ENGINEERING RESEARCH INSTITUTE
UNIVERSITY OF MICHIGAN
ANN ARBOR

COMPLETE DECODING NETS: GENERAL THEORY AND MINIMALITY

By

ARTHUR W. BURKS
ROBERT McNAUGHTON
CARL H. POLLMAR
DON W. WARREN
JESSE B. WRIGHT

Project 1828

BURROUGHS RESEARCH CENTER
PAOLI, PENNSYLVANIA

15 July 1954

engr
UMR 1020

TABLE OF CONTENTS

COMPLETE DECODING NETS: GENERAL THEORY AND MINIMALITY[1]

1. Introduction.

A digital computing circuit is a physical device with input and output wires and which functions as follows. At any moment of time each wire assumes one of a finite number of discrete states. Moreover, the state of the output wires is determined by the past history of the states of the input wires. A telephone exchange, a digital electronic computing machine, and probably the human nervous system are all digital computing circuits. For reasons of simplicity we will consider only digital computing circuits in which every wire has one of two states, i.e., is either activated or not activated.

We will be concerned in this paper exclusively with digital computing circuits without memory or useful internal delays. Specifically, we will deal with complete decoding circuits. Such a circuit may be roughly characterized as follows. Its input wires may be partitioned into two or more sets each consisting of one or more wires. Moreover, it is such that when exactly one wire of each input set is activated exactly one of the final output wires is activated, and no two different ways of activating exactly one wire of each input set cause the same

final output wire to be activated.

The present paper falls into two parts. In the first part (consisting of the following five sections) we will define precisely three important kinds of complete decoding circuits (exponential switch circuits, tree circuits, and balanced multiplicative switch circuits) and study their interrelations. In the second part (consisting of the last section) we will establish a minimality result concerning complete decoding circuits which may be approximately stated as follows: on certain assumptions concerning the relative costs of different kinds of components, the balanced multiplicative switch circuit is the most economical type of complete decoding circuit. Some of the ideas of this paper are already employed by computer engineers; what we have done is to extend, systematize, and make rigorous what is already known in this field.

We will treat the subject matter of complete decoding circuits rigorously by using the methods of symbolic logic. This is done as follows. A digital computing circuit may be (and often is) represented by a diagram or net which gives the logical structure of the circuit in abstraction from many physical features of the circuit that are not relevant to this logical structure (e.g.,

actual voltage levels, certain time delays). These diagrams may be regarded as symbols in a non-linear language, this language can be formalized in somewhat the same manner as the languages of symbolic logic, and then theorems can be proved about this formal language. This was done in a previous paper,[2] and in the present paper we will use these results to define precisely, and to prove various theorems about, complete decoding circuits. Since certain modifications need to be made in our earlier theory of logical nets to make it most convenient for the problem at hand (in particular, we do not need delay elements in the present paper) and, since we desire that the present paper be independent of the earlier one, we will construct the theory anew.

2. The net language.

The basic symbols of our formal net language consist of brackets of various sizes, the truth constants "0" and "1", and an infinite number of primitive elements. Each element consists of an enclosure or nucleus containing a function symbol representing a two-valued truth-functional logical operation of M arguments, M inwardly directed line segments or input wires, and a single outwardly directed line segment or output wire. Some examples are: the two-input alternation ("or") element (Fig. 1a), the three-input conjunction ("and") element (Fig. 1b), the stroke element (Fig. 1c), the negation ("not") element (Fig. 1d), the material implication ("if-then") element (Fig. 1e), where "A", "K", "S", "N", and "C" represent alternation, conjunction, the stroke function, negation, and material implication respectively.[3]

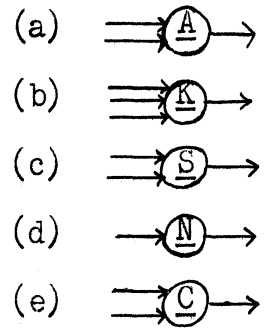


Fig. 1

The reader may find it strange to consider this system of diagrams a language. For (1) he usually thinks of a language as treating linearly arranged entities: letters are strung together to form words, words to form sentences, etc., all in one direction. And (2) he usually thinks of a language as a vehicle for expressing complete thoughts in sentences. Clearly neither of these is the case in our net language. To justify the use of the word "language" we note that as explained below the diagrams do represent objects in the same sense that some words of ordinary language (like "cat", "moon", etc.) represent objects. We also note that, although a diagram does not literally make any assertion, in effect it can tell us much about the circuit which it represents; for example, whether two wires are joined together. Also, engineers do use such diagrams to communicate with each other. The linear language introduced in Section 6 is more like ordinary language in all these respects.

The elements of our formal net language represent electronic, relay, or neural devices that perform the functions indicated by the symbols of the nuclei. Circuits constructed from these devices may be represented by symbols compounded out of the elements, as in Fig. 2.

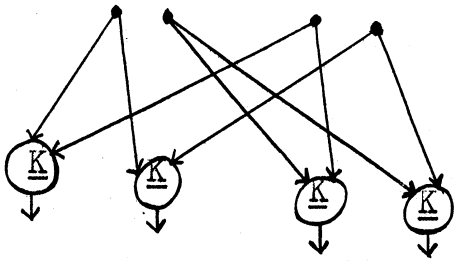


Fig. 2

Such a diagram as that of Fig. 2 is called a "plain net"; the reason for the qualification "plain" will become apparent in a moment. In some applications, such as decoding (see the next section), the inputs of a circuit are operated in groups. To represent digital computing circuits used in this way we use diagrams called "partitioned nets." The diagram of Fig. 3 is a partitioned net and represents a complete decoding circuit with two pairs of inputs and four outputs.

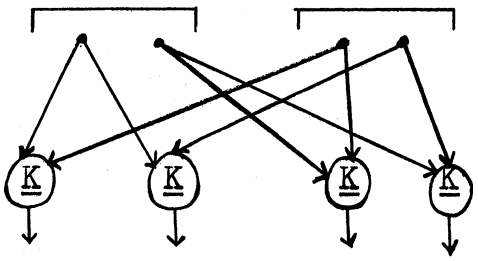


Fig. 3

Both plain nets and partitioned nets are called "nets."

It is sometimes useful to symbolize not merely a digital computing circuit but that circuit in a particular state. We do this by means of diagrams called "net states" which are derived from nets by labeling each junction with the truth constant ("0" or "1"); "0" indicates that the junction labeled is inactive, "1" that it is active. Fig. 4 is a net state of Fig. 3, and represents a complete decoding circuit with one input from each pair activated and one output activated.

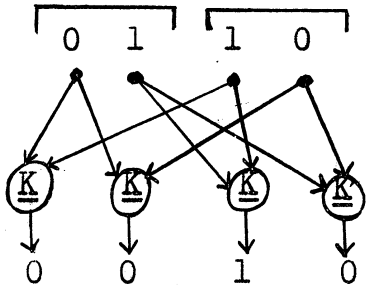


Fig. 4

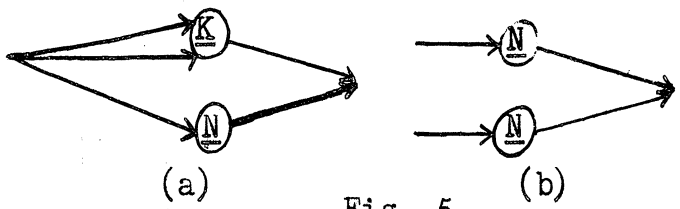
We proceed now to define formally the concepts of plain net, partitioned net, and net; the concept of net state will be defined formally in the next section.

A <u>plain net</u> consists of elements (possibly one) and a finite number of points called junctions, such that every junction is on the outer end of at least one wire of some element, and such that the outer end of any wire of

any element is on at least one junction. If there is no arrowhead at a junction, it is an input junction; otherwise, it is an output junction. An output junction to which only one wire is connected is a final output junction.

A partitioned net is a plain net plus a bracketing of all input junctions into two or more disjoint sets (called bracketed sets). A b-d net is a partitioned net with d bracketed sets each of which contains b junctions. A net is either a plain net or a partitioned net. The definitions of "junction", "input junction", "output junction", and "final output junction" are extended in the obvious way to cover partitioned nets as well as plain nets.

Not all nets are physically realizable. Fig. 5a is not, for example, for the conjunction element therein represents a component whose input and output wires are in the same state while the negation element represents a component whose input and output wires are in opposite states.



(a) (b)

Fig. 5

We will next define a certain class of nets, called "well-formed nets" ("w.f.n.") which will exclude these as well as other kinds of nets we are not interested in here, and which is sufficiently large to include representations of almost all digital computing circuits without delays.[4]

To define this class precisely we need the following notions. Any junction which is common to two or more output wires is a multiple junction. A junction α directly drives a junction β if and only if there is an element such that α is on an input wire and β is on the output wire of this element. α drives β if and only if there exists a sequence $\alpha_1,\ldots,\alpha_I$ such that α_1 is α, α_I is β, and α_i directly drives α_{i+1} for $\bigwedge_{\text{each } i}\, 1 \leqq i < I$ such that. Note that the driving relation is transitive: if α drives β and β drives γ then α drives γ. A cycle is a sequence of junctions $\alpha_0,\ldots,\alpha_{I-1}$ such that $\alpha_{i \bmod I}$ directly drives $\alpha_{(i+1) \bmod I}$, i.e., such that each junction of the cycle directly drives the next junction of the cycle.

We can now define: a net N is well-formed (w.f.) if and only if N has (1) no multiple junctions and (2) no cycles. The nets of Figs. 1, 2, 3, and 4 are w.f.n., while the nets of Fig. 5 are not w.f.n.

We will often be concerned with the following

Note: For page 10 refer to file copy.

3. Complete decoding nets.

A function that digital computing circuits are often constructed to perform is that of code conversion. We will make a few remarks about code conversion in general as a background for our discussion of complete decoding circuits.

The heart of the process of code conversion is a character by character translation or transliteration, based on a unique assignment of a character of the second code to each character of the first code. Transliterations are realized by circuits which will translate or convert any arbitrary character (sequence of digits) of the first code into the corresponding character of the second code. The code conversion of a complete message will then consist of a sequence of these character conversions plus, perhaps, other minor operations.

A standard procedure for converting a character from one code to another is first to decode and then to encode. We will illustrate this process with a conversion from one six-bit code to another. Suppose a six-bit character is represented on six input junctions going to a polarizing circuit whose output consists of six pairs of junctions, exactly one junction of each pair being activated. These six pairs of junctions are inputs to a complete decoding circuit with

sixty-four outputs, of which exactly one is activated, and such that for two different ways of activating exactly one junction of each input pair two different output junctions are activated. These sixty-four junctions in turn become the inputs to an encoding circuit with six output junctions on which are represented the converted characters.

This example makes clear the role of a complete decoding circuit in the general process of code conversion. This is only one way of making a code conversion, of course, and there may be other more direct and efficient ways, depending on the circumstances, the equipment available, etc. But this particular technique has certain advantages. The complete decoding circuit may be constructed entirely of conjunction elements (and hence may be represented by a w.f.c.n.) connected in any one of a number of systematic ways, and the encoding circuit can be constructed entirely of alternation elements (and hence may be represented by a w.f.a.n.) connected in a systematic way. Complete decoding circuits are also of interest in their own right, for in many instances a decoding is needed without a subsequent encoding.

It is evident from the preceding example that complete decoding involves the use of a circuit in a certain way: the input wires are partitioned into subsets and exactly one wire of each subset is activated. Circuits

used in this way are represented by partitioned nets; the partitioning of the input junctions of the partitioned net represents the partitioning of the input wires of the circuit. Thus the complete decoding circuit of the preceding example may be represented by a partitioned net with six bracketed sets of two junctions each. Only partitioned nets with bracketed sets of two junctions each are needed for code conversion of the type just illustrated, but we allow bracketed sets with any number of junctions in the interest of greater generality. The various states which a complete decoding circuit has when operated in this way may be symbolized by what we called "net states." We will now formally define this concept, for plain nets as well as for partitioned nets, so we may be able to symbolize the states of circuits represented by plain nets and by partitioned nets.

A net state (n.s.) of a w.f.n. N is the net N plus truth constants inserted according to the following rules:

(1a) If N is a plain net, label the input junctions of N arbitrarily with the truth constants "0" and "1".

(1b) If N is a partitioned net, label exactly one input of each bracketed set with the truth-constant "1" and label all other input junctions with the truth-constant "0".

(2) Repeat as many times as possible the following operation:

Find an element of $\underline{N}$ all of whose input wires are joined to labeled junctions. Let these labels be $\alpha_1, \ldots, \alpha_M$, respectively, counterclockwise from the output wire of the element, and let the function symbol contained in the nucleus of the element be θ^M. Label the junction to which the output wire of the element is joined with α, where $\alpha \equiv \theta^M \alpha_1 \ldots \alpha_M$.

The definition of "junction", "input junction", "output junction", and "final output junction" are extended in the obvious way to cover n.s. as well as nets. It is worth noting that we have defined n.s. only for w.f.n., and that an application of the rules of the above definition to a w.f.n. will result in a diagram in which every junction is labeled with exactly one truth-constant.[5]

We can now define precisely the kind of symbol of our net language which represents a complete decoding circuit. $\underline{N}$ is a <u>complete decoding net</u> if and only if $\underline{N}$ is a w.f. partitioned net such that for every n.s. $\underline{N}'$ of $\underline{N}$ there exists a junction (called a <u>decoding</u> junction) of $\underline{N}$ which is labeled 1 in $\underline{N}'$ and in no other n.s. of $\underline{N}$.[6] Fig. 3 is a complete decoding net.

The kind of complete decoding circuit described at the beginning of this section has the special property that the

sets into which its input wires are partitioned have the same number of elements. Complete decoding circuits of this special kind are represented by b-d decoding nets, defined as follows. N is a b-d decoding net if and only if N is a complete decoding net which is a b-d net. Fig. 3 is a 2-2 decoding net.

Let N be a b-d decoding net and let K be a set of junctions of N obtained by selecting for each n.s. N' of N exactly one decoding junction which is labeled 1 in N'. The set K has b^d elements, since there are b^d n.s. of any b-d net. Each of these n.s. can be represented by a d-digit base-b number, of which the i'th digit tells which of the b junctions of the i'th bracketed set is activated. This shows the relation of a b-d decoding net to a type of code conversion more general than that of the example given earlier in this section.

We are usually interested in b-d decoding nets in which the elements of K are all the final output junctions of the net. However, we do not require that the elements of K be final output junctions nor that a b-d decoding net have exactly b^d final output junctions. We thus broaden the scope of some of our theorems without complicating their proofs. One could also allow for extra

input junctions, but here the complexity introduced does not seem warranted by the gain in generality.

4. Multiplicative switch nets.

We will study three different species of complete decoding circuits (exponential switch circuits, tree circuits, and balanced multiplicative switch circuits) that are in common use. These turn out on careful analysis to belong to the same genus, that of multiplicative switch circuits. In this section we will define precisely a class of nets (called "multiplicative switch nets") which represent circuits of this genus and then define classes of nets (called "exponential switches", "trees", and "balanced multiplicative switch nets", respectively) representing these three species of this genus in such a way as to bring out their interrelations.

Multiplicative switch nets are compounds of multiplicative switches, so we will define this concept first.

The $i_1, i_2, \ldots, i_d$ multiplicative switch (m.s.) is the partitioned net consisting of

(1) $i_1 + i_2 + \ldots + i_d$ input junctions, bracketed into d sets of $i_1, \ldots, i_d$ members respectively;

(2) $i_1 \cdot i_2 \cdot \ldots \cdot i_d$ output junctions;

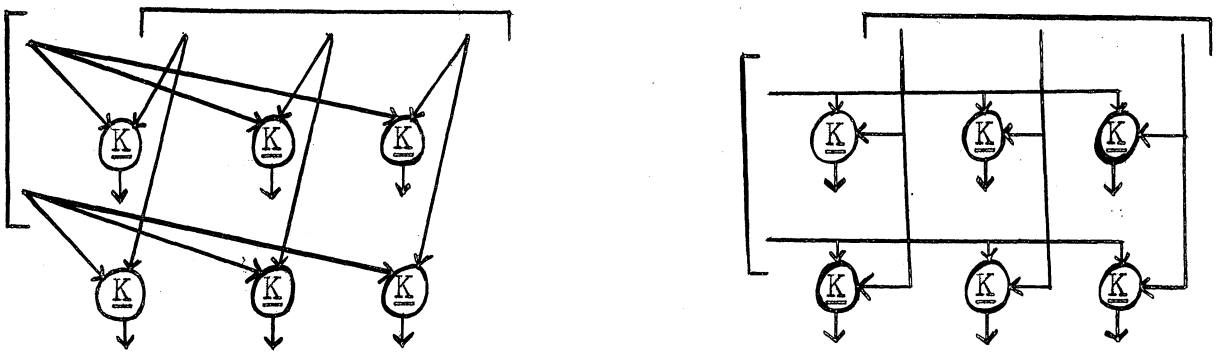
(3) $i_1 \cdot i_2 \cdot \ldots \cdot i_d$ elements, each of which is a d-input conjunction;

The connection of the wires of the elements to the junctions can be described as follows:

(4) By "representative set", we mean a set of d input junctions, one from each bracketed set; there are $i_1 \cdot \ldots \cdot i_d$ representative sets. Each representative set is the set of input junctions on the input wires of exactly one element.

(5) Every output junction is on the output wire of exactly one element.[7]

Note that an m.s. is a w.f.c.n. and also a complete decoding net. Fig. 3 is an m.s., as is also Fig. 6a.



(a) (b)

Fig. 6

The structure of the m.s. of Fig. 6a is more clearly shown in Fig. 6b. Fig. 6b is a modification of Fig. 6a. We will hereafter modify m.s. in a similar way when it is convenient to do so.

Roughly, an "m.s. net" is composed of m.s.

More precisely, a m.s. net (m.s.n.) is defined recursively as follows:

(1) An m.s. is an m.s.n.

(2) Assume $\underline{N}_1$ and $\underline{N}_2$ are disjoint m.s.n. with $\Gamma_1, \ldots, \Gamma_J$ the final output junctions of $\underline{N}_1$ and $\Delta_1, \ldots, \Delta_J$ a bracketed set of $\underline{N}_2$. Then the result of dropping the bracket around $\Delta_1, \ldots, \Delta_J$ and joining (i.e., identifying) Γ_1 to $\Delta_1, \ldots, \Gamma_J$ to Δ_J is an m.s.n.

Fig. 7 is an m.s.n. composed of two m.s.

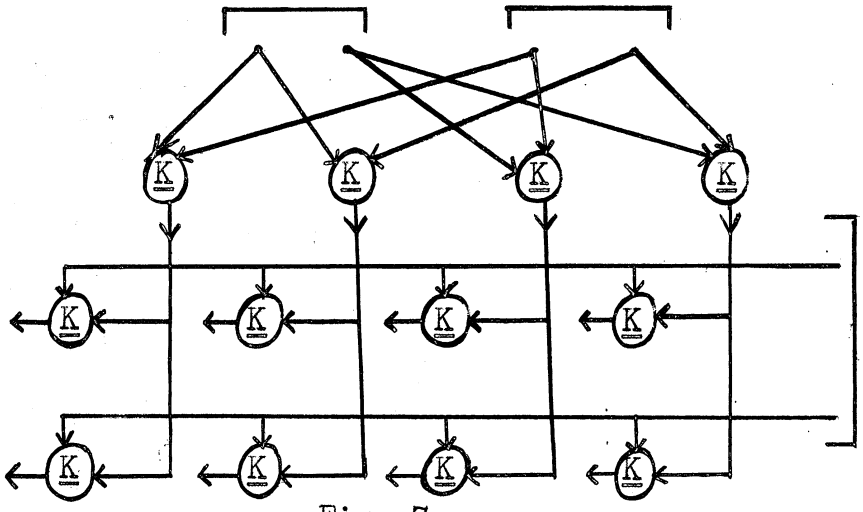


Fig. 7

An m.s.n. is a complete decoding net and hence an m.s.n. with $\underline{d}$ bracketed sets of $\underline{b}$ junctions each is a $\underline{b}$-$\underline{d}$ decoding net. An m.s.n. is a w.f.c.n.

It is obvious that an m.s.n. has sufficient regularity of structure to permit abbreviation of it. An m.s. may be abbreviated by a single symbol as follows.

An $i_1, \ldots, i_d$ m.s. symbol consists of a box with d input cables labeled $i_1, \ldots, i_d$ and a single output cable labeled $i_1 \cdot \ldots \cdot i_d$. It abbreviates an $i_1, \ldots, i_d$ m.s. Thus Fig. 8 abbreviates Fig. 3.

Fig. 8

Fig. 9 abbreviates Fig. 6.

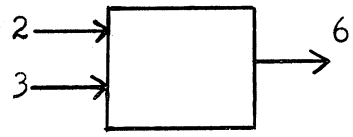


Fig. 9

An m.s. net diagram is defined recursively as follows:

(1) An m.s. symbol is an m.s. net diagram. An input (output) cable of the m.s. symbol is a net input (output) cable.

(2) If M_1 is an m.s. net diagram with a net output cable labeled x and M_2 is an m.s. net diagram with a net input cable labeled x, the result of joining the two cables is an m.s. net diagram. An input (output) cable of any m.s. symbol is a net input (output) cable if and only

if it is not joined to the output (input) cable of another m.s. symbol.

For example, in Fig. 10 below there are three input cables each labeled 2, and two output cables one labeled 4, the other 8.

It is clear that every m.s.n. is abbreviated by exactly one m.s. net diagram and that every m.s. net diagram abbreviates exactly one m.s.n. Thus Fig. 10 abbreviates Fig. 7.

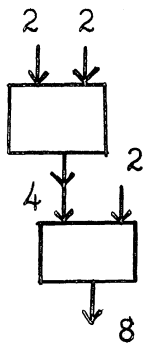


Fig. 10

A b-d m.s. net diagram is an m.s. net diagram having d net input cables each labeled b.

We are now prepared to define exponential switches, trees, and balanced multiplicative switch nets in terms of m.s.n. In what follows we shall not always make a clear distinction between the m.s.n. and the m.s. net diagram.

A b-d exponential switch is an m.s. symbol with d input cables each labeled b or a net which is abbreviated by such. Fig. 3 is a 2-2 exponential switch

while Fig. 11 is a 3-5 exponential switch.

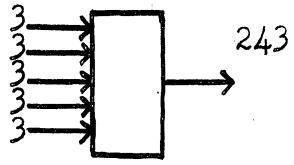


Fig. 11

The $\underline{b}$-$\underline{d}$ tree is the m.s. net diagram (or the net abbreviated by it) composed of a $\underline{b},\underline{b}$ m.s. symbol, a $\underline{b}^2,\underline{b}$ m.s. symbol, ..., and a $\underline{b}^{d-1},\underline{b}$ m.s. symbol; these $\underline{d}$-1 m.s. symbols are connected in the following way: for every $\underline{j}$, $\underline{j}$ = 1, ..., $\underline{d}$-2, the output cable having the same label of the $\underline{b}^j,\underline{b}$ m.s. symbol is joined to the input cable of the $\underline{b}^{j+1},\underline{b}$ m.s. symbol. Fig. 10 is a 2-3 tree while Fig. 12 is a 3-5 tree.

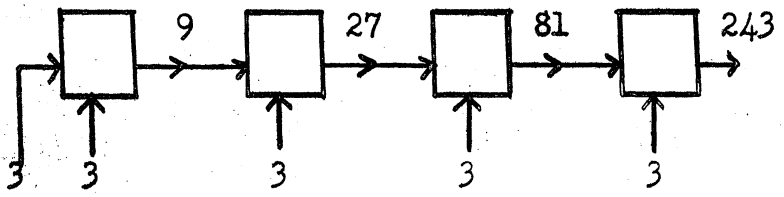


Fig. 12

It should be noted that the kind of tree defined here is the so-called standard tree as contrasted to a folded tree. A folded tree, while composed of m.s., is not an m.s.n.[8]

We now define "b-d balanced m.s.n." To do so, we make use of the commonly used notation " $[x]$ ", which means the integral part of x, i.e., the greatest integer not greater than x. The b-d balanced m.s.n. is the m.s. net diagram (or the net which it abbreviates) which can be constructed as follows:

(1) A $b^{[d/2]}, b^{[(d+1)/2]}$ m.s. symbol is drawn.

(2) The following step is iterated until all net input cables are labeled b: where there is an input cable labeled b^x, a $b^{[x/2]}, b^{[(x+1)/2]}$ m.s. symbol is drawn, its output cable joined to that input cable.[9]

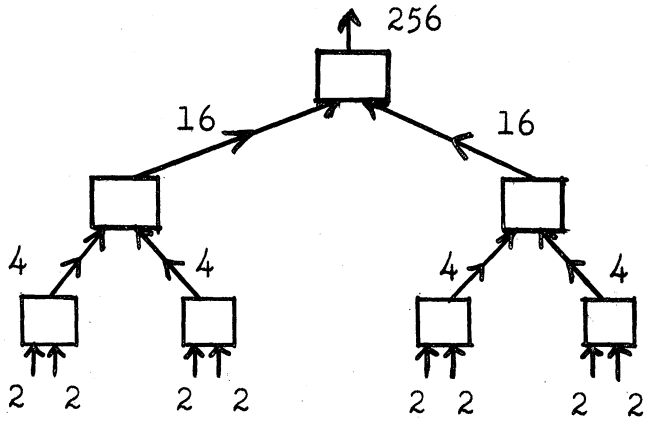


Fig. 13

Fig. 13 is a 2-8 balanced m.s.n. while Fig. 14 is a 3-5 balanced m.s.n. It should be noted that for each b and d there is a unique balanced m.s.n.

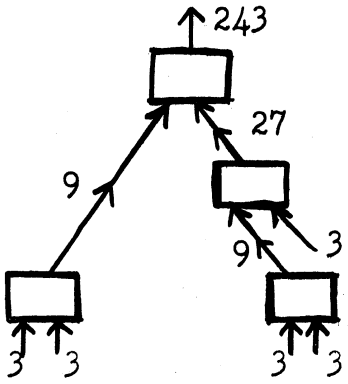


Fig. 14

b-d exponential switches, b-d trees, and b-d balanced m.s.n. are all b-d decoding nets, and hence they all represent complete decoding circuits. (Thus Figs. 11, 12, and 14 are all 3-5 decoding nets.) These three kinds of b-d decoding nets are of particular interest because of the regularity and simplicity of their structure.

note: for page 25 refer to file copy

practical engineering skill and judgment. There are nevertheless problems of pure logic capable of precise formulation and of rigorous solution which have an important practical bearing on these engineering problems. We will formulate and partially solve one such class of problems. The element-input count of a net N is the number of (element) input wires of N, i.e., the number of input wires of each element summed over all the elements of N. A net N is minimal in a given class of nets if and only if the element-input count of N is as small as the element-input count of every other net of that class. The minimality problems we are interested in are those of the form: For any class of b-d decoding nets, to find a net minimal in this class.

For the most part, this problem is quite trivial in any case where $b = 1$. For the remainder of this paper, therefore, we shall consider only cases in which $b > 1$.

We will make a few remarks to indicate the bearing of these theoretical minimality problems on practical circuit design. Certainly one factor in the cost of a complete decoding circuit is measured by the element-input count of the corresponding net. In some instances

this is a dominating factor. For example, for a circuit constructed out of crystal rectifiers connected to form logical conjunctions, the element-input count is equal to the number of rectifiers.[10] Further, in this case two cost factors which are difficult to compare need not be compared, for, as we shall show later, the b-d decoding net minimal in the class of b-d decoding w.f.c.n. is very uniform and simple in structure, and there is no problem of comparing it with a slightly more uniform net having a few more elements.

We will consider now the element-input counts of our three fundamental kinds of m.s.n. Let $C_E(b,d)$, $C_T(b,d)$, $C_B(b,d)$ be the element-input counts of b-d exponential switches, b-d trees, and b-d balanced m.s.n., respectively. These functions are defined only for b,d natural numbers ≥ 2.

(1) $C_E(b,d) = db^d$

(2) $C_T(b,d) = \frac{2b^{d+1} - 2b^2}{b - 1}$

We know of no exact, simple formula for $C_B(b,d)$, but (3) is a simple recursive formula (derived from the rules of construction for a b-d balanced m.s.n.) giving the values of this function; (3') is a non-recursive

formula for the case where $\underline{d} = 2^k$.

(3) $C_B(\underline{b},2) = 2\underline{b}^2$

$C_B(\underline{b},3) = 2\underline{b}^3 + 2\underline{b}^2$

$C_B(\underline{b},\underline{d}) = 2\underline{b}^{\underline{d}} + C_B(\underline{b}, [\underline{d}/2]) + C_B(\underline{b}, [(\underline{d}+1)/2])$ $(\underline{d} \geq 4)$,

where '$[\underline{x}]$' means 'the integral part of $\underline{x}$' as before.

(3') $C_B(\underline{b},2^k) = 2 \cdot \sum_{j=1}^{k} 2^{k-j}\underline{b}^{2^j}$.

The following minimality results may be established by algebraic and inductive proofs from these formulas. For $\underline{d} = 2$ and for $\underline{d} = 3$, $\underline{b} = 2$: $C_T(\underline{b},\underline{d}) = C_E(\underline{b},\underline{d})$. In the first case ($\underline{d} = 2$) this equality holds because a $\underline{b}$-2 tree is a $\underline{b}$-2 exponential switch. For other values of $\underline{b},\underline{d}$ (i.e., for $\underline{d} = 3$ and $\underline{b} > 2$, and for $\underline{d} > 3$) $C_T(\underline{b},\underline{d}) < C_E(\underline{b},\underline{d})$. For $\underline{d} = 2,3$ $C_B(\underline{b},\underline{d}) = C_T(\underline{b},\underline{d})$, because a $\underline{b}$-2 balanced m.s.n. is a $\underline{b}$-2 tree and a $\underline{b}$-3 balanced m.s.n. is a $\underline{b}$-3 tree. For other values of $\underline{d}$ $(\underline{d} \geq 4)$, $C_B(\underline{b},\underline{d}) < C_T(\underline{b},\underline{d})$. We can summarize these results by saying that a $\underline{b}$-$\underline{d}$ balanced m.s.n. is as small (in the sense of having as small an element-input count) as a $\underline{b}$-$\underline{d}$ tree, and in general is smaller, and a $\underline{b}$-$\underline{d}$ tree is as small as a $\underline{b}$-$\underline{d}$ exponential switch, and in general is smaller. The actual element-input count for each kind of net for $\underline{b},\underline{d} \leq 5$ is given in the following tables. We have stopped the development of the table for $C_T(\underline{b},\underline{d})$ as soon as the values become larger than those of $C_B(\underline{b},\underline{d})$, and similarly for $C_E(\underline{b},\underline{d})$ with respect to $C_T(\underline{b},\underline{d})$.

Values of $C_B(b,d)$

b \ d	2	3	4	5
2	8	24	48	96
3	18	72	198	576
4	32	160	576	2240
5	50	300	1350	6600

Table 1

Values of $C_T(b,d)$

b \ d	2	3	4	5
2	8	24	56	
3	18	72	234	
4	32	160	672	
5	50	300		

Table 2

Values of $C_E(b,d)$

b \ d	2	3	4	5
2	8	24	64	
3	18	81		
4	32			
5	50			

Table 3

Though we know of no exact, non-recursive formula for $C_B(b,d)$, it follows from the results of the preceding paragraph that

$$(4)\quad C_B(b,d) = \frac{2b^{d+1} - 2b^2}{b-1} \quad \text{for } d = 2,3.$$

Approximate information concerning $C_B(b,d)$ for larger values of d $(d > 3)$ may be gained from the following formulas. An upper bound on $C_B(b,d)$ may be found by replacing the two subnets driving the final m.s. of a b-d balanced m.s.n. by a b-$\lceil d/2 \rceil$ tree and a b-$\lceil (d+1)/2 \rceil$ tree. Hence

$$(5)\quad C_B(b,d) \leq 2b^d + \left(\frac{2b}{b-1}\right)\left(b^{\lceil d/2 \rceil} + b^{\lceil (d-1)/2 \rceil}\right) - \frac{4b^2}{b-1},$$

for $d > 3$.

A lower bound on $C_B(b,d)$ may be found by deleting all of the b-d balanced m.s.n. except the final three m.s. Hence

$$(6)\quad C_B(b,d) \geq 2b^d + 2\left(b^{\lceil d/2 \rceil} + b^{\lceil (d+1)/2 \rceil}\right) \quad \text{for } d > 3.$$

An approximate formula for $C_B(b,d)$ useful for $d > 3$ may be derived from (5) and (6) by deleting the last term of (5) and averaging the result with (6).

$$(7)\quad C_B(b,d) = 2b^d + \left(\frac{2b-1}{b-1}\right)\left(b^{\lceil d/2 \rceil} + b^{\lceil (d+1)/2 \rceil}\right)$$

for $d > 3$.

Except for some of the values of $C_B(b,d)$ given in Table 1 (and except for $d \leq 3$, of course), this formula is accurate to better than 0.6 of one per cent.

The decoded information produced by a b-d decoding net appears on b^d output junctions. A useful criterion of the "cost" of a b-d decoding net is the element-input count per used output of a b-d decoding net, defined as the element-input count divided by b^d. Let $R_E(b,d)$, $R_T(b,d)$, and $R_B(b,d)$ be these ratios for a b-d exponential switch, a b-d tree, and a b-d balanced m.s.n., respectively. The following relations may be easily established:

(8) $R_E(b,d) = d$

(9) $2 \leq R_T(b,d) < 2 + \frac{2}{b-1} \leq 4$

(10) $2 \leq R_B(b,d) \leq 3$.

It may be shown that the element-input count per output of any b-d decoding net is at least 2. The superiority of a b-d balanced m.s.n. over the other two types of b-d decoding nets is indicated by the fact that this value of 2 is approached by a balanced m.s.n. as it becomes larger and larger, but not in general by an exponential switch or tree:

(11) $\underset{d \to \infty}{\text{Limit}}\ R_T(b,d) = 2 + \frac{2}{b-1}$

(12) $\lim_{d\to\infty} R_B(b,d) = 2$.

The latter may be proved from (5).[11]

The most important result established in this section so far is the minimality of a b-d balanced m.s.n. in the class composed of a b-d balanced m.s.n., a b-d tree, and a b-d exponential switch. This leads naturally to the question: What is the broadest class of b-d decoding nets with respect to which a b-d balanced m.s.n. is minimal? We are able to prove that a b-d balanced m.s.n. is minimal in the class of conjunction b-d decoding nets.[12] The proof of this result involves some new concepts of interest in their own right, so it will be postponed to the final section. We also offer the following conjecture:

CONJECTURE: A b-d balanced m.s.n. is minimal in the class of b-d decoding nets.

It will be recalled that a b-d decoding net was defined to be a w.f.n. It is possible to extend this definition to include any net, whether w.f. or not, which performs essentially the same decoding function, and we conjecture that a b-d balanced m.s.n. is minimal even in this wider class of nets.

6. The equation language.

Any ordinary language is linear in the sense that its symbols normally occur in a linear order; in contrast, our formal net language is non-linear: its symbols normally require a two-dimensional medium. There are many cases in which the net language permits the simplest and most satisfactory description of a digital computing circuit. On the other hand, there are some situations where a precise linear language is more suitable.[13] In the present section we will introduce such a language, called the equation language.

The equation language is equal in expressive power to the net language, that is, the same information can be expressed in each, and every net or net state may be translated into a suitable expression of the equation language and vice-versa. We could give a formal definition of the equation language, and then prove this intertranslatability as a theorem, but in the interest of brevity we will give only an informal exposition of the equation language, usually proceeding by showing how to translate an expression of the net language into it, and usually working in terms of examples rather than general principles.

A plain net N of the net language may be translated into a set of equations of the equation language

as follows. Associate with each junction of $\underline{N}$ a distinct propositional variable; these variables may also be regarded as associated with the wires terminating in the junctions. Consider an element with function symbol θ (of degree $\underline{M}$), variables $\alpha_1, \ldots, \alpha_M$ associated with its input wires (numbered sequentially in a counterclockwise direction from the output wire), and variable β associated with its output wire. For this element write the equation $\beta \equiv \theta\alpha_1 \ldots \alpha_M$.[14] For example, Fig. 2 translates into the set of equations "$\underline{q}_1 \equiv \underline{K}\underline{p}_1\underline{p}_3$, $\underline{q}_2 \equiv \underline{K}\underline{p}_1\underline{p}_4$, $\underline{q}_3 \equiv \underline{K}\underline{p}_2\underline{p}_3$, $\underline{q}_4 \equiv \underline{K}\underline{p}_2\underline{p}_4$"; see Fig. 15.

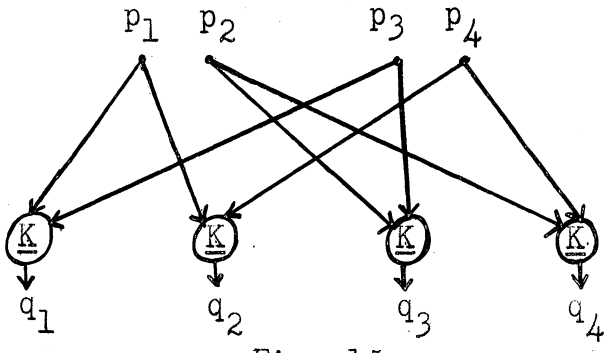


Fig. 15

A few comments on this notation are in order. The equivalence sign "$\equiv$" means the same as "if and only if." To the right of this sign we write the function symbol found in the nucleus followed by the

variables associated with the input wires in the proper order. We use here the symbolism of Łukasiewicz' parenthesis-free notation for the propositional calculus: for example, "Apq" for "p or q", "Apqr" for "p or q or r", "Kpq" for "p and q", "Kpqr" for "p and q and r", "Np" for "not-p", "Cpq" for "if p then q", and "Spq" for "either not-p or not-q." Thus "$q \equiv Kp_1p_2$" means that the junction associated with "q" is activated if and only if the two junctions associated with "p_1" and "p_2" are both activated. Note finally that Fig. 16 translates into "$q \equiv Cp_1p_2$" and not into "$q \equiv Cp_2p_1$."

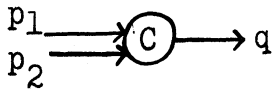


Fig. 16

In the case of a completely commutative function symbol θ the order in which the input variables follow the symbol doesn't matter (e.g., "Apq" is equivalent to "Aqp") but in other cases it does (e.g., "Cp_1p_2" is not equivalent to "Cp_2p_1"), and to be uniform we always follow the function symbol by the variables associated with the input wires of the element taken in a counterclockwise direction from the output wire.

Partitioned nets are translated into sets of equations by using both variables with subscripts and variables with superscripts and subscripts in the way indicated by the following example. Fig. 3 translates into "$\underline{q}_1 \equiv \underline{K}\underline{p}_1^1\underline{p}_2^1$, $\underline{q}_2 \equiv \underline{K}\underline{p}_1^1\underline{p}_2^2$, $\underline{q}_3 \equiv \underline{K}\underline{p}_1^2\underline{p}_2^1$, $\underline{q}_4 \equiv \underline{K}\underline{p}_1^2\underline{p}_2^2$"; see Fig. 17.

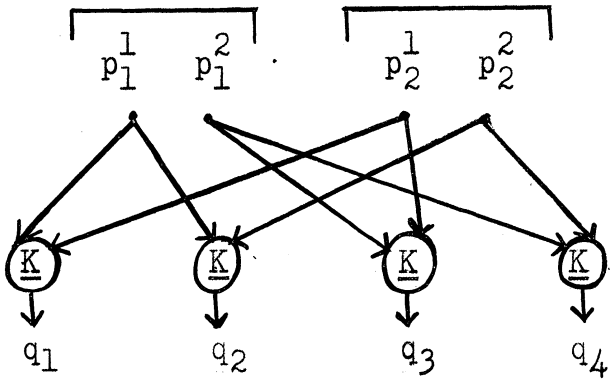


Fig. 17

Note that "$\underline{p}_y^x$" is associated with the $\underline{x}$'th member of the $\underline{y}$'th bracketed set.

A net state $\underline{N}'$ of a net $\underline{N}$ may be translated into a set of equations as follows. Translate $\underline{N}$ into a set of equations according to previous instructions. Then for each variable α associated with an input junction of $\underline{N}$, add "$\alpha \equiv 0$" or "$\alpha \equiv 1$" to this set of equations accordingly as the junction associated with α is labeled "1" or "0" in $\underline{N}'$. Thus Fig. 4 translates into "$\underline{q}_1 \equiv \underline{K}\underline{p}_1^1\underline{p}_2^1$, $\underline{q}_2 \equiv \underline{K}\underline{p}_1^1\underline{p}_2^2$, $\underline{q}_3 \equiv \underline{K}\underline{p}_1^2\underline{p}_2^1$, $\underline{q}_4 \equiv \underline{K}\underline{p}_1^2\underline{p}_2^2$, $\underline{p}_1^1 \equiv 0$, $\underline{p}_1^2 \equiv 1$, $\underline{p}_2^1 \equiv 0$, $\underline{p}_2^2 \equiv 0$."

An equation expresses the state of an output junction in terms of the states of the junctions which directly drive it. In some cases it is useful to explain the state of an output junction in terms of the states of the input junctions (of the net) which drive it. This may be accomplished for w.f.n. as follows. Let N be a w.f.n. First, associate a distinct variable with each input junction of N; if N is a partitioned net these variables must have subscripts and superscripts as indicated in the last paragraph. Second, iterate the following step as long as possible: if Γ is an output junction of N on an output wire of an element with function symbol Θ of degree M and directly driven by junctions to which are associated formulas $\Lambda_1, \ldots, \Lambda_M$ (taken in a counterclockwise order), associate $\Theta\Lambda_1 \ldots \Lambda_M$ to Γ. The junctions of the net of Fig. 18 are labeled with formulas associated with them by these rules.

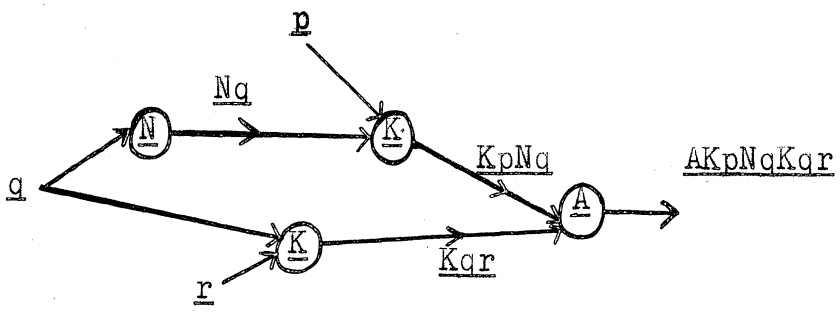


Fig. 18

If these rules are applied to a w.f.n. $\underline{N}$, every junction of $\underline{N}$ will have a formula associated with it.[15] The net may then be represented by the set of formulas associated with its final output junctions. We regard such a set of formulas as an abbreviation for the set of equations into which the net translates. Thus "$\underline{AKpNqKqr}$" abbreviates the set of equations into which Fig. 18 translates, i.e., "$\underline{r}_1 \equiv \underline{Nq}$, $\underline{r}_2 \equiv \underline{Kqr}$, $\underline{r}_3 \equiv \underline{Kpr}_1$, $\underline{r}_4 \equiv \underline{Ar}_3\underline{r}_2$."[16]

One may specify the desired structure of a net by specifying a set of formulas. In this case we must proceed in the opposite direction. For example, given "$\underline{AKpNqKqr}$" we would construct the net of Fig. 18 by starting with the input junctions associated with "$\underline{p}$", "$\underline{q}$", and "$\underline{r}$" and proceeding through the formulas "$\underline{Nq}$", "$\underline{KpNq}$", "$\underline{Kqr}$", "$\underline{AKpNqKqr}$" in that order. Similarly, "$\underline{KKp_1^1p_2^1p_3^1}$, $\underline{KKp_1^1p_2^1p_3^2}$, $\underline{KKp_1^1p_2^2p_3^1}$, $\underline{KKp_1^1p_2^2p_3^2}$, $\underline{KKp_1^2p_2^1p_3^1}$, $\underline{KKp_1^2p_2^1p_3^2}$, $\underline{KKp_1^2p_2^2p_3^1}$, $\underline{KKp_1^2p_2^2p_3^2}$" translates into Fig. 10. It should be noted that such a set of formulas always translates into a w.f.n.

The last mentioned set of formulas is very systematic in structure and consequently permits even further abbreviation. If we denote the three net input cables of Fig. 10 by "$\underset{\sim}{p}_1^2$", "$\underset{\sim}{p}_2^2$", and "$\underset{\sim}{p}_3^2$", where the superscript

"2" indicates the degree of the cable, and use "x" to denote the operation of combining cables in an m.s., then Fig. 10 translates into "$((P^2_1 x P^2_2) x P^2_3)$." We shall call such formulas construction formulas and regard them as abbreviations of sets of equations. For example, "$((P^2_1 x P^2_2) x P^2_3)$" abbreviates the set of equations into which Figs. 10 and 7 translate. Every set of equations which is a translation of an m.s.n. may be so abbreviated. It will be noted that construction formulas abbreviate sets of equations in very much the same way as m.s. net diagrams abbreviate m.s.n.

Construction formulas show the structure of our three basic kinds of m.s.n. very clearly. A construction formula for a b-d m.s. net diagram contains d "P"'s each with a different subscript and each with a superscript b. A construction formula for a b-d exponential switch contains only one pair of parentheses; e.g., "$(P^3_1 x P^3_2 x P^3_3 x P^3_4 x P^3_5)$" represents the 3-5 exponential switch of Fig. 11. Every construction formula for a b-d m.s. net diagram in which all the left-hand parentheses are together and each "P" except the first is followed by a single right-hand parenthesis represents a b-d tree, and every b-d tree may be so represented; e.g., "$((((P^3_1 x P^3_2) x P^3_3) x P^3_4) x P^3_5)$" represents the 3-5 tree of

Fig. 12. The construction formulas for b-d balanced m.s.n. may be characterized by the following conditions:

(1) It is a construction formula which represents a b-d m.s. net diagram.

(2) Every occurrence of the operator "x" in it has exactly two arguments; moreover the number of "P"'s in the first and the number of "P"'s in the second are equal or differ by one.

"$(((P_1^2 x P_2^2) x (P_3^2 x P_4^2)) x ((P_5^2 x P_6^2) x (P_7^2 x P_8^2)))$" is the construction formula for the 2-8 balanced m.s.n. of Fig. 14, while "$((P_1^3 x P_2^3) x ((P_3^3 x P_4^3) x P_5^3))$" is the construction formula for the 3-5 balanced m.s.n. of Fig. 14.

7. A proof of the minimality theorem.

In this section we shall prove that the b-d balanced m.s.n. is minimal in the class of all conjunction b-d decoding nets. It turns out that there are other minimal nets in this class; these are, however, only slight variations of the b-d balanced m.s.n. We shall define in Subsection 7.2 "quasi balanced m.s.n." denoting balanced m.s.n. and these other nets. In Subsection 7.3 we shall prove (in the main theorem) that a necessary and sufficient condition that a net be minimal in the aforementioned class is that it be a quasi balanced m.s.n. It therefore follows as a corollary that the balanced m.s.n. is minimal.

Before introducing these nets, we introduce a few helpful concepts and prove some theorems about them. It is hoped that some of these concepts, such as that of prorated cost, level, potential cost, overlap (Subsection 7.1), and quasi balanced configuration (Subsection 7.2), may be useful in the solution of other minimality problems. One such problem might be to find a conjunction b-d net with a minimal element input count serving some purpose other than that of complete decoding. For example,

we might want a net with only a certain subset of the decoding junctions as its final output junctions.

7.1. Some basic concepts.

We make use of a reduction procedure, applied to conjunction b-d decoding nets. By considering only nets to which this procedure has been applied, we limit the class somewhat, but we do not thereby exclude any minimal nets; the result is that our problem of minimality is simplified. The reduction procedure consists of steps which are applications of the following rules in any order and repeated any number of times.

Rule 1. If two junctions α and β (β not driving α) are driven by exactly the same input junctions then delete β and the element whose output wire is on β, reconnecting any input wires on β to α.

Rule 2. Delete a final output junction and the element whose output wire is on it if that junction is not a decoding junction.

Rule 3. If an element has two input wires on the same junction, then delete one of the wires (thus changing an n input conjunction to an n-1

input conjunction).

An irreducible b-d net is a conjunction b-d decoding net which cannot be reduced by any of the above rules. We prove the following.

(A) No two junctions of an irreducible net are driven by exactly the same input junctions. If there were two such junctions, at least one of them would not drive the other, for otherwise the net would contain a cycle and not be well-formed. Therefore, Rule 1 is applicable and the net is not irreducible.

(B) No junction is driven by two distinct members of the same bracketed set. Such a junction could not drive a decoding junction. It would, therefore, either be a final output junction or drive a non-decoding final output junction. In either case, Rule 2 could be applied.

(C) The element input count is decreased by an application of Rule 1, 2, or 3 to a conjunction b-d decoding net and the resulting net is also a conjunction b-d decoding net. This is obvious from the rules.

(D) Any net which is minimal in the above-mentioned class is irreducible. This follows directly from (C).

In what follows, because we are interested in minimality, we shall, in the light of (D), consider only irreducible nets. We shall sometimes make use of (A), (B), (C), and (D) without explicit reference to them.

We say that an (element) input wire W directly drives a junction α if W is part of an element whose output wire is on α. We say that an input wire W drives a junction α if the junction β directly driven by W either is α or drives α. The use of an input wire W, u(W), is the number of final output junctions driven by W. It is convenient to prorate the cost of a net C(N) among the final output junctions of the net. To this end, where α is a final output junction, the prorated cost c(α) is the sum of the reciprocals of the uses of all the wires driving α. Note the 3-2 net, a portion of which is shown in Fig. 19. W (W') drives the final output junctions β and γ. The use of W (W') is 2. The use of every other input wire shown is 1. The prorated cost of β(γ) is therefore 3.

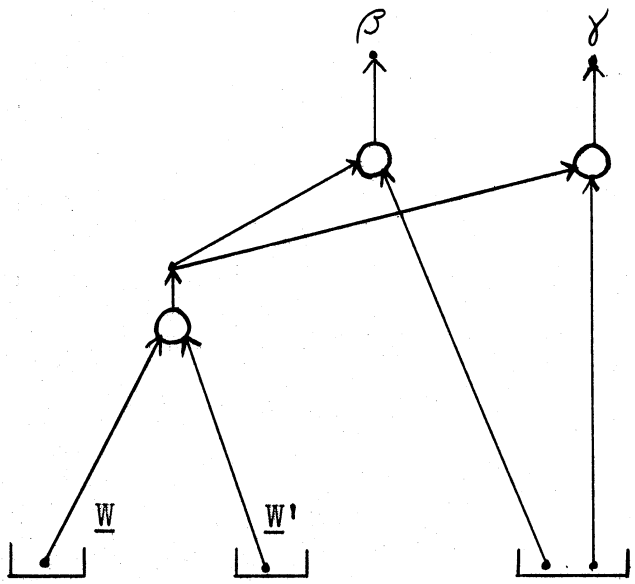


Fig. 19

THEOREM 1. $\underline{C}(\underline{N}) = \sum \underline{c}(\alpha)$, summing over all the final output junctions α in $\underline{N}$.

Proof: $\underline{C}(\underline{N})$ is the number of input wires in $\underline{N}$. Consider any input wire $\underline{W}$ in $\underline{N}$. There are $\underline{u}(\underline{W})$ final output junctions driven by $\underline{W}$. To the prorated cost of each of these $\underline{W}$ contributes $\frac{1}{u(W)}$, and to the prorated cost of all the other final output junctions $\underline{W}$ contributes nothing. Therefore, to the sum of the prorated costs of all final output junctions $\underline{W}$ contributes exactly 1, which is what it contributes

to $\underline{C}(\underline{N})$. Since this is true of all input wires $\underline{W}$, Theorem I follows.

A junction α *represents* a bracketed set when some net input junction of the bracketed set either drives, or is identical to, α. The *level* of a junction is the number of bracketed sets represented by the junction (i.e., assuming the net is irreducible, the number of input junctions driving it). The *configuration* of a junction α is the subnet consisting of α itself and all junctions driving α, together with all elements whose output wires are on such junctions. Thus, all input wires driving α will be in the configuration of α; and, conversely, any input wire in the configuration of α drives α. Any junction β in the configuration of α which is not α itself is of lower level than α; any bracketed set represented by β is also represented by α.

Corresponding to the concepts of *driving*, *use*, and *prorated cost*, we have the concepts of *potentially driving*, *potential use*, and *potential cost*. We say that a junction α *potentially drives* a junction β, when α is not β and every input junction driving α

also drives β. We say that an input wire W potentially drives a junction β, if the junction α directly driven by W either is β or potentially drives β. The potential use of an input wire W is the number of final output junctions potentially driven by W. We shall use the fact that, in an irreducible b-d net, the potential use of a wire directly driving a junction of level i is b^{d-i}. The potential cost of a final output junction α is equal to the sum of the reciprocals of the potential uses of all wires driving α. Note the portion of a 3-2 net in Fig. 20. The junction α and the wires W and W' potentially drive β and γ, but drive only γ. The use of W (W') is 1 while the potential use is 2. The prorated cost of γ is 4; the potential cost of γ is 3.

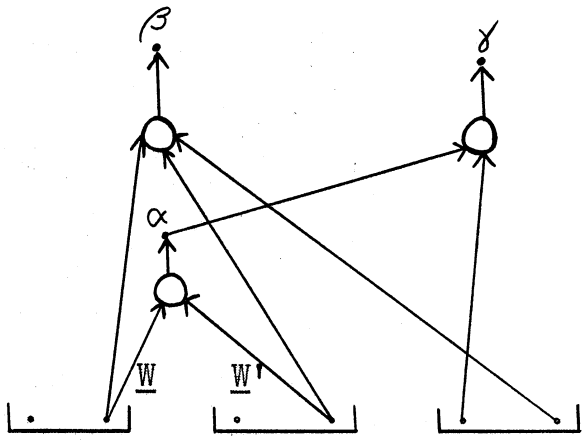


Fig. 20

THEOREM II. The prorated cost of a final output junction α equals the potential cost of α if and only if every wire in the configuration of α drives all the final output junctions which it potentially drives. Otherwise the prorated cost is greater.

The proof of this is obvious from the definitions of the concepts involved. (Note, in Fig. 20, that the prorated cost of γ is greater than the potential cost of γ because $\underline{W}$ and $\underline{W}'$, which are in the configuration of γ, potentially drive but do not drive β.)

We define $\underline{n}(\underline{i},\beta)$ to be the number of wires directly driving junctions of level $\underline{i}$ $(2 \leqq \underline{i})$ in the configuration of a junction β of level $\underline{j}$ $(2 \leqq \underline{j} \leqq \underline{d})$. Where there are no junctions of level $\underline{i}$, $\underline{n}(\underline{i},\beta) = 0$; in particular, this is so where $\underline{i} > \underline{j}$. Let $\underline{m}(\beta) = \sum_{i=2}^{j} \frac{n(i,\beta)}{b^{j-i}}$. The following theorem shows the importance of $\underline{m}(\beta)$ where β is a final output junction.

THEOREM III. For a final output junction β of an irreducible net, $\underline{m}(\beta)$ is equal to the potential cost of β.

The proof is direct from the definition of $\underline{m}(\beta)$ and the observation made above that, in an irreducible net, the potential use of any wire directly driving a junction of level $\underline{i}$ is $\underline{b}^{d-i}$.

Where β and γ are in a net, then the configurations of β and γ <u>overlap</u> if there is at least one wire $\underline{W}$ driving both β and γ. We are interested in the case where β and γ each directly drive a junction α. Note that such a situation exists in Fig. 21. (Fig. 21, and all the diagrams which follow are abbreviated. The nuclei and output wires of elements, all of which are conjunction elements, are omitted, the input wire being drawn directly from one junction to another. Junctions are represented by small circles. Arrowheads are omitted, always being understood as pointing upward. Finally, junctions of the same level are placed on the same horizontal level.)

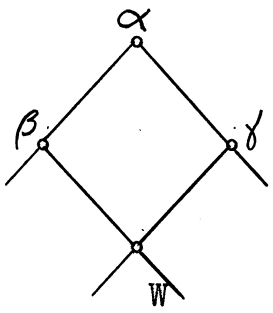


Fig. 21

THEOREM IV. If $\beta_1, \ldots, \beta_r$ of levels $j_1, \ldots, j_r$, respectively, are the r junctions directly driving a junction α of level i in an irreducible net, then

$$m(\alpha) \leqq r + \frac{m(\beta_1)}{b^{i-j_1}} + \ldots + \frac{m(\beta_r)}{b^{i-j_r}} .$$

The equality holds if and only if the configurations of no two of the β's overlap.

Proof: Suppose first that there is no such overlap. Then $n(i,\alpha) = r$. For each p, $2 \leqq p < i$, $n(p,\alpha) = n(p,\beta_1) + \ldots + n(p,\beta_r)$. Hence (by definition of $m(\alpha)$), the equality in the theorem holds. Where there is such overlap, suppose there is a wire W driving β_k and $\beta_{k'}$ $(k \neq k')$. Suppose that the level of the junction which W directly drives is p, which will be greater than 1. Then $n(p,\alpha) < n(p,\beta_1) + \ldots + n(p,\beta_r)$. Hence the inequality in the theorem holds.

7.2. Quasi balanced m.s. nets.

We now define "quasi balanced configuration." The significance of this concept is contained in Theorem VI below. A configuration is quasi balanced if every one of its output junctions β is directly

driven as follows, assuming j is the level of β : (1) if $b = 2$ and $j = 3$, either by one junction of level 1 and one of level 2 or by three junctions each of level 1; (2) in all other cases, by two junctions of level $\left[\frac{j}{2}\right]$ and $\left[\frac{j+1}{2}\right]$. Thus Fig. 22(a) is a quasi balanced configuration of β of level 3 in any net, while Fig. 22(b) is quasi balanced in a 2-d net but not in a b-d net for which $b > 2$.

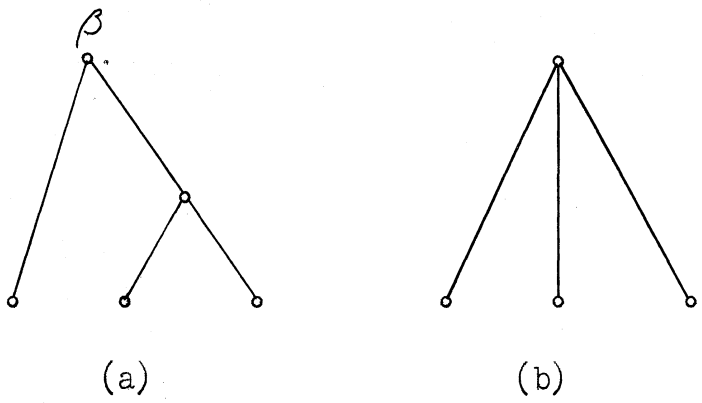


Fig. 22

(Note that a quasi balanced configuration of a junction of level j is like the configuration of a junction of level j in a balanced m.s.n. in every respect except possibly in the case of junctions driving junctions of level 3 in 2-d nets.)

Note that in a quasi balanced configuration no two configurations of junctions directly driving another junction overlap, as can be seen as follows. Suppose the configurations of γ and δ, directly driving β of level j, overlap. Then the sum of the levels of γ and δ would be greater than j, whereas in a quasi balanced configuration the sum is always equal to or less than j.

We note that, for a given b and j, for all junctions β of level j in b-d nets (regardless of d) which have quasi balanced configurations, $m(\beta)$ has a constant value. Let $m(j,b)$ be this value. The following theorem summarizes some properties of $m(j,b)$.

THEOREM V. (a) $m(2,b) = 2$.

(b) $m(3,b) = 2 + \frac{2}{b}$.

(c) For $j \geq 2$, $m(2j,b) = 2 + \frac{2m(j,b)}{b^j}$.

(d) For $j \geq 2$, $m(2j+1,b) = 2 + \frac{m(j,b)}{b^{j+1}} + \frac{m(j+1,b)}{b^j}$.

(e) $m(4,b) = 2 + \frac{4}{b^2}$.

(f) $m(5,b) = 2 + \frac{4}{b^3} + \frac{2}{b^2}$.

(g) For all $j \geq 2$, $m(j,b) \leq 3$.

(h) For $j > 2$, $m(j,b) > 2$.

(i) For all $j \geq 2$, $m(j,b) - m(j+1,b) < \frac{1}{2}$.

(a) and (b) are obvious. The proofs of (c) and (d) are direct from Theorem IV and the definition of a quasi balanced configuration, since there is no overlap in a quasi balanced configuration which would lead to an application of the inequality of Theorem IV. (e) follows directly from (a) and (c). (f) follows directly from (a), (b), and (d). (h) follows from the fact that in the configuration of any junction β of level 3 or more, there must be at least two wires directly driving β and at least one other wire somewhere in the configuration.

The proofs of (g) and (i) are given in Appendix 1.

Note that (a), (b), (c), and (d) resemble (3) of Section 5, the recursive formulas for the element-input count of the balanced m.s.n. The former characterize $\underline{m}(\underline{j},\underline{b})$ which will be the potential cost of a junction with a quasi balanced configuration. Now, as we show later, the potential cost equals the prorated cost of a final output junction of a balanced m.s.n. and every configuration of an m.s.n. is quasi balanced. Since there are $\underline{b}^{\underline{d}}$ final output junctions in a $\underline{b}$-$\underline{d}$ m.s.n., one can find the element input count from $\underline{m}(\underline{d},\underline{b})$ by multiplying by $\underline{b}^{\underline{d}}$.

THEOREM VI. For a given $\underline{b}$, if a junction β of level $\underline{j}$ does not have a quasi balanced configuration, then $\underline{m}(\beta) > \underline{m}(\underline{j},\underline{b})$.

The proof of this theorem is given in Appendix 2. It follows from this theorem that in the class of all conjunction $\underline{b}$-$\underline{d}$ nets, in order that a final output junction have a minimal potential cost, it is necessary and sufficient that it have a quasi balanced configuration.

We proceed to define "$\underline{b}$-$\underline{d}$ quasi balanced m.s.n." in several steps.

We define first "2-3 quasi balanced m.s.n." Our definition proceeds by describing the alternative ways in which the junctions of level 2 are driven. Once the junctions of level 2 have been specified, a unique 2-3 decoding net is determined as follows. Each pair of junctions, one of level 2 and the other an input junction in the bracketed set not represented by the first, is put on the inputs of a conjunction whose output wire is on a final output junction. Then each set of three input junctions, one from each bracketed set such that no two directly drive a junction of level 2, is put on the input wires of a three input conjunction whose output wire is on a final output

junction. (Note that Fig. 26 below is determined in this way from Fig. 25.) Supposing that α_1, α_1' of the first bracketed set, α_2, α_2' of the second, and α_3, α_3' of the third are the input junctions of a net $\underline{N}$, $\underline{N}$ is a 2-3 quasi balanced m.s.n. if and only if it is determined by the following.

(1) There are four junctions β, β', β'', and β''' such that β is driven by α_1 and α_2, β' by α_1 and α_2', β'' by α_1' and α_2, and β''' by α_1' and α_2'. (See Fig. 23.)

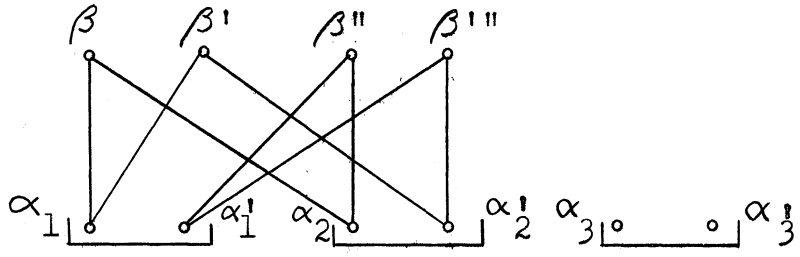


Fig. 23

(2) There are four junctions β_1, β_1', β_3, β_3' such that β_1 is driven by α_1 and α_2, β_1' by α_1' and α_2, β_3 by α_2' and α_3, β_3' by α_2' and α_3'. (See Fig. 24.)

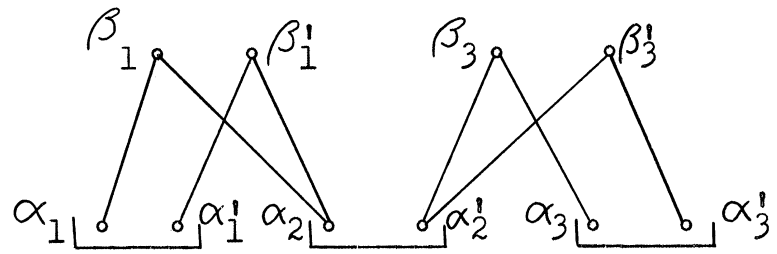


Fig. 24

(3) There are three junctions β_3, β_2, β_1 such that β_3 is driven by α_1' and α_2', β_2 by α_1 and α_3', and β_1 by α_2 and α_3. (See Fig. 25.)

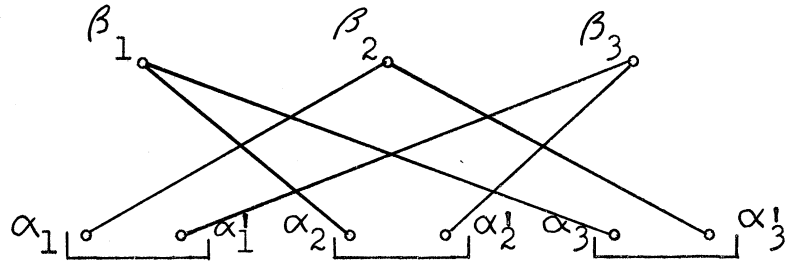


Fig. 25

(4) The same as (1), (2), or (3) except that one or more (possibly all) the junctions of level 2 are deleted.

The reader, with some effort, can ascertain that there are exactly ten 2-3 quasi balanced m.s.n. The net determined by (1) is the 2-3 balanced m.s.n. The net determined from (1), (2), or (3) by deleting all junctions of level 2 is the 2-3 exponential switch. The net determined from (3) is pictured in Fig. 26.

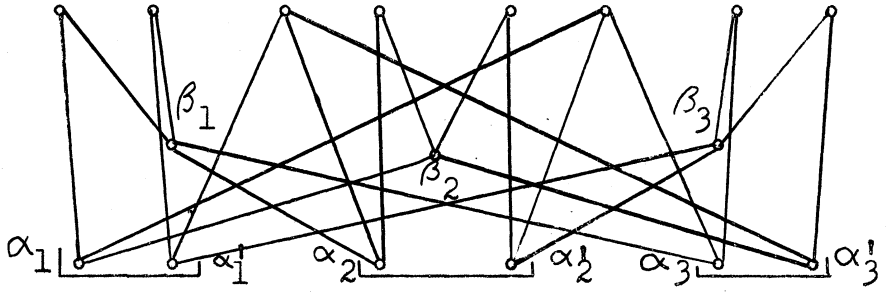


Fig. 26

A $\underline{b}$-3 net $\underline{N}$ ($\underline{b} > 2$) is a quasi balanced m.s.n. if and only if it satisfies one of the following conditions. (1) $\underline{N}$ is a $\underline{b}$-3 balanced m.s.n. (2) $\underline{N}$ is constructed as follows. Let the three bracketed sets be $\underline{I}_1$, $\underline{I}_2$, and $\underline{I}_3$. Take any subset $\underline{K}$ of $\underline{I}_2$ of cardinality $\underline{k}$, $1 \leqq \underline{k} \leqq \left[\frac{\underline{b}}{2}\right]$. Construct the m.s. which has as its bracketed sets $\underline{I}_1$ and $\underline{K}$. Construct the m.s. which has as its bracketed sets $\underline{I}_2$-$\underline{K}$ and $\underline{I}_3$. Let $\underline{O}_1$ and $\underline{O}_2$

be, respectively, the sets of outputs of these m.s. Construct a third m.s. whose bracketed sets of inputs are $\underline{O}_1$ and $\underline{I}_3$, and a fourth m.s. whose bracketed sets of inputs are $\underline{O}_2$ and $\underline{I}_1$. The outputs of these two m.s. are the final outputs of $\underline{N}$ in which $\underline{I}_2$ is bracketed and superfluous brackets are deleted. (Cf. the 3-3 net of Fig. 27.)

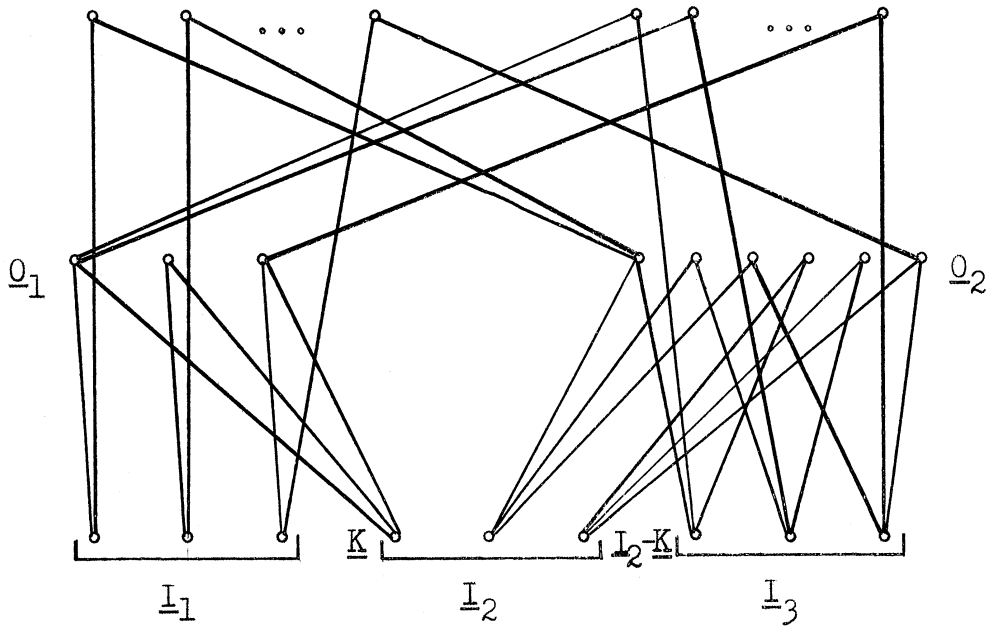


Fig. 27

Note that the radical difference between the definition of "$\underline{b}$-3 quasi balanced m.s.n." for $\underline{b} = 2$ and the definition for $\underline{b} > 2$ is related to the difference between the case in which $\underline{b} = 2$ and that in which $\underline{b} > 2$ in the definition of "quasi balanced configuration."

A net $\underline{N}$ is a $\underline{b}$-5 quasi balanced m.s.n. if and only if it satisfies one of the following conditions. (1) Where $\underline{N}_1$ is a $\underline{b}$-3 quasi balanced m.s.n. and $\underline{N}_2$ is a $\underline{b}$-2 m.s. with output sets $\underline{O}_1$ and $\underline{O}_2$, respectively, and where $\underline{N}_3$ is an m.s. having $\underline{O}_1$ and $\underline{O}_2$ as its bracketed sets of inputs, the net $\underline{N}$ consists of $\underline{N}_1$, $\underline{N}_2$, and $\underline{N}_3$ with superfluous brackets deleted. (2) $\underline{N}$ has five bracketed sets of inputs $\underline{I}_1, \ldots, \underline{I}_5$, each with $\underline{b}$ junctions, $\underline{I}_3$ consisting of $\underline{K}$ and $\underline{I}_3$-K. $\underline{N}$ is the net consisting of $\underline{N}_1, \ldots, \underline{N}_6$ (described as follows) with superfluous brackets deleted. $\underline{N}_1$ is the m.s. having $\underline{I}_1$ and $\underline{I}_2$ as its bracketed sets of inputs. $\underline{N}_2$ is the m.s. having $\underline{I}_4$ and $\underline{I}_5$ as its bracketed sets of inputs. Where $\underline{O}_1$ and $\underline{O}_2$ are the output sets of $\underline{N}_1$, N_2, respectively, $\underline{N}_3$ is the m.s. having $\underline{O}_1$ and $\underline{K}$ as its bracketed sets of inputs, and $\underline{N}_4$ is the m.s. having $\underline{O}_2$ and $\underline{I}_3$-$\underline{K}$ as its bracketed sets

of inputs. Where $\underline{O}_3$ and $\underline{O}_4$ are the output sets of $\underline{N}_3$, $\underline{N}_4$ respectively, $\underline{N}_5$ is the m.s. having as its bracketed sets of inputs $\underline{O}_3$ and $\underline{O}_2$, and $\underline{N}_6$ is the m.s. having as its bracketed sets of inputs $\underline{O}_1$ and $\underline{O}_4$. The outputs of $\underline{N}$ are, of course, those of $\underline{N}_5$ together with $\underline{N}_6$. (Cf. the 2-5 net of Fig. 28.)

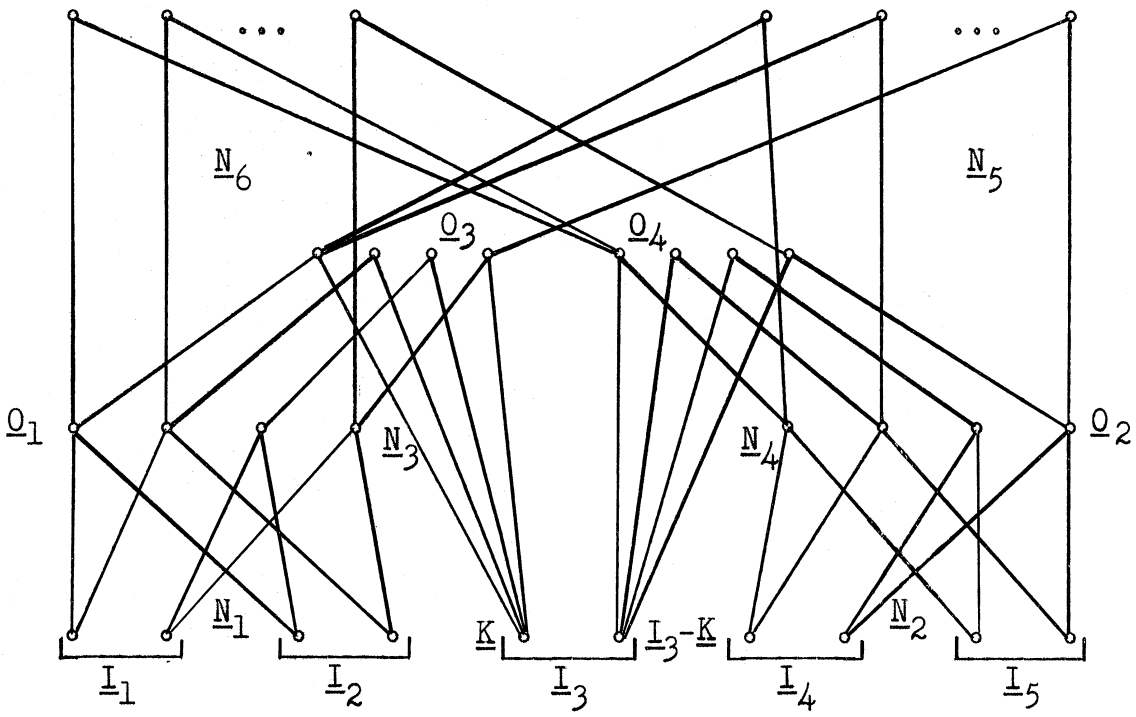


Fig. 28

A b-d balanced m.s.n. N may have a b-3 or b-5 balanced m.s.n. N' as a part. N' has in common with the remainder of N just the set of b^3, or b^5, junctions which are the final output junctions of N'. (See the definition of "balanced m.s.n." in Section 4.) The operation which consists of replacing N' by another b-3 or b-5 quasi balanced m.s.n. is used in the definition below.

We define b-d quasi balanced m.s.n. for $d \neq 3$ and $d \neq 5$ as follows:

(1) The b-d balanced m.s.n. is a b-d quasi balanced m.s.n.

(2) The result of replacing any number of b-3 and b-5 balanced m.s.n. by b-3 and b-5 (respectively) quasi balanced m.s.n. in the b-d balanced m.s.n. is a b-d quasi balanced m.s.n.

Note that a b-d quasi balanced m.s.n., if it is not a b-d balanced m.s.n., differs from it only in junctions of level 5 or less. Thus, we are justified in saying that a quasi balanced m.s.n. is either a balanced m.s.n. or a slight variation thereof. It is also interesting to note that a b-d quasi balanced m.s.n. where d is a power of 2 must be a balanced m.s.n., since in this case there

are no b-3 or b-5 subnets, all junctions in such a balanced m.s.n. being of level 2^r, for some non-negative integer r. Note, finally, that a quasi balanced m.s.n. is not necessarily an m.s.n., as defined in Section 4. (The word, "quasi," is used as an adjective modifying the phrase, "balanced m.s.n.," not as an adverb modifying "balanced.")

7.3. Proof of the main theorem.

MAIN THEOREM. A net is minimal in the class of conjunction b-d decoding nets if and only if it is a b-d quasi balanced m.s.n.

This theorem follows directly from Lemmas 2 and 7 below since we know that minimal nets must be irreducible and since any quasi balanced m.s.n. is irreducible. The series of lemmas refers to the presence and absence in nets of two important properties. The first minimizing property of a net is that every final output junction has a quasi balanced configuration. We note that, by Theorems III and VI, a net has the first minimizing property if and only if each final output junction has a minimal potential cost. The second minimizing property of a net is that

the potential cost of every final output junction equals the prorated cost. Before presenting the lemmas we state an important corollary of the main theorem which depends on the fact that the balanced m.s.n. is a quasi balanced m.s.n.

COROLLARY. The b-d balanced m.s.n. is minimal in the class of conjunction b-d decoding nets.

Note that the corollary requires only the first two of the following lemmas, since a balanced m.s.n. is a quasi balanced m.s.n. and is irreducible. Most of the complication in the proofs of Lemmas 3 through 7, as well as that in the definition of "quasi balanced m.s.n." itself, is included in this paper because we wished to give a condition which is necessary, as well as sufficient, for minimality.

LEMMA 1. A quasi balanced m.s.n. has both minimizing properties.

Proof: It is possible to check all the cases to see that each quasi balanced m.s.n. has, for every final output junction, a quasi balanced configuration. It is also possible to check that every junction (and

thus every wire) in such a net drives every final output junction which it potentially drives. Therefore, by Theorem II, a quasi balanced m.s.n. also has the second minimizing property.

LEMMA 2. A net is minimal in the class of irreducible b-d nets if and only if it has both minimizing properties.

Proof: Lemma 2 can be established if we can establish the following: (1) there exists a b-d net having both minimizing properties; (2) if $\underline{N}$ has both minimizing properties, then $\underline{C}(\underline{N}) = \underline{b}^{\underline{d}}\underline{m}(\underline{d},\underline{b})$; (3) if $\underline{N}$ lacks either of the two properties, then $\underline{C}(\underline{N}) > \underline{b}^{\underline{d}}\underline{m}(\underline{d},\underline{b})$. (1) follows from Lemma 1. Suppose $\underline{N}$ has both minimizing properties; consider any final output junction α. $\underline{c}(\alpha)$ equals the potential cost of α, by the second minimizing condition, which equals $\underline{m}(\alpha)$, by Theorem II, which equals $\underline{m}(\underline{d},\underline{b})$ by the first minimizing condition. Hence, (2) follows by Theorem I. Suppose that $\underline{N}$ lacks the first minimizing property. Then there is at least one junction β which does not have a quasi balanced configuration. Then $\underline{c}(\beta) \geq \underline{m}(\beta)$, by Theorem II, $> \underline{m}(\underline{d},\underline{b})$ by Theorem VI. For every other junction

α of $\underline{N}$, $\underline{c}(\alpha) \geq \underline{m}(\underline{d},\underline{b})$, by similar reasoning. Hence, $\underline{C}(\underline{N}) > \underline{b}^{\underline{d}}\underline{m}(\underline{d},\underline{b})$. If $\underline{N}$ lacks the second minimizing property, then there is a junction β whose potential cost is not equal to $\underline{c}(\beta)$. By Theorem II $\underline{c}(\beta)$ is greater than the potential cost, which is $\geq \underline{m}(\underline{d},\underline{b})$ by Theorem VI. Again, by Theorem I, $\underline{C}(\underline{N}) > \underline{b}^{\underline{d}}\underline{m}(\underline{d},\underline{b})$. Hence (3) follows.

LEMMA 3. In an irreducible $\underline{b}$-$\underline{d}$ net $\underline{N}$ ($\underline{d} \geq 3$) having both minimizing properties the following is impossible: a junction β is driven by α and α', and β'' by α' and α'', where α, α', α'' are input junctions of distinct bracketed sets, and β and β'' are of level 2.

Proof: α, α', and α'' together drive at least one final output junction γ. β and β'' potentially drive γ. Therefore, we can assume that β and β'' must each drive γ, since $\underline{N}$ has the second minimizing property. But then the configuration of γ would not be quasi balanced, which can be seen as follows. There must be a junction δ in the configuration of γ which is driven by both β and β'', such that there are no junctions driven by β and β'' which drive δ. (δ may be γ itself.) δ is directly driven by a junction η which is driven by,

or identical to, β but not β'' and a junction η'' driven by, or identical to, β'' but not β. Now if there is at least one other junction directly driving δ, then the configuration of γ would not be quasi balanced (for no junction in a quasi balanced configuration is driven by more than two junctions, unless it is a junction of level 3 in a 2-d net driven by three junctions of level 1). We can, therefore, assume that η and η'' are the only junctions directly driving δ. For the configuration of γ to be quasi balanced, the sum of the levels of η and η'' must be equal to the level of δ. But this is impossible here, since α' drives both η and η''.

LEMMA 4. If an irreducible 2-3 net has both minimizing properties, then it is a quasi balanced m.s.n.

Proof. In defining "2-3 quasi balanced m.s.n." we characterized the net by enumerating the junctions of level 2, and specifying how they were to be driven. The manner in which the eight final output junctions were to be driven was completely determined by the condition that a junction of level 2 drive every

final output junction which it potentially drives. Clearly a net can be determined in this manner if and only if it has the second minimizing property. For the purposes of this proof we can restrict ourselves to such nets. Consider nets $\underline{N}_1$ and $\underline{N}_2$ such that $\underline{N}_2$ contains all the junctions of level 2 that $\underline{N}_1$ has and one more besides. If $\underline{N}_1$ has the first (as well as the second) minimizing property, and if every such $\underline{N}_2$ lacks the first minimizing property, we say that $\underline{N}_1$ is a <u>maximal</u> net having both minimizing properties. It is not difficult to see that if $\underline{N}_2$ has the first minimizing property then so does $\underline{N}_1$. And, finally, if $\underline{N}_2$ is a quasi balanced m.s.n., then so is $\underline{N}_1$. It follows that if there is a net which has both minimizing properties which is not quasi balanced then there is a maximal such net.

Consider now any maximal net $\underline{N}$ having both minimizing properties. It is not difficult to see that the following three cases are exhaustive.

Case I: There are two bracketed sets $\underline{I}_1$ and $\underline{I}_2$ of $\underline{N}$ such that all junctions of level 2 represent only these. The reader can check that the only maximal such case is (1) of the definition of

"2-3 quasi balanced m.s.n." Case II: All bracketed sets are represented by junctions of level 2, but there is an $\underline{I}_2$ which is represented by all such junctions. Making use of Lemma 3, the reader can check that the only maximal such case is (2) of the definition. Case III: Every pair of bracketed sets is represented by a junction of level 3. Again, by Lemma 3, one can see that the only maximal such case is (3) of the definition.

We say that two junctions α and β in a conjunction net $\underline{N}$ are conjoined when there is a junction γ such that α and β directly drive γ, and no other junction directly drives γ.

LEMMA 5. For $\underline{b} > 2$, if an irreducible $\underline{b}$-3 net has both minimizing properties, then it is a quasi balanced m.s.n.

Proof: For $\underline{b} > 2$, since every junction of level 3 must have a quasi balanced configuration, there must be junctions of level 2. Either there are only two bracketed sets which are represented by the junctions of level 2 (Case I) or all bracketed sets are so represented (Case II).

Case I: Suppose that $\underline{I}_1$ and $\underline{I}_2$ are the

bracketed sets. It is easy to see in this case that every pair α_1 and α_2 of junctions $\underline{I}_1$ and $\underline{I}_2$, respectively, are conjoined. For take any junction α_3 of $\underline{I}_3$: α_1, α_2, and α_3 must drive a final output junction γ; since γ has a quasi balanced configuration, there must be a junction β of level 2 in it; but β cannot be driven by α_3, so it must be driven by α_1 and α_2. Since every such pair drives a junction of level 2, and since there are no other junctions of level 2, the net is a balanced m.s.n. and a fortiori a quasi balanced m.s.n.

Case II: There must be at least one bracketed set, say $\underline{I}_2$, containing a junction α_2 conjoined with a junction α_1 of $\underline{I}_1$ and containing a junction α_2' conjoined with a junction α_3' of $\underline{I}_3$. It is then impossible that any junction α_1' of $\underline{I}_1$ be conjoined with a junction α_3 of $\underline{I}_3$. For consider the junction γ driven by α_1', α_2, and α_3': α_1 and α_1' (α_2 and α_2') (α_3 and α_3') would be distinct by Lemma 3; neither α_1' and α_2, nor α_1' and α_3', nor α_2 and α_3' could be conjoined, by Lemma 3; and, therefore, γ would not have a quasi balanced configuration.

Now let $\underline{K}$ be the set of all junctions of $\underline{I}_2$ which are conjoined with a junction of $\underline{I}_1$. Let α_1, α_2, α_2', and α_3 be any four junctions of $\underline{I}_1$, $\underline{K}$, $\underline{I}_2$-$\underline{K}$, and $\underline{I}_3$, respectively. Let γ (γ') be the junction driven by α_1, α_2 (α_2'), and α_3. The configuration of γ (γ') must have a junction β (β') of level 2 in order to be a quasi balanced configuration. β must be driven by α_1 and α_2; the other possibilities are excluded by the definition of "$\underline{K}$", Lemma 3, and what was established in the above paragraph. Likewise β' must be driven by α_2' and α_3. Since the α's were arbitrary junctions, it follows that there are m.s., one whose bracketed sets are $\underline{I}_1$ and $\underline{K}$, the other whose bracketed sets are $\underline{I}_3$ and $\underline{I}_2$-$\underline{K}$. It is easy to see then that $\underline{N}$ fulfills (2) of the definition of "$\underline{b}$-3 quasi balanced m.s.n." ($\underline{b} \geq 3$). This completes the proof of Lemma 5.

LEMMA 6. An irreducible $\underline{b}$-5 net $\underline{N}$ with both minimizing properties is a quasi balanced m.s.n.

Proof: $\underline{N}$ must have junctions of level 3, in order that its final output junctions have quasi balanced

configurations. Let β be such a junction and suppose it is driven by α_1 of $\underline{I}_1$, α_2 of $\underline{I}_2$, and α_3 of $\underline{I}_3$. Consider any α_4 of $\underline{I}_4$ and α_5 of $\underline{I}_5$. The final output junction δ driven by all the α's mentioned must be driven by the wires driving β, and thus by β itself, because of the second minimizing property (by Theorem II). Since the configuration of δ is quasi balanced, β must be conjoined with a junction of level 2. Therefore, α_4 and α_5 must be conjoined. But α_4 and α_5 were arbitrarily chosen; it follows that there is an m.s. whose bracketed sets of inputs are $\underline{I}_4$ and $\underline{I}_5$. Now either all junctions of level 3 represent $\underline{I}_1$, $\underline{I}_2$, and $\underline{I}_3$ (Case I) or there is a junction of level 3 which represents either $\underline{I}_4$ or $\underline{I}_5$ (Case II).

Case I: For any $\alpha_1, \ldots, \alpha_5$, the final output junction δ driven by all these must have a junction β of level 3 in its quasi balanced configuration, which is driven by α_1, α_2, and α_3. This means (since α_1, α_2, and α_3 are arbitrarily chosen) that the subnet $\underline{N}_1$ of $\underline{N}$ consisting of all such junctions β and junctions which drive

them is a $\underline{b}$-3 decoding net. To show that $\underline{N}$ fulfills (1) of the definition of "$\underline{b}$-5 quasi balanced m.s.n." it suffices to prove that $\underline{N}_1$ is a $\underline{b}$-3 quasi balanced m.s.n. But every configuration of a junction in a quasi balanced configuration is quasi balanced; hence $\underline{N}_1$ has the first minimizing property because $\underline{N}$ has it. But by Theorem II it is easy to see that, because $\underline{N}$ has the second minimizing property, $\underline{N}_1$ also has it. By Lemmas 4 and 5 then $\underline{N}_1$ must be a $\underline{b}$-3 quasi balanced m.s.n.

Case II: Let β' be a junction of level 3 representing, say, $\underline{I}_4$. By reasoning similar to the first paragraph of the present proof there is an m.s. whose bracketed sets of inputs are the two bracketed sets not represented by β'. Neither of these can be $\underline{I}_5$, by Lemma 3, since $\underline{I}_4$ and $\underline{I}_5$ are the bracketed sets of an m.s. So β' represents $\underline{I}_4$, $\underline{I}_5$ and, say, $\underline{I}_3$. And there is an m.s. whose bracketed sets of inputs are $\underline{I}_1$ and $\underline{I}_2$. Let the set of outputs of this m.s. be $\underline{O}_1$ and the set of outputs of the m.s. of the first paragraph be $\underline{O}_2$.

Let $\underline{K}$ be the set of all junctions of $\underline{I}_3$ which are conjoined with a member of $\underline{O}_1$. It is easy to

see that no member of $\underline{K}$ is conjoined with a member β_2 of $\underline{O}_2$. This can be seen as follows. Suppose α_3 of $\underline{K}$ were. Then there would be an α_1 and α_2 of $\underline{I}_1$ and $\underline{I}_2$ driving β_1 of $\underline{O}_1$ which together with α_3 drives γ_1 of level 3. And there would be an α_4 and α_5 of $\underline{I}_4$ and $\underline{I}_5$ driving β_2 of $\underline{O}_2$ which together with α_3 drives γ_2 of level 3. The final output junction δ driven by α_1, α_2, α_3, α_4, and α_5, by the second minimizing property, would have to be driven by β_1 and β_2; but then the configuration of δ would not be quasi balanced.

Now in order that the configurations of $\underline{N}$ be quasi balanced every member of $\underline{K}$ must be conjoined with every member of $\underline{O}_1$, and every member of $\underline{I}_3$-$\underline{K}$ must be conjoined with every member of $\underline{O}_2$. (No member of $\underline{I}_1$ or $\underline{I}_2$ can be conjoined with any member of $\underline{I}_4$ or $\underline{I}_5$, by Lemma 3.) It is now easy to see that $\underline{N}$ must fulfill (2) of the definition of "$\underline{b}$-5 quasi balanced m.s.n." This completes the proof of Lemma 6.

LEMMA 7. For all b and d ($b,d \geq 2$), an irreducible b-d net has both minimizing properties if and only if it is a quasi balanced m.s.n.

Proof: By Lemma 1 it suffices to prove that any such net having both minimizing properties is a quasi balanced m.s.n. We prove this by induction on d. For each b, there is only one irreducible b-2 decoding net, within isomorphism. That net is obviously a quasi balanced m.s.n. and has both minimizing properties. We assume, as an inductive hypothesis, that any b-d decoding net, for $d < s$, having both minimizing properties is a quasi balanced m.s.n. We must show from this that any b-s decoding net N having both minimizing properties is a quasi balanced m.s.n. We distinguish two cases according to whether s is even (Case I) or odd (Case II).

Case I: Put $r = \frac{s}{2}$. Take any final output junction γ of N. Since the configuration of γ is quasi balanced, γ must be directly driven by two junctions β_1 and β_2 each of level r. Let S_1 (S_2) be the set of junctions conjoined with β_2 (β_1). By the second minimizing property S_1 (S_2) has b^r junctions, otherwise β_2 (β_1)

would not drive all the junctions of level $2\underline{r}$ which it potentially drives. For, since there can be no junctions whose level is between $\underline{r}$ and $2\underline{r}$, in order for $\mathcal{B}_2$ ($\mathcal{B}_1$) to drive a final output junction, it has to directly drive it. By the first minimizing property and definition of "quasi balanced configuration", all the junctions of $\underline{S}_1$ ($\underline{S}_2$) are of level $\underline{r}$. Every junction of $\underline{S}_1$ ($\underline{S}_2$) must represent just those $\underline{r}$ bracketed sets of inputs of $\underline{N}$ which $\mathcal{B}_2$ ($\mathcal{B}_1$) does not represent. It is easy to see that every junction of $\underline{S}_1$ is conjoined with every junction of $\underline{S}_2$, otherwise some junctions would not drive all the final output junctions they potentially drive. Furthermore, such account for all the $\underline{b}^{2r}$ final output junctions. The two subnets $\underline{N}_1$ and $\underline{N}_2$ whose final output junctions are $\underline{S}_1$ and $\underline{S}_2$, respectively, must account for all the junctions of $\underline{N}$, except $\underline{N}$'s final output junctions. By definition of "balanced m.s.n." (Section 4) and "quasi balanced m.s.n.", we can prove that $\underline{N}$ is a quasi balanced m.s.n. if we can prove that $\underline{N}_1$ and $\underline{N}_2$ are. But since $\underline{N}$ has both minimizing properties, $\underline{N}_1$ and $\underline{N}_2$ must have them (cf. Case I of Lemma 6).

By inductive hypothesis $\underline{N}_1$ and $\underline{N}_2$ must therefore be quasi balanced m.s.n.

Case II: Put $\underline{r} = \frac{s-1}{2}$. For $\underline{r} = 1$ and $\underline{r} = 2$, Case II has already been proved in Lemmas 4, 5, and 6. We can assume, therefore, that $\underline{r} \geq 3$. Paralleling Case I, take any final output junction γ of $\underline{N}$. Since $\underline{N}$ has the first minimizing property, γ must be directly driven by β_1 of level $\underline{r}$ and β_2 of level $\underline{r}+1$, by definition of "quasi balanced configuration." Let $\underline{S}_1$ ($\underline{S}_2$) be the set of all junctions conjoined with β_2 (β_1). Neither β_1 nor β_2 can drive any junctions except final output junctions, otherwise some configuration would not be quasi balanced. (Note that for $\underline{r} = 2$ or 1, β_1 could drive a junction of level $\underline{r}+1$; this is what makes $\underline{b}-3$ nets and $\underline{b}-5$ nets exceptional.) Thus, by the second minimizing property, $\underline{S}_1$ ($\underline{S}_2$) has $\underline{b}^r$ (b^{r+1}) junctions. By the first minimizing property, and by definition of "quasi balanced configuration," all the junctions of $\underline{S}_1$ ($\underline{S}_2$) are of level $\underline{r}$ ($\underline{r}+1$). From here on, the proof is similar to the proof of Case I. This completes the proof of Lemma 7.

Arthur W. Burks
Robert McNaughton
Carl H. Pollmar
Don W. Warren
Jesse B. Wright

University of Michigan

Footnotes

1. The writing of this paper and the research which it reports were done under the sponsorship of the Burroughs Corporation.

The authors wish to thank Dr. J. Richard Büchi for many helpful suggestions.

2. Burks, Arthur W., and Jesse B. Wright, "Theory of Logical Nets," Proceedings I.R.E., Vol. 41 (1953), pp. 1357-1365.

3. The reader unfamiliar with the notation of symbolic logic may find it helpful to read part of the section, "The Equation Language," at this point.

4. Burks and Wright, op. cit., pp. 1361-1365.

5. Burks and Wright, op. cit., p. 1362.

6. Note that, according to this definition, there are complete decoding nets which have bracketed sets of only one junction. Although this is contrary to ordinary usage, the inclusion of these nets facilitates the development of the theory. Note that the decoding junctions need not be final output junctions. Note also that there may be, for a net state N', more than one decoding junction labeled 1 in N'.

7. The word "multiplicative" derives from the set-theoretic definition of multiplication in which the

cardinality of the Cartesian Product (here the collection of representative sets) of a number of given sets is the product $(i_1 \cdot \ldots \cdot i_d)$ of the cardinalities of the given sets.

8. Shannon, C. E., "The Synthesis of Two-Terminal Switching Circuits," Bell System Technical Journal, Vol. 28 (No. 1, January, 1949), pp. 59-98.

9. Brown, D. R., and N. Rochester, "Rectifier Networks for Multiposition Switching," Proceedings I.R.E., Vol. 37 (1949), pp. 139-147, give definitions of 2-d m.s.n. and 2-d balanced m.s.n.; their definitions are in terms of crystal rectifiers and are more complicated than the ones we give.

Synthesis of Electronic Computing and Control Circuits, by the staff of the Computation Laboratory (Harvard University), p. 137, gives some informal rules of construction for the realization of a 2-d balanced m.s.n. by vacuum tubes or crystal rectifiers that are similar to the above definition.

10. Brown and Rochester, op. cit.

11. Synthesis of Electronic Computing and Control Circuits gives (11) and (12) for the case $b = 2$.

12. Brown and Rochester, op. cit., state and prove the theorem that a 2-d balanced m.s.n. is minimal in the class of 2-d m.s.n.

13. One possible use of a linear language is in computing minimality on a digital computer, and for this purpose a linear language would be much more convenient than the net language. The process by which this could be accomplished is roughly as follows. The machine is instructed to construct all sets of equations (or sets of formulas -- see below) of a given type, and then to determine for each such set the element-input count of the circuit it represents and whether the circuit it represents realizes the desired transformations.

14. The elements of a set of equations are equation-tokens rather than equation-types, in the sense of Peirce (The Collected Papers of Charles Sanders Peirce, edited by Charles Hartshorne and Paul Weiss, Vol. 4 (1933), paragraph 537). Thus, a network composed of two two-input conjunctions connected in parallel from junctions with associated variables "p" and "q" to junction with associated variable "r" translates into "$r \equiv Kpq$, $r \equiv Kpq$."

15. In some unimportant cases two or more junctions of a w.f.n. may have the same formula associated with them; in what follows we assume that these formulas are distinguished by appropriate uses of superscripts on the function symbols in them.

16. It should be noted that a net does not translate into a unique set of equations inasmuch as different variables may be associated with the junctions of the net. For a similar reason, a given set of formulas in general abbreviates many sets of equations.

APPENDIX 1

Proof of (g) of Theorem V. By (a), (b), (e), and (f), (g) holds for $j < 6$. By induction we prove it for all j. Assume that, for any $j \geq 6$, (g) holds for all numbers less than j. Case I: $j = 2i$. Here, $i \geq 3$. By (c) $m(j,b) = 2 + \frac{2m(i,b)}{b^i} \leq 2 + \frac{2 \cdot 3}{2^3} < 3$. Case II: $j = 2i + 1$. Again, $i \geq 3$. By (d), $m(j,b) = 2 + \frac{m(i,b)}{b^{i+1}} + \frac{m(i+1,b)}{b^i} \leq 2 + \frac{3}{2^4} + \frac{3}{2^3} < 3$. (Note that the equality of (g) holds only for $b = 2$ and $j = 3$, 4, or 5.)

Proof of (i) of Theorem V. Case I: $j = 2$, 3, or 4. Here (i) can be established directly using (a), (b), (e), and (f) of Theorem V. Case II: $j = 2i$, $i \geq 3$. By (c) and (d), $m(2i,b) - m(2i+1,b) = \frac{2bm(i,b) - m(i,b) - bm(i+1,b)}{b^{i+1}} < \frac{4b-2}{b^{i+1}}$, applying (g) and (h) of Theorem V. The latter is easily shown to be less than 1/2. Case III: $j = 2i + 1$, $i \geq 2$. By (c) and (d), $m(2i+1,b) - m(2i+2,b) = \frac{m(i,b) + (b-2)m(i+1,b)}{b^{i+1}} \leq \frac{3b-3}{b^{i+1}}$, by applying (g) of Theorem V. The latter is easily shown to be less than 1/2.

APPENDIX 2

Proof of Theorem VI. The proof is by induction on j. We prove Theorem VI first for $j = 2$ and $j = 3$. Then we show that the inductive hypothesis that (for $j \geq 4$) Theorem VI holds for all $j' < j$ implies that it holds for j.

Proof for $j = 2$. Here Theorem VI holds because a junction of level 2 in an irreducible net cannot have a configuration which is not quasi balanced.

Proof for $j = 3$. Suppose first that $b = 2$. By (b) of Theorem V, $\underline{m}(3,2) = 3$. With reference to the definition of "quasi balanced configuration", we know that β is not directly driven by exactly one junction of level 2 and exactly one junction of level 1; nor is it directly driven by exactly three junctions of level 1. The following possibilities remain, since β is of level 3: (1) β is directly driven by exactly one junction of level 2 and more than one junction of level 1; (2) β is directly driven by more than one junction of level 2 (and possibly junctions of level one). In either of these cases, it is easy to see, $\underline{m}(\beta) > 3$.

We now consider the case in which $\underline{b} > 2$. By (b) of Theorem V, $\underline{m}(3,\underline{b}) = 2 + \frac{2}{\underline{b}}$. For β there are the two possibilities of the preceding paragraph and another, namely, (3) β is directly driven by exactly three junctions of level 1. In any of the three cases, it is easy to see that $\underline{m}(\beta) > 2 + \frac{2}{\underline{b}}$.

Inductive proof. Case I: $\underline{j} = 2\underline{n}$ $(\underline{n} \geqq 2)$. We divide this into four subcases, according to whether (Ia) there are exactly two junctions γ and δ directly driving β each of level $\underline{n}$, (Ib) the same but γ is of level $\underline{n}+1$ and δ is of level $\underline{n}-\underline{i}+\underline{k}$ $(\underline{n} > \underline{i} > 0$ and $\underline{n}+\underline{i}-1 \geqq \underline{k} \geqq 0)$, the configurations of γ and δ not overlapping, (Ic) the same as Ib except that the configurations do overlap $(\underline{k} > 0)$, or (Id) there are three or more junctions directly driving β. That the first three cases are exhaustive when there are exactly two junctions is implied by the fact that, in order that $2\underline{n}$ bracketed sets be represented by β, the sum of the levels of the two junctions must be at least $2\underline{n}$; apart from this, the two junctions may be of any two levels, each less than $2\underline{n}$.

Subcase Ia: Since γ and δ are each of level $\underline{n}$, and β is of level $2\underline{n}$, the configurations of γ

and δ do not overlap. If the configurations of γ and δ were both quasi balanced then β would have a quasi balanced configuration. Therefore, either γ or δ does not have a quasi balanced configuration. By inductive hypothesis either $\underline{m}(\gamma) > \underline{m}(\underline{n},\underline{b})$ or $\underline{m}(\delta) > \underline{m}(\underline{n},\underline{b})$, and inequality in the reverse direction holds for neither. But $\underline{m}(2\underline{n},\underline{b}) = 2 + \frac{2m(n,b)}{b^n}$, by (c) of Theorem V. Hence, by Theorem IV, $\underline{m}(\beta) = 2 + \frac{m(\gamma)}{b^n} + \frac{m(\delta)}{b^n} > m(2\underline{n},\underline{b})$.

Subcase Ib: Here, by inductive hypothesis, $\underline{m}(\gamma) \geq \underline{m}(2\underline{n}+\underline{i},\underline{b})$ and $\underline{m}(\delta) \geq \underline{m}(\underline{n}-\underline{i}+\underline{k},\underline{b})$. Now, by (a) of Theorem V and Theorem IV, $\underline{m}(2\underline{n},\underline{b}) = 2 + \frac{2m(n,b)}{b^n}$, and $\underline{m}(\beta) = 2 + \frac{m(\gamma)}{b^{n-i}} + \frac{m(\delta)}{b^{n+i-k}}$. Therefore, in order to prove that $\underline{m}(\beta) > \underline{m}(2\underline{n},\underline{b})$, it will be sufficient to prove that $\frac{2m(n,b)}{b^n} < \frac{m(n+i,b)}{b^{n-i}} + \frac{m(n-i+k,b)}{b^{n+i-k}}$. It suffices, therefore, to prove that $2\underline{m}(\underline{n},\underline{b}) - \underline{b}^{\underline{i}}\underline{m}(\underline{n}+\underline{i},\underline{b}) < \frac{m(n-i+k,b)}{b^{i-k}}$.

For $\underline{i} \geq 2$, the left side must be negative since $\underline{b} \geq 2$, $\underline{m}(\underline{n},\underline{b}) \leq 3$ (by (g) of Theorem V) and $\underline{m}(\underline{n}+\underline{i},\underline{b}) > 2$ (by (h) of Theorem V). The inequality

in that case holds because the right side is positive. For $\underline{i} = 1$ and $\underline{b} \geq 3$, the left side for similar reasons is at most 0. For $\underline{i} = 1$, $\underline{b} = 2$, $\underline{n} = 2$, the left side is negative by (a) and (b) of Theorem V. It remains to consider the case where $\underline{i} = 1$, $\underline{b} = 2$, and $\underline{n} \geq 3$; here the inequality becomes $2(\underline{m}(\underline{n},2) - \underline{m}(\underline{n}+1,2)) < \frac{\underline{m}(\underline{n}-1,\underline{b})}{2^{1-k}}$. By (i) of Theorem V, the left side must be less than 1, and, by (h), the right side must be greater than 1, since $\underline{n}-1 \geq 2$, and $\underline{k} \geq 0$.

Subcase Ic: This subcase is disposed of by showing that it can be transformed into Subcase Ia, Subcase Ib, or a quasi balanced configuration; and in all of these possibilities β will have a smaller potential cost as a result of the transformation. The transformation will proceed in several steps. Since the configurations of γ and δ overlap, there is a wire $\underline{V}$ which drives both γ and δ; $\underline{V}$ must directly drive a junction α of level 2 or more. Take any wire $\underline{W}$ which does not drive γ but which is on α. Delete $\underline{W}$. If $\underline{W}$ is part of a 3-input (or more) element, then this element is changed into a 2-input (or more, correspondingly) element. If it is part of a 2-input element, then delete the element

and the output junction α' of this element, reconnecting the wires originally on α' to the junction α'' other than α which, originally, directly drove the element. It is not difficult to see that under this first step of the transformation the configuration is still well-formed, each junction drives β, and no input junctions are deleted. Furthermore, wires are either deleted or drive junctions of lower level, so that the estimated cost will decrease. This step is repeated for each junction α of level 2 or more in the configurations of both γ and δ.

Subcase Id: Since there are three wires directly driving β, $\underline{m}(\beta) > 3$. Thus, $\underline{m}(\beta) > \underline{m}(2\underline{n},\underline{b})$, by (g) of Theorem V.

Case II: $\underline{j} = 2\underline{n} + 1$ $(\underline{n} \geq 2)$. Again there are four subcases, according to whether (IIa) there are exactly two junctions γ and δ directly driving β such that γ is of level $\underline{n}+1$ and δ is of level $\underline{n}$, (IIb) the same, but γ is of level $\underline{n}+1+\underline{i}$ and δ is $\underline{n}-\underline{i}+\underline{k}$ $(\underline{n} > \underline{i} > 0$ and $\underline{n}+\underline{i} \geq \underline{k} \geq 0)$, the configurations of γ and δ not overlapping, (IIc) the same as IIb except that the configurations do overlap

($\underline{k} > 0$) and (IId) there are three or more junctions directly driving β.

Subcase IIa: Reasoning as in Subcase Ia, either $\underline{m}(\gamma) > \underline{m}(\underline{n}+1,\underline{b})$ or $\underline{m}(\delta) > \underline{m}(\underline{n},\underline{b})$, and inequality in the reverse direction holds for neither. Since $\underline{m}(2\underline{n}+1,\underline{b}) = 2 + \frac{\underline{m}(\underline{n}+1,\underline{b})}{\underline{b}^{\underline{n}}} + \frac{\underline{m}(\underline{n},\underline{b})}{\underline{b}^{\underline{n}+1}}$, $\underline{m}(\beta) = 2 + \frac{\underline{m}(\gamma)}{\underline{b}^{\underline{n}}} + \frac{\underline{m}(\delta)}{\underline{b}^{\underline{n}+1}} > \underline{m}(2\underline{n}+1,\underline{b})$.

Subcase IIb: Here $\underline{m}(\gamma) \geq \underline{m}(\underline{n}+1+\underline{i},\underline{b})$, and $\underline{m}(\delta) \geq \underline{m}(\underline{n}-\underline{i}+\underline{k},\underline{b})$. But, $\underline{m}(2\underline{n}+1,\underline{b}) = 2 + \frac{\underline{m}(\underline{n}+1,\underline{b})}{\underline{b}^{\underline{n}}} + \frac{\underline{m}(\underline{n},\underline{b})}{\underline{b}^{\underline{n}+1}}$, and $\underline{m}(\beta) = 2 + \frac{\underline{m}(\gamma)}{\underline{b}^{\underline{n}-\underline{i}}} + \frac{\underline{m}(\delta)}{\underline{b}^{\underline{n}+1+\underline{i}-\underline{k}}}$. Therefore, in order to prove $\underline{m}(\beta) > \underline{m}(2\underline{n}+1,\underline{b})$, it will be sufficient to prove that $\frac{\underline{m}(\underline{n}+1,\underline{b})}{\underline{b}^{\underline{n}}} + \frac{\underline{m}(\underline{n},\underline{b})}{\underline{b}^{\underline{n}+1}} < \frac{\underline{m}(\underline{n}+1+\underline{i},\underline{b})}{\underline{b}^{\underline{n}-\underline{i}}} + \frac{\underline{m}(\underline{n}-\underline{i}+\underline{k},\underline{b})}{\underline{b}^{\underline{n}+1+\underline{i}-\underline{k}}}$.

It suffices, therefore, to prove that $\underline{b}\underline{m}(\underline{n}+1,\underline{b}) + \underline{m}(\underline{n},\underline{b}) - \underline{b}^{\underline{i}+1}\underline{m}(\underline{n}+1+\underline{i},\underline{b}) < \frac{\underline{m}(\underline{n}-\underline{i}+\underline{k},\underline{b})}{\underline{b}^{\underline{i}-\underline{k}}}$. By (g) and (h) of Theorem V, the left side is negative if $\underline{b} \geq 3$ and it is negative if $\underline{i} \geq 2$. In these cases the inequality holds because the right side is positive. Also, if $\underline{n} = 2$, $\underline{i} = 1$, $\underline{b} = 2$, the left side is negative. It remains to consider the case where $\underline{i} = 1$, $\underline{b} = 2$, and

$\underline{n} \geq 3$; here the inequality becomes $2\underline{m}(\underline{n}+1,2) + \underline{m}(\underline{n},2) - 4\underline{m}(\underline{n}+2,2) < \frac{\underline{m}(n-1+k,2)}{2^{1-k}}$. Here the left side is less than 1, by (g) and (h) of Theorem V. The right side is greater than 1, by (h) of Theorem V, since $\underline{k} \geq 0$.

Subcase IIc and IId: The proofs here are exactly the same as in Subcases Ic and Id, respectively.

This completes the proof of Theorem VI.