 1$  for every graph  $G$  accepted by  $\Phi_F(X)$ , and
- (b)  $\Phi_F(X)$  accepts all graphs from  $\mathcal{G}(n, k)$ .

To show (a), suppose that  $\Phi_F(X)$  accepts some graph  $G$ . Then this graph must be accepted by some sub-circuit  $K_{f,H}$  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  $\{i, j\}$ , i.e., if  $S \cap \{i, j\} \neq \emptyset$ , then the set  $f(S)$  must cover the edge  $\{a, b\}$ . This means that, for any vertex cover  $S$  in  $G$ , the set  $f(S)$  is a vertex cover in  $H$ . Taking a minimal vertex cover  $S$  in  $G$  we obtain  $\tau(G) = |S| \geq |f(S)| \geq \tau(H) = k + 1$ .

To show (b), take an arbitrary graph  $G = ([n], E)$  in  $\mathcal{G}(n, k)$ , and let  $U$  be the set of its non-isolated vertices. By Claim 1,  $|U| \leq 2\tau(G) = 2(k + 1)$ . By the definition of  $F$ , some function  $f : [n] \rightarrow [r]$  must be one-to-one on these vertices. For  $i, j \in U$  join  $a = f(i)$  and  $b = f(j)$  by an edge iff  $\{i, j\} \in E$ . Since  $G \in \mathcal{G}(n, k)$  and  $f$  is one-to-one on all non-isolated vertices of  $G$ , the resulting 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-circuit  $K_{f,H}(X)$  of  $\Phi_F(X)$ , and hence, the circuit  $\Phi_F(X)$  itself must accept  $G$ .

It remains to estimate the number of fanin-2 AND gates in  $\Phi_F(X)$ . If we allow unbounded fanin, then each sub-circuit  $K_{f,H}$  contributes just one AND gate. Hence,  $\Phi_F(X)$  has at most  $|\mathcal{G}(r, k)| + |F|$  unbounded fanin AND gates. Since the fanin of each such gate is actually bounded by the number  $|E(H)|$  of edges in the corresponding graph  $H \in \mathcal{G}(r, k)$ , all these AND gates can be realized by at most  $N = (|\mathcal{G}(r, k)| + |F|) \cdot M$ , fanin-2 AND gates, where  $M$  is the largest number of edges in a graph  $H \in \mathcal{G}(r, k)$ . Erdős, Hajnal and Moon [5] prove that every  $\tau$ -critical graph has at most  $\binom{\tau(H)+1}{2}$  edges. Hence,  $M \leq \binom{k+2}{2} = O(k^2)$  and  $|\mathcal{G}(r, k)| \leq (r/k)^{O(M)} \leq (r/k)^{O(k^2)}$ .

This completes the proof of Theorem 2.  $\square$

**Remark 1.** Observe that the circuit  $\Phi_F(X)$  for  $\text{VC}(n, k)$  is multilinear, i.e. inputs to each its AND gate are computed from disjoint sets of variables. On the other hand, Krieger [7] shows that *every* monotone multilinear circuit for the dual function  $\text{CLIQUE}(n, n - k)$  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 On the explicitness

Using *explicit* perfect hash families we can obtain explicit circuits. 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 [11] using algebraic curves over finite fields. With this construction Theorem 2 achieves the upper bound stated in Theorem 1 by explicit monotone circuits.

**Theorem 3.** *For every constant  $k$ , the function  $\text{CLIQUE}(n, n - k)$  can be computed by an explicit monotone circuit using only  $O(\log n)$  number of fanin-2 OR gates.*

The construction in [11] is almost optimal (the family has only *logarithmic* number of functions), but is somewhat involved. On the other hand, perfect hash families of *polylogarithmic* size can be constructed very easily.

Let  $s \geq 1$  be a fixed integer and  $r = 2^s$ . 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  $h_1, \dots, h_b$  be the family of functions  $h_i : [n] \rightarrow \{0, 1\}$  determined by the columns of  $M$ ; hence,  $h_i(a) = m_{a,i}$ . Let also  $g : \{0, 1\}^s \rightarrow [r]$  be defined by  $g(x) = \sum_{i=1}^s x_i 2^{i-1}$ .

By Bondy's theorem [2], the projections of any set of  $s + 1$  distinct binary vectors on some set of  $s$  coordinates must all be distinct. Hence, for any set  $a_1, \dots, a_{s+1}$  of  $s + 1$  rows there exist  $s$  columns  $h_{i_1}, \dots, h_{i_s}$  such that all  $s + 1$  vectors  $(h_{i_1}(a_j), \dots, h_{i_s}(a_j))$ ,  $j = 1, \dots, s + 1$  are distinct. Therefore, the function  $f(x) = g(h_{i_1}(x), \dots, h_{i_s}(x))$  takes different values on all  $s + 1$  points  $a_1, \dots, a_{s+1}$ . Thus, taking the superposition of  $g$  with  $\binom{b}{s} \leq \log^s n$   $s$ -tuples of functions  $h_1, \dots, h_b$ , we obtain a family  $F$  of  $|F| \leq \log^s n$  functions  $f : [n] \rightarrow [r]$  which is  $(s + 1)$ -perfect.

#### 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] D. Coppersmith, S. Winograd, On the asymptotic complexity of matrix multiplication, *SIAM J. on Comput.*, 11 (1982) 472–492.
- [4] F. Chung, R. M. Karp, in: Open problems proposed at the NSF Conf. on Complexity Theory, Eugene, Oregon 1984, *SIGACT News* 16 (1984), 46.
- [5] P. Erdős, A. Hajnal, J. Moon, *Mathematical notes*, *Am. Math. Monthly*, 71 (1964), 1107–1110.
- [6] A. Hajnal, A theorem on  $k$ -saturated graphs, *Canad. Math. J.* 17 (1965) 720–724.
- [7] M. Krieger, On the incompressibility 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*.

- [8] 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.
- [9] J. Nešetřil, S. Poljak, On the complexity of the subgraph problem, CMUC 26 (1985) 415–419.
- [10] A. A. Razborov, Lower bounds on the monotone complexity of some Boolean functions, Soviet Math. Doklady 31 (1985) 354–357.
- [11] H. Wang, C. Xang, Explicit constructions of perfect hash families from algebraic curves over finite fields, J. Comb. Theory (A) 93 (2001) 112–124.
- [12] I. Wegener, The Complexity of Boolean Functions, Wiley-Teubner, 1987.

## A Appendix

**Proposition 1.** *For  $k \leq n/2$ , every monotone circuit computing  $\text{CLIQUE}(n, n - k)$  requires  $2^{\Omega(k^{1/3})}$  gates.*

*Proof.* Fix the integer  $m$  with  $m - s = k$  where  $s = \lfloor (m/\log m)^{2/3}/4 \rfloor$ ; hence  $s = \Omega(k^{2/3})$ . Then  $\text{CLIQUE}(m, s)$  is a sub-function of (i.e. can be obtained by setting to 1 some variables in)  $\text{CLIQUE}(n, n - k)$ : 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). On the other hand, according to the lower bound Alon and Boppana [1] (mentioned in the introduction) the function  $\text{CLIQUE}(m, s)$ , and hence, also the function  $\text{CLIQUE}(n, n - k)$  requires monotone circuits of size  $2^{\Omega(\sqrt{s})} = 2^{\Omega(k^{1/3})}$ .  $\square$

## B Appendix

**Theorem 4** (Hajnal [6]). *In a  $\tau$ -critical graph without isolated vertices every independent set  $S$  has at least  $|S|$  neighbors.*

*Proof.* Let  $G = (V, E)$  be a  $\tau$ -critical graph without isolated vertices. Then  $G$  is also  $\alpha$ -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  $\alpha$ -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. □