Computing 16, 99—111 (1976)

Monotone Switching Circuits and Boolean Matrix Product

K. Mehlhorn, Saarbrücken, and **Z. Galil*,** Ithaca, New York

Received April 25, 1975

Abstract — Zusammenfassung

Monotone Switching Circuits and Boolean Matrix Product. We explore the concept of local transformations of monotone switching circuits, i.e. what kind of local changes in a network leave the input/output behavior invariant. We obtain several general theorems in this direction. We apply these results to boolean matrix product and prove that the school-method for matrix multiplication yields the unique monotone circuit**.

Monotone Schaltkreise und Multiplikation Boolescher Matrizen. Wir untersuchen den Effekt lokaler Änderungen auf das Eingabe/Ausgabe-Verhalten monotoner Netzwerke. Wir wenden die Ergebnisse auf die Multiplikation Boolescher Matrizen an und zeigen, daß die Schulmethode für die Matrizenmultiplikation den alleinigen optimalen Schaltkreis liefert.

I. Introduction

The complexity of monotone switching networks, i.e. circuits consisting of and- and or-gates only, received considerable attention recently (Lamagna and Savage, Mehlhorn, Paterson, Pratt, Schnorr). The concept of local transformations is basic to all of these papers, i.e all of these papers use the fact that certain local changes in a monotone network leave the input/output of the circuit invariant. In section III we prove several general theorems in this direction. All transformations used in the papers mentioned above are special cases of these theorems.

In sections IV, V and VI we combine our results on local transformations with the concepts of stepwise simplification and necessary gates and obtain:

Theorem:

a) *Every monotone circuit for boolean matrix product of $n \times n$ matrices uses at least n^3 and-gates and $n^3 - n^2$ or-gates.*

b) *Only the "simple-minded" circuit yielded by the school method for matrix multiplication — compute the n^3 products $a_{ik}b_{kj}$ and sum them up — actually achieves these lower bounds.*

---

* The second author was partially supported by ONR Grant N00014-67-A-0077-0021.

** Related results were obtained independently by Mike Paterson.

II. Notation

A (monotone) boolean circuit is a finite directed acyclic graph with the following properties:

(1) The in-degree of every node is either 0 or 2,

(2) the source nodes (nodes with in-degree 0) are labelled by some set V of variables,

(3) the nodes of in-degree 2 are labelled by elements of the set {and, or}.

Every circuit β associates boolean functions with its edges (wires) in an obvious way. We denote the function associated with wire v in circuit β by $\text{res}_{\beta, v}$. A circuit β realizes the set F of boolean functions if for every $f \in F$ there is some wire v with $f = \text{res}_{\beta, v}$.

$K = \{0, 1\}$ is the set of truth values. We order K by $0 \leq 1$, K^n is ordered componentwise. A function $f: K^n \to K$ is monotone if it preserves this ordering. Monotone circuits realize monotone functions and vice versa.

A monome is a product of variables (no negations). The prime implicants of monotone functions are monomes; we denote the set of prime implicants of a function f by prime (f). If t_1 and t_2 are monomes then $t_1 \cap t_2$ denotes the monome consisting of the variables common to t_1 and t_2. ε denotes the empty monome, i.e. the monome consisting of no variables. $t_1 \subseteq t_2$ if $t_1 = t_1 \cap t_2$.

Let t be any monome. $\psi(t)$ is the function which is equal to 1 only if all variables in t are equal to 1. $\psi(t_1, \ldots, t_n)$ is equal to $\psi(t_1) \vee \ldots \vee \psi(t_n)$.

Let F be a set of boolean functions. Then $C_{AND}(F)$ $(C_{OR}(F))$ is the minimal number of and-gates (or-gates) in any monotone circuit realizing F.

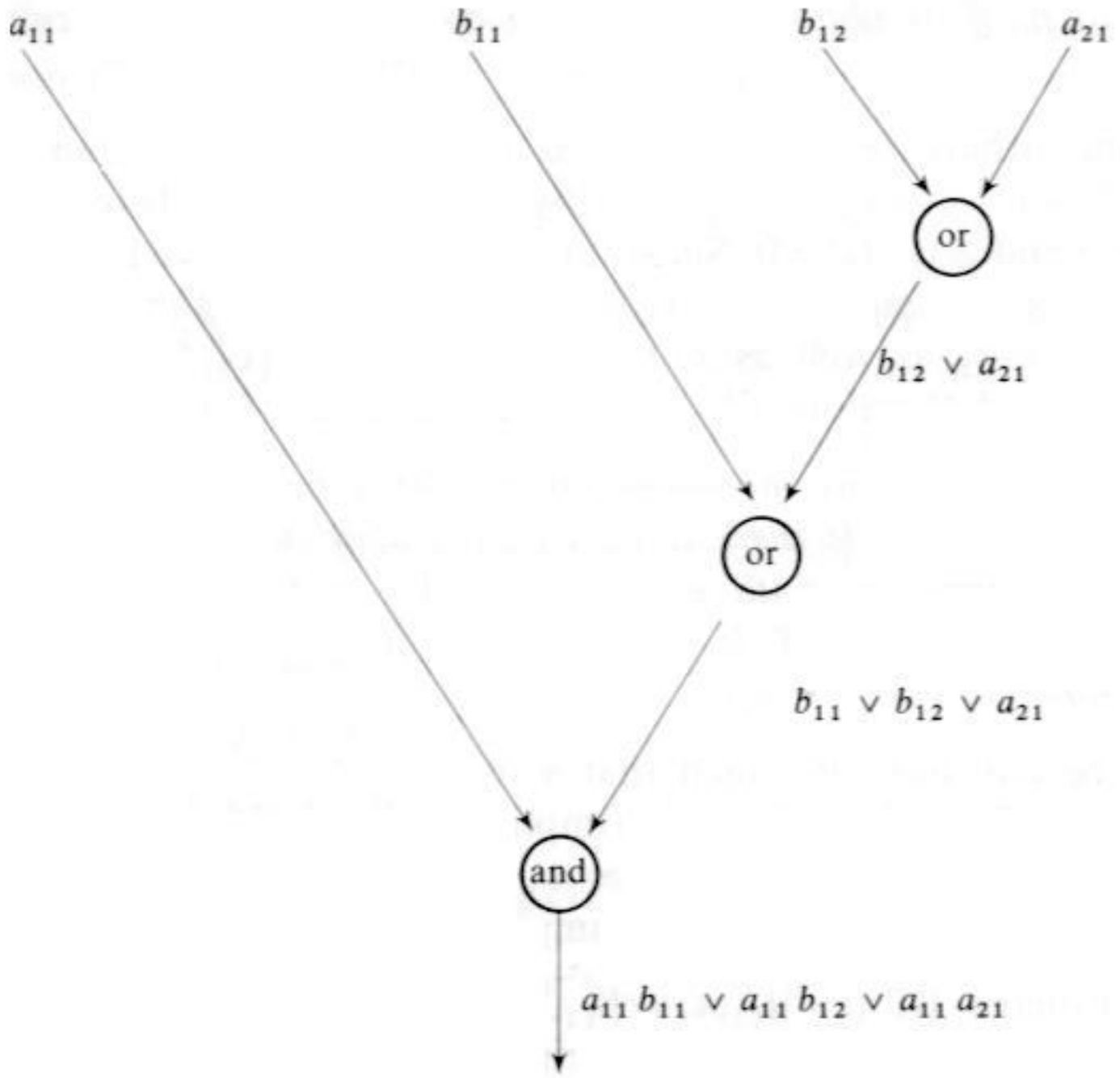
III. Local Transformations of Monotone Networks

In this section we investigate the effect of local changes in a monotone network on the input/output behavior of the network. The theorems below formalize the idea that merging of information is an irreversible process in monotone computations.

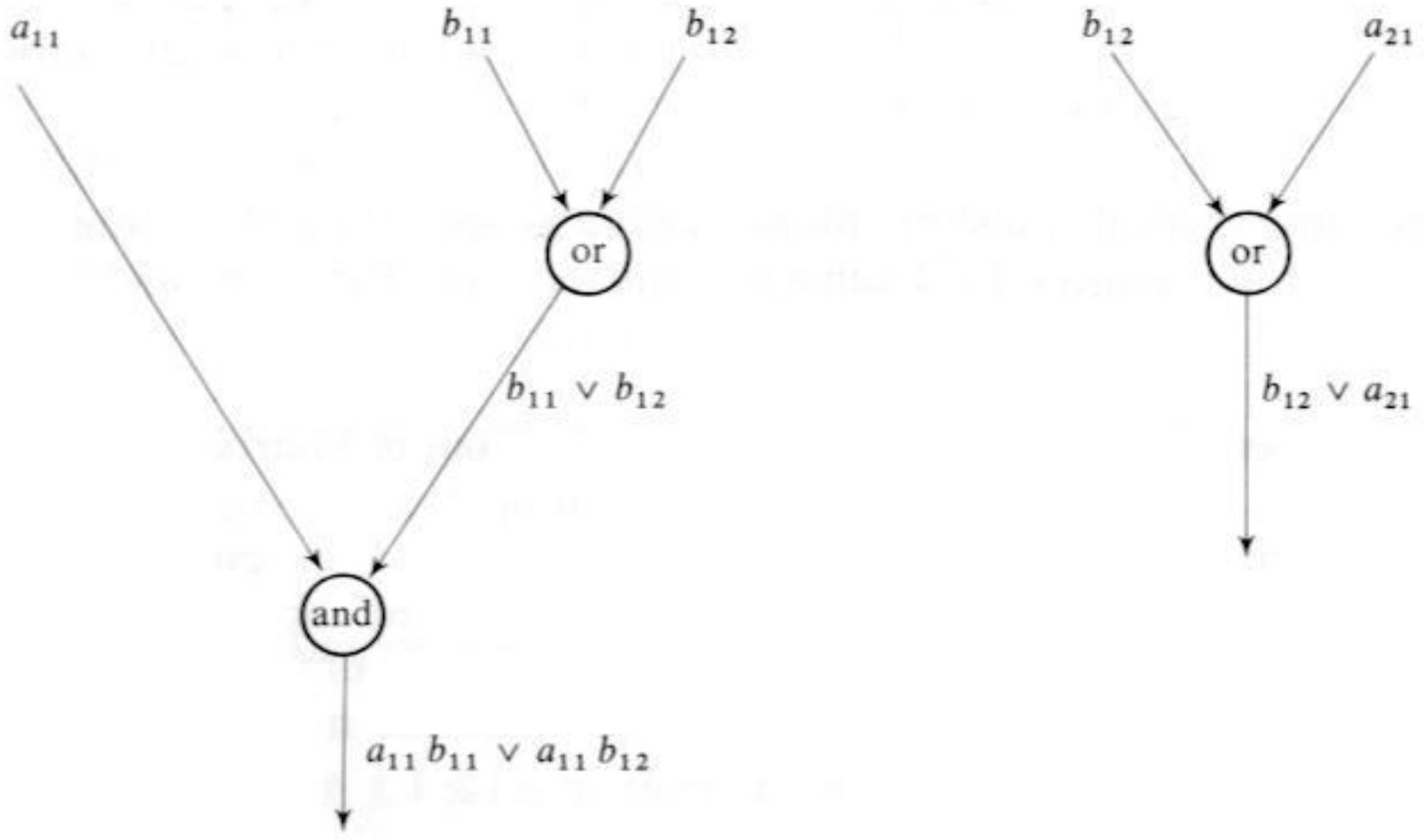
Theorem 1: *Let β be any monotone circuit, let v be any wire in β, and let prime $(res_{\beta, v}) = \{t_0, t_1, \ldots, t_k\}$. If there is no monome t with $t \cdot t_0 \in$ prime (f) for some output f of β then the following circuit β' is equivalent to β. β' is obtained from β by connecting v to a circuit realizing $\psi(t_1, \ldots, t_k)$.*

Proof: Assume otherwise. Then there is some choice of inputs such that $f^{\beta} = 1$ but $f^{\beta'} = 0$ for some output f (by monotonicity). Moreover $\psi(t_1, \ldots, t_k) = 0$, but $\psi(t_0) = 1$. Hence turning off any variable in t_0 will send the output f from 1 to 0. Thus $t \cdot t_0 \in$ prime (f) for some monome t. ■

Suppose that the following circuit is part of a monotone network for boolean matrix product.



$a_{11} a_{21}$ is prime implicant of the output wire of the and-gate. $a_{11} a_{21}$ is not part of any prime implicant $a_{ik} b_{kj}$ of any of the functions $f_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$. Therefore we may replace the subcircuit shown above by the following circuit.



Theorem II: *Let v be any wire in a monotone circuit β and let $t \cdot t_1$ and $t \cdot t_2 \in prime\,(res_{\beta, v})$. If for all outputs f of β and monomes s, $s \cdot t$ is an implicant of f whenever $s \cdot t \cdot t_1$ and $s \cdot t \cdot t_2$ are implicants of f then the following circuit β'*

is equivalent to β. β' is obtained from β by connecting v to a circuit realizing $res_{\beta,v} \vee \psi(t)$.

Proof: Assume otherwise. Then there is some choice of inputs such that $f^{\beta}=0$, but $f^{\beta'}=1$ for some output f (by monotonicity). Moreover $\psi(t)=1$, yet $\psi(t \cdot t_1)=0$ and $\psi(t \cdot t_2)=0$. Since turning off any variable in t will send $f^{\beta'}$ to 0 there is a prime implicant $s \cdot t$ of $f^{\beta'}$. We infer from the structure of the network that $s \cdot t \cdot t_1$ as well as $s \cdot t \cdot t_2$ are implicants of f. Hence $s \cdot t$ is an implicant of f^{β} and therefore $f^{\beta}=1$. This contradicts $f^{\beta}=0$. ■

Corollary III: *Let β be any monotone circuit, let v be any wire in β, and let $t \cdot t_1, t \cdot t_2 \in prime(res_{\beta,v})$. If there is no output f of β such that there are prime implicants $e_1, e_2 \in prime(f)$ with $e_1 \cap t_1 \neq \varepsilon$ and $e_2 \cap t_2 \neq \varepsilon$ then the following circuit β' is equivalent to β. β' is obtained from β by connecting v to a circuit realizing the function $res_{\beta,v} \vee \psi(t)$.*

Proof: Let s be any monome such that $s \cdot t \cdot t_1$ and $s \cdot t \cdot t_2$ are implicants of some output f of β. Then there are prime implicants e_1, e_2 of f such that $e_1 \subseteq s \cdot t \cdot t_1$ and $e_2 \subseteq s \cdot t \cdot t_2$. By assumption either $e_1 \cap t_1 = \varepsilon$ or $e_2 \cap t_2 = \varepsilon$. Thus either $e_1 \subseteq s \cdot t$ or $e_2 \subseteq s \cdot t$. Therefore $s \cdot t$ is implicant of f. Apply theorem II. ■

We apply corollary III ($t=a_{11}, t_1=b_{11}, t_2=b_{12}$) to the output wire of the and-gate. It allows us to add a_{11} as prime implicant of the output wire, ie. we may replace the output wire by the input a_{11} and eliminate the and-gate. We apply theorem II to the output wire of the right or-gate ($t=\varepsilon, t_1=a_{21}, t_2=b_{12}$). It allows to set the output wire to the constant 1. So we end up with the following circuit.

Our theorems on local transformations include as special cases all transformations used in the papers by Lagmagna and Savage, Paterson, and Pratt.

IV. The Number of And-Gates in Monotone Realizations of Matrix Multiplication

In this section we show that n^3 and-gates are used in every monotone realization of matrix multiplication. The "simple-minded" approach to matrix multiplication – compute the n^3 products $a_{ik} b_{kj}$ and sum them up – uses exactly this number of and-gates. Thus it optimizes the number of and-gates in monotone computations.

Boolean matrix multiplication corresponds to computing the following system F of n^2 functions f_{ij}, $1 \leq i,j \leq n$, in the $2n^2$ variables a_{ij}, b_{ij}, $1 \leq i,j \leq n$:

$$f_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

We will also consider systems F^J for $J \subseteq \{1, \ldots, n\}^2$ (this idea is due to Schnorr). The function f_{ij}^J is obtained from f_{ij} by setting $a_{ik}=1$ for $(i,k) \in J$, i.e.

$$f_{ij}^J = \sum_{k=1}^{n} \hat{a}_{ik}\, b_{kj}$$

where $\hat{a}_{ik}=1$ for $(i,k) \in J$ and $\hat{a}_{ik}=a_{ik}$ otherwise. The system F^J is the simpler the larger the set J is. F^J is equal to F for $J=\emptyset$, F^J is trivial for $J=\{1, \ldots, n\}^2$.

Lemma 1: *Let* $J=\{1, \ldots, n\}^2$. *Then* $C_{AND}(F^J)=0$.

Proof: Let $J=\{1, \ldots, n)^2$. Then $f_{ij}^J = \sum_{k=1}^{n} b_{kj}$ and hence $C_{AND}(F^J)=0$. ■

Consider now any sequence

$$\emptyset = J_0 \subset J_1 \subset \ldots \subset J_{n^2-1} \subset J_{n^2} = \{1, \ldots, n\}^2$$

with $|J_{k+1} - J_k| = 1$ for $k=0, 1, \ldots, n^2-1$. By lemma 1

$$C_{AND}(F) = C_{AND}(F^{J_0}) - C_{AND}(F^{J_{n^2}})$$
$$= \sum_{k=0}^{n^2-1} \left[C_{AND}(F^{J_k}) - C_{AND}(F^{J_{k+1}})\right].$$

Lemma 2: *Let* $J \subseteq \{1, \ldots, n\}^2$ *and* $(i,k) \notin J$. *Then*

$$C_{AND}(F^J) - C_{AND}(F^{J \cup \{(i,k)\}}) \geq n$$

Suppose we proved lemma 2. Then $C_{AND}(F^{J_k}) - C_{AND}(F^{J_{k+1}}) \geq n$ *for* $k=0, \ldots, n^2-1$ *and hence* $C_{AND}(F) \geq n \cdot n^2 = n^3$.

Proof of lemma 2: Let $J \subset \{1, \ldots, n\}^2$ and $(i,k) \notin J$. The n products $a_{ik}\, b_{kj}$, $1 \leq j \leq n$, are prime implicants of the functions f_{ij}^J, but they are not prime implicants of the corresponding functions $f_{ij}^{J \cup \{(i,k)\}}$. Remember, we set $a_{ik}=1$. We will show that every monotone circuit β for F^J uses n and-gates to compute these products and that we can obtain a circuit for $F^{J \cup \{(i,k)\}}$ by eliminating these n and-gates.

Let β be a monotone circuit for F^J such that:

(1) the number of and-gates in β is equal to $C_{AND}(F^J)$,

(2) among the circuits with property (1), β also uses a minimal number of or-gates.

We consider the computation of the products $a_{ik}\, b_{kj}$ in β (this idea is due to Pratt).

Lemma 3: *For every* j, $1 \leq j \leq n$, *there exists an and-gate* G_j *in* β *such that*

(1) $a_{ik}\, b_{kj}$ *is a prime implicant of (the function associated with) the output wire of* G_j,

(2) a_{ik} *is prime implicant of some input wire of* G_j *and every other prime implicant of this wire contains* b_{lj} *for some* l, $1 \leq l \leq n$.

Proof: Assume otherwise, say gate G_j does not exist. We show: If $a_{ik}b_{kj}$ is prime implicant of $\text{res}_{\beta,v}$ for wire v then $\text{res}_{\beta,v}$ also has a prime implicant containing neither a_{ik} nor any b_{lj} with $l \neq k$. Call this property (A). (A) contradicts the definition of f_{ij}^J.

β is a directed graph. We can order the wires of β in the obvious way, the input wires being the minimal elements. We prove (A) by induction on this ordering of the wires in β.

Let v be an input wire of β starting at node v, say. Then v is the only prime implicant of $\text{res}_{\beta,v}$ and hence (A) holds for all input wires.

Consider now output wires of or-gates. If $a_{ik}b_{kj}$ is prime implicant of such a wire then it is also prime implicant of one of the input wires. (A) holds for this input wire and hence it also holds for the output wire.

It remains to consider the output wires of and-gates. If $a_{ik}b_{kj}$ is prime implicant of the output wire then either

(1) a_{ik} is prime implicant of the "left" input wire and b_{kj} is prime implicant of the "right" input wire or

(2) a_{ik} is prime implicant of the left input wire and $a_{ik}b_{kj}$ is prime implicant of the right input wire or

(3) b_{kj} is prime implicant of the left input wire and $a_{ik}b_{kj}$ is prime implicant of the right input wire or

(4) $a_{ik}b_{kj}$ is prime implicant of both input wires.

We discuss only the cases (2) and (3), the others being similar.

Case 2: Since a_{ik} is prime implicant of the left input wire there is a prime implicant t_1 of this wire containing neither a_{ik} nor any b_{lj} with $1 \leq l \leq n$. Since $a_{ik}b_{kj}$ is prime implicant of the right input wire and (A) holds for this wire there is a prime implicant t_2 of this wire containing neither a_{ik} nor any b_{lj} with $l \neq k$. $t_1 \cdot t_2$ is implicant of the output wire and contains neither a_{ik} nor any b_{lj} with $l \neq k$. A monome $t \subseteq t_1 \cdot t_2$ is prime implicant of the output wire. Hence (A) holds for the output wire of the and-gate under consideration.

Case 3: Since $a_{ik}b_{kj}$ is the prime implicant of the right input wire and (A) holds for this wire, there is a prime implicant t_1 of this wire which does not contain either a_{ik} or b_{lj} with $l \neq k$. $A\, t \subseteq b_{kj} \cdot t_1$ is prime implicant of the output wire. Hence (A) holds for the output wire. ■

For every $j, 1 \leq j \leq n$, let G_j be the and-gate whose existence is ensured by lemma 3.

Lemma 4: *The gates $G_j, 1 \leq j \leq n$, are pairwise distinct.*

Proof: Assume otherwise, say $G = G_{j_1} = G_{j_2}$ for $j_1 \neq j_2$. The products $a_{ik}b_{kj_1}$ and $a_{ik}b_{kj_2}$ are prime implicants of the output wire og G, hence a_{ik} is prime implicant of just one input wire. Say, a_{ik} is prime implicant of the right input

wire. Let t be any other prime implicant of the right input wire of G. By property (2) of the gates G_{j_1} and G_{j_2} t contains $b_{l_1 j_1}$ as well as $b_{l_2 j_2}$ for some l_1, l_2. Theorem I implies that we can obtain a circuit β' which is equivalent to β if we simply feed a_{ik} into the right input wire of G. This transformation does not increase the number of either type of gate. Therefore we may assume w.l.o.g. that a_{ik} is the sole prime implicant of the right input wire. We apply now theorem II to the output wire of G. Let $t = a_{ik}$, $t_1 = b_{kj_1}$, and $t_2 = b_{kj_2}$. No output f_{ij}^J of β depends on b_{kj_1} as well as b_{kj_2}. Hence we may add a_{ik} as prime implicant of the output wire of G. Since a_{ik} is the sole prime implicant of the right input wire of G, a_{ik} is contained in every prime implicant of the output wire of G. Thus the following circuit β' is equivalent to β. Simply feed a_{ik} into the output wire of G and drop G. This contradicts the optimality of β. Thus we derived a contradiction to $G_{j_1} = G_{j_2}$ for $j_1 \neq j_2$. ■

If we set $a_{ik} = 1$ the right input wire of the and-gates G_j, $1 \leq j \leq n$, becomes constant. We can eliminate these and-gates by directly connecting their left input wire with their output wire. This transformation produces a circuit for $F^{J \cup \{(i,k)\}}$ which has n and-gates less than β. Thus

$$C_{AND}(F^{J \cup \{(i,k)\}}) \leq (\text{number of and-gates in } \beta) - n = C_{AND}(F^J) - n.$$

This ends the proof of lemma 2 and theorem II. ■■

Theorem IV: *Every monotone circuit for matrix multiplication contains at least n^3 and-gates.*

Corollary V: *n^3 and-gates are necessary and sufficient to realize matrix multiplication using only and- and or-gates.*

V. The Number of Or-Gates in Monotone Realizations of Matrix Multiplication

In this section we determine the number of or-gates used in monotone realizations of matrix multiplication: $n^3 - n^2$ or-gates are necessary and sufficient. Combined with the result of the preceding section this shows the optimality of the "simple-minded" approach to matrix multiplication.

Again we consider systems $F^{J\,1}$. The Function f_{ij}^J is obtained from f_{ij} by setting $a_{ik} = 0$ for $(i, k) \in J$, i.e.

$$f_{ij}^J = \sum_{k=1}^{n} \hat{a}_{ik} b_{kj},$$

where $\hat{a}_{ik} = 0$ for $(i, k) \in J$ and $\hat{a}_{ik} = a_{ik}$ otherwise.

We prove a lemma similar to lemma 2.

Lemma 5: *Let $J \subset \{1, \ldots, n\}^2$, $(i, k) \notin J$ and $(i, k') \notin J$. Then*

$$C_{OR}(F^J) - C_{OR}(F^{J \cup \{(i,k)\}}) \geq n.$$

---

[1] F^J and G_j below are not those of Section IV, but they have the same role.

Suppose we proved lemma 5. Let

$$\emptyset = J_0 \subset J_1 \subset \ldots \subset J_{n^2} = \{1, \ldots, n\}^2$$

be a sequence with $|J_{k+1} - J_k| = 1$ for $k = 0, 1, \ldots, n^2 - 1$. By lemma 5 $C_{OR}(F^{J_K}) - C_{OR}(F^{J_{k+1}}) \geq n$ for $n^2 - n$ k's with $0 \leq k \leq n^2 - 1$. Hence $C_{OR}(F) \geq n(n^2 - n) = n^3 - n^2$.

Proof of lemma 5: Let $J \subset \{1, \ldots, n\}^2$, $(i, k) \notin J$ and $(i, k') \notin J$ with $k \neq k'$. Therefore f_{ij}^J has at least two prime implicants, namely $a_{ik} b_{kj}$ and $a_{ik'}, b_{k'j}$. We will exhibit an or-gate which computes their disjunction.

Let β be a monotone circuit for F^J such that:

(1) the number of or-gates in β is equal to $C_{OR}(F^J)$,

(2) among the circuits with property (1), β also uses a minimal number of and-gates.

Lemma 6: *For every* $j, 1 \leq j \leq n$, *there exists an or-gate* G_j *such that*

either

(1) a_{ik} *is sole prime implicant of one input wire,*

(2) *every prime implicant of the other input wire contains either* a_{ik} *or some* b_{lj} *with* $l \neq k$,

(3) *there is a prime implicant of the other input wire which does not contain* a_{ik},

or

(1′) $a_{ik} b_{kj}$ *is sole prime implicant of one input wire,*

(2′) *every prime implicant of the other input wire contains either* $a_{ik} b_{kj}$ *or* $a_{ik} a_{op}$ *with* $(o, p) \neq (i, k)$ *or some* b_{lj} *with* $l \neq k$,

(3′) *there is some prime implicant of the other input wire which does not contain* $a_{ik} b_{kj}$.

Proof: Assume otherwise. Say, G_j does not exist. We show: If a_{ik} or $a_{ik} b_{kj}$ is prime implicant of $\text{res}_{\beta, v}$ for some wire v and it is not sole prime implicant of that wire then $\text{res}_{\beta, v}$ also has a prime implicant containing neither a_{ik} nor any b_{lj} with $l \neq k$. Call this property (A). (A) contradicts the definition of f_{ij}^J. We prove (A) by induction on the ordering of the wires which was defined in the proof of lemma 3. Let v be any wire of β. If v is an input wire then (A) holds trivially.

Assume now that v is the output wire of an or-gate, that a_{ik} or $a_{ik} b_{kj}$ is prime implicant of this wire and that v has other prime implicants. a_{ik} $(a_{ik} b_{kj})$ is prime implicant of at least one of the input wires.

Suppose it is sole prime implicant. Since there does not exist an or-gate G_j (by assumption) either (2) [(2′)] or (3) [(3′)] is not satisfied. If (3) [(3′)] is

not fulfilled then a_{ik} $(a_{ik}b_{kj})$ is subterm of every prime implicant of the other input wire and hence a_{ik} $(a_{ik}b_{kj})$ is sole prime implicant of the output wire. Thus we can eliminate this or-gate. This contradicts the minimality of β. Assume now that (3) [(3′)] is satisfied, but (2) [(2′)] is not.

Case 1: a_{ik} is sole prime implicant of the left input wire. Since (2) is not satisfied there is a prime implicant t of the right input wire containing neither a_{ik} nor any b_{lj} with $l \neq k$. t is prime implicant of the output wire.

Case 2: $a_{ik}b_{kj}$ is sole prime implicant of the left input wire and there is some prime implicant t of the right input wire containing neither $a_{ik}b_{kj}$ nor any $a_{ik}a_{op}$ with $(i,k) \neq (o,p)$ mor any b_{lj} with $l \neq k$. t is also prime implicant of the output wire. If (A) does not hold for the output wire then t contains either a_{ik} or some b_{lj} with $l \neq k$. Thus either $t = a_{ik}$ or $t = a_{ik}b_{o_1p_1} \ldots b_{o_mp_m}$ with $p_1 \neq j, \ldots, p_m \neq j$. In either case we obtain a simpler yet equivalent (by theorem II) circuit if we simply feed a_{ik} into the left input wire of the orgate under consideration.

It remains to consider the case that a_{ik} $(a_{ik}b_{kj})$ is not sole prime implicant of either input wire. The reader can verify that in this case the claim is an immediate consequence of the induction hypothesis.

Finally assume that v is the output wire of an and-gate, that a_{ik} or $a_{ik}b_{kj}$ is prime implicant of that wire and that v has other prime implicants. If a_{ik} is prime implicant of the output wire then a_{ik} is prime implicant of both input wires. Moreover, it is not sole prime implicant of either input wire (otherwise a_{ik} would be sole prime implicant of the output wire). Since (A) holds for both input wires there are prime implicants t_1 and t_2 of the left and right input wire respectively which do not contain either a_{ik} or some b_{lj} with $l \neq k$. A subterm of $t_1 t_2$ is prime implicant of the output wire.

If $a_{ik}b_{kj}$ is prime implicant of the output wire then we have to consider four cases (see proof of lemma 3).

Case 1: a_{ik} is prime implicant of the left input wire, b_{kj} is prime implicant of the right input wire. If a_{ik} is not sole prime implicant of the left input wire then the left input wire has a prime implicant t containing neither a_{ik} nor any b_{lj} with $l \neq k$. A subterm of b_{kj} t is prime implicant of the output wire. On the other hand, suppose that a_{ik} is sole prime implicant of the left input wire. Then the right input wire has prime implicants $t_1, \ldots, t_m$ different from b_{kj} (otherwise $a_{ik}b_{kj}$ would be sole prime implicant of the output wire). There is some t_1 such that $a_{ik}t_1$ is subterm of some prime implicant of some output function f_{ip}^J of β. Otherwise we could obtain a simpler, yet equivalent (by theorem I) circuit by simply feeding b_{kj} into the right input wire. Hence $t_1 = a_{ik}b_{kp}$ or $t_1 = b_{kp}$ with $p \neq j$. Note that t_1 and b_{kj} are distinct prime implicants of the right input wire. Theorem II allows us to add a_{ik} as prime implicant of the output wire. Therefore we can delete the and-gate. Contradiction!

Case 2: a_{ik} is prime implicant of the left input wire and $a_{ik}b_{kj}$ is prime implicant of the right input wire.

If $a_{ik}b_{kj}$ is sole prime implicant of the right input wire then $a_{ik}b_{kj}$ is sole prime implicant of the output wire and we have nothing to show. Otherwise there is a prime implicant t_2 of the right input wire containing neither a_{ik} nor any b_{lj} with $l \neq k$.

If a_{ik} is not sole prime implicant of the left input wire then there is a prime implicant t_1 of the left input wire containing neither a_{ik} nor any b_{lj} with $l \neq k$. A subterm of $t_1 t_2$ is prime implicant of the output wire.

Suppose now that a_{ik} is sole prime implicant of the left input wire. Let $a_{ik}b_{kj}, t_2, \ldots, t_m$ be the prime implicants of the right input wire (pairwise distinct). If directly feeding b_{kj} into the right input wire of the and-gate under consideration does not result in an equivalent circuit then (by theorem I) there has to be some $t_r (2 \leq r \leq m)$ such that $a_{ik}t_r$ is subterm of a prime implicant of some f^J_{ip}. Hence $t_r = a_{ik}$ or $t_r = b_{kp}$ or $t_r = a_{ik}b_{kp}$. If $t_r = a_{ik}$ or $t_r = b_{kp}$ or $t_r = a_{ik}b_{kp}$ and $p = j$ then either $a_{ik}b_{kj}$ is not prime implicant of the right input wire or $a_{ik}b_{kj}, t_2, \ldots, t_m$ are not pairwise distinct. In either case we derived a contradiction. Hence $p \neq j$. Then theorem II allows us to add a_{ik} as prime implicant of the output wire and thus to eliminate the and-gate under consideration. This contradicts the minimality of β.

Case 3: b_{kj} is prime implicant of the left input wire and $a_{ik}b_{kj}$ is prime implicant of the right input wire. If $a_{ik}b_{kj}$ is sole prime implicant of the right input wire then $a_{ik}b_{kj}$ is sole prime implicant of the output wire and we have nothing to show. Otherwise there is a prime implicant t_2 of the right input wire containing neither a_{ik} nor any b_{lj} with $l \neq k$. A subterm of $b_{kj}t_2$ is prime implicant of the output wire.

Case 4: $a_{ik}b_{kj}$ is prime implicant of both input wires. Argue as in case 3.

This ends the proof of lemma 6. ■

For every j, $1 \leq j \leq n$, let G_j be the or-gate whose existence is guaranteed by lemma 6.

Lemma 7: *The gates G_j are pairwise distinct.*

Proof: Assume otherwise, say $G = G_{j_1} = G_{j_2}$ with $j_1 \neq j_2$. Call G_j a type 1 or-gate if a_{ik} is sole prime implicant of either input wire and call it a type 2 or-gate if $a_{ik}b_{kj}$ is sole prime implicant of either input wire.

Case 1: G_{j_1} and G_{j_2} are both type 1 gates. If a_{ik} is sole prime implicant of both input wires then a_{ik} is sole prime implicant of the output wire and we can eliminate the or-gate G. Thus a_{ik} is prime implicant of just one input wire, say the left one. Let $t_1, \ldots, t_m$ be the prime implicants of the right input wire which do not contain a_{ik}. By property (2) each of them contains $b_{l_1 j_1}$ as well as $b_{l_2 j_2}$ for some l_1, l_2. Theorem I allows us to drop these prime implicants and thus to eliminate G. This contradicts the minimality of β.

Case 2: G_{j_1} is a type 1 gate and G_{j_2} is a type 2 gate. Property (1) implies that a_{ik} is sole prime implicant of the left input wire and that $a_{ik}b_{kj_2}$ is sole prime implicant of the right input wire. This contradicts (3).

Case 3: G_{j1} and G_{j2} are both type 2 gates. Argue as in case 2. ■

If we set $a_{ik}=0$ one input wire of the n distinct or-gates G_j, $1 \leq j \leq n$, becomes constant. Thus we can eliminate these gates. This shows

$$C_{OR}(F^{J \cup \{(i,k)\}}) \leq C_{OR}(F^J) - n.$$

This ends the proof of lemma 5 and theorem. ■■

Theorem VI: *Every monotone realization of matrix multiplication uses at least* $n^3 - n^2$ *or-gates.*

Theorem VII: $n^3 - n^2$ *or-gates are necessary and sufficient to realize boolean matrix multiplication using only or- and and-gates.*

Theorem VIII: *The "simple-minded" approach to matrix multiplication is optimal.*

VI. Uniqueness of the Optimal Circuit

In this section we will show that the simple-minded circuit for boolean matrix product is the unique optimal circuit (up to associativity and commutativity of disjunction).

Let β be any optimal circuit, i.e. β consists of n^3 and-gates and $n^3 - n^2$ or-gates. Reconsider section IV. There we eliminated and-gates by successively turning on variables. Since the order in which the variables were selected was inessential we may assume w.l.o.g. that a_{ik} was chosen first, i.e. $J_1 = \{(i,k)\}$, for any pair (i,k). Lemma 3 ensures the existence of n distinct and-gates G_j, $1 \leq j \leq n$, in β such that a_{ik} is prime implicant of one of the input wires of G_j and $a_{ik} b_{kj}$ is prime implicant of the output wire of G_j.

Lemma 8: *G_j satisfies the following conditions:*

(1) *$a_{ik} b_{kj}$ is prime implicant of the output wire of G_j,*

(2) *a_{ik} is prime implicant of the left input wire of G_j, every other prime implicant of this wire contains some b_{lj}, and if b_{lj} or $a_{i'l} b_{lj'}$ is prime implicant of that wire then $l \neq k$,*

(3) *b_{kj} is prime implicant of the right input wire of G_j, every other prime implicant of this wire contains some a_{il}, and if a_{il} or $a_{il} b_{lj'}$ is prime implicant of this wire then $l \neq k$.*

Proof: It is easy to see that for every triple (i,k,j) there has to be an and-gate that a_{ik} is prime implicant of the left input wire and b_{kj} is prime implicant of the right input wire. Setting a_{ik} to 1 eliminates all and-gates such that a_{ik} is prime implicant of one of the input wires. There can be no more than n of these, otherwise β would contain more than n^3 and-gates by the proof of theorem IV. Since the n and-gates G_j, $1 \leq j \leq n$, are pairwise distinct, setting a_{ik} to 1 eliminates exactly the gates G_j. Thus, the gate described above has to be some $G_{j'}$.

Suppose $j \neq j'$. Then $b_{kj'}$ or $a_{ik} b_{kj'}$ and b_{kj} are prime implicants of the right input wire. Theorem II ($t = \varepsilon$, $t_1 = b_{kj}$, or $a_{ik} b_{kj'}$, $t_2 = b_{kj}$) allows us to add ε as

prime implicant of the right input wire, i.e. to eliminate the gate. This contradicts the minimality of β.

Assume now that (2) is violated, i.e. either b_{kj} or $a_{i'k}b_{kj}$ is prime implicant of the left input wire. Then theorem II ($t=\varepsilon$, $t_1=a_{ik}$, $t_2=b_{kj}$ or $a_{i'k}b_{kj}$) allows us to add ε as prime implicant of the left input wire. Hence we may eliminate the and-gate.

The remaining assertions follow from the fact that matrix multiplication is completely symmetric in the a's and the b's. ■

Consider now the structure of the prime implicants of the output wire of G_j. We infer from properties (2) and (3) of G_j that $a_{ik}b_{kj}$ is the only prime implicant containing either a_{ik} or b_{kj} which can not be deleted by means of theorem I.

Suppose now that either input wire of G_j has more than one prime implicant. We may assume w.l.o.g. that it is the left one. Consider the "history" of the left input wire. At some point there has to be an or-gate such that a_{ik} is sole prime implicant of one of the input wires of this or-gate and every prime implicant of the other input wire contains some b_{lj} (note that there can be no and-gate having a_{ik} as prime implicant of both input wires). The output wire of this or-gate is connected with only one and-gate, namely G_j. Assume otherwise, say it is also connected to $G_{j'}$ with $j \neq j'$. Then every prime implicant of the output wire of the or-gate must contain some b_{lj} as well as some $b_{l'j'}$. Theorem I allows us to replace the output wire of the or-gate by the input a_{ik}. This contradicts the minimality of β. Thus the or-gate is connected with exactly one and-gate, namely G_j. We call this or-gate the left superfluous or-gate.

If b_{kj} is sole prime implicant of the right input wire of G_j then we delete the left superfluous or-gate and feed only a_{ik} into the left input of G_j.

If b_{kj} is not sole prime implicant of the right input wire of G_j then we determine analogously the right superfluous or-gate. We eliminate both superfluous or-gates by deleting the singleton inputs a_{ik} and b_{kj} respectively, we add a new and-gate computing the product $a_{ik}b_{kj}$ and we or the product $a_{ik}b_{kj}$ with the output of gate G_j. This transformation reduces the number of or-gates by 1.

In either case, the remark following the proof of lemma 8 shows that the transformed circuit is equivalent to the original one, yet is contains one less or-gate. This contradicts the minimality of β.

Thus a_{ik} is sole prime implicant of the left input wire and b_{kj} is sole prime implicant of the right input wire of gate G_j. Since (i,k) was arbitrary we conclude that then n^3 and-gates have argument pairs (a_{ik}, b_{kj}) for $1 \leq i,k,j \leq n$.

It is now easy to see that the n^3-n^2 or-gates are used to sum up the products $a_{ik}b_{kj}$ to form $\sum_{k=1}^{n} a_{ik}b_{kj}$ for $1 \leq i,j \leq n$.

Theorem IX: *The simple-minded approach to matrix multiplication yields the unique optimal circuit.*

VII. Conclusion

We prove the unique optimality of the simple-minded approach to boolean matrix multiplication in the realm of monotone computations. This result was obtained independently by Mike Paterson. Furthermore we establish several general theorems about local transformations in monotone networks. They may be a first step towards a theory of monotone computations.

References

Lamagna, Savage: Combinational Complexity of Some Monotone Functions. 15th SWAT Conference, New Orleans, 1974, 140—144.

Mehlhorn: On the Complexity of Monotone Realizations of Matrix Multiplication. Universität des Saarlandes, TR 74 – 11 (1974).

Paterson: Complexity of Monotone Networks for Boolean Matrix Product. University of Warwick, TR 2, 1974. (To appear in Theoretical Computer Science.)

Pratt: The Power of Negative Thinking in Multiplying Boolean Matrices. 6th ACM Conference on Theory of Computing, Seattle, 1974, 80—83.

Schnorr: Lower Bounds on the Complexity of Monotone Networks (unpublished memo).

Kurt Mehlhorn
Fachbereich 10
Universität des Saarlandes
D-6600 Saarbrücken
Federal Republic of Germany

Zvi Galil
Dept. of Computer Science
Cornell University
Ithaca, NY 14853
U.S.A.

Printed in Austria
Druck: Paul Gerin, A-1021 Wien