

# How Circuits Work

---

**Johan De Kleer**

*Xerox PARC, Intelligent Systems Laboratory, Palo Alto, CA 94304,  
U.S.A.*

---

## ABSTRACT

*This paper presents a theory of commonsense understanding of the behavior of electronic circuits. It is based on the intuitive qualitative reasoning electrical engineers use when they analyze circuits. This intuitive reasoning provides a great deal of important information about the operation of the circuit, which although qualitative in nature, describes important quantitative aspects of circuit functioning (feedback paths, stability, impedance and gain estimates, etc.).*

*One aspect of the theory, causal analysis, describes how the behavior of the individual components can be combined to explain the behavior of composite systems. Another aspect of the theory, teleological analysis, describes how the notion that the system has a purpose can be used to structure and aid this causal analysis. The theory is implemented in a computer program, EQUAL, which, given a circuit topology, can construct by qualitative causal analysis a description of the mechanism by which the circuit operates. This mechanism is then parsed by a grammar for circuit functions.*

---

## 1. Introduction

This theory explains how the function of a circuit (i.e., its purpose) is related to its structure (i.e., its schematic). This issue is explored by showing how the function of a circuit is derived from its structure. The intermediate point between structure and function is behavior. Structure is what the device is, and function is what the device is for, but behavior is what the device does. Causal reasoning analyzes how disturbances from an operating point propagate through a circuit. It thereby determines the qualitative behavior of a circuit from its structure. In addition, unlike quantitative predictions it produces intuitive, causal explanations for the behavioral predictions. These behavioral predictions, combined with their explanations, form the basis for reasoning which explains how the purposes of the circuit are achieved by that behavior.

The research reported here contributes to the field of qualitative reasoning in six ways:

(1) *Qualitative physics.* The causal reasoning used for analyzing circuits is based on earlier work on qualitative physics [6–8]. This paper restates only the

barest essentials of qualitative physics. Those papers are concerned with the broad fundamental issues and (purposely) neither present any detailed examples nor explore any domain in particular. This paper contains a detailed qualitative physics for a particular domain: electronics. It shows how the concepts and distinctions of qualitative physics are used to analyze a wide variety of electronic circuits. The power of these techniques, which seem artificial when applied to the simple examples of the earlier papers, can thus be illustrated. In addition, it is possible to be explicit about AI applications for a qualitative theory of circuit behavior.

(2) *Teleological reasoning*. Although qualitative physics describes the behavior of physical artifacts, it says little about their function. Devices are designed by man to achieve some purpose. These purposes provide an alternative method for understanding the behaviors of physical systems. This paper presents a theory of teleology for physical systems and develops the details of one for electronic circuits.

(3) *Structure and function*. Qualitative physics integrates with other kinds of knowledge. This paper presents a theory for how the function of a device is related to its structure. The basic idea is that the qualitative physics relates structure to behavior and teleology relates behavior to function. Thus it is able to construct a complete account for how a particular device achieves its intended function.

(4) *Qualitative vs. quantitative*. One of the central driving intuitions behind qualitative physics is the idea that in reasoning about an artifact one first obtains a deep intuitive understanding of how the artifact works and then uses this deep understanding to guide further, perhaps quantitative analysis. Most of the potential applications of qualitative reasoning to electronics discussed in the conclusions are based on this idea.

(5) *Architecture*. The computation which uses the qualitative physics to produce accounts for behavior is surprisingly complex. This paper includes discussion of some of the critical architectural decisions involved in implementing the system. The necessity for the complications only becomes clear by considering more complex examples than those discussed in the earlier qualitative physics papers.

(6) *Electronics*. All the major contributions of this paper require working through the details of a particular qualitative physics. I have chosen electronics, primarily because of my familiarity with that subject matter. As a consequence, some of the sections may be difficult to follow. Section 2 on the qualitative physics of circuits is sufficiently elementary that if the reader did not know how a transistor worked before he or she can learn that from reading the paper. However, Section 6 on teleological reasoning may be harder to follow. This should not be surprising: teleology relates behavior to function and the function is expressed within the technical jargon of electrical engineering. If it were otherwise, my project would be a failure. The research contribution here

is not for electrical engineers—they already know their own jargon. The contribution is to AI: a technique for constructing accounts in the language of the experts (i.e., electrical engineers).

Although this paper concentrates on electronics, the underlying qualitative physics, the framework of teleological analysis, and the algorithms are all domain-independent. Only the component models, component configurations, and some of the teleological parsing rules are idiosyncratically electrical. For example, the ideas and techniques apply to hydraulic, acoustic, thermal as well as electrical systems.

### 1.1. Methodology

One of the ultimate goals of qualitative physics research is to develop a sufficiently complete account of reasoning about designed artifacts (such as electrical circuits) so that it can be automated. This research programme presumes that there is a coherent and parsimonious collection of theories which underlie expert human performance which needs to be uncovered before any reasonable fragment of an engineer's skill can be automated. These different theories are hard to identify, and must be studied in isolation in order to uncover their power and limitations.

This paper presents two related theories of the many engineers must use to understand circuit behavior. This is not intended to be a psychological theory of how human engineers actually work. Engineers use a variety of techniques for reasoning about circuits, most of which are not discussed in this paper. Perhaps, the most relevant of these involves geometry and teleology. Presented with a circuit, an engineer will recognize familiar circuit fragments and recall their function. In fact, engineers have explicit geometric conventions for drawing well-known circuit fragments. Furthermore, an engineer usually has a pretty good idea of what the new circuit is for, and he uses this teleological information to help guide his analysis of the circuit.

Studying causal reasoning in isolation is not taking an easy road. Causal reasoning encounters severe complexities which can usually be resolved by incorporating teleological or geometrical knowledge. If I were interested in building a performance program, the temptation for including this extra knowledge would be overwhelming. However, that would be shortsighted. To understand what causal reasoning, or teleological reasoning is, one must study it in isolation uncorrupted by other forms of reasoning. Otherwise one has merged two types of reasoning without ever identifying either one individually. In addition little scientific progress is made and we are not much closer to the ultimate goal as the limitations of the resulting system are completely unclear and extending the system may prove impossible. To achieve robust performance, the underlying theories must be identified.

This methodology stands in sharp contradistinction with the popular expert-

systems methodology. Expert systems are aimed at producing what performance is possible in the short term without consideration of the longer term. Typically this is achieved by recording as many of the heuristics and rules of thumb that experts actually use in practice, as possible. This is misguided. The reasoning of experts is based on underlying theories that must be teased out. The expert systems approach can be caricatured as a stimulus-response model—good for some purposes, but ineffective in the long run.

The primary reason for initially starting with causal and teleological reasoning is that these two are basic to most of the other forms of reasoning about artifacts. For example, although recognition plays various roles in reasoning about circuits, its major task is to reduce the new device to something previously analyzed. Thus recognition is parasitic on other forms of reasoning. An additional pragmatic reason for isolating teleological reasoning from causal reasoning is that teleology is an external point of view established by the designer on the circuit. The circuit may or may not behave as intended, therefore it is important for causal reasoning to be able to identify all the possibilities independent of teleology.

## 1.2. EQUAL

Both the causal reasoning and teleological reasoning components have been completely implemented and run on hundreds of circuits. The program, EQUAL, takes as input the schematic for the circuit. As output it produces a qualitative prediction of the circuit's behavior, an explanation for that behavior, and a teleological parse which relates every component to the purposes of the overall circuit.

EQUAL need not be told the purpose of the circuit, as it can infer this itself. This is not as difficult a task as it might seem as the circuit is presumed to be a power-supply, logic-gate, amplifier or multivibrator where AC affects are unimportant.

In performing its causal analysis, EQUAL utilizes a component model library which describes the behavior of every type of electrical component. The teleological reasoning, on the other hand, utilizes a grammar of basic mechanisms (*not schematic fragments*), to parse the circuit's behavior.

## 2. The Basics of a Qualitative Physics for Circuits

Most circuits are designed to deal with changing inputs or loads. For example, an amplifier must amplify changes in its input, digital circuits must switch their internal states as applied signals change, and power-supplies must provide constant current or voltage in the face of changing loads and power sources. The purposes of these kinds of circuits is best understood by examining how they respond to change. Thus, causal analysis is primarily concerned with how circuits, in some equilibrium state, respond to input perturbations.

## 2.1. Causal analysis

When an electrical engineer is asked to explain the operation of an electrical system he will often describe it in terms of a sequence of events each of which is 'caused' by previous events. Each event is an assertion about some behavioral parameter of some constituent of the system (e.g., current through a resistor). Sequential descriptions are ubiquitous in engineers' verbal and text-book explanations. Consider the Schmitt trigger (Fig. 1). The explanation reads as if a time flow has been imposed on it.

"... An increase in  $v_1$  augments the forward bias on the emitter junction of the first transistor, thereby causing an incremental increase in the collector current,  $i_{C1}$  of that transistor. Consequently both the collector-to-ground voltage  $v_1$  of the first transistor, and the base-to-ground voltage of the second transistor  $v_3$ , decrease. The second transistor operates as an emitter follower which has an additional load resistor on the collector. Therefore, there is a decrease in the emitter-to-ground voltage  $v_2$ . This decrease in  $v_2$  causes the forward bias at the emitter of the first transistor to increase even more than would occur as a consequence of the initial increase in  $v_1$  alone..." [19, p. 68]

The theories of qualitative, causal, and teleological reasoning provide the inference mechanisms to support this kind of explanation. The basic characteristics of the theory are evident in the preceding quote.

*The usual notion of variable.* The variables mentioned, e.g.,  $v_1$ ,  $v_1$ , etc., are the same ones used in quantitative analysis.

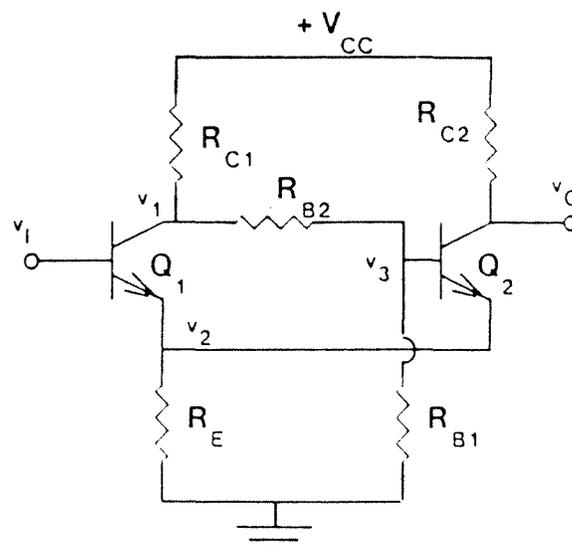


FIG. 1. The Schmitt trigger.

*The explanation focuses on change.* The most significant arena of behavior is how the circuit responds to input perturbations such as an increase in input.

*Change is described qualitatively.* All circuit quantities are described qualitatively: increasing, decreasing or unchanging.

*The explanation contains unstated assumptions.* Why does the  $v_1$  increment appear across  $Q_1$  instead of  $R_E$ ? Why does the voltage  $v_1$  drop, as  $Q_2$ 's turning-off should raise it? Why is the current contributed by  $Q_2$ 's turning-off more than the current taken by  $Q_1$ 's turning-on? All of these undesired possibilities occur if the component parameters are chosen inappropriately. Thus, they truly are assumptions. Although the actual values of the parameters can be used to validate assumptions, engineers rarely do so. Rather teleological reasoning is used to differentiate among the possibilities.

*Flow of causality.* The explanation is in terms of a sequence of events each caused by preceding events. The increase in  $v_1$  causes the increase in forward bias, which in turn causes an increase in  $i_{C1}$ , etc.

*Prediction and explanation.* The description of the Schmitt trigger both predicts qualitative values for circuit quantities and explains these values. Causal analysis is just one of many ways for explaining qualitative behavior. A causal explanation is a qualitative description of the equilibrating process that ensues when a signal is applied to the circuit. Before examining causal explanation, we must consider qualitative values and qualitative component models.

## 2.2. Forms of causality

Circuit theory is based on the notion of equilibrium conditions (see Section 3.3 for more details), and not on any notion of causality. Causality is a framework externally imposed by electrical engineers. One view of this causality is that it is a crude rendition of what actually happens when the circuit is examined closely (i.e., propagation of electrical fields) and that causality arises out of the solution of partial differential equations. Another view (and not necessarily inconsistent with the first), upon which this paper is based, is that causality arises out of viewing the system has a collection of information processors. Causation arises out of the passing of information between the parts of the system. Therefore we begin by examining the language for describing the behavior of system components.

## 2.3. Qualitative variables

The qualitative value of the expression  $x$  is denoted  $[x]$ . For brevity sake,  $[d^n x/dt^n]$  is abbreviated  $\partial^n x$  and  $\partial^0 x$  sometimes used for  $[x]$ . A qualitative variable can only take on one of a small number of values. These values correspond to intervals separated by points at which transitions of interest

occur. One quantity space [16] of particular interest is the sign of a quantity:  $[x] = +$  iff  $x > 0$ ,  $[x] = 0$  iff  $x = 0$  and  $[x] = -$  iff  $x < 0$ .

The behavior of components is described by qualitative equations (called confluences). Qualitative multiplication is easily defined for this quantity space as  $[x][y] = [xy]$ . Unfortunately, addition is more complicated.  $[x] + [y] = [x + y]$  unless  $[x] = -[y] \neq 0$  in which case addition is undefined.

#### 2.4. Qualitative quiescent and incremental analysis

Qualitative analysis utilizes a generalization of the electrical engineering distinction between quiescent and incremental analysis. The quiescent value of the quantity refers to the operating-point value (or steady-state value after the transients have died out). The incremental value is the deviation from the quiescent value. Using this distinction, circuit analysis is subdivided into two relatively independent subtasks: quiescent analysis with quiescent models and incremental analysis with incremental models.

Qualitative analysis generalizes this approach in two ways. First, the quiescent value of a quantity is defined to be its qualitative value and the incremental value of a quantity is defined to be the qualitative value of its rate of change.<sup>1</sup> Second, instead of having only quiescent and incremental models, qualitative analysis has models for every derivative order. The zero-order models are the quiescent models and the first-order models are the incremental models. As change is of primary interest, this paper concentrates on incremental qualitative analysis.

#### 2.5. Qualitative models for simple components

As in conventional electrical engineering we must take care in constructing the component models. The no-function-in-structure principle of qualitative physics states that each component of the same type must be modeled in the same way: we may not choose an idiosyncratic component model to handle a behavior of a particular circuit. Said differently, the component model must characterize the essential generic behavior of the component type. Otherwise, the structure-to-function analysis would be of little value because the structural choices already contained functional information.

Admittedly, a human engineer will use different quantitative models for different tasks, but he will only do so because he already understands how the circuit works and which approximations are valid. Presumably an analogous

<sup>1</sup>This does some violence to the usual electrical engineering notions. Usually, the total variable is defined to be the sum of the quiescent and incremental values for some small deviations from the operating point. In qualitative analysis, the quiescent value is defined to be the qualitative value of the total variable. The only reason this makes sense with regard to the usual engineering conventions is that the coarseness of the qualitative-value space makes deviations in a unit time interval and derivatives indistinguishable.

process occurs in selecting the qualitative models. However, the validity of the qualitative predictions directly depend on the quality of the component models he chooses. Unsubstantiated violations of the no-function-in-structure principle will lead to a faulty understanding of circuit behavior and function.

Constructing component models is a surprisingly difficult task. The standard engineering models that describe the behavior of electrical components are widely agreed upon. However, the causal qualitative models that people use to reason about circuits are not. These models are rarely articulated, even though the tacit models that underlie people's arguments appear to be very similar. One of the contributions of this research is a specification of these qualitative models. The approach for constructing the models is to start with the classical constraint models [13, 18] and reformulate them in qualitative terms. For simple components this process is relatively direct. Kirchhoff's laws are particularly easy to model qualitatively.

### 2.5.1. Kirchhoff's current law

The most basic network constraint is Kirchhoff's current law (KCL). It states that the sum of the currents  $i_j$  flowing into a component or node (or set of components and nodes) must be zero:

$$\sum i_j = 0, \quad (1)$$

where  $i_j$  are the currents in the terminals of the node, component or the boundary terminals of a set of nodes and components. In qualitative terms KCL is:

$$\sum [i_j] = 0. \quad (2)$$

However, for causal analysis the change in the quantities is more important than the quantities themselves. Differentiating equation (1) gives:

$$\sum di_j/dt = 0. \quad (3)$$

Thus, qualitatively:

$$\sum \hat{d}i_j = 0. \quad (4)$$

Note that while equation (1) implies (3), equation (2) does not imply (4). An unfortunate fact of the qualitative algebra is that  $d[x]/dt$  cannot be defined in terms of  $[x]$ , but rather as  $d[x]/dt = [dx/dt]$ . Just because  $[x] = [y]$  does not imply that  $[dx/dt] = [dy/dt]$ . For example,  $[1] = [1 + x^2]$ , but  $[0] \neq [2x]$ . Thus, although the qualitative calculus has the notion of derivative, it has no general

definition of differentiation. As a consequence, in general, a separate qualitative equation must be written each time a higher-order derivative is introduced. Fortunately, if a component is linear, all of its higher-order models have the same form. KCL is linear, thus the complete KCL model is:

$$\begin{aligned} i_0 &= \text{current in terminal } 0, \\ &\cdot \\ &\cdot \\ &\cdot \\ i_n &= \text{current in terminal } n, \\ \sum \partial^k i_j &= 0 \quad \text{for } k = 0, 1, \dots \end{aligned}$$

A law like Kirchhoff's current law is needed for many physical domains. The analogous law for fluid systems is conservation of matter, for thermal systems, conservation of energy, for mechanical systems, Newton's law for equilibrium of forces at a point.<sup>2</sup>

### 2.5.2. Kirchhoff's voltage law

Kirchhoff's voltage law states the sum of all voltages around any loop is zero. Qualitatively, for every loop of  $k$  nodes  $n_1, \dots, n_k, n_{k-1} = n_1$ :

$$\sum_{i=1}^k \partial^m v_{n_i, n_{i+1}} = 0 \quad \text{for } m = 0, 1, \dots$$

A law like Kirchhoff's voltage law is needed for many physical domains. The analogous law for fluid systems is that pressures must add (e.g., the sum of the pressure from point A to point B and the pressure from point B to point C must equal the pressure from point A to point C). For thermal and mechanical systems, temperatures and velocities must sum similarly.<sup>3</sup>

### 2.5.3. Ohm's law

Ohm's law has a particularly simple formulation. Assuming that resistors have fixed positive resistance, Ohm's law  $v = iR$  is expressed qualitatively as  $[v] = [i]$ . Currents are defined to flow into components away from nodes. The complete resistor model is:

$$\begin{aligned} i &= \text{the current in terminal } 1, \\ v &= \text{the voltage from terminal } 1 \text{ to terminal } 2, \\ \partial^k v &= \partial^k i. \end{aligned}$$

<sup>2</sup>These are all instances of the continuity condition of system dynamics [27] which is incorporated into qualitative physics [6].

<sup>3</sup>These are all instances of the compatibility condition of system dynamics [27], which is incorporated into qualitative physics [6].

### 2.5.4. Batteries

Batteries are also modeled simply. An ideal battery supplies constant voltage:

$v$  = voltage from terminal 1 to terminal 2 ,

$$[v] = + , \quad \partial^n v = 0 , \quad n > 0 .$$

### 2.5.5. Bipolar transistors

Transistors<sup>4</sup> have the incremental model

$\partial v_{b,e}$  = voltage from the base to emitter ,

$\partial i_c$  = current flowing into the collector ,

$\partial i_e$  = current flowing into the emitter ,

$\partial i_b$  = current flowing into the base ,

$$\partial v_{b,e} \Rightarrow \partial i_c , \quad \partial v_{b,e} - \Rightarrow \partial i_e , \quad \partial v_{b,e} \Rightarrow \partial i_b .$$

$\partial v_{b,e} \Rightarrow \partial i_c$  indicates that  $\partial v_{b,e} = \partial i_c$ , but  $\partial v_{b,e}$  causes  $\partial i_c$ , not vice versa. Likewise,  $\partial v_{b,e} - \Rightarrow \partial i_e$  indicates that  $\partial v_{b,e} = -\partial i_e$ , but that the change in  $v_{b,e}$  causes the change in  $\partial i_e$ . (The model for the transistor's other operating regions is presented in Section 9.)

## 2.6. Sign conventions

To readers familiar with the problems of sign conventions, this section is hopelessly pedantic and can be ignored.

There is an immense amount of confusion surrounding sign conventions. This is not helped by the fact that engineers are often rather cavalier about choosing them and often the precise meanings of variables are never made explicit. However, to formulate the component models and present the arguments constructed by EQUAL the sign conventions must be consistent. Admittedly, there is no significant information content in the sign conventions of variables, but without a precise definition of the quantities involved, the models and explanations are meaningless. The voltage from node 1 to 2 is notated  $v_{1,2}$ . The current flowing into R1 through terminal #1 is notated  $i_{\rightarrow \#1(R1)}$  (which is sometimes abbreviated  $i_{\#1(R1)}$ ) and the current out of R1,  $i_{\leftarrow \#1(R1)}$ .

As a consequence every circuit quantity can be expressed in two ways:  $v_{1,2} = -v_{2,1}$  and  $i_{\rightarrow \#1(R1)} = -i_{\leftarrow \#1(R1)}$ . In causal analysis, these choices are irrelevant. However, for engineers they are very significant. For example, ground (G) is often taken as a common node to which all voltages are references. Thus  $v_{1,G}$  is much preferred over  $v_{G,1}$ . When the voltage is referenced to ground, ground is dropped:

<sup>4</sup> NPN or PNP.

$v_{1,G} = v_1$ . If  $v_{1,2}$  is negative it is much more intuitive to use  $v_{2,1}$  because it refers to a change in a positive quantity. An NPN transistor has positive base and collector currents but negative emitter currents. As a consequence, engineers will often define  $i_C$  and  $i_B$  to be the currents flowing into the transistor, and  $i_E$  to be the current flowing out of the transistor. I notate these  $i_{\rightarrow C(Q)}$ ,  $i_{\rightarrow B(Q)}$  and  $i_{\leftarrow E(Q)}$ .

### 3. Incremental Causal Analysis

The central aspect of how a circuit works is its response to change. Thus, in this section, I ignore quiescent models, higher-order models, and how they are used. Likewise, the determination of the state of a component or its state transitions is discussed later. Changing variables over time (i.e., integration) is also left until later. All components are assumed to behave linearly within a state. Finally, qualitative analysis is assumed to be unambiguous. All these assumptions are important and are dealt with later, but are unnecessary to understand the basic structure-to-function framework.

#### 3.1. Causal propagation

The essential intuition behind causal analysis is that each component acts as an individual information processor which only performs an action when a signal reaches it from one of its neighbors. Then it passes along the signal to its neighbors according to its component model.

This notion of causality is based on two fundamental tenets: locality and directionality. Each component acts only when it is acted upon by neighboring components, and then it only acts on its neighbors. Thus, each interaction only has access to local information (i.e., from its neighboring components), and produces only local information. Finally, each component propagates information in one and only one direction. Thus, negotiation between component processors is impossible.

Causal analysis produces a causal argument which is a qualitative description of how the circuit equilibrates—how it responds to perturbations from its equilibrium. This description is, in effect, a simulation of the circuit's equilibration. This kind of explanation is often what people mean by a description of how something 'works'. Not surprisingly, the causal analysis process itself uses simulation in producing its analyses (complications which are introduced later require a great deal of additional problem-solving effort).

Consider as an example the circuit of Fig. 2. This circuit contains a transistor (Q1), a resistor (R1), and a battery (B1). These are connected by four groups of wires (i.e., nodes) IN, OUT, VCC, and G.

Components interact with each other by placing currents and voltages on their adjacent terminals. Thus, two components can only interact if there is a wire connecting them. The topology of potential causal interactions can be laid out as a topology of alternating component models and variables as is illus-

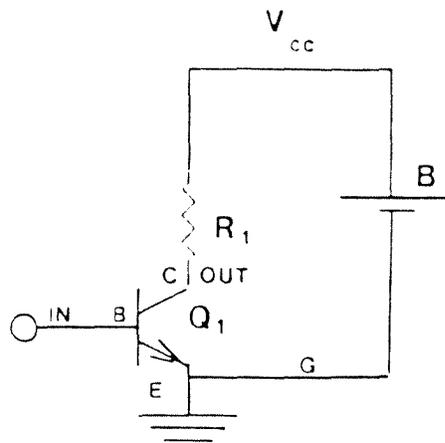


FIG. 2. Simple amplifier.

trated in Fig. 3. Each circle indicates a component model which can only communicate information through shared variables. Causal analysis proceeds by propagating an input disturbance through this information topology. In this example, the propagation is very simple. Fig. 4 illustrates the causal topology for the simple amplifier when an increase is applied. Notice that the edges in the causal topology are directed corresponding to the direction of propagation and that only a subset of the potential causal interactions are utilized. The fundamental task of causal analysis is to determine that subset of potential interactions that 'best' describes the functioning of the circuit. We will see in a moment that for simple circuits such as that of Fig. 2 this task is relatively simple, but fundamental complexities arise for more complex circuits.

The causal argument for an output consists of a sequence of steps, starting with the input disturbance, each depending on previous steps in the sequence

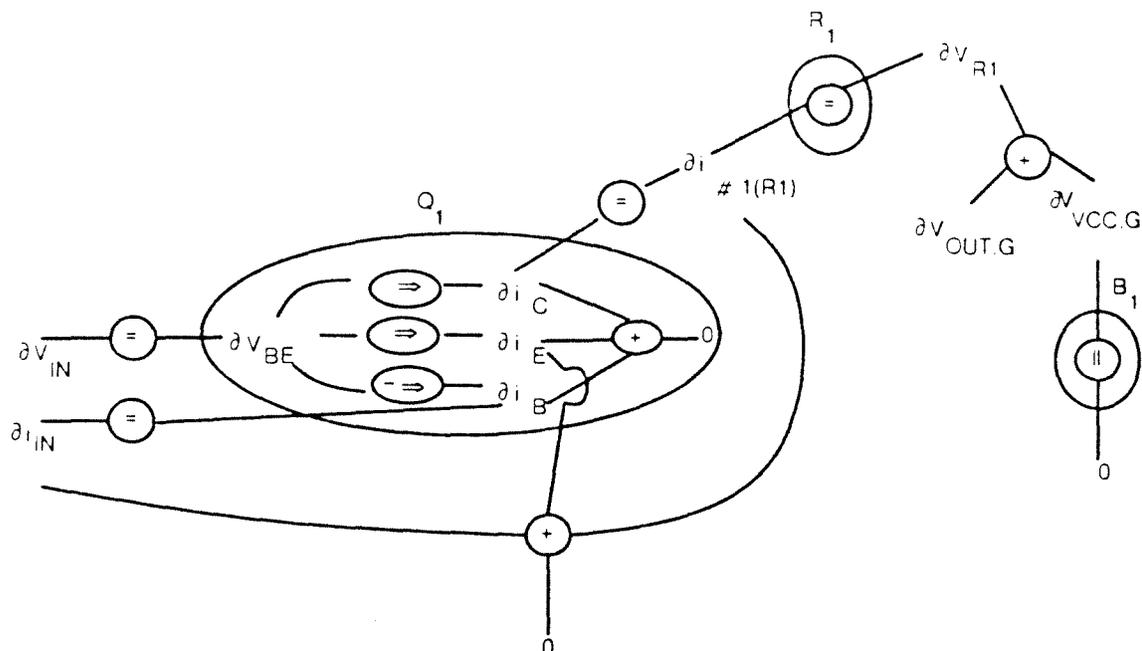


FIG. 3. Potential causal topology of Fig. 2.

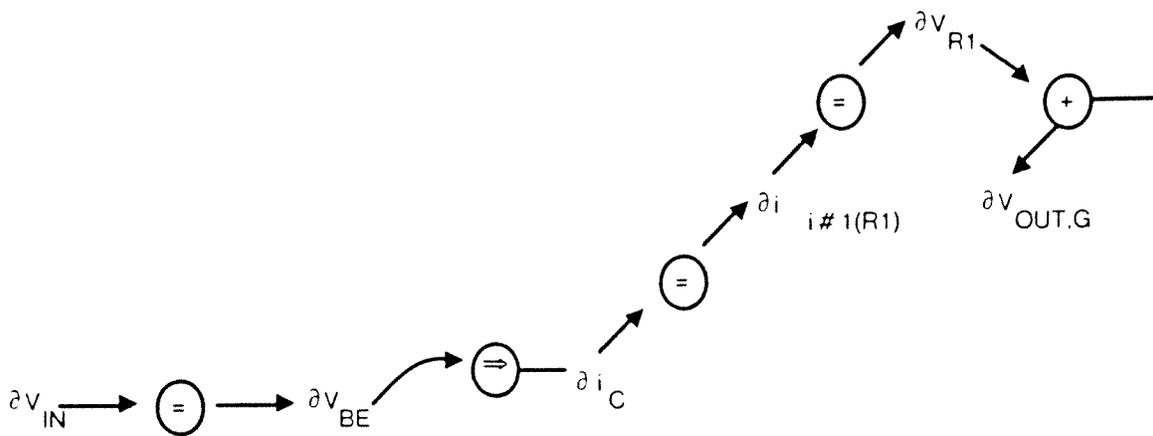


FIG. 4. Causal topology of Fig. 2.

and terminating in the output. The following is EQUAL's formulation of the causal argument of the simple amplifier. Each step in the argument is placed on a single line. The steps are organized into four columns. The first column indicates the antecedents to the step. If the second column is blank, that step is a premise (e.g., an input disturbance).  $\Rightarrow$  indicates that the step follows directly from its antecedents. If none are listed in the first column and the second column is  $\Rightarrow$  the step's antecedent is the previous line. The third column indicates the variable assigned a value in that step. The final column provides the reason which always refers to some component model. Note that Ohm's law is defined with respect to the #1 terminal of a resistor. As #1 is the top terminal in this circuit, this current must be propagated by KCL from #2 which is connected to the collector to #1 which is connected to the battery. As quiescent current is flowing down through the resistor to the collector, EQUAL has chosen sign conventions such that current flows out of the bottom of R1 and into the tops of R1 and Q1.<sup>5</sup>

Antecedents	Event	Reason
	$\partial v_{IN} = +$	Given
	$\Rightarrow \partial i_{C(Q1)} = +$	$\partial v_{B,E} \Rightarrow \partial i_C$ for NPN-transistor Q1
	$\Rightarrow \partial i_{\#2(R1)} = +$	KCL for node OUT
	$\Rightarrow \partial i_{\#1(R1)} = +$	KCL for resistor R1
	$\Rightarrow \partial v_{VCC,OUT} = +$	$\partial i_{\#1} \Rightarrow \partial v_{\#1,\#2}$ for resistor R1
	$\partial v_{VCC} = 0$	$\partial v_{+,+} = 0$ for battery B1
$\partial v_{VCC} = 0, \partial v_{VCC,OUT} = +$	$\Rightarrow \partial v_{OUT} = -$	KVL applied to nodes OUT, VCC, GROUND.

<sup>5</sup>EQUAL tries to determine the signs of the quantities and all the causal arguments presented in this paper adopt sign conventions such that the underlying quantity is positive. In a few cases, EQUAL cannot determine the sign of the underlying quantity, in which case the voltage choice is arbitrary and currents are chosen to flow into components. These sign conventions do not affect the chain of causal interactions produced by the disturbance, and are thus irrelevant to teleological analysis. The necessary quiescent analysis is discussed later. However, without adopting these sign conventions the causal explanations would be nearly incomprehensible.

In English: The increasing voltage at node IN is applied to the base-emitter junction of Q1. By the transistor law  $\partial v_{BE} \Rightarrow \partial i_C$  this causes an increase in current from node OUT into the collector of Q1. By Kirchhoff's current law, this same current is flowing out of the bottom of R1 into node OUT. By Ohm's law, this causes an increased voltage across it. Thus, the voltage from node VCC to OUT increases. Batteries supply constant voltage so the voltage between VCC and GROUND is constant. Thus, the increase in the voltage from VCC to OUT also produces an increase in the voltage from GROUND to OUT. Hence, the voltage at OUT drops with respect to GROUND.

### 3.1.1. *Nonlocality of KCL and KVL*

The formulations of KCL (Section 2.5.1) and KVL (Section 2.5.2) violate the locality tenet of causality. Fortunately, no important information is lost if KCL is restricted to be local (i.e., only applies to single components and nodes). However, KVL is inherently nonlocal:  $v_{ij} + v_{jk} = v_{ik}$  where nodes  $i$ ,  $j$  and  $k$  may range over all the nodes in the circuit. If KVL were restricted to be purely local, i.e., only apply if nodes  $i$ ,  $j$  and  $k$  are all connected to the same component, the causal propagation of the previous section could not go through.

The nonlocality of KVL is inescapable as it arises from the definition of voltage. The identical problem arises with temperatures, velocities, pressures, etc. The KVL-like law arises out of the very definition of voltage, pressure, temperature or velocity. This is illustrated by the more commonsense quantity of temperature. If the temperature difference between points A and B is  $t_1$  and the temperature difference between points B and C is  $t_2$  then the temperature difference between points A and C is  $t_1 + t_2$ . This non-local inference arises out of the basic definition of temperature.<sup>6</sup>

## 3.2. Implementation

Causal analysis uses propagation of constraints [10,28,29,30] to determine circuit values. The propagator is designed to find the simplest explanation first. KCL and KVL deductions are considered simpler than component models, thus all KCL and KVL deductions are done first. On rare occasions the same cell value can be found in two different ways. EQUAL adopts the simple heuristic of picking the value utilizing fewest components in its argument. The breadth-first phase of the propagations is implemented using a queue to hold newly assigned values which have not yet been propagated.

<sup>6</sup>Under the approximations of conventional circuit theory, Maxwell's equations can be solved by a scalar potential, this potential is the definition of voltage. A scalar potential has the property and its change from point A to point B is independent of the path taken to reach B from A. This is exactly KVL.

The data structure that propagation operates on is a set of cells representing the voltages and currents of the circuit and a set of constraints linking those cells. A new cell is allocated for each derivative of every variable. This data structure is constructed from the schematic by creating a voltage cell for every pair of nodes and a current cell for every component terminal. Then the component models, KCL and KVL are instantiated for each component and node by linking each confluence to their corresponding cells. The result is a data structure of cells and constraints linking those cells.

Each cell may have a value. This value may come from outside the circuit or may be deduced from values in other cells via the constraint confluences. When a cell is assigned a value each constraint it participates in is examined to determine if enough information is available to enable it to use that constraint to deduce a value for another cell. Discovering a new value may thus determine yet other values, thus 'propagating the constraints'. Associated with each value is a justification indicating how it derives from values in other cells. A complete audit trail of a value's propagation is constructed by tracing this justification path back to the initial inputs. The audit trail is the causal argument.

Sometimes two values are found for the same cell. This does not mean there are two competing tendencies on the cell: causality demands that once a cell receives a value it remains at that value. If the two values of such a coincidence conflict, the component models are faulty (e.g., a transistor model with a voltage source on the collector instead of a current source), the circuit cannot exist (e.g., two current sources in series) or the inputs are impossible (e.g., specifying that the current into the #1 terminal of a resistor is + and that the current into the #2 terminal is also +). If the two values of a coincidence corroborate, propagation of the second value is halted as no additional information can be gained by propagating it. The trivial cases of corroborations are avoided by a refraction rule disallowing the result of a constraint to trigger itself.

It is important to discover conflicts as early as possible. It makes no sense to create a long propagation chain only to discover that some simple antecedent is contradictory. This is why the propagation proceeds breadth-first.

Unilateral constraints such as  $\partial v_{b,e} \Rightarrow \partial i_c$  for transistors require special handling. However, for well-designed circuits and correct models this unilateral restriction is never necessary and no analysis in this paper requires it.

KVL presents two sorts of problems. The first, discussed in more depth later, concerns how many KVL confluences to use. Including a confluence for *every* loop is extremely redundant. Redundancy is undesirable not because it introduces inefficiency, but rather that it gives rise to multiple syntactically different explanations for the same behavior. Using the conventional quantitative methods for formulating KVL confluences produces insufficient confluences. EQUAL includes a KVL confluence for every set of three nodes:

For every set of three nodes:  $i, j$ , and  $k$

$v_{ij}$  = voltage from node  $i$  to node  $j$ ,

$v_{j,k}$  = voltage from node  $j$  to node  $k$ ,

$v_{i,k}$  = voltage from node  $i$  to node  $k$ ,

$$\partial^n v_{ij} + \partial^n v_{j,k} = \partial^n v_{i,k} \quad \text{for } n = 0, 1, \dots$$

However, this formulation is still a source of redundancy. EQUAL implements KVL as a procedure which is invoked whenever a new voltage is discovered. It only 'creates' three-node KVL constraints when necessary, thus increasing efficiency by avoiding explicitly encoding all the KVL confluences and reducing redundant corroborations by not introducing KVL confluences when they are not needed. Unfortunately, it is not possible to avoid all redundancy.

### 3.3. What is causality?

The notion of causality is not that well understood or agreed upon. As the arguments of this paper depend crucially on a particular notion of causality, this section briefly outlines the basic intuitions behind my notion of causality. Consider these steps of the causal argument for the circuit of Fig. 2:

Antecedents	Event	Reason
	$\Rightarrow \partial i_{\leftarrow \#2(R1)} = +$	KCL for node OUT
	$\Rightarrow \partial i_{\rightarrow \#1(R1)} = +$	KCL for resistor R1
	$\Rightarrow \partial v_{VCC,OUT} = +$	$\partial i_{\#1} \Rightarrow \partial v_{\#1,\#2}$ for resistor R1

This argument is written as if the change in current in R1 somehow causes the change in voltage across R1. This is absurd, as according to Ohm's law  $v = iR$  any change in  $v$  must co-occur with any change in  $i$ . So what is causality?

There is no notion of delay in equilibrium models—all signals propagate in zero time and all circuit quantities change simultaneously when an input is applied. The network laws, KCL, KVL, etc., are all equilibrium conditions. They are quasistatic approximations which only apply if the wavelengths of the signals are long enough, energy levels of the signals are low enough and observation times of the signals long enough. Mathematically the behavior of the lumped-parameter system is described by *ordinary* differential equations.

On the other hand our intuitive notion of causality is quite different. The change in input causes a disequilibrium in the components near the input. This disequilibrium propagates through the circuit at finite speed until the circuit reaches a new overall equilibrium state. This equilibrating process is described by causal analysis. Note that during the equilibrating process, circuit laws do not apply.

In a way, this intuition corresponds to what actually happens if we examine the circuit closely. A far more accurate quantitative model is based on

Maxwell's field equations. The lumped-parameter circuit model is an idealization and simplification of the behavior of the electromagnetic fields in and around the circuit components. As changes in these fields propagate at finite speeds, this process takes a certain amount of time. The differential equations of the lumped-parameter circuit model cannot account for what happens during this period of disequilibrium. Within this period the changing fields of the input signals propagate until global equilibrium is reached. This propagation can be viewed as a kind of causal flow: the input field changes and propagates to other materials causing further fields to change. These changes can be partially ordered in a time sequence in which each change is caused by changes earlier in the sequence and earlier in time.

The quantitative calculation of the causal flow that happens during the period of disequilibrium is intractable. Although the electromagnetic laws that govern the physics are known, there is no practical way to quantitatively describe this causal process. This means that the electrical engineer can never completely analyze a circuit.

When an electrical engineer reasons about a circuit, he reintroduces a kind of causality that the lumped-circuit model throws away. He does this by using local causal models and imposing a time flow on the changes in circuit quantities. Only by creating local causal models that throw away much of the detail that of the quantitative view is he able to make the causal analysis tractable. The models are not based on Maxwell's equations, rather each circuit component is viewed as an active agent which attempts to re-establish a local equilibrium within its own vicinity. The engineer's qualitative theory of circuit causality explains the period of equilibration by introducing finite time flow and permitting the circuit to be in disequilibrium.

The formal linkage between the causal models derived earlier and Maxwell's equations is unexplored. However, the sense of causality presented in this paper seems to be an accurate description of engineer's reasoning and hence connects with the conventional language of electrical engineering. The models are sufficient to produce accurate predictions and explanations for circuits.

### 3.4. Causal heuristics

The component models alone are insufficient to construct causal analyses of some circuits. Consider the simple circuit fragment illustrated in Fig. 5. This circuit is like that of Fig. 2 except that an extra resistor is placed in the emitter and unspecified circuitry is attached to the emitter and collector. In the present framework there is no causal analysis for this circuit fragment.

The propagator can get no further than the voltage at the input. The only component that even references the input variable and hence has any hope of propagating the value is Q1. But Q1 responds to voltage across its base and emitter, not the voltage between its base and ground. Hence the propagation gets stuck. (In quantitative analysis this would be a point at which to introduce a new unknown variable to propagate around the loop.)

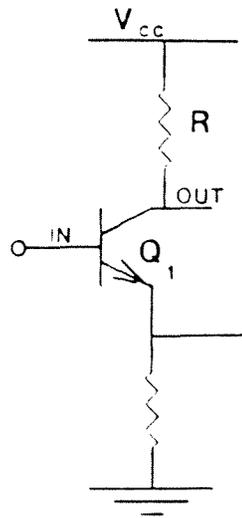


FIG. 5. Amplifier input.

### 3.4.1. *The three causal heuristics*

Propagation impasses can only occur at sums of three or more quantities. The propagation cannot halt at  $x = y$  since if  $x$  is known  $y$  can be computed and vice versa. Sums of three or more quantities arise from KVL, KCL and (rarely) the component models themselves. EQUAL incorporates three heuristics to circumvent impasses at each of the types of sums: the KVL-heuristic, KCL-heuristic, and the confluence heuristic.

These heuristics are domain-independent. No matter what the domain, propagation will encounter impasses. The general solution is the introduction of three domain-independent heuristics: the component heuristic, the conduit heuristic and the confluence heuristic which are identical to the electrical KVL-, KCL- and confluence heuristics.

Unlike KVL, KCL and component-confluence deductions, the conclusions of heuristics are necessarily tentative being based on assumptions about circuit functioning that may later be found to be incorrect.

### 3.4.2. *The KVL-heuristic*

The KVL-heuristic handles the case where the propagation reaches an impasse at the KVL confluence  $\partial v_{n,G} = \partial v_{n,m} + \partial v_{m,G}$  where  $n$  and  $m$  are nodes connected to different terminals of the same component and  $\partial v_{n,G}$  is + or -. The propagator always employs the rule  $\partial v_{n,G} \Rightarrow \partial v_{n,m}$ . In addition, if  $\partial v_{n,G} = \partial v_{m,G}$  the propagator also employs  $\partial v_{m,n} = 0$ .

It is now possible to propagate past the initial impasse of Fig. 2:

$$\begin{array}{ll}
 \partial v_{IN} = + & \text{Given} \\
 \Rightarrow \partial v_{IN,E1} = + & \partial v_B \Rightarrow \partial v_{B,E} \text{ KVL-heuristic for NPN-transistor Q1} \\
 \Rightarrow \partial i_{-C(Q1)} = + & \partial v_{B,E} \Rightarrow \partial i_C \text{ for NPN-transistor Q1.}
 \end{array}$$

The intuitive rationale for this heuristic is as follows. Consider the simple case where  $\partial v_{n,m}$  and  $\partial v_{m,G}$  are unknown. No causal path has yet been found which affects  $\partial v_{m,G}$ . Thus,  $\partial v_{n,G}$  can be viewed as causing the changes in  $v_{n,m}$  and  $v_{m,G}$ . The KVL-heuristic assumes that the change of voltage at node  $n$  produces an equal-valued change in the voltage across the component and that the voltage at  $m$  is insignificant (i.e.,  $v_{m,G} \approx 0$ ). It is very likely that, as a consequence,  $v_{m,G}$  is affected as well:  $\partial v_{n,G} \Rightarrow \partial v_{n,m} \Rightarrow \dots \Rightarrow \partial v_{m,G}$ . Thus the KVL-heuristic is based on the presupposition that the change in  $v_{m,G}$  produced by the change in  $v_{n,G}$  cannot swamp the change in  $v_{m,G}$ .

In the case where  $\partial v_{n,G} = \partial v_{m,G} \neq 0$ , the KVL-heuristic will already have produced two conflicting propagations:  $\partial v_{n,G} \Rightarrow \partial v_{n,m}$  and  $\partial v_{m,G} \Rightarrow \partial v_{m,n}$ . However, the two changes might exactly cancel requiring the consideration of the third possibility:  $\partial v_{n,m} = 0$ . (No one would ever design a circuit that depended on this, but nevertheless it is a mathematical possibility.)

From a network theory point of view the KVL-heuristic is based on the presupposition that all incremental impedances are positive. If the situation can be modeled as illustrated in Fig. 6 a positive component impedance in series with an unknown composite positive impedance, then KVL-heuristic is correct.

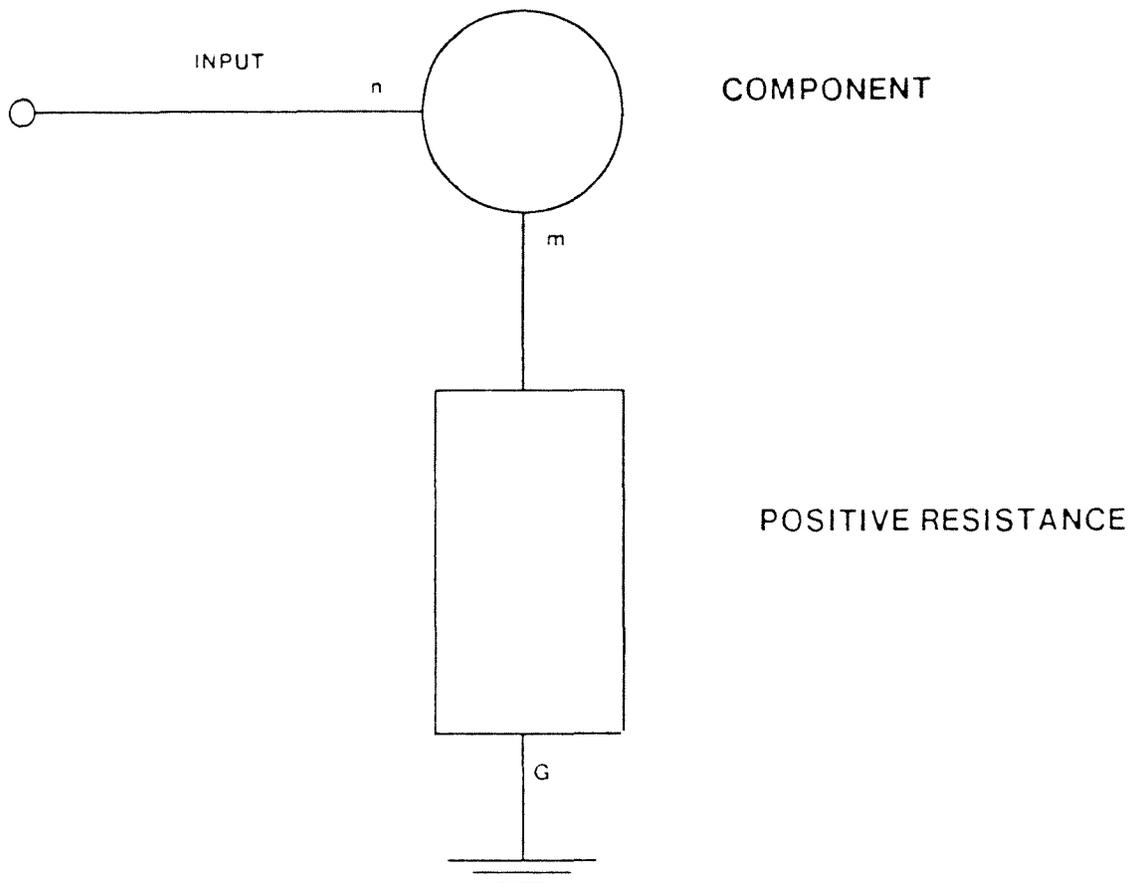


FIG. 6. Rationale for KVL-heuristic.

### 3.4.3. The KCL-heuristic

The KCL-heuristic handles the case where the propagation reaches an impasse at a KCL confluence for some node (which can only happen if the node has three or more component terminals connected to it). The rule is that if components directly cause terminal currents of a node, the voltage at the node is determined by the equation  $\sum \partial i_{jk} \Rightarrow \partial v$ . In addition, if some  $\partial i_j = -\partial i_n$  the propagator employs  $\partial v = 0$ .

It is now possible to complete the causal argument for the output of Fig. 2:

$$\begin{array}{ll}
 \partial v_{IN} = + & \text{Given} \\
 \Rightarrow \partial v_{IN,E1} = + & \partial v_B \Rightarrow \partial v_{B,E} \text{ KVL-heuristic for NPN-transistor Q1} \\
 \Rightarrow \partial i_{-C(Q1)} = + & \partial v_{B,E} \Rightarrow \partial i_C \text{ for NPN-transistor Q1} \\
 \Rightarrow \partial v_{OUT} = - & \partial i_{-C(Q1)} \Rightarrow \partial v_{OUT} \text{ KCL-heuristic for node OUT.}
 \end{array}$$

The KCL-heuristic is extremely common in an engineer's analyses: "Increasing current *out of* the node pulls it down, increasing current *into* the node forces it to rise."

The KCL-heuristic makes the presupposition that terminals which are causing current flow into nodes can be modeled as the output terminal of a current source. This current source sees an incremental positive impedance to ground. Fig. 7 illustrates this situation. If this is an accurate depiction of the situation, the voltage at the node can be determined directly from Ohm's law.

### 3.4.4. The confluence heuristic

The final source of three or more variable confluences is the component models themselves. However, such confluences are relatively rare. One such case occurs when the transistor is saturated. When the transistor is saturated, the change in base current is controlled by the combined action of the two junctions:  $\partial v_{b,e} - \partial v_{c,b} \Rightarrow \partial i_b$ . The ' $\Rightarrow$ ' in this context indicates the  $\partial v_{b,e} - \partial v_{c,b} = \partial i_b$ , but the change in voltages cause the change in the current. If the propagation reaches an impasse on this confluence, the confluence heuristic, is invoked. Two other cases are a valve where the area available for flow controls both flow and pressure [6] and a mosfet [34].

All the cases I have encountered where this heuristic is used are of the form

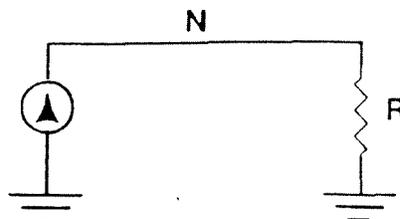


FIG. 7. Rationale for KCL-heuristic.

$\partial x + \partial y \Rightarrow \partial z$  where  $\partial x$  or  $\partial y$  are known. In such circumstances the propagation proceeds with  $\partial z \Rightarrow \partial x$  or  $\partial z \Rightarrow \partial y$  leaving the unknown  $\partial y$  or  $\partial x$  alone.<sup>7</sup>

### 3.4.5. Interestingness

These three heuristics are designed to find values for the important (i.e., interesting) circuit variables. The heuristics will not attempt to find a value for a voltage unless it is a component input, or it is a voltage which is (a) with respect to ground, and (b) has current flowing into the non-ground node. The heuristics will never directly attempt to find a value for a current unless it is part of a component confluence of more than two variables. The remaining variables are presumed to be of no causal importance, or automatically follow from important variables so that their values can be safely ignored. EQUAL adopts the convention that uncaused uninteresting variables are zero.

## 3.5. Causal analysis of the Schmitt trigger

We are now in a position to completely analyze the behavior of the Schmitt trigger. When an increase is applied to the input of the Schmitt trigger (Fig. 8), the voltage at the emitter drops. EQUAL produces the following causal argument for the drop in emitter voltage:

$\partial v_{B1} = +$	Given
$\Rightarrow \partial v_{B1,E1} = +$	$\partial v_B \Rightarrow \partial v_{B,E}$ KVL-heuristic for NPN-transistor Q1
$\Rightarrow \partial i_{\rightarrow C(Q1)} = +$	$\partial v_{B,E} \Rightarrow \partial i_C$ for NPN-transistor Q1
$\Rightarrow \partial v_{C1} = -$	$\partial i_{\rightarrow C(Q1)} - \Rightarrow \partial v_{C1}$ KCL-heuristic for node C1
$\Rightarrow \partial v_{C1,B2} = -$	$\partial v_{\#1} \Rightarrow \partial v_{\#1,\#2}$ KVL-heuristic for resistor RB2
$\Rightarrow \partial i_{\rightarrow \#1(RB2)} = -$	$\partial v_{\#1,\#2} \Rightarrow \partial i_{\#1}$ for resistor RB2
$\Rightarrow \partial i_{\rightarrow \#2(RB2)} = +$	KCL for resistor RB2
$\Rightarrow \partial v_{B2} = -$	$\partial i_{\rightarrow \#2(RB2)} - \Rightarrow \partial v_{B2}$ KCL-heuristic for node B2
$\Rightarrow \partial v_{B2,E1} = -$	$\partial v_B \Rightarrow \partial v_{B,E}$ KVL-heuristic for NPN-transistor Q2
$\Rightarrow \partial i_{\leftarrow 2E(Q2)} = -$	$\partial v_{B,E} \Rightarrow \partial i_E$ for NPN-transistor Q2
$\Rightarrow \partial i_{\leftarrow 2E(Q2)} = -$	$\partial i_{\leftarrow E(Q2)} \Rightarrow \partial v_{E1}$ KCL-heuristic for node E1

Of course, this argument only holds if the transistors are in their correct operating regions, the resistances and gains are just right, and that the input level is high enough—these issues are all discussed in later sections. If the conditions are right for the Schmitt trigger to operate as intended, it behaves as described by the preceding argument.

<sup>7</sup>The generalization is straightforward. The confluence is of the form  $\Sigma \partial x \Rightarrow \Sigma \partial y$ . The heuristic computes a value for the left-hand side assuming all unknown  $\partial x$  are zero. Then it hypothesizes all possible assignments to the unknown  $\partial y$  which are consistent with the computed value for the right-hand side.

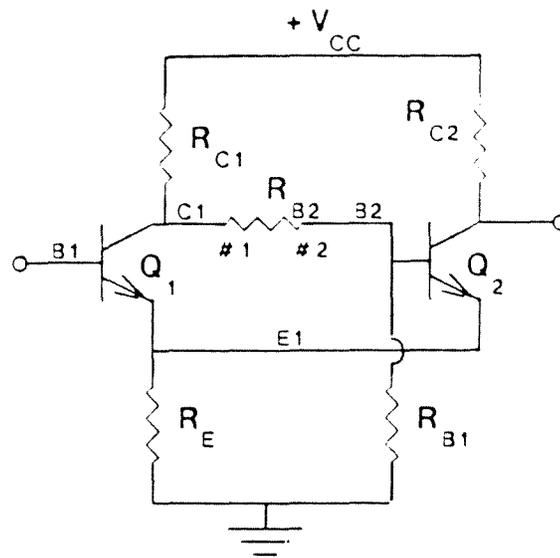


FIG. 8. Schmitt trigger.

### 3.6. Implementation

As all of the rules and heuristics are order-insensitive, a heuristic may be applied at any time in the propagation process. As an application of a heuristic may later be shown to be invalid or superfluous, an explicit assumption is recorded for each use of a heuristic. This assumption may need to be retracted if subsequent contradictions are encountered. For example, when the KVL-heuristic notices a voltage at the base of a transistor, it triggers the transistor model on the base-emitter voltage. In doing so, the heuristic makes the assumption that the emitter voltage's effect is negligible compared to that of the base voltage. If it is discovered later that the emitter voltage is significant, the assumption and all its ensuing propagations need to be retracted.

As the signal propagates through the circuit it may accumulate a number of such assumptions. Although the assumptions of a value can be computed from its justification, it is more convenient to propagate the assumptions of a value along as well. Thus, the assumptions of a newly deduced value is computed by constructing the set union of the assumptions of its antecedents. More precisely, EQUAL propagates value triples of the form  $\langle qvalue, \{assumptions\}, justification(s) \rangle$  (where *qvalue* is one of +, - or 0). For example, if  $\partial x = \partial y + \partial z$ , and  $\partial y = \alpha: \langle +, A_\alpha \rangle$  and  $\partial z = \beta: \langle +, B_\beta \rangle$ , then the propagator would deduce  $\partial x = \gamma: \langle +, A_\alpha \cup B_\beta, +(\alpha, \beta) \rangle$ .

Although the ordering of the propagation steps is irrelevant to completeness, it has significant effect on efficiency. There is no point propagating a value for a cell which is subsequently replaced by a simpler value or contradicted. EQUAL orders its propagations such that it finds contradictions as early as possible and finds the simplest causal argument for each value. EQUAL has three priorities for propagation, first KCL and KVL deductions, second component model deduc-

tions, and third KVL- and KCL-heuristics. The highest priority value is always propagated first. For example, if there are two possible independent KCL-heuristics pending and one is run, the next propagation will be KVL deductions from the new voltage, not the other KCL-heuristic. All three priorities are propagated breadth-first.

In order to distinguish values under different assumptions the notation  $[C x]$  is used to indicate the assumption that  $x$  is the dominant input to  $C$ . The assumption that  $v_B$  is the dominant input to  $Q$  is written  $[Q v_B]$ . Assumptions  $[Q v_E]$  and  $[Q v_B]$  can both be active, potentially producing contradictory values for  $i_C$ . Assumptions introduced by the KCL-heuristic are notated analogously. For example, the KCL-heuristic used in analyzing Fig. 5 is notated  $[\text{OUT } Q1]$  as  $Q1$  is the dominant input to node  $\text{OUT}$ .

#### 4. Feedback

Feedback, the idea of a device sensing its own output to control its own behavior, is one of the most powerful strategies for designing and analyzing circuits, mechanical systems, economic systems, biological systems, etc. Feedback is as important and prevalent to machine design as loop constructs are to programming. The benefits of negative feedback to circuit design are varied and extremely important. Feedback is primarily used to compensate for non-ideal components or operating conditions. In using feedback, the designer embeds the circuit fragment whose behavior is to be modified within a feedback loop, thereby mitigating undesirable behavior and enhancing the desirable behavior. Feedback can be used to stabilize circuit gain against variations in circuit parameters due to manufacturing or temperature. It allows the designer to modify the input and output impedances of the circuit, reduce distortion by improving linearity, and increase the bandwidth characteristic of amplifiers. For these reasons the use of feedback is ubiquitous in circuit design.

Feedback is also critical for understanding circuit behavior. It is the basic building block for designing analog circuits. It is also extremely important as a design strategy. Troubleshooting circuits with feedback requires understanding the feedback action. In addition to detecting it, the feedback must be characterized in the standard conventions of electrical engineers. This section presents how causal analysis is used to locally detect and analyze feedback.

##### 4.1. Definition of feedback

Feedback cannot be directly observed in the voltage and current values. The quantitative behavior of a circuit is determined by solving the simultaneous set of equations obtained from characteristics of each component, KCL and KVL using either the node or mesh method [13]. Whether the circuit contains feedback or not has no effect on this process. It is impossible to tell from the solution whether the circuit utilizes feedback. But an engineer examining a

circuit can determine whether it uses feedback. He does not do this using conventional network theory, but relies on something quite different—the topology or causality of the circuit.

A feedback amplifier can be viewed as a basic amplifier surrounded by extra circuitry to modify its behavior. Fig. 9 illustrates the basic parts of a feedback system. The input signal passes through the summing network, is amplified and subsequently output. However, as well as being connected to the overall circuit output, the main amplifier is connected to another network. The signal passes through this network and is subtracted from the input signal. Thus a fraction of the output is sampled and subtracted from the input. The advantage of this arrangement can be seen by considering what happens if the gain of the basic amplifier is suddenly disturbed. As the gain is reduced, the output signal drops. Thus less signal is feedback and less is subtracted from the input signal. As there is now more input to the main amplifier, the output increases again, thereby correcting for the disturbance.

When an engineer recognizes feedback in a circuit, he is noticing that the circuit somehow abstractly matches Fig. 9. He could try to recognize the basic amplifier, network and summing and sampling points. Recognizing these modules requires considerably more expertise than just knowledge of network theory. For example, most topological loops in the circuit schematic do not correspond to instances of feedback loops, let alone important ones.

Feedback is ultimately a property of the circuit's function (i.e., how it works), not its behavior (i.e., what it does). Although quantitative analysis (i.e., using network theory) cannot be used to identify feedback, causal analysis can. That is because causal analysis explains the *how* as well as the *what* of circuit behavior.

Feedback is an elusive concept. Devices based on feedback have been with mankind since antiquity [35] but feedback was not appreciated as an explicit design tool until James Watt's invention of the centrifugal governor in the late

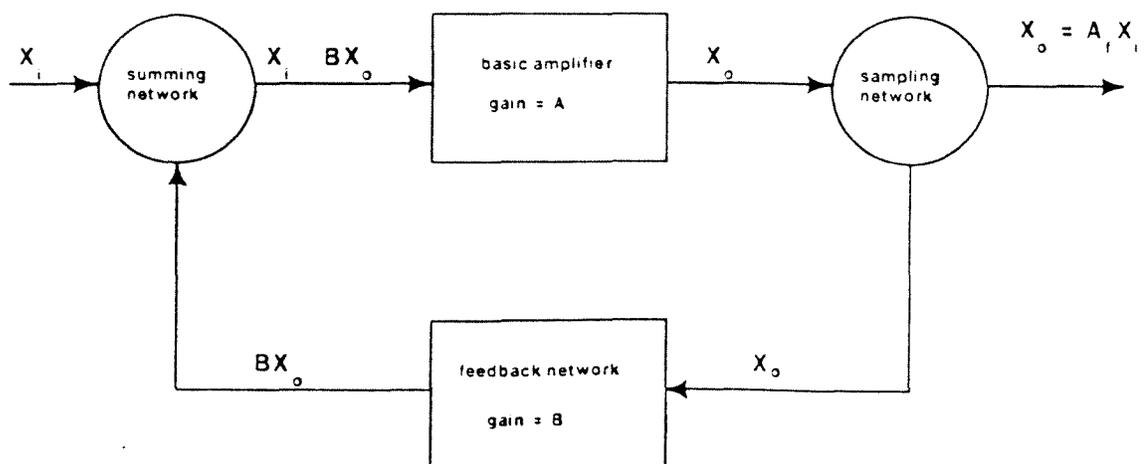


FIG. 9. Block diagram of a feedback system.

18th century. The abstraction of Fig. 9 is an invention of the twentieth century. Mayr's book "The Origins of Feedback Control" takes some pains to define the criteria for feedback in order to identify which machines in antiquity used feedback. Analogous care must be taken in causal analysis, for otherwise, feedback is seen everywhere. Intuitively, feedback occurs when a chain of causal interactions falls back on itself (as is the case with the Schmitt trigger) producing an endless chain of causal interactions. However, by that definition two resistors in series manifest feedback! Mayr introduces a very important criteria "The system includes a sensing element and a comparator, at least one of which can be distinguished as a physically separate element."

#### 4.2. Feedback detection

The heuristics provide a clean way to detect feedback. All occurrences of the heuristics are at sums of three or more variables. Each occurrence where two or more of the variables are unknown are potentially points at which feedback occurs. The variables at each such point can be divided into three classes: inputs, outputs and insignificants. For example, given  $\partial x + \partial y + \partial z = 0$ , the heuristic might have, given  $\partial x$ , propagated  $\partial x \Rightarrow \partial y$  assuming  $\partial z$  to be insignificant. *Feedback is present if, as a consequence of propagating the output variable(s), a value is propagated (perhaps after many steps) for the variables assumed to be insignificant.* For example, if  $\partial x \Rightarrow \partial y \Rightarrow \dots \Rightarrow \partial z$ , the circuit contains a feedback loop from  $\partial x$  through  $\partial y$  to  $\partial z$ . In particular, the propagator has discovered that  $\partial z = f(\partial y)$  and thus that  $\partial y = -f(\partial y) - \partial x$ , a clear indication of feedback.

Consider the Schmitt trigger. Section 3.6 presented the causal argument for why the emitter voltage dropped in response to an increase in input:

$$\begin{array}{ll} \partial v_{B1} = + & \text{Given} \\ \Rightarrow \partial v_{B1,E1} = + & \partial v_B \Rightarrow \partial v_{B,E} \text{ KVL-heuristic for NPN-transistor Q1} \\ & \vdots \\ \Rightarrow \partial v_{E1} = - & \partial i_{-E(Q2)} \Rightarrow \partial v_{E1} \text{ KCL-heuristic for node E1} \\ \Rightarrow \partial v_{E1} = -\partial v_{B1} & \text{Positive feedback at NPN-transistor Q1.} \end{array}$$

As  $\partial v_{E1}$  was assumed insignificant in the KVL-heuristic propagation for  $\partial v_{B1,E1}$ , the propagation indicates a feedback loop starting with  $\partial v_{B1}$ , through  $\partial v_{B1,E1}$  back to  $\partial v_{E1}$ . Thus,  $\partial v_{B1,E1} = f(\partial v_{B1,E1}) + \partial v_{E1}$ .<sup>8</sup>

<sup>8</sup>The delay between the initial input and the emitter drop is mythical. As there are no delay elements in the loop, formally, the input rises and emitter drops at exactly the same time. In the causal analysis, the first causes the second, but no time elapses. As there are no delay elements, there cannot be any overshoot or ringing. If this time delay is important, it should be included by adding capacitors or inductors in the loop. These additional components cause explicit time delays, resulting in overshoot and ringing (as discussed later).

### 4.3. Feedback configurations

Electrical engineering has developed a tremendous amount of theory concerning feedback [32]. There are a variety of different nomenclatures for describing feedback configurations. The terminology used here is from [18]. Table 1 outlines the 8 possible configurations, each important to electrical engineering. The type of feedback exhibited by a circuit is classified according its sign, comparison type, and sampling type. The sign of the feedback indicates whether the feedback signal acts with or against the input signal. For example, the Schmitt trigger exhibits positive feedback. The point at which the feedback is detected is the *comparison point* because it compares an input signal with a feedback signal to produce a composite of the two. The other distinguished point of a feedback loop is the last value which causally affects a circuit output. At this point a split occurs and one signal continues to the output of the circuit while the other is fed back. This value is called the *sampling point* as most of the signal goes on to circuit output, while rest (i.e., a sample) is fed back to the circuit input.

The presence of feedback, its sign, the comparison point and its type, and the sampling point and its type can all be determined from the causal analysis. The sign is computed directly from comparison the feedback signal to the input. Each application of a heuristic introduces a comparison point. For feedback detected via the KCL-heuristic this point is a node in the circuit. For feedback detected via the KVL-heuristic this point is a component. The sampling point is the last shared point on the causal argument for the output and feedback signals. The details (and definitions) for determining the type of the comparison point and sampling point can be found in [11].

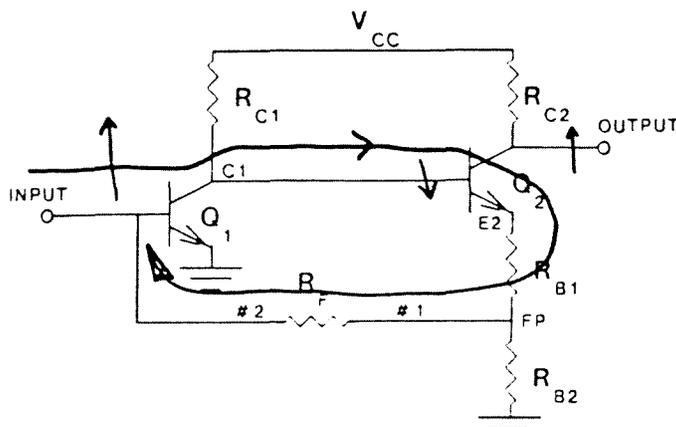
Not every occurrence of feedback is on a path to the output. For example, feedback loops may be nested so that a feedback path itself contains feedback. For these loops the notion of sampling point must be redefined to be that last vertex of the loop that either directly affects the output or is a member of another feedback loop whose comparison point affects the output (recursively).

TABLE 1. Taxonomy of feedback types

Sign	Comparison type	Sampling type	Figure
-	Node	Node	
-	Node	Loop	10
-	Loop	Node	11
-	Loop	Loop	
+	Node	Node	
+	Node	Loop	
+	Loop	Node	
+	Loop	Loop	8

4.4. Two examples

Fig. 10 illustrates a circuit which contains negative feedback recognized using the KCL-heuristic. The input signal is an increase in current into the input. By the KCL-heuristic, the voltage at the input rises as well. This assumes the currents in the base of Q1 and RF are not relevant. However, the signal loops back through Q1, Q2, RB1 and RF producing an additional current into the input node at which the KCL-heuristic applied. Hence, feedback is present. As this current is of opposite sign to the input, the feedback is negative. (Note that the negative feedback signal cannot completely cancel or dominate the input, for it is the input itself which causes the feedback signal.)



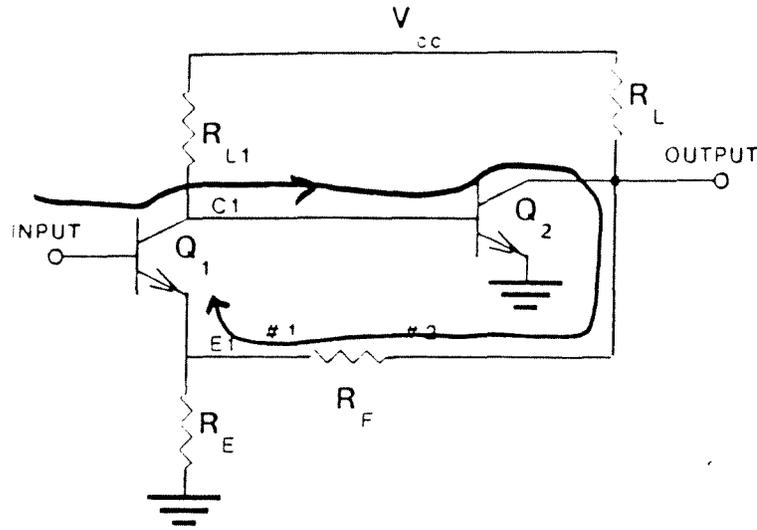
Except for  $R_F$   
# 1 is top terminal  
and # 2 is bottom terminal.

$\partial i_{\rightarrow \text{INPUT}} = +$	Given
$\Rightarrow \partial v_{\text{INPUT}} = +$	$\partial i_{\rightarrow \text{INPUT}} \Rightarrow \partial v_{\text{INPUT}}$ KCL-heuristic for node INPUT
$\Rightarrow \partial i_{\rightarrow C(Q1)} = +$	$\partial v_{B,E} \Rightarrow \partial i_C$ for NPN-transistor Q1
$\Rightarrow \partial v_{C1} = -$	$\partial i_{\rightarrow C(Q1)} \Rightarrow \partial v_{C1}$ KCL-heuristic for node C1
$\Rightarrow \partial v_{C1,E2} = -$	$\partial v_B \Rightarrow \partial v_{B,E}$ KVL-heuristic for NPN-transistor Q2
$\Rightarrow \partial i_{\rightarrow E(Q2)} = -$	$\partial v_{B,E} \Rightarrow \partial i_E$ for NPN-transistor Q2
$\Rightarrow \partial i_{\rightarrow \#1(RB1)} = -$	KCL for node E2
$\Rightarrow \partial i_{\rightarrow \#2(RB1)} = -$	KCL for resistor RB1
$\Rightarrow \partial v_{FP} = -$	$\partial i_{\rightarrow \#2(RB1)} \Rightarrow \partial v_{FP}$ KCL-heuristic for node FP
$\partial v_{\text{INPUT}} = +, \partial v_{FP} = -$	$\Rightarrow \partial v_{FP, \text{INPUT}} = -$ KVL applied to nodes FP, GROUND, INPUT
$\Rightarrow \partial i_{\rightarrow \#1(RF)} = -$	$\partial v_{\#1, \#2} \Rightarrow \partial i_{\#1}$ for resistor RF
$\Rightarrow \partial i_{\rightarrow \#2(RF)} = +$	KCL for resistor RF
$\Rightarrow \partial i_{\rightarrow \#2(RF)} = \partial i_{\rightarrow \text{INPUT}}$	Negative feedback at node INPUT.

FIG. 10. Feedback amplifier (CE-FEEDBACK).

Fig. 11 illustrates a negative feedback amplifier with node sampling. The

sampling point is  $v_{\text{OUTPUT}} = +$  and it causes both  $v_{\text{OUTPUT,E1}}$  and the overall circuit output.



	$\partial v_{\text{INPUT}} = +$	Given
	$\Rightarrow \partial v_{\text{INPUT,E1}} = +$	$\partial v_{\text{B}} \Rightarrow \partial v_{\text{B,E}}$ KVL-heuristic for NPN-transistor Q1
	$\Rightarrow \partial i_{\rightarrow \text{C(Q1)}} = +$	$\partial v_{\text{B,E}} \Rightarrow \partial i_{\text{C}}$ for NPN-transistor Q1
	$\Rightarrow \partial v_{\text{C1}} = -$	$\partial i_{\rightarrow \text{C(Q1)}} \Rightarrow \partial v_{\text{C1}}$ KCL-heuristic for node C1
	$\Rightarrow \partial i_{\rightarrow \text{C(Q2)}} = -$	$\partial v_{\text{B,E}} \Rightarrow \partial i_{\text{C}}$ for NPN-transistor Q2
	$\Rightarrow \partial v_{\text{OUTPUT}} = +$	$\partial i_{\rightarrow \text{C(Q2)}} \Rightarrow \partial v_{\text{OUTPUT}}$ KCL-heuristic for node OUTPUT
	$\Rightarrow \partial v_{\text{OUTPUT,E1}} = +$	$\partial v_{\#1,\#2} \Rightarrow \partial v_{\#1,\#2}$ KVL-heuristic for resistor RF
	$\Rightarrow \partial i_{\rightarrow \#1(\text{RF})} = -$	$\partial v_{\#1,\#2} \Rightarrow \partial i_{\#1}$ for resistor RF
$\partial v_{\text{INPUT,E1}} = +$	$\Rightarrow \partial i_{\leftarrow \text{E(Q1)}} = +$	$\partial v_{\text{B,E}} \Rightarrow \partial i_{\text{E}}$ for NPN-transistor Q1
$\partial i_{\rightarrow \#1(\text{RF})} = -, \partial i_{\leftarrow \text{E(Q1)}} = +$	$\Rightarrow \partial i_{\rightarrow \#1(\text{RE})} = +$	KCL for node E1
	$\Rightarrow \partial v_{\text{E1}} = +$	$\partial i_{\#1} \Rightarrow \partial v_{\#1,\#2}$ for resistor RE
	$\Rightarrow \partial v_{\text{E1}} = \partial v_{\text{INPUT}}$	Negative feedback at NPN-transistor Q1.

FIG. 11. Loop-node amplifier.

### 4.5. Local feedback

Electrical engineers distinguish a subclass of feedback called local feedback, in which the sampled signal is immediately fed back. Consider the first few steps of the analysis of the circuit of Fig. 11:

$\partial v_{\text{INPUT}} = +$	Given
$\Rightarrow \partial v_{\text{INPUT,E1}} = +$	$\partial v_{\text{B}} \Rightarrow \partial v_{\text{B,E}}$ KVL-heuristic for NPN-transistor Q1
$\Rightarrow \partial i_{\rightarrow \text{C(Q1)}} = +$	$\partial v_{\text{B,E}} \Rightarrow \partial i_{\text{C}}$ for NPN-transistor Q1
$\Rightarrow \partial v_{\text{C1}} = -$	$\partial i_{\rightarrow \text{C(Q1)}} \Rightarrow \partial v_{\text{C1}}$ KCL-heuristic for node C1.

The causal argument for RL1's current continues:

$$\begin{array}{ll}
 \Rightarrow \partial v_{C1} = - & \partial i_{\rightarrow C(Q1)} \Rightarrow \partial v_{C1} \text{ KCL-heuristic for node C1} \\
 \partial v_{VCC} = 0 & \partial v_{+,-} = 0 \text{ for battery B} \\
 \partial v_{VCC} = 0, \partial v_{C1} = - & \Rightarrow \partial v_{C1,VCC} = - \text{ KVL applied to nodes C1,GROUND,VCC} \\
 \Rightarrow \partial i_{\rightarrow \#1(RL1)} = + & \partial v_{\#1,\#2} \Rightarrow \partial i_{\#1} \text{ for resistor RL1} \\
 \Rightarrow \partial i_{\leftarrow \#2(RL1)} = + & \text{KCL for resistor RL1.}
 \end{array}$$

Thus, as a consequence of utilizing a KCL-heuristic at the collector of Q1, an additional current is propagated into the node. By the criteria of Section 4.1 this is evidence for feedback.

An analogous situation arises with the KVL-heuristic. Consider the simple amplifier of Fig. 12. An increase in input voltage causes (via the KVL-heuristic) an increased base-emitter voltage across Q1. As a consequence, the transistor conducts more current, increasing the flow out of its emitter and into the emitter resistor. The increased current causes increased emitter voltage, negatively influencing the base-emitter voltage.

Both of these examples meet the strict dictionary definition of feedback, and are detected by the mechanism of the previous section. However, the feedback is local because the feedback signal path does not form a loop in the circuit schematic (but, of course, a loop in the potential causal topology). Local feedback, although it exhibits all of the characteristics of feedback, can usually be ignored in analyzing the global behavior of the circuit. Local feedback does, however, perform an important validation function. A local feedback loop represents the loading effect of a load on a source. Thus, every KCL- or KVL-heuristic should have a local feedback (i.e., a reflection) as the heuristics presume the presence of a load. If there is no load, the heuristic is invalid.

#### 4.6. Negative impedances

The KCL- and KVL-heuristics assume the circuitry connected to the component or node acts as a positive impedance. Thus, if the circuit fragment contains a negative impedance, any heuristic propagating a signal into it will get a reflected or feedback signal opposite to what it expected. Negative

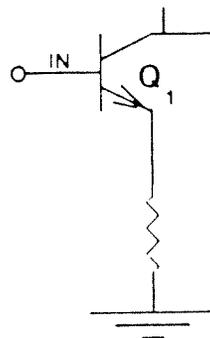


FIG. 12. Common-emitter amplifier.

impedance appears as a positive reflection. (It is important to note that even though none of the components individually may manifest negative resistance, the overall circuit might e.g., the Schmitt trigger can provide a negative impedance load to its input signal.) As a consequence it is always possible to tell whether the circuit contains negative impedances as well as which particular nodes see the negative impedances. The KCL- and KVL-heuristics are therefore implemented to only propagate a negative impedance signal after they have received a positive reflection assuming positive impedances. Thus causal analysis applies to negative impedances as well.<sup>9</sup>

### 5. The Mechanism Graph

With the inclusion of feedback detection, causal analysis is complete. Sections 5 and 6 discuss how the information constructed by the causal analysis process is analyzed to construct more abstract descriptions of the behavior and to relate this description to the ultimate purpose of the circuit. The *mechanism graph* provides a primitive representation for describing the mechanism by which a circuit achieves its input/output behavior. In Section 6 I describe how the mechanism graph is parsed into higher-level descriptions.

Fig. 13 illustrates the mechanism graph for the circuit of Fig. 10. A vertex in this graph represents information contained in a cell. An edge represents the fact that the information in the vertex adjacent from the edge contributes to the information in the vertex adjacent to the edge. Thus, the mechanism graph describes how information in one cell affects the information in other cells (i.e., causality).

As change is of interest, the mechanism graph only contains vertices which represent changing quantities. The only uncaused vertices are inputs and thus are the vertices with in-degree zero. The mechanism graph only includes quantities which causally affect circuit outputs (the black vertices). An edge is dashed if it depends on a heuristic, solid if it is deduced via a component rule. Circle vertices represent currents and square vertices represent voltages. In order to avoid cluttering the mechanism graphs, only a minimal amount of annotation is included on the graphs. If the quantity is a current, it is labeled by

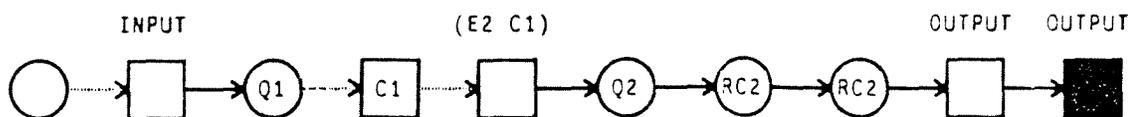


FIG. 13. Mechanism graph for the circuit of Fig. 10's output.

<sup>9</sup>With one exception, I haven't figured out what to do if the circuit contains both positive feedback involving positive impedances and negative feedback using negative impedances—I doubt that any circuit in existence depends on it.



FIG. 14. Splits and joins.

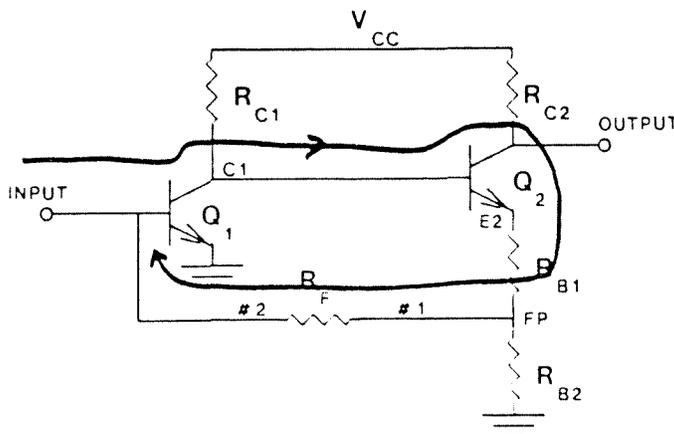
the component it is a current of. If the quantity is a voltage with respect to ground, it is labeled by the name of the non-ground node. Otherwise the voltage vertex is labeled by its pair of circuit nodes. If there is room, the label is printed inside the vertex, otherwise it is printed above it.

Not all causal graphs are straight lines. A causal graph may contain *splits* and *joins* (Fig. 14). A split occurs at a vertex with out-degree greater than one. Such a vertex represents a quantity that directly causes two other quantities to change, both of which eventually contribute to the output. A join occurs at a vertex with in-degree greater than one. In this case the change in the quantity is caused by the simultaneous change in all the antecedent quantities. A graph represents a single argument. Most splits and joins are the direct result of particular circuit mechanisms, but some splits and joins are necessary to account for the difference in number of inputs and outputs. *Feedforward* is used to describe the situation where two paths originating at a split recombine at a join.

### 5.1. Feedback in the mechanism graph

In order to represent feedback, the mechanism graph is augmented to include the graphs for the error signals, where extra dashed edges are included to indicate where signals feed back into comparison points. As a consequence of this modification, the mechanism graph is a digraph containing cycles.

Fig. 16 is the complete mechanism graph for Fig. 15. The feedback signal is



Except for  $R_F$   
 # 1 is top terminal  
 and # 2 is bottom terminal.

FIG. 15. Feedback amplifier (CE-FEEDBACK).

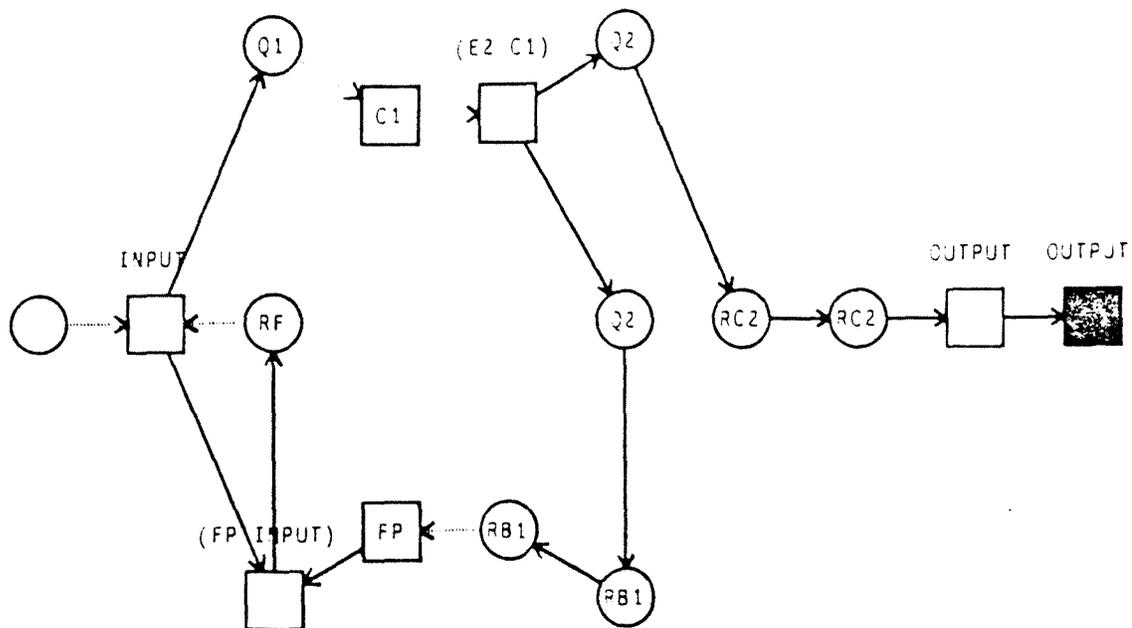


FIG. 16. Complete mechanism graph for Fig. 15.

the current flowing from resistor RF into node INPUT. This current is deduced by applying Ohm's law to the voltage determined by applying KVL to nodes INPUT, FP and GROUND. So both the voltage at INPUT and the voltage at FP contribute to the feedback signal.

## 6. Teleological Analysis

Circuits are designed and manufactured to achieve specific functions. As these artifacts have to be conveniently designed, efficiently manufactured and easily maintained, the designer attempts to make his circuits as simple as possible. These desiderata dictate that every component must contribute in some way to the ultimate purpose of the circuit. For designed artifacts, every component can be related to the ultimate function of the device. This is the *teleological* perspective.

Teleology plays such a fundamental role in understanding circuits, that electrical engineering has agreed upon a language for discussing it. Causal analysis provides a method for automatically determining the teleology of a circuit in the nomenclature of electrical engineers.

### 6.1. Configurations

The question, "What is the purpose of R1?" can be answered at many levels. A very simple kind of teleology can be defined using the mechanism graph of a circuit. The mechanism graph is a global description of how the circuit achieves its input/output behavior, and the purpose of an individual component is defined by how it contributes to this mechanism. In particular, a taxonomy of

roles a component can play defines all the possible local causal patterns (called configurations) in which it can appear in the mechanism graph. By local, I mean propagations made using the component models and their immediate antecedents and consequents. The taxonomy proposed here is an expanded and formalized version of the language electrical engineers use. A great amount of lore, rules of thumb and advice is associated with each component configuration. The configuration provides a kind of index key into the data base of knowledge available about the component.

Consider the resistor as an example. The resistor is a relatively simple passive component and electrical engineers do not have a formal vocabulary for describing its functions. Nevertheless, its simplicity makes it easy to see how causal analysis distinguishes roles for components. A similar analysis can be done for each component type. The basic resistor law is  $\partial v = \partial i$ . This formulation suggests two basic causal patterns for the resistor: one in which the voltage causes the current (i.e.,  $\partial v \Rightarrow \partial i$ ) and one in which the current causes the voltage (i.e.,  $\partial i \Rightarrow \partial v$ ). By definition, within any particular mechanism graph the resistor is used in one of these two ways. The resistor may contribute to the mechanism graph as a voltage-to-current converter or as a current-to-voltage converter. In this view a resistor has two possible functions. However, closer examination shows that the resistor has far more possible roles.

The simplicity of the equation  $\partial v = \partial i$  hides most of the options. In actual fact, the resistor has 18 possible configurations. As KCL applies to resistors, we must also take into account the equation  $\partial i_{\#1} = \partial i_{\#2}$ . Thus, a resistor might convert an input voltage to a current in terminal #1 which is then propagated out, or to a current in terminal #2. Analogously, the input to the resistor might be one of those two currents. These considerations result in two types of current-to-voltage conversion, two types of voltage-to-current conversion, and two types of current propagation (i.e., KCL alone).

Although a component voltage intrinsically refers to two terminals of a component, it is often possible to identify a distinguished voltage terminal. If a KVL deduction is used to determine the input voltage, i.e.,  $\partial v_{\#1,\#2} = \partial v_{\#1,x} + \partial v_{x,\#2}$ , where  $\partial v_{\#1,x} = 0$  or  $\partial v_{x,\#2} = 0$ , then the input terminal is #2 or #1 respectively. If a KVL-heuristic is applied to a component, the terminal connected to the node at which the KVL-heuristic is applied is considered the input terminal. If the output current of the resistor is an immediate antecedent to a KCL-heuristic, the output is really a voltage-to-ground at the terminal.

Tables 2 and 3 illustrate the 18 primitive causal patterns for a resistor (as the resistor is symmetric, #2 inputs are left out).

Examine the mechanism graph of Fig. 17. It mentions resistors RF, RC2 and RB1. As can be seen from the diagram, the two inputs to RF are the voltages at its two ends and the output is a current. Hence, it is functioning in a V-SENSOR configuration. The input to RC2 is a current and the output is a voltage at the same node, thus it is functioning in an I-LOAD configuration. On

TABLE 2. Resistor causal patterns

Input	Output	
	$i_{\#1}$	$i_{\#2}$
$i_{\#1}$	*	BIAS
$v_{\#1}$	V-LOAD	V-TO-I-COUPLE
$v_{\#1,\#2}$	V-SENSOR	V-SENSOR

TABLE 3. Resistor causal patterns

Input	Output		
	$v_{\#1}$	$v_{\#2}$	$v_{\#1,\#2}$
$i_{\#1}$	I-LOAD	I-TO-V-COUPLE	I-SENSOR
$v_{\#1}$	*	DIVIDER	DIVIDER

the other hand, the input to RB1 is a current and the output is a voltage so RB1 is functioning in the I-TO-V-COUPLE configuration.

Components that only participate in local feedback loops do not explicitly appear in the mechanism graph. The causal configurations of these components are determined by applying the standard configuration patterns as if the reflected

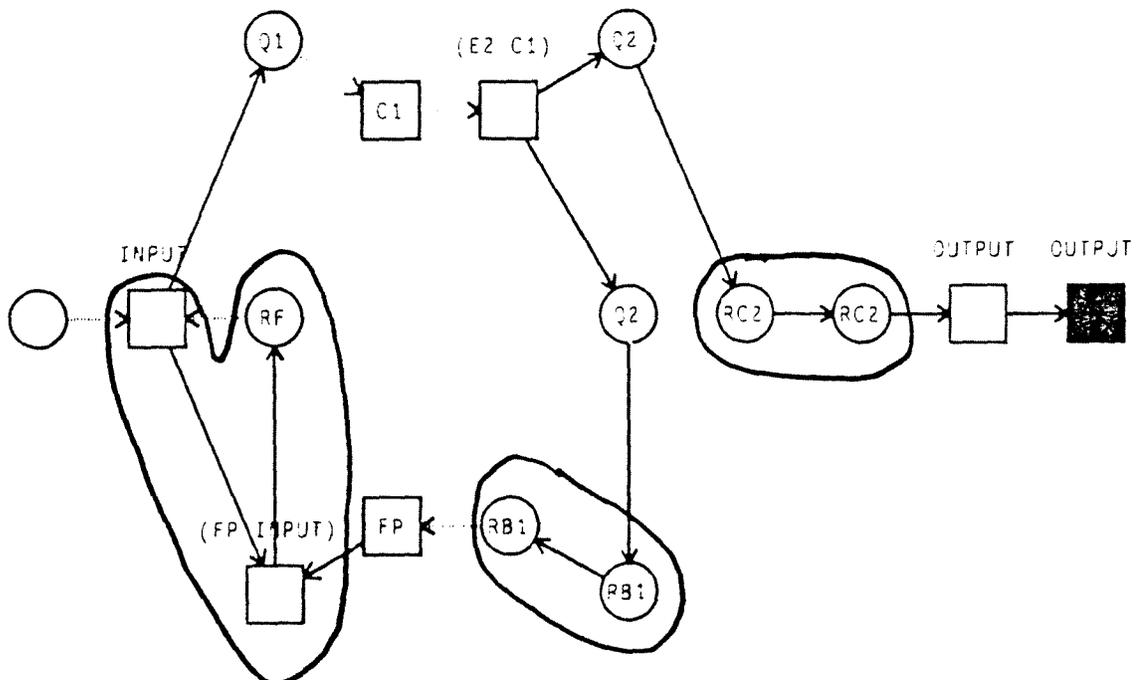


FIG. 17. Mechanism graph with resistor configurations.

signal were the input, and the heuristically deduced value the output. For example, RC1 does not appear in the mechanism graph but it contributes to a reflected current substantiating a KCL-heuristic, and thus causally it functions as an I-LOAD. So does RB2.

It is important to note that the same component may contribute in more than one way to the circuit mechanism. A resistor may be exhibiting more than one of the configurations listed in Tables 2 and 3. For example, a resistor may be converting a voltage to two different currents which propagate through different regions of the circuit. However, none of the resistors of the circuit of Fig. 10 exhibits a dual purpose.

## 6.2. No-function-in-structure

One of the central tenets of qualitative physics is the no-function-in-structure principle (NFIS): the laws of the parts of a component may not presume the functioning of the whole [6]. I have not stressed this issue earlier as it is implicit in electrical engineering. The definition of causal configuration just presented provides a simple means to motivate the concept by illustrating how easy it is to violate this principle.

Electrical engineering incorporates NFIS implicitly in its choice of conventions and notations. Likewise, the notation of qualitative physics is chosen to make it difficult to violate NFIS. However, if we were more liberal in writing components models, it would be quite easy to violate NFIS.

Tables 2 and 3 present 18 different causal configurations for resistor functioning. Given a more general notation, each of these 18 different causal configurations could be encoded as separate component models. For example, a V-LOAD model is  $\partial v_i \Rightarrow \partial i_i$  and a V-TO-I-COUPLE is  $\partial v_{\#1} \Rightarrow \partial i_{\#2}$ . If causal analysis used these 18 models (using similarly developed models for the other component types), there would be no need for the heuristics or the causally ambiguous Ohm's law. The task of causal analysis would be to identify which of the possible models for a component were applicable in the given circuit being analyzed. The attraction of this approach is that the causal heuristics become superfluous, as a choice of component model implicitly involves making an assumption.

This approach unfortunately leads to incomplete models. In qualitative analysis generally there is a seductive tendency to confuse function with structure. For example, one might analyze a few circuits, notice that resistors were either BIAS's, V-TO-I-COUPLE's or I-LOAD's and build a model which incorporated only those possibilities. This model would be a gross violation of the no-function-in-structure principle as it would presume that the resistor does not ever function as, say, a V-SENSOR. The circuit of Fig. 10 could not be analyzed correctly. However, those circuits which only used resistors as BIAS's, V-TO-I-COUPLE's or I-LOAD's would be analyzed correctly.

Qualitative physics attempts to achieve no-function-in-structure by insisting that the models are as general and context-free as possible. Thus, for example, although the KCL- and KVL-heuristics could be incorporated into the component models, they are not. Instead, each model and law is general, local and parsimonious. A side benefit of this approach is that it is now possible to identify systematically all the possible functional roles of a resistor by looking at all the possible mechanism graph contexts in which the resistor can appear.

### 6.3. Transistor configurations

Electrical engineering distinguishes three configurations in which the transistor is capable of providing useful amplification. Each configuration has a unique combination of input impedance, output impedance, voltage gain, and current gain. The most familiar transistor configuration is the common-emitter (CE) in which the input is applied to the base and the output is taken off of the collector. The emitter provides the common terminal between input and output. In the common-base (CB) configuration the input is applied to the emitter and the output is taken off of the collector. The base provides the common terminal. In the common-collector configuration the input is applied to the base and the output is taken off of the emitter. This configuration is also called the emitter-follower configuration.

The correct transistor configuration is determined by the input and output variables of the transistor and these are determined in the same way as for resistor configurations. However, in laying out the possibilities many more transistor configurations come to the fore. A transistor can be used to compare two independent signals (the SUM configuration). As it also can be viewed as two back-to-back diodes it has many of the analogous configurations as a resistor (i.e., loading, sensing, and coupling). Table 4 summarizes the possible transistor configurations.

TABLE 4. Transistor configurations for forward active region

Input	Output			
	$i_E$	$i_B$	$i_C$	$v_{B,E}$
$v_E$	LOAD-Q	COUPLE-Q	CB	
$v_B$	CC	LOAD-Q	CE	SENSE-Q
$v_{B,E}$	SENSE-Q	SENSE-Q	SUM	*

Table 5 summarizes all the configurations for the circuit of Fig. 10. '(C)' indicates a causal (i.e., part of mechanism graph) configuration; '(F)' indicates a (non-local) feedback (i.e., part of mechanism graph but on a feedback path); and '(R)'

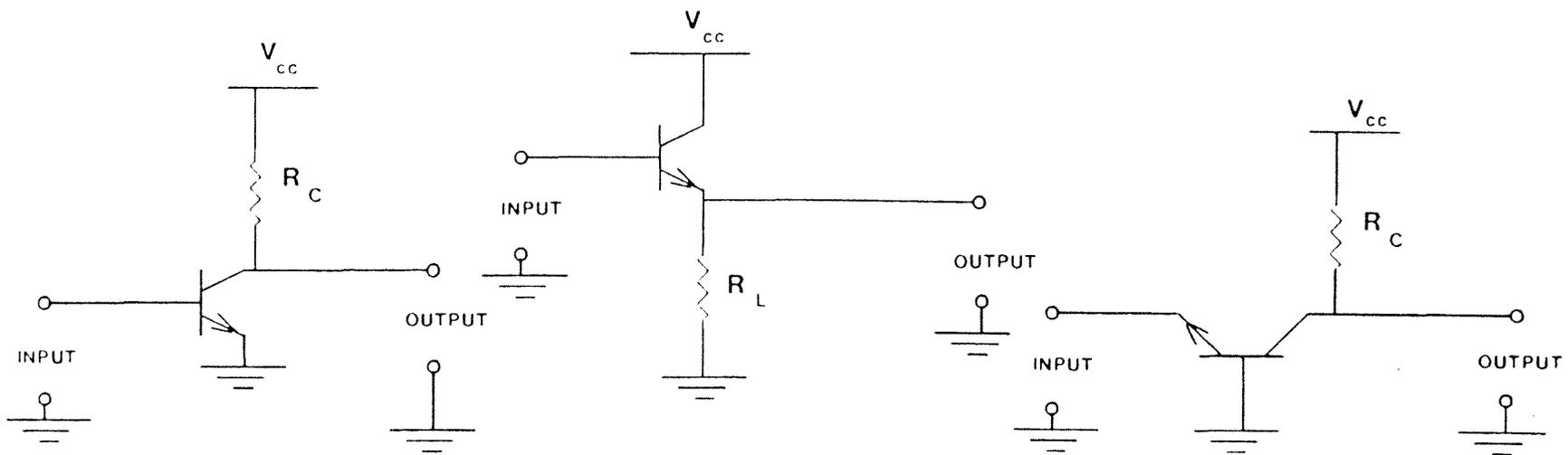


FIG. 18. Common-emitter, common-collector, and common-base configurations.

TABLE 5. Configurations for CE-FEEDBACK

Component	Configuration
Q1	CE(C) LOAD-Q(R)
Q2	CE(C) CC(F) LOAD-Q(R)
RC1	I-LOAD(R)
RC2	I-LOAD(C)
RB2	I-LOAD(R)
RB1	I-TO-V-COUPLE(F)
RF	V-SENSOR(F) I-LOAD(R)

indicates a reflection (i.e., supporting a heuristic propagation in the mechanism graph) configuration. Q1 functions in the common-emitter configuration on the signal path and LOAD-Q configuration in a reflection to the input signal. Q2 functions as a common-emitter stage to the output and as a common-collector stage feeding the signal back to the input. RC1 is functioning as an I-LOAD in a reflection configuration, RC2 is functioning as an I-LOAD on the causal path to the circuit output. RF is functioning as a V-SENSOR for the feedback path and as a reflection for the input signal. RB1 (with RB2) converts the transistor's output current to a voltage to be compared and fed back.

#### 6.4. Fragment graph

The taxonomy of component configurations developed in Sections 6.1 and 6.4 can be used to parse the mechanism graph. Each component functioning in a particular configuration characterizes a subgraph of connected edges and vertices in the mechanism graph. These are the *primitive teleological fragments* of the circuit. As the output of every component is connected to the input of some other component, the fragments can be connected together in a topology homomorphic to that of the original mechanism graph. The result of this process is an abstract description of the functioning of the circuit in terms of the basic roles of the components. Fig. 19 is the fragment graph corresponding to the mechanism graph of Fig. 16.

Vertices corresponding to the component roles are labeled by their type. The different configurations are grouped into eight different classes which delineate the generic actions of fragments (e.g., input/output, coupling and amplification stage). Most classes are broken down further into types (e.g., the common-emitter, common-collector and common-base transistor configurations). The complete ontology used by EQUAL is given in Table 6.

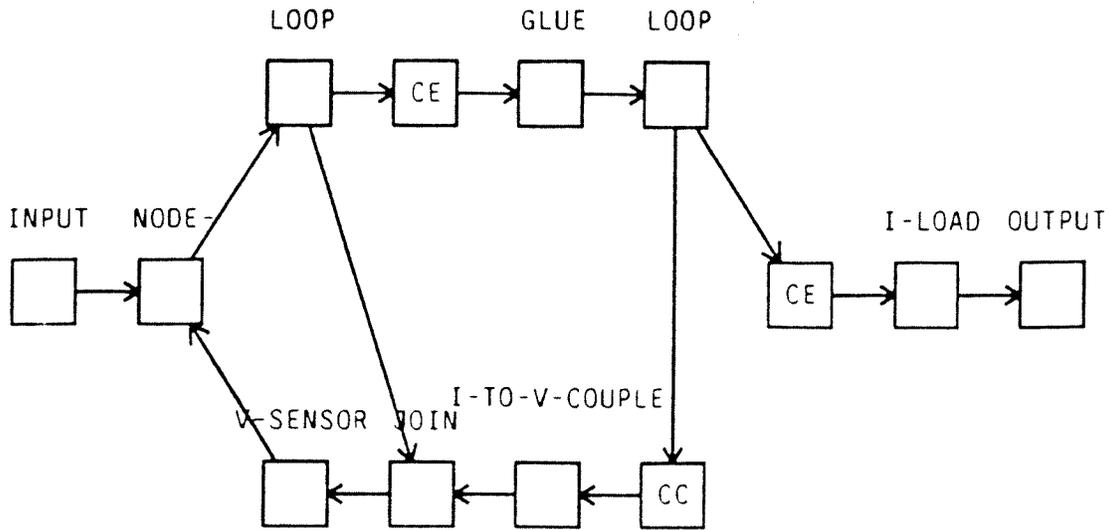


FIG. 19. Primitive fragments of mechanism graph for CE-FEEDBACK.

Not all these fragments types occur in the circuit CE-FEEDBACK or the component configuration tables. Fragment types SPLIT and JOIN are used to describe feedforward mechanisms where two signals split and later rejoin. CASCADE and FEEDBACK configurations are abstract fragments introduced in the Section 6.5.

As in the mechanism graph, the same component may fulfill multiple roles. Hence, for example, Q2 appears twice in the fragment graph: once in the common-emitter configuration in the main signal path and once in the common-collector configuration on the feedback path.

### 6.5. An amplifier parser

In the first two steps of the recognition process causal analysis applied to the

TABLE 6

Class	Type	Description
IO	INPUT.OUTPUT	signals on boundary
SAMPLING	NODE.LOOP	feedback sampling point
COMPARISON	NODE.LOOP	feedback comparison point
SPLIT	unused	signal splits <i>n</i> ways
JOIN	VOLTAGES.CURRENTS	<i>n</i> signals combine
STAGE	CE.CC.CB.CASCADE.FEEDBACK	amplifying stage
DIFF-2-1	SUM	differential amplifier
COUPLING	GLUE.V-LOAD.WIRE.I-LOAD, V-SENSOR.V-TO-I-COUPPING, I-TO-V-COUPPING.SENSE-Q.LOAD-Q, COUPLE-Q.BIAS.LEVEL	signal couplers

circuit's topology constructs a representation of the circuit's abstract mechanism (Fig. 16) and the role each component plays in this mechanism (Fig. 19). The final step performs topological manipulations on this representation of circuit mechanism in order to construct a hierarchical decomposition of the mechanism with the ultimate purpose at the most abstract level using the fragment graph as its base.

COUPLING fragments only couple signals from one part of the circuit to another. As they perform no useful amplification, fan-in or fan-out roles the first parsing step is to incorporate all of them into their adjacent fragments. The result is Fig. 20.

Although pure reflections are never included in the mechanism graph, feedback which is both local and global must be included. A parsing rule is required to remove the local feedback loop, leaving the global one in place. The substitution rule is given in Fig. 21. The resultant fragment graph is Fig. 22.

The circuit can now be (correctly) parsed as a feedback amplifier followed by a common-emitter stage. However, there is an alternate parsing of the circuit which gives a better account of the gains (a topic discussed in more detail in [11]). EQUAL is also concerned with determining the input/output characteristics of each circuit fragment in terms of input impedance, output impedance, transresistance, current gain, voltage gain, and transconductance. In particular, it determines approximate ranges for their values and how stable they are. The fragment graph of Fig. 22 implies that the feedback action does not stabilize the gain of Q2's CE configuration. However, Q2's collector and emitter currents are identical, thus the feedback action also stabilizes the gain of Q2's configuration. For this reason EQUAL employs an additional parsing rule to rewrite the fragment graph (and the gains of the fragments—although that is

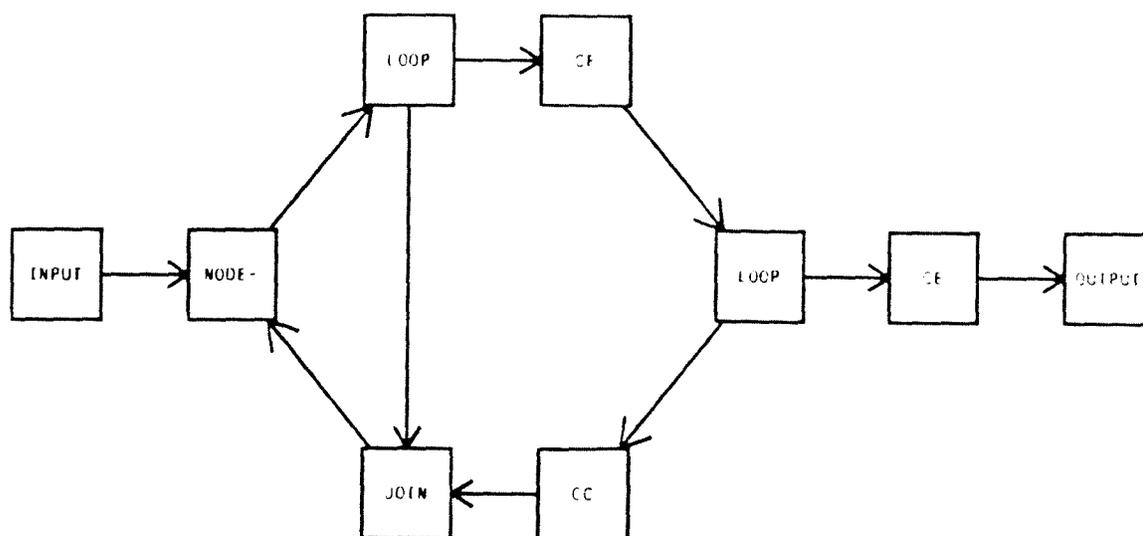


FIG. 20. Parse of CE-FEEDBACK after removing COUPLING.

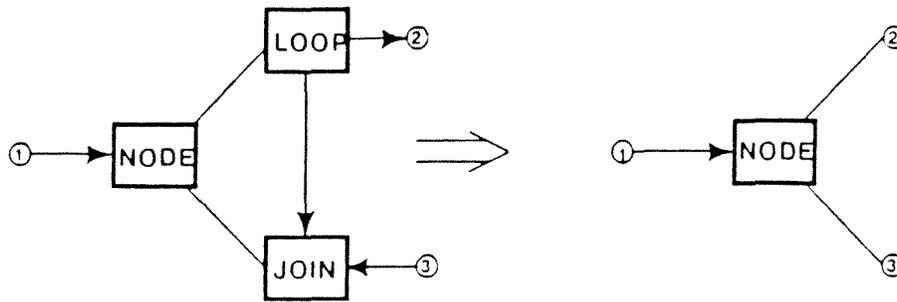


FIG. 21. Local feedback substitution rule.

not discussed here). Fig. 23 illustrates this rule. The resultant fragment graph is Fig. 24.

Two active component configurations (stages) can be combined into a single fragment called a **CASCADE**. After applying this cascade rule the fragment graph is Fig. 25. All feedback loops are of the form of Fig. 26 and may be summarized. The result of applying this rule is Fig. 27. The parse terminates at the point as the pattern of input-amplifier-output is a valid amplifier.

The purpose of a fragment is given by the role it plays in the more abstract fragments in which it participates. The role name is provided by the rewrite rule. For example, the two antecedent stages for the cascade rule are called **STAGE1** and **STAGE2**. The rewrite rules in order of application are (the antecedent roles, if any, are listed in parentheses):

- 1 FLUSH-GLUE( )  
Remove all *COUPLING* and *FEEDBACK COUPLING*
- 2 CASCADE(STAGE1,STAGE2)  
Combine successive stages

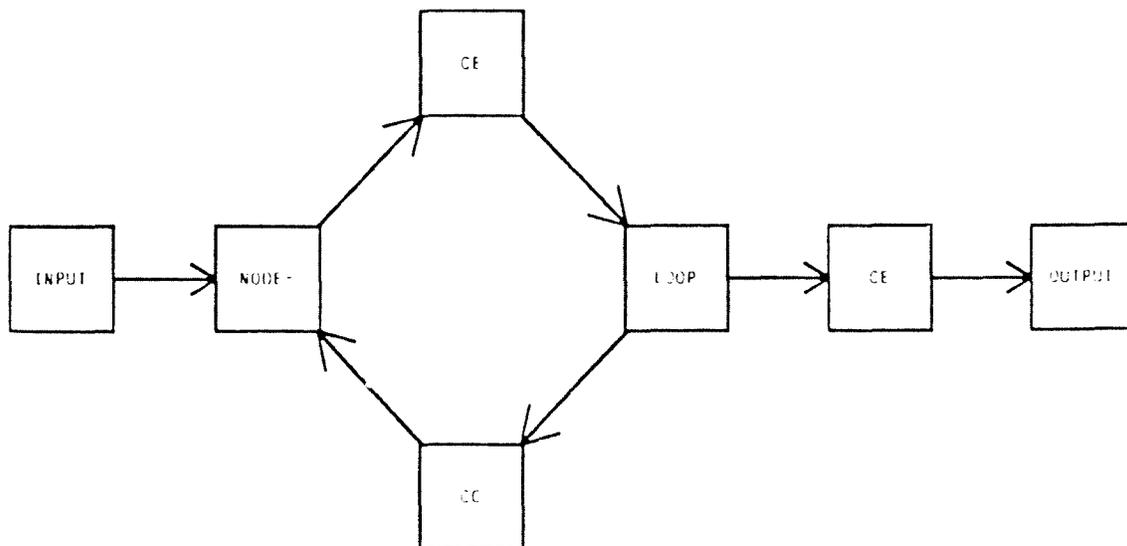


FIG. 22. Parse of CE-FEEDBACK after removing local feedback.

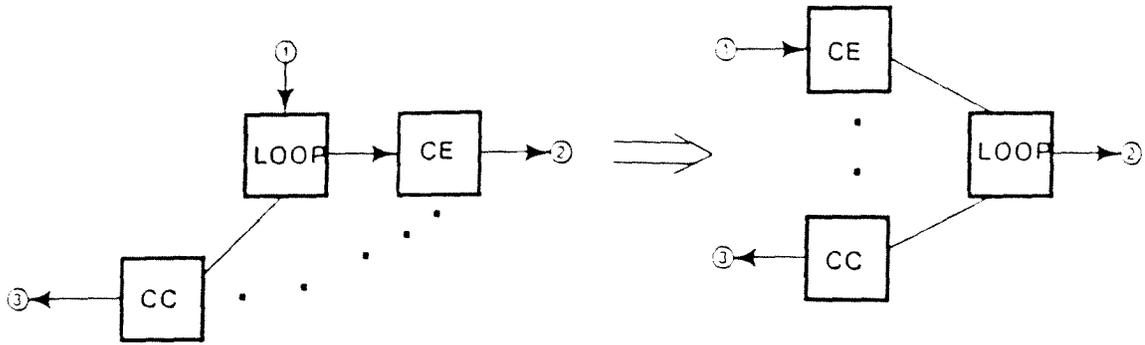


FIG. 23. Sampling rewrite rule.

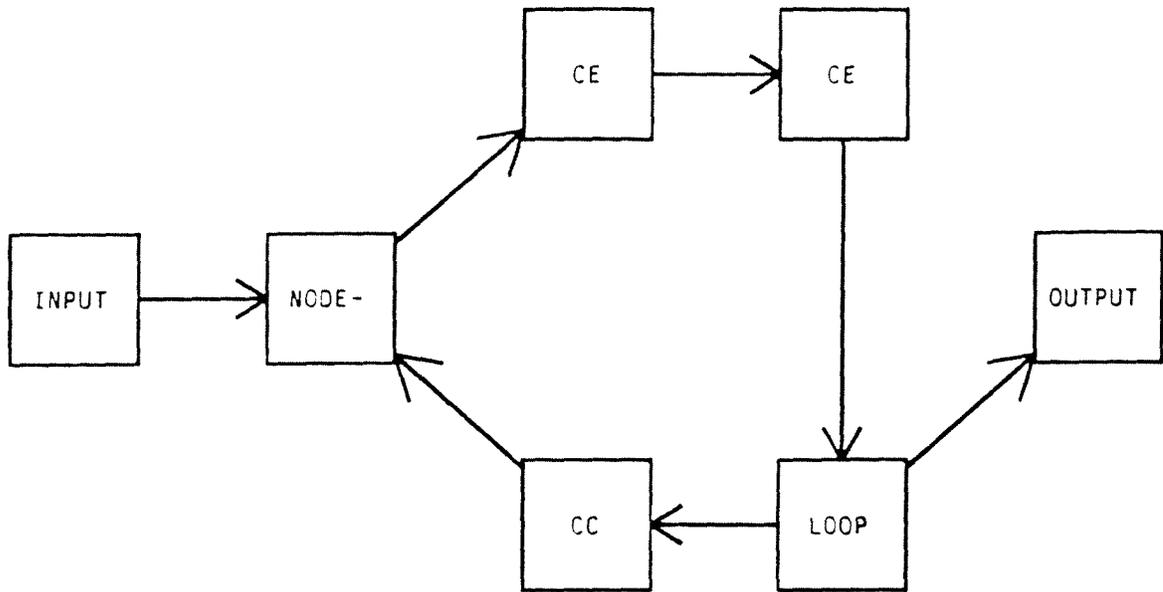


FIG. 24. Parse of CE-FEEDBACK after applying sampling rewrite rule.

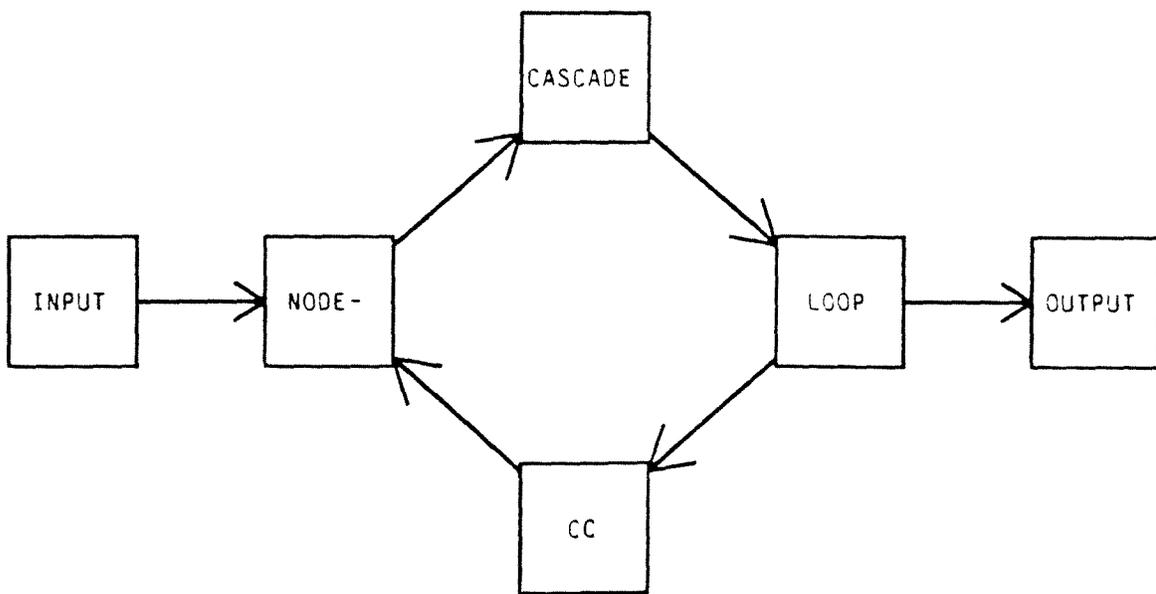


FIG. 25. Parse of CE-FEEDBACK after cascading stages.

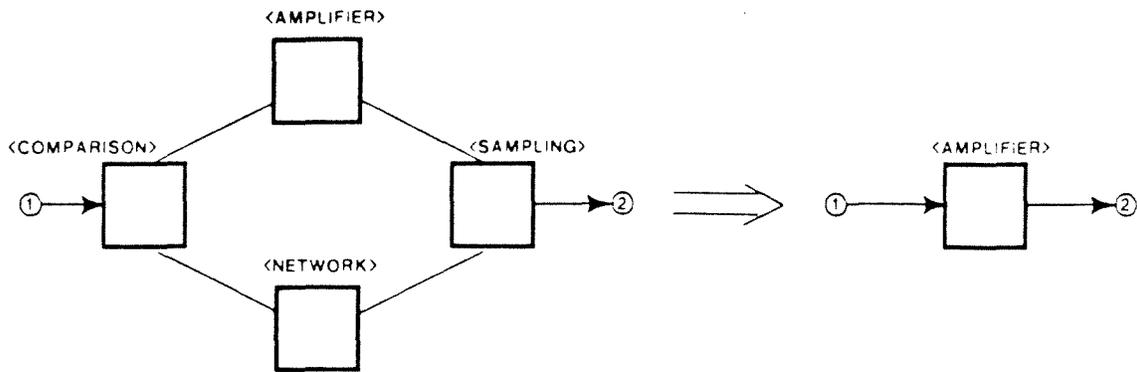


FIG. 26. Feedback rewrite rule.

- 3 LOCAL-FEEDBACK( )  
*Remove local feedback*
- 4 UNLOOPS( )  
*Move sampling point backward*
- 5 FEEDBACK(COMPARISON.BASIC-AMPLIFIER.SAMPLING.FEEDBACK-NETWORK)  
*Feedback amplifier*
- 6 TOP-LEVEL(INPUT-NETWORK.STAGE.OUTPUT-NETWORK)  
*Termination*

The ordering of these rules is important. For example, interchanging the rules UNLOOPS and FEEDBACK causes faulty analyses.

The hierarchical description generated by the fragment rewrite rules provides a detailed description of the role of each component of the circuit. The following is EQUAL's explanation for the purpose of each of the components in CE-FEEDBACK. Reflection configurations are left out unless and component is functioning only in a reflection configuration (i.e., LOAD-Q for Q2 is left out, while I-LOAD for RC1 is included). Each rewrite rule assigns role names to each fragment on its right-hand sides. A separate explanation is included for each configuration of each component. The first line of each component's description is its configuration. This is typically followed by a sequence of lines of the form "Which is <role> or <rule>" which indicates how the fragments the component is part of are parsed. Passive components in load configurations

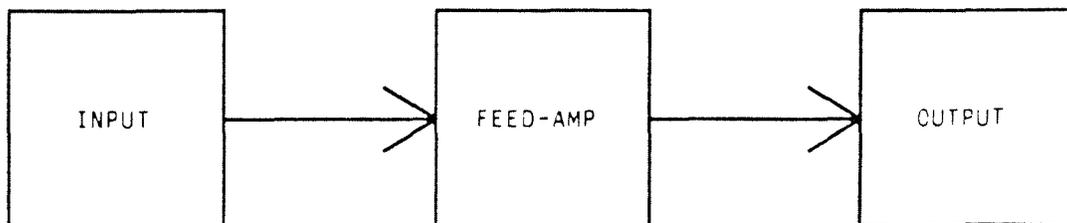


FIG. 27. Parse of CE-FEEDBACK after eliminating feedback.

(which are otherwise removed by the FLUSH-GLUE rule) are related to the active components they load.

Q1 is functioning in CE configuration.  
Which is STAGE1 of CASCADE  
Which is BASIC-AMPLIFIER of FEEDBACK  
Which is STAGE of TOP-LEVEL

Q2 is functioning in LOOP configuration.  
Which is SAMPLING of FEEDBACK  
Which is STAGE of TOP-LEVEL  
And,

Q2 is functioning in CE configuration.  
Which is STAGE2 of CASCADE  
Which is BASIC-AMPLIFIER of FEEDBACK  
Which is STAGE of TOP-LEVEL  
And,

Q2 is functioning in CC configuration.  
Which is FEEDBACK-COUPPLING of FEEDBACK-NETWORK  
Which is FEEDBACK-NETWORK of FEEDBACK  
Which is STAGE of TOP-LEVEL

RC1 is functioning in I-LOAD configuration.  
For Q1 functioning in CE configuration.  
Which is STAGE1 of CASCADE  
Which is BASIC-AMPLIFIER of FEEDBACK  
Which is STAGE of TOP-LEVEL

RC2 is functioning in I-LOAD configuration.  
Which is COUPLING of OUTPUT-NETWORK  
Which is OUTPUT-NETWORK of TOP-LEVEL

RB1 is functioning in I-TO-V-COUPLE configuration.  
Which is FEEDBACK-COUPPLING of FEEDBACK-NETWORK  
Which is FEEDBACK-NETWORK of FEEDBACK  
Which is STAGE of TOP-LEVEL

RB2 is functioning in I-LOAD configuration.  
For Q2 functioning in CC configuration.  
Which is FEEDBACK-COUPPLING of FEEDBACK-NETWORK  
Which is FEEDBACK-NETWORK of FEEDBACK  
Which is STAGE of TOP-LEVEL

RF is functioning in V-SENSOR configuration.  
Which is FEEDBACK-COUPPLING of FEEDBACK-NETWORK  
Which is FEEDBACK-NETWORK of FEEDBACK  
Which is STAGE of TOP-LEVEL

Much more work can be gotten out of teleology than is presented here. First, it is possible to define the set of rewrite rules to cover a wide class of amplifiers and power-supplies. Second, it is possible to include estimates of the gain and impedances of stages and adjust these in the parsing process. For example, feedback has a dramatic effect on gains and impedances. Third, voltage and

current sources need to be analyzed with different types of rules. Another report [11] extends the outline presented in this paper into these three areas.

### 6.6. The power of the teleological parse

The teleological analysis constructs a canonical form for circuit functioning. This has two immediate advantages. First, it identifies the role each component plays in the composite device. Electrical engineers associate with each configuration a large body of knowledge of how it is used, what its peculiar characteristics are, what its failure modes are, etc. Thus, the parse provides an index into this body of knowledge. For example, once it is recognized that the transistor is acting as a common-emitter amplifier, knowledge about its problems, frequency response, etc. can be accessed. Second, canonicalization helps identify similar circuits. For example, two circuits may have radically different schematics, but at a given level of abstraction they may both be series-regulated power-supplies. The techniques outlined here apply to fluid, hydraulic, mechanical, etc. systems as well, so the language of mechanism is sufficiently powerful to notice a mechanical analog of some electrical system, because at some level they have the same fragment graph.

## 7. Ambiguity

Sections 3–6 developed the structure-to-function analysis very quickly, glossing over many details. In these concluding sections, we step back and examine some of these glossed-over issues in more detail. The advantage of showing the role of teleology first, is that it now provides a motivation and organization for dealing with some of these harder problems.

Qualitative analysis is fundamentally ambiguous. It is often possible to find more than one qualitative value or more than one causal argument for a circuit quantity. When this occurs EQUAL considers both possibilities simultaneously and produces multiple causal arguments for the circuit's behavior(s). Other, non-causal, reasoning must be used to select amongst the possibilities. However, before dealing with the mechanics of selection it is important to consider the origin of ambiguity and the generation of the possibilities.

### 7.1. The origin of ambiguity

Due to the coarse distinctions of the qualitative value set, qualitative analysis is ambiguous compared to quantitative analysis. However, one would have hoped that qualitative analysis were unambiguous with respect to itself, i.e., that given a qualitative description of a situation it produces a distinct qualitative description of the behavior. Unfortunately, this is not possible and, even qualitatively, behavior is ambiguous. The root cause of this ambiguity is that the qualitative

operators do not define a field and hence most of the usual theorems of linear algebra and network theory upon which our intuitions are based fail to hold. For a more detailed discussion see [6].

It is important to note that the ambiguity does not result from the implementation choices discussed in Section 3.2. Adding KCL or KVL confluences does not remove any ambiguity.

## 7.2. Global ambiguity

The task of causal analysis is to identify all possible globally consistent behaviors. The only way more than one behavior can arise is if there are some ambiguities in the propagation. It is important to note that a single heuristic, by itself, does not introduce any ambiguity. Rather, an ambiguity is only introduced if at some point in the propagation two or more heuristics apply. The ambiguity is whether either one or both of the heuristics are valid. However, it is not possible to tell, locally, which is the case. Only additional propagation can distinguish among the possibilities.

The only mechanism for disambiguation is contradiction detection. Suppose heuristics A and B apply at some impasse. There is now a potential ambiguity between  $\{A\}$ ,  $\{B\}$  and  $\{A, B\}$ . A contradiction involving only A, rules out  $\{A\}$  and  $\{A, B\}$  leaving B as the only option. Likewise a contradiction involving only B, leaves A as the only option. Both these contradictions completely remove the local ambiguity (and hence the global ambiguity). However, there is a third case which we haven't yet considered, neither A nor B may individually contradict, however their combination  $\{A, B\}$  might. In this situation either A or B holds, but not at once. Then, and only then, is there a global ambiguity.

## 7.3. Implementation

The algorithm explores all possibilities at once, never backtracking (which is unnecessary if the underlying algorithm will find all the possibilities). The propagator is essentially the same as the one used earlier propagating KCL and KVL first, then component rules, and finally heuristics.

The implementation makes no commitments about global ambiguity until propagation finishes. As each value triple (i.e.,  $\langle \text{value}, \{\text{assumptions}\}, \text{justifications}(s) \rangle$ ) carries with it the set of assumptions under which it holds, each value carries with it the ambiguities that matter to it. This design decision has other important benefits. It is not necessary to have a completely separate causal argument structure for each global ambiguity. For example, a multi-stage amplifier may have no ambiguity until the last stage. All of the cells for the earlier stages will have only one value which applies to all global ambiguities. The causal argument for any value in the last stage only differs within that final part of the argument within the last stage. Aside from the

obvious efficiency advantages of this approach, with this representation it is trivial to see where two global ambiguities differ.

In the simple propagator each cell representing a circuit quantity could have at most one value. When a second value was discovered, one of the two would be chosen and propagated further. Now each cell may have any number of value triples and each propagated. It is important to note that when a cell receives a second value, the other values in the cell are not changed. In particular, the justification tree of a propagation does not change even if other values are discovered for the cells of its antecedent values. As a consequence, the set of assumptions underlying a propagated value never change.

Care must be taken to identify all contradictions and ensure that only the simplest argument is propagated. Both of these goals are achieved by the algorithm which adds a new value to a cell. If the qualitative expression of a new value triple differs from the qualitative expression of an existing value triple, the union of the set of assumptions of the old and new value triples is marked as contradictory. (And all superset assumption sets are also marked contradictory and all other value triples depending on these assumption sets are removed from the cells and EQUAL's propagation queues.) If the new value triple is equal to an old value triple (i.e., same qualitative expression, same justification, and same assumption set) it is ignored. Simpler value triples have fewer assumptions. If the assumption set of the new value triple is a superset of an old value triple's, the new value triple is likewise discarded. If the assumption set of the new value triple is a subset of an old value triple's, the old value triple is discarded instead and all its consequents are removed from the cells and the queues. Only if the new value triple passes these tests is it added to the cell and allowed to propagate further. Thus, the simplest causal argument is found for each cell, and no contradictory value triples are propagated.

The implementation for the constraints themselves must also be modified to accept multiple-valued antecedents and to run multiple times. Naively, if a rule has two antecedents the first having  $n$  values, the second  $m$  values, the rule is applied  $nm$  times producing  $nm$  results. Unfortunately, the situation is far more subtle: this simple scheme is extremely inefficient and produces faulty propagations. It is unnecessary to apply the rule to cell values present the last time the rule was run. The new values can be easily distinguished from old values: the new values are on the queue of pending propagations. In addition, it is not worth the effort of generating the result if it is known to be contradictory. This is done by precomputing the union of the antecedent assumption sets to determine whether they are contradictory (i.e., contain a contradiction as a subset).

Suppose a particular heuristic is used twice during a causal analysis with the same antecedent cells, but different values of those cells. And furthermore, suppose the ensuing propagations from the two assumptions interact. This case,

which occurs rarely, is clearly wrong by the preceding argument. This is avoided by treating each instance of a heuristic, even those having the same antecedent cells as mutually contradictory.

Applications of heuristics require special care. A heuristic usefully applies only if there is an impasse in the causal propagation, but if cells contain multiple value triples it is harder to see whether an impasse has been reached within some assumption set. The algorithm used is complex, but it is equivalent to an efficient implementation of the following simple and inefficient scheme. As usual run heuristic rules last, and just assume that every value is at an impasse. The pruning algorithm which applies when multiple values are added to a cell eliminates at the unnecessary heuristic applications. The actual implementation of the algorithm looks ahead to see whether the pruning algorithm would immediately eliminate the heuristic.

As manipulation of sets of assumptions is the most fundamental and common type of operation of causal analysis, some care must be taken to implement these potentially complex algorithms. EQUAL uses a subset of a full TMS algorithm (see [5] for a description) which is highly optimized for the operations needed for qualitative reasoning. Each assumption and set of assumptions is made unique and hashed into a lattice structure. This lattice structure directly represents superset and subset relationships among sets. Contradictions are automatically inherited along superset links and the data base of values is updated. The data structure is optimized for the two most common operations: adding an assumption to a set and computing the union of two sets. Aside from being optimized for causal analysis, this simple TMS has another significant advantage over Doyle's and McAllester's [14, 25]. Those TMS's force the data base to hold one consistent set of facts in at a time. Every contradiction brings different facts in and changes justifications for deductions. In EQUAL's scheme, multiple inconsistent views can be represented and compared at the same time, and additionally the justifications for deductions are never changed. The latter property is critical for causal analysis, and the former property is critical for comparing alternative causal arguments for a circuit quantity (discussed in Section 8.2).

## 8. Interpretations

An *interpretation* is a global point of view for circuit behavior which selects a single value for every cell, a coherent overall framework for circuit behavior, and a single mechanism graph. If the analysis is ambiguous, the circuit will have multiple interpretations. A fundamental success criteria of qualitative analysis is that the set of interpretations must be *complete* and *realizable*. Every possible behavior of every possible ideal circuit must be described by some interpretation (*completeness*) and every interpretation generated by causal reasoning must correspond to an ideal physically realizable circuit (*realizabil-*

ity). Put concretely, each of the six interpretations for CE-FEEDBACK (see Table 7) corresponds to the behavior of a version of CE-FEEDBACK with parameter values picked appropriately, and the behavior produced by any assignment of parameter values for CE-FEEDBACK is described by one of the interpretations.

Table 7 includes every circuit variable except the second current for two-terminal components. '#' indicates that no causal value is found for these quantities. Causal propagation only finds values for the interesting circuit variables.

As realizability and completeness are fundamental to qualitative reasoning, I want to be very precise by what I mean by 'ideal': an ideal circuit behaves in accordance with the component models of electrical engineers. Although these quantitative models are extremely accurate and almost always adequate, they may not completely describe the precise behavior of real physical components. Realizability and completeness can only be applied to the idealizations of electrical engineering. (Admittedly, I would like these criteria to also hold with respect to the real physical world but that would force this research into an empirical methodology.) Of course, electrical engineers sometimes use different quantitative models and therefore realizability and completeness should be stated differently: given *any* set of quantitative models, causal analysis working with the corresponding qualitative models should produce realizable and complete interpretations with respect to those quantitative models.

TABLE 7. Causal interpretations.

	1	2	3	4	5	6		1	2	3	4	5	6
$\partial i_{-INPUT} =$	+	+	+	+	+	+	$\partial v_{C1,VCC} =$	+	0	-	-	-	-
$\partial i_{-B1} =$	+	+	#	#	#	#	$\partial v_{C1} =$	+	0	-	-	-	-
$\partial i_{-I(RC2)} =$	-	-	-	-	-	-	$\partial v_{E2,OUTPUT} =$	#	#	-	#	-	-
$\partial i_{-I(RF)} =$	+	+	+	+	+	+	$\partial v_{E2,VCC} =$	+	+	-	+	-	-
$\partial i_{-I(RB2)} =$	-	-	-	-	0	+	$\partial v_{E2} =$	+	+	-	+	-	-
$\partial i_{-I(RB1)} =$	-	-	-	-	-	-	$\partial v_{E2,C1} =$	+	+	+	+	+	+
$\partial i_{-I(RC1)} =$	-	0	+	+	+	+	$\partial v_{INPUT,OUTPUT} =$	#	#	#	#	#	#
$\partial i_{-C(Q2)} =$	-	-	-	-	-	-	$\partial v_{INPUT,VCC} =$	+	+	+	+	+	+
$\partial i_{-B(Q2)} =$	-	-	-	-	-	-	$\partial v_{INPUT} =$	+	+	+	+	+	+
$\partial i_{-E(Q2)} =$	+	+	+	+	+	+	$\partial v_{INPUT,C1} =$	+	+	+	+	+	+
$\partial i_{-C(Q1)} =$	+	+	+	+	+	+	$\partial v_{INPUT,E2} =$	+	+	+	+	+	+
$\partial i_{-B(Q1)} =$	+	+	+	+	+	+	$\partial v_{FP,OUTPUT} =$	#	#	#	#	-	-
$\partial i_{-E(Q1)} =$	-	-	-	-	-	-	$\partial v_{FP,VCC} =$	+	+	+	+	0	-
$\partial v_{VCC,OUTPUT} =$	-	-	-	-	-	-	$\partial v_{FP} =$	+	+	+	+	0	-
$\partial v_{OUTPUT} =$	+	+	+	+	+	+	$\partial v_{FP,C1} =$	+	+	+	+	+	+
$\partial v_{VCC} =$	0	0	0	0	0	0	$\partial v_{FP,E2} =$	+	+	+	+	+	+
$\partial v_{C1,OUTPUT} =$	#	-	-	-	-	-	$\partial v_{FP,INPUT} =$	-	-	-	-	-	-

### 8.1. Constraint satisfaction vs. causal analysis

Although realizability and completeness define the goal of causal analysis, they are very hard to attain in practice. However, it is possible to apply them partially and this partial technique was used to debug EQUAL, the component models and the heuristics. The ultimate test of qualitative analysis can be broken down into three steps: (1) quantitative models describe physical components, (2) qualitative models describe 'ideal' quantitative models, and (3) causal arguments describe each qualitative behavior. We avoid the first step by simply leaving quantitative modeling to the electrical engineers and physicists. The correctness of the second step is achieved by being systematic in constructing the component models (see [6]). The correctness of the last step can be evaluated explicitly. In particular, every qualitatively possible behavior should have a causal argument. This can be tested by explicitly stating every qualitative equation that causal propagation uses and solving the system of equations thus produced with a conventional constraint satisfaction technique which is unencumbered with producing causal arguments. The constraint satisfaction algorithm is equivalent to generate and test so it is guaranteed to find every possible solution. Every possible qualitative behavior should have a corresponding causal analysis and every possible causal analysis should correspond to a possible qualitative behavior.

For example, CE-FEEDBACK is described by 64 qualitative equations in 41 variables (adding more equations neither improves, or eliminates interpretations).

Each transistor is modeled by three qualitative equations:

$$\begin{aligned} \partial v_{E2,C1} + \partial i_{\rightarrow C(Q2)} &= 0, & \partial i_{\rightarrow E(Q2)} - \partial v_{E2,C1} &= 0, \\ \partial v_{E2,C1} + \partial i_{\rightarrow B(Q2)} &= 0, & \partial v_{INPUT} - \partial i_{\rightarrow C(Q1)} &= 0, \\ \partial v_{INPUT} + \partial i_{\rightarrow E(Q1)} &= 0, & \partial v_{INPUT} - \partial i_{\rightarrow B(Q1)} &= 0. \end{aligned}$$

Ohm's law for each resistor is:

$$\begin{aligned} \partial v_{C1,VCC} + \partial i_{\rightarrow \#1(RC1)} &= 0, & \partial v_{FP,E2} + \partial i_{\rightarrow \#1(RB1)} &= 0, \\ \partial v_{FP} + \partial i_{\rightarrow \#1(RB2)} &= 0, & \partial v_{FP,INPUT} + \partial i_{\rightarrow \#1(RF)} &= 0, \\ \partial i_{\rightarrow \#1(RC2)} - \partial v_{VCC,OUTPUT} &= 0. \end{aligned}$$

The battery model is:

$$\partial v_{VCC} = 0.$$

The input signal is:

$$\partial i_{\rightarrow INPUT} = +.$$

A KCL confluence is required for each component and node:

$$\partial i_{\rightarrow \text{INPUT}} - \partial i_{\rightarrow \#1(\text{RF})} - \partial i_{\rightarrow \text{B(Q1)}} = 0,$$

$$\partial i_{\rightarrow +(B1)} + \partial i_{\rightarrow -(B1)} = 0,$$

$$\partial i_{\rightarrow +(B1)} + \partial i_{\rightarrow \#1(\text{RC2})} + \partial i_{\rightarrow \#1(\text{RC1})} = 0, \quad \text{etc.}$$

A KVL equation is required for every three circuit nodes (35 equations).

$$\partial v_{\text{FP,E2}} - \partial v_{\text{FP,INPUT}} - \partial v_{\text{INPUT,E2}} = 0,$$

$$\partial v_{\text{FP,C1}} - \partial v_{\text{FP,INPUT}} - \partial v_{\text{INPUT,C1}} = 0,$$

$$\partial v_{\text{FP}} - \partial v_{\text{FP,INPUT}} - \partial v_{\text{INPUT}} = 0, \quad \text{etc.}$$

In quantitative analysis such a set of equations has a unique solution. However, constraint satisfaction reveals that the above set of confluences has 85 solutions! Many of these interpretations result because some variable is completely unconstrained. If totally unconstrained quantities are eliminated there are 41 interpretations (See Table 8; '?' is used to indicate the totally unconstrained variables; for brevity, quantities with constant value are left out and only one terminal current is included for two-terminal components). Constraint satisfaction discovers far more interpretations than causal analysis because values for causally unimportant variables are included (as defined in Section 3.4.5). Table 7 can be collapsed by merging causally unimportant variables. In particular, if two interpretations differ only in a causally unimportant variable, the interpretations are merged and a '\*' substituted for the uninteresting variable. Note that the values for uninteresting variables which need not be merged are left unchanged. The result after merging is five interpretations (Table 9).

Table 9 matches Table 7 as follows: 1 to 4, 2 to 5, 3 to 3, 4 to 3, 5 to 2, and 6 to 1. A comparison of the results of causal analysis and constraint satisfaction reveals two curiosities. As expected, constraint satisfaction discovers many more values, but always for causally uninteresting variables. Second, there are fewer interpretations generated by constraint satisfaction than with causal analysis. Constraint satisfaction interpretation 3 (Table 9) corresponds to causal interpretations 3 and 4 (Table 7). Causal interpretations 3 and 4 differ only in uninteresting variables  $\partial v_{\text{E2,OUTPUT}}$ ,  $\partial v_{\text{E2,VCC}}$  and  $\partial v_{\text{E2}}$ . Causal analysis produced differing values for these uninteresting quantities because these two interpretations have differing mechanism graphs. Note that causal analysis does not reliably produce all values for uninteresting quantities. For example, the full 41 interpretation constraint satisfaction (Table 8) shows that  $\partial v_{\text{E2,VCC}}$  and  $\partial v_{\text{E2}}$  can both be zero, while no causal interpretation shows this.

It is more difficult to collapse causal interpretations than constraint satisfaction interpretations because the causal arguments and mechanism graphs need to be merged as well. However, on the basis of qualitative values alone,

TABLE 8

Interpretation	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41				
$\partial i_{\rightarrow+(B1)} =$	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
$\partial i_{\rightarrow\#1(RF)} =$	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
$\partial i_{\rightarrow\#(RB2)} =$	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	+	-	-	-	-	-	-	-	-	-	0	+	-	-	-	-	-	-	-	-	-	-	-	0	+		
$\partial i_{\rightarrow\#1(RC1)} =$	-	-	+	0	-	+	+	+	0	-	+	+	+	0	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
$\partial v_{C1,OUTPUT} =$	+	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
$\partial v_{C1,VCC} =$	+	+	-	0	+	-	-	-	0	+	-	-	-	0	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
$\partial v_{C1} =$	+	+	-	0	+	-	-	-	0	+	-	-	-	0	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
$\partial v_{E2,OUTPUT} =$	+	+	?	?	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
$\partial v_{E2,VCC} =$	+	+	+	+	+	0	-	+	+	+	0	-	+	+	+	0	-	+	-	+	0	-	+	0	-	+	0	-	+	0	-	-	-	+	0	-	+	0	-	+	0	-	-	-	-
$\partial v_{E2} =$	+	+	+	+	+	0	-	+	+	+	0	-	+	+	+	0	-	-	-	+	0	-	+	0	-	+	0	-	-	-	+	0	-	+	0	-	+	0	-	+	0	-	-	-	-
$\partial v_{INPUT,OUTPUT} =$	+	+	+	+	+	+	+	+	+	+	+	?	?	?	?	?	?	?	?	+	+	+	+	+	+	?	?	?	?	?	?	+	+	+	+	+	+	+	?	?	?	?	?	?	?
$\partial v_{FP,OUTPUT} =$	+	+	+	+	+	+	+	0	0	0	0	0	-	-	-	-	-	-	-	+	+	+	0	0	0	-	-	-	-	-	+	+	+	0	0	0	-	-	-	-	-	-	-	-	-
$\partial v_{FP,VCC} =$	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0	-	+	+	+	+	+	+	+	+	+	+	0	-	+	+	+	+	+	+	+	+	+	+	+	0	-	
$\partial v_{FP} =$	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0	-	+	+	+	+	+	+	+	+	+	+	0	-	+	+	+	+	+	+	+	+	+	+	+	0	-	

TABLE 9. Satisfaction interpretations

	1	2	3	4	5		1	2	3	4	5
$\partial i \rightarrow \text{INPUT} =$	+	+	+	+	+	$\partial v_{C1.VCC} =$	-	-	-	+	0
$\partial i \rightarrow +(B1) =$	*	*	*	+	+	$\partial v_{C1} =$	-	-	-	+	0
$\partial i \rightarrow \#1(RC2) =$	-	-	-	-	-	$\partial v_{E2.OUTPUT} =$	-	-	*	*	*
$\partial i \rightarrow \#1(RF) =$	+	+	+	+	+	$\partial v_{E2.VCC} =$	-	-	*	+	+
$\partial i \rightarrow \#1(RB2) =$	+	0	-	-	-	$\partial v_{E2} =$	-	-	*	+	+
$\partial i \rightarrow \#1(RB1) =$	-	-	-	-	-	$\partial v_{E2.C1} =$	+	+	+	+	+
$\partial i \rightarrow \#1(RC1) =$	+	+	+	-	0	$\partial v_{INPUT.OUTPUT} =$	?	?	*	*	*
$\partial i \rightarrow C(Q2) =$	-	-	-	-	-	$\partial v_{INPUT.VCC} =$	+	+	+	+	+
$\partial i \rightarrow B(Q2) =$	-	-	-	-	-	$\partial v_{INPUT} =$	+	+	+	+	+
$\partial i \rightarrow E(Q2) =$	+	+	+	+	+	$\partial v_{INPUT.C1} =$	+	+	+	+	+
$\partial i \rightarrow C(Q1) =$	+	+	+	+	+	$\partial v_{INPUT.E2} =$	+	+	+	+	+
$\partial i \rightarrow B(Q1) =$	+	+	+	+	+	$\partial v_{FP.OUTPUT} =$	-	-	*	*	*
$\partial i \rightarrow E(Q1) =$	-	-	-	-	-	$\partial v_{FP.VCC} =$	-	0	+	+	+
$\partial v_{VCC.OUTPUT} =$	-	-	-	-	-	$\partial v_{FP} =$	-	0	+	+	+
$\partial v_{OUTPUT} =$	+	+	+	+	+	$\partial v_{FP.C1} =$	+	+	+	+	+
$\partial v_{VCC} =$	0	0	0	0	0	$\partial v_{FP.E2} =$	+	+	+	+	+
$\partial v_{C1.OUTPUT} =$	-	-	-	*	-	$\partial v_{FP.INPUT} =$	-	-	-	-	-

causal analysis and constraint satisfaction predict the same set of behaviors. Thus, for CE-FEEDBACK every possible behavior has a causal argument and every possible causal argument corresponds to a possible behavior. The desideratum is that this be true for every possible circuit. Note, however, that the number of interpretations need not always match. Sometimes, as is the case with CE-FEEDBACK the same constraint satisfaction analysis may have multiple causal analyses. This also occurs in the few rare instances where the identical behavior (interesting and uninteresting quantities) has multiple causal explanations. There will never be more constraint satisfaction interpretations than causal interpretations due to the merging of uninteresting variables.

The technique of constraint satisfaction is used extensively in debugging EQUAL and its models as well as experimenting with differing heuristics. EQUAL has been run on hundreds of examples. For each, the causal and constraint satisfaction interpretations have been compared (via another program) to confirm that every possible behavior has a causal argument. The evidence on the adequacy of the KVL and KCL equation formulation was obtained using these means.

## 8.2. Causal interpretations

Although causal analysis finds values for each circuit quantity, it does not separate out which values belong to which causal argument or global behavior. This is the role of causal interpretations. Unlike the notion of constraint-

satisfaction interpretations which arise out of differences in qualitative values, causal interpretations arise out of differences in causal assumptions. A causal interpretation is defined by a set of assumptions and selects those value triples whose assumption set is a subset of the interpretation assumption set. Each interpretation is intended to correspond to a single global behavior and the values for a single well-formed mechanism graph.

The remainder of this section presents criteria for selecting among these interpretations to identify which one best describes the 'normal' behavior of the circuit. The set of tentative interpretations is generated by finding all sets of assumptions which (1) could be a valid assumption set for a value; imagine a final circuit value whose antecedents were every circuit quantity: the assumption set must be contradiction-free, (2) is maximal, i.e., the inclusion of any additional assumption either introduces a contradiction or adds no information, and (3) is minimal, i.e., the removal of any assumption removes all values from some circuit quantity. CE-FEEDBACK has 6 causal interpretations:

- 1: {[INPUT IN][FP RB1][Q2  $v_B$ ][C1 Q1]}
- 2: {[INPUT IN][FP RF RB1][Q2  $v_B$ ][C1 Q1][RF  $v_{\#1}$ ]}
- 3: {[INPUT IN][C1 Q1][E2 RB1][RB1  $v_{\#2}$ ][FP RF][RF  $v_{\#1}$ ]}
- 4: {[INPUT IN][E2 Q2][Q2  $v_B$ ][C1 Q1][FP RF][RF  $v_{\#1}$ ]}
- 5: {[INPUT IN][C1 Q2 Q1][Q2  $v_E$ ][E2 RB1][RB1  $v_{\#2}$ ][FP RF][RF  $v_{\#1}$ ]}
- 6: {[INPUT IN][C1 Q2][Q2  $v_E$ ][E2 RB1][RB1  $v_{\#2}$ ][FP RF][RF  $v_{\#1}$ ]}

This numbering corresponds to that of Table 7.

Coincidentally, Interpretation 1 is the one the designer probably intended. Its mechanism graph corresponds to the one we have been using earlier (Fig. 16). Interpretations 2 and 4 arise when there is no feedback in the circuit, but the signal is being amplified normally. Interpretations 2 and 4 differ in the behaviors of circuit quantities (around node FP) not on the main signal path so they have the same mechanism graph. Fig. 13 illustrates the mechanism graph. Interpretation 3 arises when the signal travels forward instead of backward on the feedback path and combines with the main signal at the base-emitter of Q2. Its mechanism graph is Fig. 28 and the flow is illustrated by Fig. 29.

Interpretations 5 and 6 occur when the signal reversing the feedback path dominates and the gain along the normal signal path is non-existent. The differences between Interpretations 5 and 6 (around node C1) do not affect the output value so their mechanism graphs are the same. Fig. 30 is the mechanism graph and Fig. 31 illustrates the causal flow.

All but one of these interpretations are nonsensical. The second and fourth interpretations have a useless RF. The third interpretation feeds forward a signal that is orders of magnitude smaller than the signal it is added to at Q2. The smaller signal through RF serves no purpose. The fifth and sixth interpretations

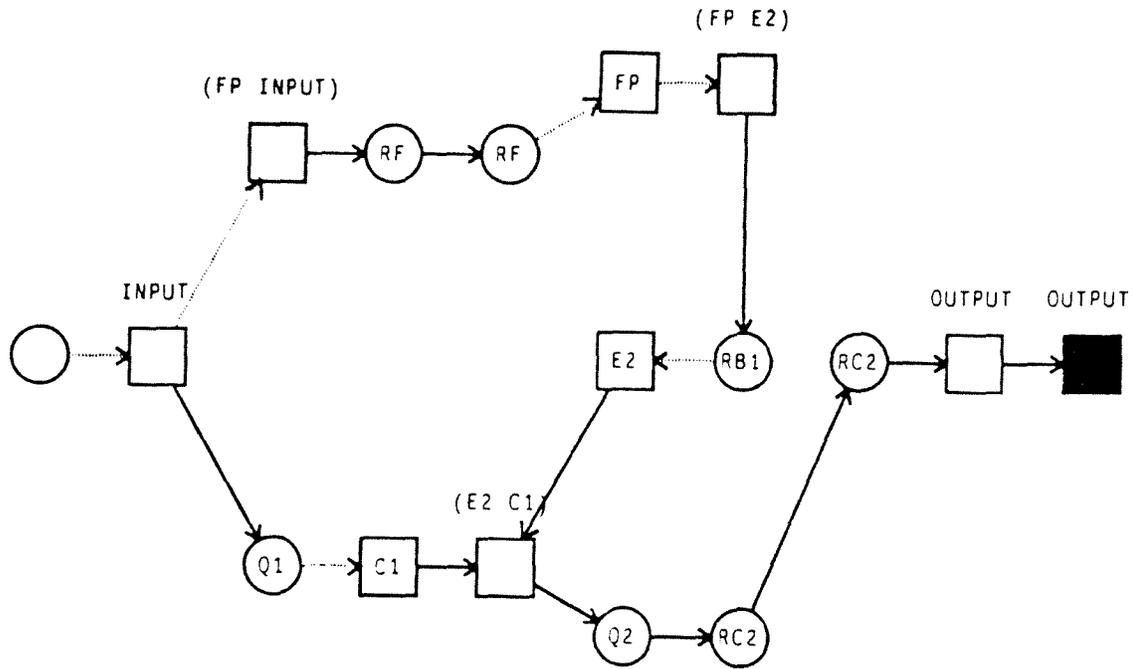


FIG. 28. Interpretation 3 (Feedforward) of CE-FEEDBACK.

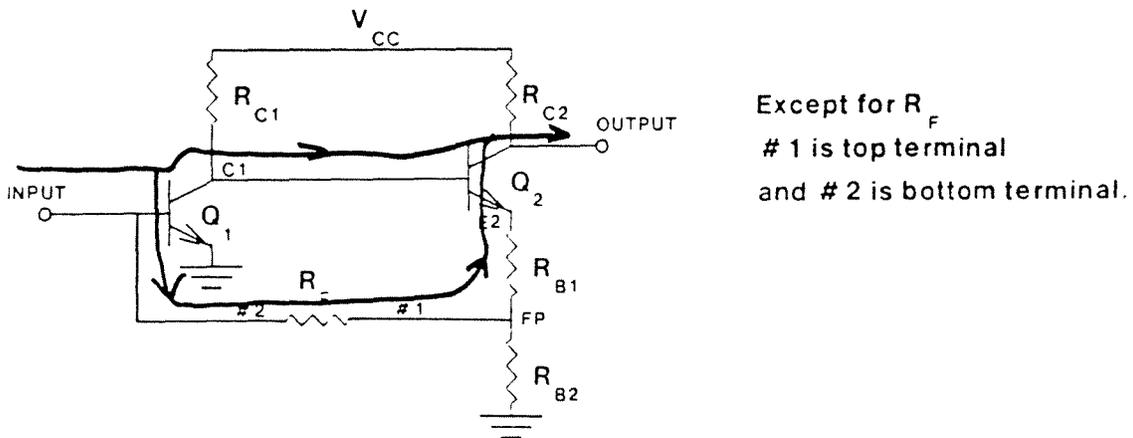


FIG. 29. CE-FEEDBACK.

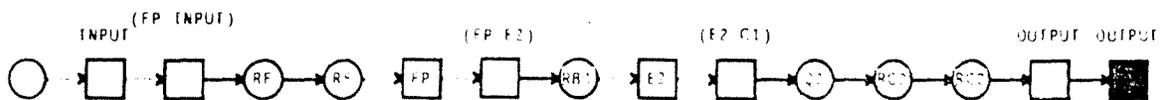
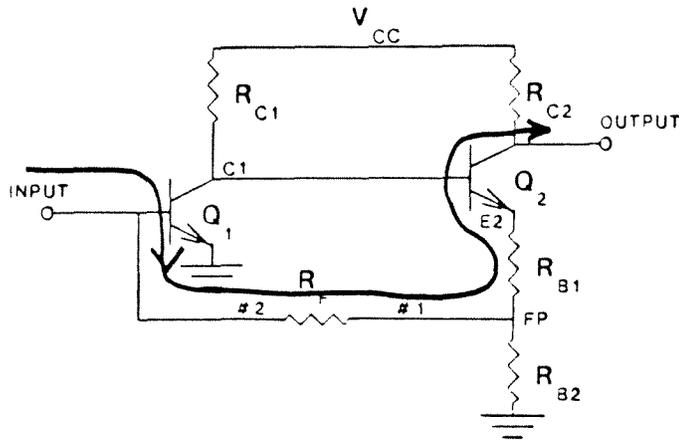


FIG. 30. Interpretations 5 and 6 (Unity Gain) of CE-FEEDBACK.

have no gain and Q1 and RC1 need not even be part of the circuit. These kinds of judgments for choosing the 'correct' interpretation are made by EQUAL's teleological reasoning. The task of causal analysis is to produce all the theoretical possibilities. With appropriate choices of resistor values and transistor gains each of these interpretations could arise.



Except for  $R_F$   
 # 1 is top terminal  
 and # 2 is bottom terminal.

FIG. 31. CE-FEEDBACK.

### 8.3. Quantitative criteria

If the circuit is linear, given the numerical values of circuit parameters and inputs, quantitative analysis produces a single solution for the circuit. The signs of the variables describe the correct interpretation. As a consequence a particular linear circuit can exhibit only one of its interpretations and this interpretation can be determined using numerical analysis or symbolic algebra. Both of these quantitative techniques can be avoided, as a small set of easily applied rules usually selects the correct interpretation.

### 8.4. Teleological criteria

The causal interpretations define all the physically realizable behaviors of the circuit. For most the circuits, one of these interpretations can often be identified as the one the designer originally intended. Knowing the ultimate purpose of the circuit makes it easy to identify the correct interpretation. Surprisingly, the correct interpretation can be identified without knowing the circuit's ultimate purpose. The conventions for circuit design are so stylized and the designer is under such tight economic constraints that one interpretation usually stands out as the one the designer intended. This section presents a number of selection criteria for identifying the correct interpretation.

These selection criteria all depend on the fact that the circuit is well-designed in the first place using functional criteria alone. If an engineer adds a resistor because it makes the schematic look like a picture of his house, or adds a useless transistor so he can advertise it as a six instead of five transistor radio, the criteria may fail. The same is true for badly designed circuits or circuits which cannot possibly work.

The first criterion is that the circuit has a non-zero (i.e., changing) output signal. The designer cannot have intended the circuit to have no output signal. CE-FEEDBACK has no such interpretations.

The second criterion is that the circuit does not propagate a zero-value assuming two currents or voltages exactly cancel. Such situations can exist only if the signals perfectly balance, which is impossible in practice given that they come from different circuit paths and, in any case, can only exist momentarily in the face of a changing input signal. This criterion rules out Interpretations 2 and 6 of CE-FEEDBACK. Interpretation 2 had the current from RF perfectly balancing the current from RB1. In Interpretation 6 the collector current of Q1 exactly equaled the base current of Q2. Note that both of these are mathematical possibilities, but no circuit could utilize such a perfect balance.

Third, each component should have a known causal configuration. Fourth, causal or feedback configurations are preferred over reflection configurations which are preferred over no configurations at all. Finally, there are some configurations which are known, a priori, to be unlikely. These three criteria implement the intuition that the correct interpretation is the one which has the 'best' teleological explanation for each component of the circuit. This is based on two presuppositions. First, causal analysis finds every possible behavior. Second, that the taxonomy of component configurations is complete or *or at least corresponds to the ones design engineers are actually using*.

The preference criteria are best illustrated by an example. Table 10 lists all the component configurations. CE-FEEDBACK has four interpretations: correct (Fig. 16, Interpretation 1), feedbackless (Fig. 13, Interpretation 4), feedforward (Fig. 28, Interpretation 3) and unity gain (Fig. 30, Interpretation 5). '(C)', '(F)' and '(R)' indicate causal, feedback and reflection configurations.

The interpretation which assigns the best purpose to each component is preferred. As a causal or feedback configuration contributes directly to a circuit

TABLE 10. Interpretation configurations for CE-FEEDBACK

Component	Configuration			
	Correct (1)	Feedbackless (4)	Feedforward (3)	Unity gain (5)
Q1	CE(C) LOAD-Q(R)	CE(C) LOAD-Q(R)	CE(C) LOAD-Q(R)	LOAD-Q(R)
Q2	CE(C) CC(F) LOAD-Q(R)	CE(C) LOAD-Q(R)	SUM(C) SENSE-Q(R)	CB(C)
RC1	I-LOAD(R)	I-LOAD(R)	I-LOAD(R)	I-LOAD(R)
RC2	I-LOAD(C)	I-LOAD(C)	I-LOAD(C)	I-LOAD(C)
RB2	I-LOAD(R)	I-LOAD(R)	I-LOAD(R)	I-LOAD(R)
RB1	I-TO-V-COUPLE(F)	I-LOAD(R)	DIVIDER(C) I-LOAD(R)	DIVIDER(C) I-LOAD(R)
RF	V-SENSOR(F) I-LOAD(R)	I-LOAD(R)	DIVIDER(C) I-LOAD(R)	DIVIDER(C) I-LOAD(R)

output it is preferred over a reflection configuration. A reflection configuration is preferred over no configuration at all. This preference ordering on individual component configurations places a partial order on interpretations which can be represented as a lattice. The maximal elements of the resulting lattice are the feedforward and the correct interpretations. The feedforward interpretation is preferred over the unity gain interpretation because it assigns a causal purpose to Q1 (they both assign the same purpose types to the remaining components, e.g., Q2 functions in a causal configuration in both interpretations). The correct interpretation is preferred over the feedbackless interpretation because it uses RF in a causal configuration.

Certain component configurations are unlikely *a priori*. For example, it is very unlikely that a transistor is functioning in a SUM configuration (adding the voltages at its emitter and base) because the gains and input impedances of the base and emitter signal paths are so different. Also, it is unlikely that a transistor is only functioning in LOAD-Q, COUPLING-Q, or SENSE-Q configurations. All these roles can be achieved far more cheaply with a diode. This preference establishes another partial order which almost always has a single maximum. In the case of CE-FEEDBACK, the correct interpretation is preferred over the feedforward configuration as it is unlikely Q2 is functioning in a SUM configuration.

#### 8.4.1. *More selection criteria*

If the circuit has more than one input or output (perhaps both), then that interpretation which has each input affecting as many outputs as possible is preferred.

Only the feedback path from a single sampling point with the most active components along it is included in the mechanism graph used to generate configurations and select interpretations. This final criterion applies when the circuit has multiple intertwined feedback loops (in particular multiple comparison points from a single sampling point). With the rules so far, the wrong interpretation would be selected.

The only cases where these criteria have been observed to fail are circuits which have some kind of symmetry (e.g., a push-pull output stage). For such circuits, either the two interpretations are identical and selecting either one would produce a correct analysis, or the circuit switches interpretations in its normal functioning (a situation which cannot be represented in EQUAL).

It is not hard to design a circuit for which they fail, but it is surprisingly difficult to design a circuit for which it both fails and would sensibly be designed by an experienced analog circuit designer. To construct a counter-example, one must come up with a circuit whose desired behavior was not expected by an electrical engineer *and* whose design cannot be simplified (i.e., badly

designed circuits in which only a subset of the components were performing any useful function).

### 9. Operating Regions

Electrical engineers describe the behavior of nonlinear components with piecewise linear approximations. The entire region of operation is broken down into a collection of subregions, or states, within which the behavior is approximately linear. For example, the diode model is

$$i = I_s(e^{qv/kT} - 1),$$

and its behavior is modeled by two regions. If  $v$  is small,  $e^{qv/kT}$  is nearly one and the current is nearly zero. If  $v$  is large, the exponential is nearly a straight line and can be modeled by a simple resistor. These are called the *on* and *off* regions.

The qualitative model uses the same regions:

$$\begin{aligned} i_A &= \text{current into anode,} \\ v_{A,C} &= \text{voltage from anode to cathode,} \\ \text{OFF: } &(v_{A,C} \leq v_T), \quad \partial i_A = 0, \\ \text{ON: } &(v_{A,C} > v_T), \quad \partial v_{A,C} \Rightarrow \partial i_A. \end{aligned}$$

The parenthesized conditions define the regions. Thus, if the voltage across the diode is less than some threshold, the diode is off and its current constant. For silicon diodes  $v_T = 0.55$  volts, for germanium diodes the threshold is much less.

In qualitative analysis, nonlinearity is not as much of a problem as for quantitative analysis. For example, the equation  $\partial v_{A,C} = \partial i_A$  equally well describes the behavior of a linear resistance or an exponential diode effect. Component states are still important in qualitative analysis when behavior is qualitatively different in different regions, however, it is not necessary to assume linearity. Thus, qualitative modeling has not precluded the analysis of nonlinear systems.

The transistor is typically modeled by three regions: on, off and saturated. However, there is a fourth, often ignored, region important to digital circuitry called the reverse-active (R-ON in the model) region. The following is a model for the NPN transistor (the PNP model is identical except that the state specifications are all reversed):

ON:

$$\begin{aligned} &(v_{b,c} < v_T, v_{b,e} \geq v_T), \\ &\partial v_{b,e} \Rightarrow \partial i_c, \quad \partial v_{b,e} \not\Rightarrow \partial i_e, \quad \partial v_{b,e} \Rightarrow \partial i_b; \end{aligned}$$

R-ON:

$$(v_{b,c} \geq v_T, v_{b,e} < v_T),$$

$$\partial v_{c,b} \Rightarrow \partial i_c, \quad \partial v_{c,b} - \Rightarrow \partial i_e = 0, \quad \partial v_{c,b} \Rightarrow \partial i_b = 0;$$

OFF:

$$(v_{b,c} < v_T, v_{b,e} < v_T),$$

$$\partial i_c = 0;$$

SATURATED:

$$(v_{b,e} \geq v_T, v_{b,c} \geq v_T),$$

$$\partial v_{b,e} - \partial v_{c,b} \Rightarrow \partial i_b.$$

### 9.1. Composite circuit state

The composite state of the overall circuit is defined in terms of the states of its individual components. All of the analyses of CE-FEEDBACK presumed that both its transistors were in their forward-active regions. This composite circuit state is represented as {(Q1 ON) (Q2 ON)}. Using the above transistor model it can be seen that the CE-FEEDBACK has 16 possible circuit states. For many circuits, it is quite reasonable to assume that unless shown to the contrary, all transistors are in their forward-active region. However, it is not necessary to depend on this presupposition as the 'normal' composite state can usually be determined directly. First, some composite states are contradictory in that they have no causal interpretations at all. Second, many composite states have contradictory quiescent models or violate the boundary conditions. Third, of the remaining possible states, the same teleological techniques used to select the correct interpretation are used to select the correct state.

### 9.2. Implementation

The introduction of component states requires little additional machinery. Component states are treated exactly as assumptions. The different regions of operation for a component are incompatible assumptions. This has the computational advantage that composite states share values and that any set of contradictory component states immediately rules out any superset component state set. Typically, in larger circuits, the state of one transistor determines the states of its neighbors. As a consequence, the set of circuit states needed to be explored only grows linearly in the number of active components.

### 9.3. Quiescent analysis

Quiescent analysis, unlike causal analysis, determines the values of the quantities, not how they change. Neither causal analysis, teleological analysis nor

interpretation selection utilizes quiescent analysis. This should not be too surprising, as most circuits achieve their function by responding to changing inputs or producing changing outputs. Quiescent analysis is important for determining the possible composite states of a circuit and for establishing sign conventions for the printout of causal arguments.

For linear components, the quiescent and incremental models are identical. Thus most of the quiescent models were presented in the beginning the paper. The transistor, a nonlinear component, has a quite different quiescent model. The values of the voltages can often be determined from the state specifications. This is the NPN transistor model.

ON:

$$(v_{b,c} < v_T, v_{b,e} \geq v_T),$$

$$[i_b] = +, [i_c] = +, [i_e] = -, [v_{b,e}] = +, [v_{c,e}] = +;$$

R-ON:

$$(v_{b,c} \geq v_T, v_{b,e} < v_T),$$

$$[i_b] = +, [i_c] = -, [i_e] = +, [v_{c,e}] = -, [v_{c,b}] = -;$$

OFF:

$$(v_{b,c} < v_T, v_{b,e} < v_T),$$

$$[i_b] = 0, [i_c] = 0, [i_e] = 0;$$

SATURATED:

$$(v_{b,e} \geq v_T, v_{b,c} \geq v_T),$$

$$[i_b] = +, [v_{b,e}] = +, [v_{c,b}] = -.$$

The propagation algorithm of causal analysis can be used to determine the quiescent values, but the heuristics make little sense and the propagation may reach an impasse. As causal explanations are unimportant for quiescent analysis, brute-force constraint satisfaction suffices to obtain the values (however, the logical framework presented in [6] will provide explanations for the conclusions, but they are just not causal explanations).

Constraint satisfaction on CE-FEEDBACK uses 66 confluences in 39 unknowns. The transistor confluences are:

$$\begin{array}{lll} [i_{\rightarrow B(Q2)}] = + & [i_{\rightarrow C(Q2)}] = + & [i_{\rightarrow E(Q2)}] = - \\ [v_{E2,C1}] = - & [v_{E2,OUTPUT}] = - & [i_{\rightarrow B(Q1)}] = + \\ [i_{\rightarrow C(Q1)}] = + & [i_{\rightarrow E(Q1)}] = - & [v_{INPUT}] = + \quad [v_{C1}] = + \end{array}$$

The quiescent model for the battery is:

$$[v_{VCC}] = +.$$

The remaining quiescent confluences are analogous to the incremental confluences presented earlier. Ohm's law, KCL and KVL are all linear constraints and thus the quiescent confluences are the same as the incremental confluences. There is one less KCL confluence as the input port is presumed to supply zero bias.

These confluences have three solutions (only because the uninteresting quantity  $v_{C1,OUTPUT}$  can be +, 0 or -).

$$\begin{array}{lll}
 [i_{\rightarrow(B1)}] = + & [i_{\rightarrow B(Q2)}] = + & [v_{E2}] = + \\
 [i_{\rightarrow(B1)}] = - & [i_{\rightarrow E(Q2)}] = - & [v_{E2,C1}] = - \\
 [i_{\rightarrow\#2(R2)}] = - & [i_{\rightarrow C(Q1)}] = + & [v_{INPUT,OUTPUT}] = - \\
 [i_{\rightarrow\#1(RC2)}] = + & [i_{\rightarrow B(Q1)}] = + & [v_{INPUT,VCC}] = - \\
 [i_{\rightarrow\#2(RF)}] = + & [i_{\rightarrow E(Q1)}] = - & [v_{INPUT}] = + \\
 [i_{\rightarrow\#1(RF)}] = - & [v_{VCC,OUTPUT}] = + & [v_{INPUT,C1}] = - \\
 [i_{\rightarrow\#2(RB2)}] = + & [v_{OUTPUT}] = + & [v_{INPUT,E2}] = - \\
 [i_{\rightarrow\#1(RB2)}] = - & [v_{VCC}] = + & [v_{FP,OUTPUT}] = - \\
 [i_{\rightarrow\#2(RB1)}] = - & [v_{C1,OUTPUT}] = ? & [v_{FP,VCC}] = - \\
 [i_{\rightarrow\#1(RB1)}] = + & [v_{C1,VCC}] = - & [v_{FP}] = + \\
 [i_{\rightarrow\#2(RC1)}] = - & [v_{C1}] = + & [v_{FP,C1}] = - \\
 [i_{\rightarrow\#1(RC1)}] = + & [v_{E2,OUTPUT}] = - & [v_{FP,E2}] = - \\
 [i_{\rightarrow C(Q2)}] = + & [v_{E2,VCC}] = - & [v_{FP,INPUT}] = +
 \end{array}$$

Quiescent analysis shows that 31 of the possible 36 composite states of CE-FEEDBACK are contradictory. For example, Q1 cannot be ON if Q2 is OFF. For Q1 to be ON, it must be supplied a bias, this bias can only come through RF, node FP can only be pulled up from RB1, but if Q2 is OFF no current can be flowing through RB1 pulling up FP and thus Q1 cannot be receiving any bias. Similar, proof-by-contradiction arguments eliminate the remaining 30 circuit states. The 5 possible consistent circuit states are:

$$\begin{array}{lll}
 \{(Q1\ OFF)(Q2\ SAT)\} & \{(Q1\ ON)(Q2\ SAT)\} & \{(Q1\ OFF)(Q2\ OFF)\} \\
 \{(Q1\ OFF)(Q2\ ON)\} & \{(Q1\ ON)(Q2\ ON)\} &
 \end{array}$$

Only the last two states allow the input to propagate all the way to the output, producing an interesting behavior. Using the interpretation selection rules, state  $\{(Q1\ OFF)(Q2\ ON)\}$  is an unlikely choice because it doesn't use Q1 in any useful configuration at all. Hence, the normal operating state of CE-FEEDBACK is  $\{(Q1\ ON)(Q2\ ON)\}$ .

#### 9.4. Energy storage elements

With the models presented thus far quiescent ambiguity within a state has no effect on the causal analysis. It only matters whether a state is contradictory or not. This is because no incremental confluence references a quiescent variable (or vice versa). This is not the case for models described by differential equations (e.g., inductors and capacitors). The behavior of a capacitor is described quantitatively by the differential equation

$$\begin{aligned} i &= \text{current flowing into the capacitor,} \\ v &= \text{voltage across the capacitor,} \\ C &= \text{capacitance which is a fixed positive quantity,} \\ i &= C \, dv/dt. \end{aligned}$$

In qualitative terms this is

$$[i] = \left[ C \frac{dv}{dt} \right] = [C] \left[ \frac{dv}{dt} \right] = \left[ \frac{dv}{dt} \right] = \partial v.$$

Or more generally, as it is a linear element:

$$\partial^{n-1} i = \partial^n v.$$

Analogously, the inductor modeled quantitatively as

$$v = L \, di/dt$$

has the model:

$$\begin{aligned} i &= \text{current flowing into the inductor,} \\ v &= \text{voltage across the inductor,} \\ [v] &= \partial i. \end{aligned}$$

Or more generally:

$$\partial^{n-1} v = \partial^n i.$$

Consider the simple RC circuit of Fig. 32. This circuit is modeled with 5 quiescent confluences, one Ohm's law confluence, four KCL confluences and no KVL confluences. Note that the capacitor introduces no quiescent confluence<sup>10</sup> other than a KCL. A capacitor places no constraints on the quiescent behavior, that is left to the rest of the circuit. However, the quiescent possibilities that result, directly affect the causal analysis.

<sup>10</sup>Assuming its charge is not known.

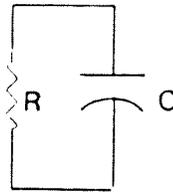


FIG. 32. Simple RC circuit.

$$\begin{aligned}
 [i_{\rightarrow \#1(R)}] - [v_N] &= 0, \\
 [i_{\rightarrow \#2(C)}] + [i_{\rightarrow \#1(C)}] &= 0, \\
 [i_{\rightarrow \#2(R)}] + [i_{\rightarrow \#2(C)}] &= 0, \\
 [i_{\rightarrow \#1(R)}] + [i_{\rightarrow \#1(C)}] &= 0, \\
 [i_{\rightarrow \#2(R)}] + [i_{\rightarrow \#1(R)}] &= 0.
 \end{aligned}$$

Not surprisingly there are 3 possible quiescent interpretations, each corresponding to one of the possible qualitative capacitor currents, see Table 11. Each one of the quiescent interpretations yields a different causal analysis, see Table 12.

Notice that unlike circuits considered thus far, a single RC network can exhibit all these interpretations without changing circuit conditions. Not one interpretation can be said to be the correct one. This is a consequence of the presence of energy storage elements. As  $[x]\partial x$  is  $-$  for all non-zero circuit quantities, they are moving towards zero.

Thus quiescent interpretations select incremental models in very much the same way as composite circuit state selects the incremental models. For the purpose of analysis, EQUAL considers the quiescent interpretations as introducing additional circuit states. However, it is only necessary to distinguish quiescent interpretations differing in the variables of energy storage elements. Thus, the three quiescent interpretations of CE-FEEDBACK differing in  $[v_{C1,OUTPUT}]$  collapse as no incremental confluence depends on its value. Said differently, capacitor currents and inductor voltages are the only interesting

TABLE 11

Interpretation	1	2	3
$[i_{\rightarrow \#1(C)}] =$	+	0	-
$[i_{\rightarrow \#2(C)}] =$	-	0	+
$[i_{\rightarrow \#1(R)}] =$	-	0	+
$[i_{\rightarrow \#2(R)}] =$	+	0	-
$[v_N] =$	-	0	+

TABLE 12

Solution	1	2	3
$\partial i_{\rightarrow \#1(C)} =$	-	0	+
$\partial i_{\rightarrow \#2(C)} =$	+	0	-
$\partial i_{\rightarrow \#1(R)} =$	+	0	-
$\partial i_{\rightarrow \#2(R)} =$	-	0	+
$\partial v_N =$	+	0	-

variables of quiescent interpretations contributing to the composite circuit state.

### 10. Time

As time passes the behavior and the causality of the circuit may change, either as a consequence of a new input behavior or of the internal workings of the circuit. For circuits such as CE-FEEDBACK, not much will change. The qualitative values continue to hold as long as the input signal is present. In other words the causal analysis (the qualitative values and the mechanism graph) holds for an extensive interval of time. This is not surprising, as CE-FEEDBACK has only one useful interpretation and state of behavior. The input signal may vary, first increasing then decreasing, but that just flips all the qualitative values, the interpretation and mechanism graph remain unchanged. Change induced by new inputs is uninteresting.

In general, two significant things can happen as time passes. First, qualitative integration takes place. Any non-zero incremental variable will cause changes in the quiescent variables.  $\partial x = +$  means  $x$  increases. If a significant amount of time passes this increase in  $x$  may become significant and qualitatively noticeable. For example, suppose a transistor is in the OFF state.  $\partial v_{b,e} = +$  causes  $v_{b,e}$  to increase which in turn causes  $v_{b,e}$  to become greater than the threshold  $v_T$ . As a consequence the transistor changes to state ON and the incremental behavior changes.

Second, the behavioral interpretations can change. If there is more than one quiescent interpretation, the non-zero incremental quantities may cause the quiescent interpretation to change. This can only happen if the circuit is either nonlinear or contains energy storage elements (otherwise the interpretation is uniquely determined by circuit conditions (component parameters and input signals).) In the simple RC circuit considered earlier the capacitor's current  $[i]$  is ambiguous producing three quiescent interpretations each having one incremental interpretation. In this example when  $[i] = +$   $\partial i = -$ , so the quiescent interpretation may change to one in which  $[i] = 0$ . The RC and transistor examples illustrate a fundamental problem of qualitative analysis. Although it is possible to identify which transitions might happen, it is difficult to determine whether they indeed do happen. Just because  $\partial v_{b,e} = +$  does not mean  $v_{b,e}$  will ever pass threshold  $v_T$ . In fact, for the RC circuit we know (from solving the differential equations) that even though  $\partial i = -$ ,  $i$  never reaches zero: it dies out exponentially. In fact, no matter what interpretation it starts in, it will approach Interpretation 2 asymptotically. More generally, a linear circuit will approach the same interpretation no matter what the initial conditions.

A circuit may also change its incremental interpretation. The Schmitt trigger

is an example of such a circuit. There is very little that can be said concerning changes between incremental interpretations without doing a second-order analysis [4]. After all, it is  $\partial^2 i$  which is causing the change in  $\partial i$ . To determine the movement between causal interpretations one needs to solve for the second-order derivatives. For linear circuits, this is easy as the models do not change and thus it is not necessary to redo the analysis, but for nonlinear circuits the higher-order models can be completely different and require an entire reanalysis.

### 10.1. Interstate behavior

Fortunately, state changes and quiescent interpretation shifts obey numerous constraints which can be used to generate the state-transition diagram for the circuit. The state-transition diagram describes the interstate behavior of the circuit, while causal analysis and the mechanism graph describes the intrastate behavior of the circuit. Fig. 33 illustrates the state-transition diagram for CE-FEEDBACK in response to a rising input. Unless the circuit contains negative resistances uncovered by the heuristics, the state diagram for a falling input is identical except for change in direction of arrowhead.

The state-transition diagram is constructed using a very basic algorithm to generate all the possibilities and then apply constraints to prune down the space of possibilities. The basic envisioning algorithm proceeds as follows. The

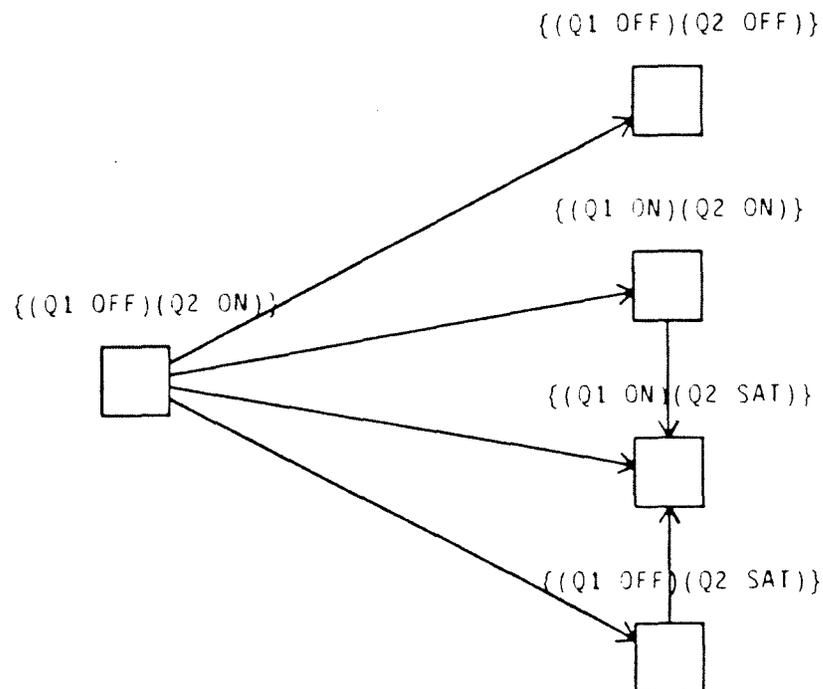


FIG. 33. State-transition diagram for CE-FEEDBACK.

circuit is presumed to be in some given initial condition. In this state, causal analysis identifies the values of all the circuit quantities. Identify all the quiescent quantities which are approaching some threshold: these are the causes for possible state transitions. The state transitions define a set of possible next states (usually there is only one). If this state has not been considered before, analyze it. If there is more than one next state, explore state space breadth-first.

The above algorithm includes three kinds of ambiguities. First, the incremental variable may be ambiguous, so we do not know whether the variable is approaching its threshold. Second, even though the variable may be approaching its threshold, we do not know whether it will reach it. Finally, more than one variable may be approaching thresholds (the case of the same variable approaching two different thresholds is handled by presuming that all inequality relationships between constants are known), and it is hard to tell whether one crosses its threshold before the other or whether they both cross simultaneously. Without additional constraints, all possible combinations of threshold crossings are possible. For example, in state  $\{(Q1\ OFF)(Q2\ ON)\}$ , the changing quantities tend to turn Q1 ON and Q2 OFF. But Q1 can turn ON before Q2 turns OFF or Q1 turns OFF before Q2 turns ON, or they both change state simultaneously.

The following is the list of legal states of CE-FEEDBACK and an indication of which threshold is being approached. This list and state diagram of Fig. 33 presumes all incremental interpretations are possible, thus some circuit values are ambiguous.

In State 1,  $\{(Q1\ OFF)(Q2\ ON)\}$

The value of  $\partial v_{C1,OUTPUT}$  is ambiguous.

If  $\partial v_{C1,OUTPUT} = -$ , Q2 may change state to SAT because  $v_{CB}$  may drop past threshold  $-v_T$ .

Because  $\partial v_{E2,C1} = +$ , Q2 may change state to OFF because  $v_{BE}$  may drop past threshold  $v_T$ .

Because  $\partial v_{INPUT,C1} = +$ , Q1 may change state to R-ON because  $v_{CB}$  may drop past threshold  $-v_T$ .

Because  $\partial v_{INPUT} = -$ , Q1 may change state to ON because  $v_{BE}$  may rise past threshold  $v_T$ .

In State 3,  $\{(Q1\ OFF)(Q2\ OFF)\}$

Because  $\partial v_{INPUT,C1} = +$ , Q1 may change state to R-ON because  $v_{CB}$  may drop past threshold  $-v_T$ .

Because  $\partial v_{INPUT} = -$ , Q1 may change state to ON because  $v_{BE}$  may rise past threshold  $v_T$ .

In State 2,  $\{(Q1\ ON)(Q2\ ON)\}$

The value of  $\partial v_{C1,OUTPUT}$  is ambiguous.

If  $\partial v_{C1,OUTPUT} = -$ , Q2 may change state to SAT because  $v_{CB}$  may drop past threshold  $-v_T$ .

Because  $\partial v_{E2,C1} = -$ , Q2 may change state to OFF because  $v_{BE}$  may drop past threshold  $v_T$ .

Because  $\partial v_{INPUT,C1} = +$ , Q1 may change state to SAT because  $v_{CB}$  may drop past threshold  $-v_T$ .

In State 5,  $\{(Q1\ ON)(Q2\ SAT)\}$

Because  $\partial v_{E2,C1} = +$ , Q2 may change state to R-ON because  $v_{B,E}$  may drop past threshold  $v_T$ .

Because  $\partial v_{INPUT,C1} = +$ , Q1 may change state to SAT because  $v_{C,B}$  may drop past threshold  $-v_T$ .

The value of  $\partial v_{C1,OUTPUT}$  is ambiguous.

If  $\partial v_{C1,OUTPUT} = -$ , Q2 may change state to ON because  $v_{C,B}$  may rise past threshold  $-v_T$ .

In State 4,  $\{(Q1\ OFF)(Q2\ SAT)\}$

Because  $\partial v_{E2,C1} = +$ , Q2 may change state to R-ON because  $v_{B,E}$  may drop past threshold  $v_T$ .

Because  $\partial v_{INPUT,C1} = +$ , Q1 may change state to R-ON because  $v_{C,B}$  may drop past threshold  $-v_T$ .

Because  $\partial v_{INPUT} = +$ , Q1 may change state to ON because  $v_{B,E}$  may rise past threshold  $v_T$ .

Most of these transitions cannot happen individually because the resulting state is not consistent with the confluences. For example, in state  $\{(Q1\ OFF)(Q2\ ON)\}$  Q1 tends to turn ON while Q2 turns OFF, but this target state is inconsistent and therefore it cannot become the next circuit state.

The second constraint on state diagrams is that variables must change value continuously. This restriction applies to both quiescent and incremental variables. For example,  $x$  may not change from  $+$  to  $-$  without an intervening state within which it is 0. This rule is used only twice in the CE-FEEDBACK's state diagram. If a state has more than one incremental interpretation, that individual interpretation which is causing the state change must be continuous with some interpretation of the target state. For example, in state  $\{(Q1\ ON)(Q2\ SAT)\}$ , Q2 may turn ON, causing a circuit state change to  $\{(Q1\ ON)(Q2\ ON)\}$ . However, this does not make much sense as the circuit would immediately tend to make Q2 saturate again. More formally, this transition does not happen because the derivatives would have to vary discontinuously [34]. In state  $\{(Q1\ ON)(Q2\ SAT)\}$   $\partial i_{\rightarrow \#2(RC2)} = -$ , in  $\{(Q1\ ON)(Q2\ ON)\}$   $\partial i_{\rightarrow \#2(RC2)} = +$ , which requires a discontinuous transition. The reason it is possible for the state change to happen in the reverse direction is that there is an interpretation of  $\{(Q1\ ON)(Q2\ SAT)\}$  in which  $\partial i_{\rightarrow \#2(RC2)} = +$  or 0 (this interpretation is continuous with one of  $\{(Q1\ ON)(Q2\ ON)\}$ ), but no circuit quantity is heading towards a threshold which would cause a transition to  $\{(Q1\ ON)(Q2\ ON)\}$ .

These are all the rules required to construct the state diagram for CE-FEEDBACK. Here is EQUAL's analysis of the preceding possibilities.

In State 1,  $\{(Q1\ OFF)(Q2\ ON)\}$

But some combinations lead to contradictory states:  $\{(Q1\ R-ON)(Q2\ SAT)\}^*$ ,  $\{(Q1\ ON)(Q2\ OFF)\}^*$ ,  $\{(Q1\ R-ON)(Q2\ OFF)\}^*$  or  $\{(Q1\ R-ON)(Q2\ ON)\}^*$ .

Therefore the device may change state to one of:

4:  $\{(Q1\ OFF)(Q2\ SAT)\}$ , 5:  $\{(Q1\ ON)(Q2\ SAT)\}$

3:  $\{(Q1\ OFF)(Q2\ OFF)\}$ , 2:  $\{(Q1\ ON)(Q2\ ON)\}$ .

In State 3, {(Q1 OFF)(Q2 OFF)}

But all combinations lead to contradictory states: {(Q1 R-ON)(Q2 OFF)}\* or {(Q1 ON)(Q2 OFF)}\*.

Therefore there are no transitions out of this state.

In State 2, {(Q1 ON)(Q2 ON)}

But some combinations lead to contradictory states: {(Q1 SAT)(Q2 SAT)}\*, {(Q1 ON)(Q2 OFF)}\*, {(Q1 SAT)(Q2 OFF)}\* or {(Q1 SAT)(Q2 ON)}\*.

The device may change state to 5: {(Q1 ON)(Q2 SAT)}

In State 5, {(Q1 ON)(Q2 SAT)}

But some combinations lead to contradictory states: {(Q1 SAT)(Q2 ON)}\*, {(Q1 ON)(Q2 R-ON)}\*, {(Q1 SAT)(Q2 R-ON)}\* or {(Q1 SAT)(Q2 SAT)}\*.

$\partial i_{\rightarrow \#1(C)}$  changes discontinuously from + to - to State 2: {(Q1 ON)(Q2 ON)}

$\partial i_{\rightarrow \#2(RC)}$  changes discontinuously from - to + to State 2: {(Q1 ON)(Q2 ON)}

Therefore there are no transitions out of this state.

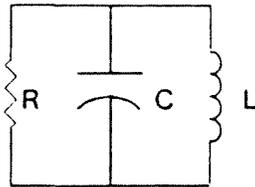
In State 4, {(Q1 OFF)(Q2 SAT)}

But some combinations lead to contradictory states: {(Q1 OFF)(Q2 R-ON)}\*, {(Q1 ON)(Q2 R-ON)}\*, {(Q1 R-ON)(Q2 R-ON)}\* or {(Q1 R-ON)(Q2 SAT)}\*.

The device may change state to 5: {(Q1 ON)(Q2 SAT)}

### 10.2. Ontology for time

This section is a direct application of the rules of [4, 6] to electrical circuits. So far we have assumed that every operating region or interpretation exists for



Episode	1	2	3	4	5	6	7	8	9
Interpretation	1	1	1	1 2 3	1	1	1	1 2 3	1
$\partial i_{\rightarrow \#1(C)} =$	0	-	-	- 0 +	+	-	+	- 0 +	-
$\partial i_{\rightarrow \#1(R)} =$	0	+	0	---	-	-	0	+++	+
$\partial i_{\rightarrow \#1(L)} =$	0	+	+	+++	0	-	-	---	0
$\partial v_N =$	0	+	0	---	-	-	0	+++	+

Episode	1	2	3	4	5	6	7	8	9
Interpretation	1	1	1	1 2 3	1	1	1	1 2 3	1
$[i_{\rightarrow \#1(C)}] =$	0	+	0	---	-	-	0	+++	+
$[i_{\rightarrow \#1(R)}] =$	0	+	+	+++	0	-	-	---	0
$[i_{\rightarrow \#1(L)}] =$	0	-	-	- 0 +	+	+	+	- 0 +	-
$[v_N] =$	0	+	+	+++	0	-	-	---	0

FIG. 34. RLC circuit.

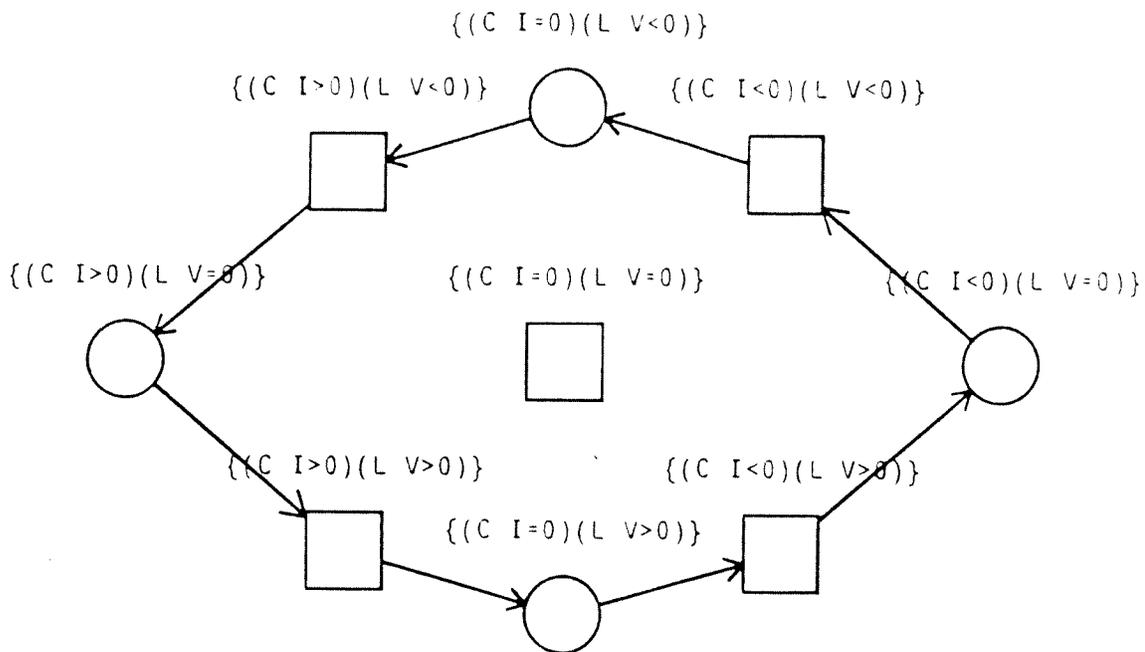


FIG. 35. State diagram for RLC circuit.

some interval time. However, some operating regions and interpretations may only exist momentarily. An operating region for which  $x = a$  and  $\partial x = -$  will only exist momentarily because  $x$  will immediately drop below  $a$ . Similarly a quiescent interpretation with  $x = 0$  and  $\partial x = +$  exists only for an instant. It is important to distinguish instants from intervals, because instants do not take any time and are always guaranteed to end, unlike intervals. This new ontology is illustrated by a simple RLC circuit (Fig. 34).

The set of confluences have 13 solutions, but these can be collapsed into 9 as  $v_N$  and  $i_{\rightarrow \neq 1(C)}$  are the only interesting variables. The corresponding state diagram is Fig. 35. Circles represent states that exist momentarily, squares represent states that may exist over a time interval. The state diagram embodies two additional constraints. First, every quantity changes from zero before any quantity changes to zero. Second, computing higher-order derivatives shows that any transition to state  $\{(C I = 0)(L V = 0)\}$  is impossible (see [4] for an explanation of this rule).

## 11. Applications

The goal of qualitative physics is to predict and explain the important behavior of physical systems without recourse to quantitative methods. This paper has presented a qualitative physics for circuits which achieves this goal. In addition, it is possible to identify the mechanism by which the circuit functions in the standard language used by electrical engineers. This concluding section briefly outlines a few of the potential applications of this theory to analysis, design, troubleshooting and training.

### 11.1. Quantitative analysis

The electrical engineer's most fundamental task is circuit analysis. He can perform this analysis either symbolically [10] or numerically [26]. A good engineer will first understand how the circuit functions qualitatively. Then based on this understanding, he will select component models, important variables, approximations and integration step sizes, then perform the necessary symbolic or numerical computation, and finally interpret the results in terms of his commonsensical understanding of how the circuit works. In other words, an engineer does not perform a quantitative analysis unless he first understands the circuit at a qualitative level. In [10] we constructed a program SYN which analyzed circuits symbolically in order to determine circuit parameters. Many of the problems of SYN were due to the fact that it has no understanding of the behavior of the circuit it was analyzing.

SYN can only work on very simple circuits because the algebraic expressions generated quickly fill up program memory. On the other hand, an engineer can analyze the circuit quickly just using pencil and paper. Comparing the engineer to SYN reveals four interesting facts. First, the engineer chooses approximations which simplify the algebra enormously without significantly changing the answer. Second, the engineer chooses models which are algebraically more tractable. Third, the engineer chooses variable substitutions and equation formulations which keep equation size relatively small. Finally, he uses standard rules of thumb to analyze well-known circuit configurations. Qualitative analysis addresses all four of these.

#### 11.1.1. Approximations

As SYN does not know how a particular component is being used, it must always use the most complex model available. In most situations simpler models are sufficient which lead to far simpler algebra. EQUAL can give advice to SYN as to how a component is being used so that it can choose an

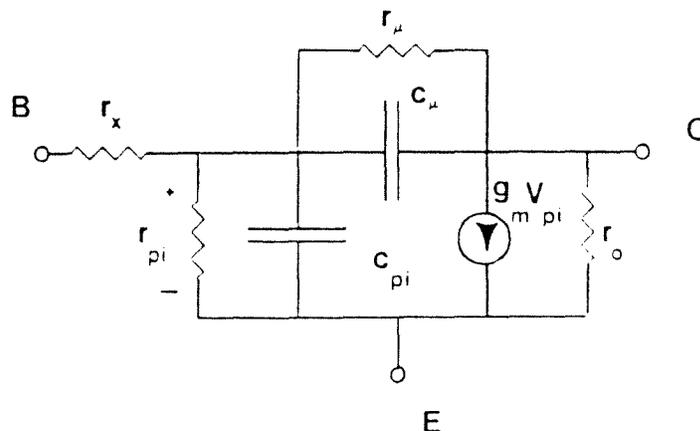


FIG. 36. Complete hybrid- $\pi$  model.

appropriate simplification of the complex model. For example, the complete hybrid- $\pi$  model for a transistor is shown in Fig. 36. (The circle containing a downward arrow represents a source which produces a downwards current of  $g_m V_\pi$ .) At reasonable currents,  $r_x \ll r_\pi$  so  $r_x$  can be ignored unless the circuit is driven by an extremely low source impedance. Analogously  $r_\mu$  and  $r_o$  can be ignored unless the load is of extremely high impedance. As only the common-emitter configuration is ever used with extremely high loads,  $r_\mu$  and  $r_o$  can be ignored in the common-base and common-collector configurations.  $c_\mu$  can be ignored unless the circuit is of high impedance. When the circuit is being driven by a very low impedance,  $c_\pi$  can be ignored. In the common-base configuration  $c_\mu$  is usually not important.

#### 11.1.2. *Choosing tractable models*

The hybrid- $\pi$  model is particularly useful for analyzing the common-emitter and common-collector configurations. In the common-base configuration a different model—the T model—makes the analysis easier. Although the T model is quantitatively equivalent to the hybrid- $\pi$  model, the resulting symbolic equations are more tractable because the source does not appear across the input. In many cases, the output signal can be computed directly from the input without having to solve simultaneous equations.

#### 11.1.3. *Choosing tractable formulations*

Causal analysis helps choose the important variables in which to formulate circuit quantities. The causal analysis identifies the important circuit feedback paths and the instances of local feedback. These feedback paths appear as simultaneities in the symbolic analysis. The causal heuristics of the correct interpretation indicate places where simultaneity must be broken. Choosing these variables as the places to break feedback loops simplifies the form of the equations and formulates intermediate results in terms familiar to the electrical engineer.

#### 11.1.4. *Rules of thumb*

Common combinations such as cascode, differential pair, and Darlington have well-known rules of thumb for making their quantitative analysis easier. EQUAL can recognize these situations for SYN. Certain combinations of transistors occur so frequently that it is useful to consider them as single fragments. A common-collector stage followed by a common-emitter stage forms an amplifier with high input impedance and moderate output impedance. This combination is known as a Darlington pair. The common-emitter/common-base combination is called a cascode and has very good frequency response. The common-collector/common-base connection forms a circuit widely used in

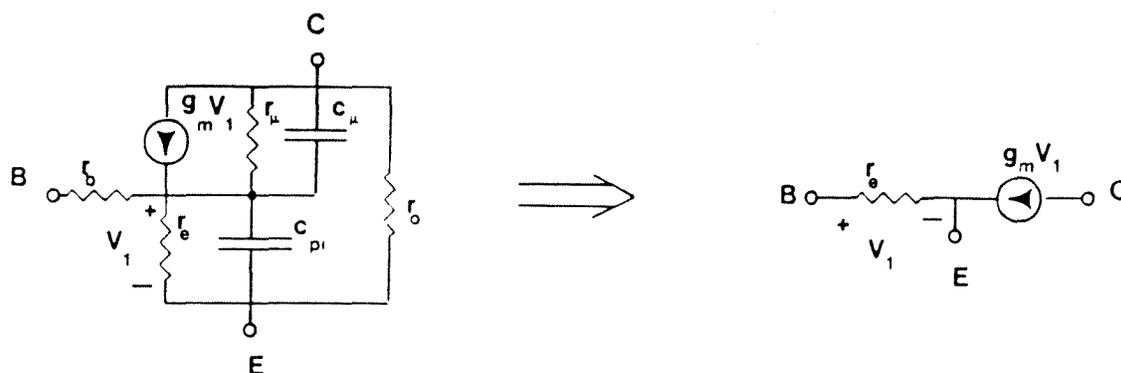


FIG. 37. T models.

operational amplifiers. The emitter-coupled pair provides differential outputs and can be direct coupled. The recognition of these combinations is not critical to EQUAL since it can calculate the impedance-gain specifications of these combinations with its composition rules, but it is of use to SYN.

The type of the feedback configuration provides direct advice about how to analytically determine the behavior of the circuit. For example, the following advice can be found in most textbooks on feedback amplifiers. In order to analyze a loop-loop configuration you should:

- (1) Use  $z$  parameters to model the two-ports.
- (2) Calculate the gain of the feedback network by driving the feedback network with a current and determining the voltage produced into an open circuit.
- (3) Calculate feedback loading at the amplifier input by open circuiting the output node.
- (4) Calculate feedback loading at the amplifier output by open circuiting the input node.

This simplifying technique can only be applied if feedback is known to be present, and only causal analysis not quantitative analysis can do that.

## 11.2. Design

Analysis is a key component of design and thus the techniques of the previous section apply to the design task as well. This observation lies at the heart of SYN's synthesis-by-analysis approach. Thus SYN can both construct a symbolic transfer function of a circuit, and determine values of circuit parameters given desiderata for circuit behavior.

Interpretation construction is also an important design tool. Causal analysis identifies all possible interpretations, including behaviors the designer does not intend. By identifying all the possible interpretations, causal analysis identifies possible pathological modes of circuit behavior that the designer must avoid.

In many ways the design task is the inverse of the recognition task around which this paper is organized. Recognition starts with a schematic of a particular circuit and constructs an abstract account for its behavior. Design, on the other hand, starts with an abstract description of the desired behavior and

constructs a schematic which achieves this abstract description. The algorithms for recognition, are thus entirely different from those of design, nevertheless the languages for describing the levels of abstraction are probably quite similar.

### 11.3. Troubleshooting

Troubleshooting involves determining why a particular correctly designed circuit is not functioning as intended, the explanation for the faulty behavior being that the particular instance of that circuit under consideration is at variance in some way with its design. Starting with the external symptom, the troubleshooter makes a series of measurements in order to localize the fault. Quantitative analysis is ruled out as a troubleshooting strategy both because it is impractical and the circuit parameters are not precisely known. Thus, qualitative reasoning plays a key role in troubleshooting.

Causal analysis can quickly evaluate the plausibility of hypothetical faults. One simply substitutes a faulty component model for the correct one and performs the qualitative analysis again to see whether the predicted results are consistent with measurements made thus far in the troubleshooting session.

Fortunately, causal analysis can be used for much more than evaluating hypotheses. The causal interpretation for the intended behavior also provides a causal explanation for how the outputs are caused by the inputs. The components mentioned in this explanation are prime candidates for possible faults and the fault modes can be determined by examining the argument. The resulting hypotheses can be evaluated to determine which faults in which of these components are consistent with the symptoms. This technique works well if some component has shifted in value, but fails if circuit behavior has radically changed. The interpretation may have changed in the presence of the fault. If a causal assumption is violated, the entire argument may be invalidated. Since the designer never intended that the circuit behave in that way, no appeal can be made to the original intention. When this happens, the troubleshooter must first make measurements to determine the unintended interpretation within which the circuit is operating.

Causal analysis can also be run backwards if the models and the heuristics are changed. This technique can be used to identify circuit faults. The input is the difference between measured and intended outputs and the propagation engine works identically. Now, the propagation identifies "what could have caused  $x$ " instead of "What does  $x$  cause"—the only difference between these two questions is that time flow is reversed.

As part of the SOPHIE project [1] we built a general-purpose troubleshooting expert. Although the troubleshooting expert was very general, it was not powerful enough to troubleshoot using the schematic alone. It had to be provided a small collection of rules which characterized circuit behavior and with those it could troubleshoot every possible fault of the power-supply. Each of these rules was based on an abstract, qualitative analysis of how the circuit functioned. Causal analysis is powerful enough to construct some, if not all, of

these rules. Thus, there is now a possibility of constructing a troubleshooter which can troubleshoot circuits which it has never seen before given only their schematic. This troubleshooter would operate by first doing a causal and teleological analysis of the circuit and then use this analysis to construct the needed abstract rules. The SOPHIE's troubleshooter could use these rules in combination with its other inference mechanisms to troubleshoot the circuit.

One of the original motivations for the no-function-in-structure principle and for studying causal reasoning in isolation was to achieve robustness in a computer-based troubleshooter. When a circuit is faulted, the circuit is no longer working as originally intended by the designer. A reasoner which presumes the use of teleological assumptions in choosing the component models or disambiguating interpretations will be of no value because the circuit functioning is radically changed. Thus, for troubleshooting, it is extremely important to identify all the possible interpretations.

#### 11.4. ICAI

Computer-based instructional systems for electronics and electronic troubleshooting need to explain their conclusions in terms that the student understands in order to enable him to build his own qualitative models of circuit functioning. Causal analysis provides a theory of explanation for humans and thus provides the basis for an explanation facility in an ICAI system. SOPHIE was first and foremost an ICAI system; the expert troubleshooter was constructed, not because it was itself important, but rather to provide explanations for troubleshooting interferences that SOPHIE itself could not give. For example, SOPHIE evaluated students hypotheses by running a numerical simulation with the modified model to see if it was consistent with observed measurements. This strategy evaluates hypotheses, but produces no explanation for why a particular hypothesis is valid or invalid. Causal reasoning provides a way to both evaluate hypotheses and produce explanations for their plausibility.

#### ACKNOWLEDGMENT

I thank John Seely Brown and Gerald Jay Sussman for continuing guidance in this research. Harry Barrow, Dan Bobrow, Randy Davis, Jeff Shrager, and Brian Williams provided many useful comments on the paper.

#### REFERENCES

1. Brown, J. S., Burton, R.R. and De Kleer, J., Pedagogical, natural language and knowledge engineering techniques in SOPHIE I, II and III, in: D. Sleeman and J.S. Brown (Eds.), *Intelligent Tutoring Systems* (Academic Press, New York, 1982) 227-282.
2. Cochin, I., *Analysis and Design of Dynamic Systems* (Harper and Row, New York, 1980).
3. Davis, R., Shrobe, H., Hamscher, W., Wieckert, K., Shirley, M. and Polit, S., Diagnosis based on description of structure and function, in: *Proceedings National Conference on Artificial Intelligence*, Pittsburgh, PA (August, 1982) 137-142.
4. De Kleer, J. and Bobrow, D.G., Qualitative reasoning with higher-order qualitative derivatives, in: *Proceedings National Conference on Artificial Intelligence*, Austin, TX, August, 1984

5. De Kleer, J., Choices without backtracking, in: *Proceedings National Conference on Artificial Intelligence*, Austin, TX, August, 1984.
6. De Kleer, J. and Brown, J.S., A qualitative physics based on confluences, *Artificial Intelligence* **24** (1984) this volume.
7. De Kleer, J. and Brown, J.S., Assumptions and ambiguities in mechanistic mental models, in: D. Gentner and A.S. Stevens (Eds.), *Mental Models* (Erlbaum, Hillsdale, NJ, 1983) 155–190.
8. De Kleer, J. and Brown, J.S., Mental models of physical mechanisms and their acquisition, in: J.R. Anderson (Ed.), *Cognitive Skills and their Acquisition* (Erlbaum, Hillsdale, NJ, 1981) 285–309.
9. De Kleer, J. and Brown, J.S., Mental models of physical mechanisms, CIS-3 Cognitive and Instructional Sciences, Xerox PARC, Palo Alto, CA, 1981.
10. De Kleer, J. and Sussman, G.J., Propagation of constraints applied to circuit synthesis, *Circuit Theory and Applications* **8** (1980) 127–144.
11. De Kleer, J., Causal and teleological reasoning in circuit recognition, Artificial Intelligence Laboratory, TR-529, MIT, Cambridge, MA, 1979.
12. De Kleer, J., Qualitative and quantitative knowledge in classical mechanics, Artificial Intelligence Laboratory, TR-352, MIT, Cambridge, MA, 1975.
13. Desoer, C.A. and Kuh, E.S., *Basic Circuit Theory* (McGraw-Hill, New York, 1969).
14. Doyle, J., A truth maintenance system, *Artificial Intelligence* **12** (1979) 231–272.
15. Forbus, K.D. and Stevens, A., Using qualitative simulation to generate explanations, Rept. No. 4490, Bolt Beranek and Newman, Cambridge, MA, 1981.
16. Forbus, K.D., Qualitative process theory, *Artificial Intelligence* **24** (1984) this volume.
17. Forbus, K.D., Qualitative reasoning about physical processes, in: *Proceedings Seventh International Joint Conference on Artificial Intelligence*, Vancouver, BC (August, 1981) 326–330.
18. Gray, P.E. and Searle, C.L., *Electronic Principles: Physics, Models, and Circuits* (Wiley, New York, 1969).
19. Harris, J.N., Gray, P.E. and Searle, C.L., *Digital Transistor Circuits* (Wiley, New York, 1966).
20. Hayes, P.J., The naive physics manifesto, in: D. Michie (Ed.), *Expert Systems in the Micro-electronic Age* (Edinburgh University Press, Edinburgh, 1979).
21. Hewitt, P.G., *Conceptual Physics* (Little and Brown, Boston, MA, 1974).
22. Karnopp, D. and Rosenberg, R., *System Dynamics: A Unified Approach* (Wiley, New York, 1975).
23. Kuipers, B., Getting the envisionment right, in: *Proceedings National Conference on Artificial Intelligence*, Pittsburgh, PA (August, 1982) 209–212.
24. Kuipers, B., Commonsense reasoning about causality: Deriving behavior from structure, *Artificial Intelligence* **24** (1984) this volume.
25. McAllester, D., An outlook on truth maintenance, Artificial Intelligence Laboratory, AIM-551, MIT, Cambridge, MA, 1980.
26. Nagel, L.W. and Pederson, D.O., Simulation program with integrated circuit emphasis, in: *Proceedings Sixteenth Midwest Symposium on Circuit Theory*, Waterloo, Canada, 1973.
27. Shearer, J.L., Murphy, A.T. and Richardson, H.H., *Introduction to System Dynamics* (Addison-Wesley, Reading, MA, 1971).
28. Stallman, R.M. and Sussman, G.J., Forward reasoning and dependency-directed backtracking in a system for computer-aided circuit analysis, *Artificial Intelligence* **9** (1977) 135–196.
29. Steele, G.L., The definition and implementation of a computer programming language based on constraints, Artificial Intelligence Laboratory, TR-595, MIT, Cambridge, MA, 1980.
30. Steele, G.L. and Sussman, G.J. CONSTRAINTS – A language for expressing almost-hierarchical descriptions, *Artificial Intelligence* **14** (1980) 1–39.
31. Suppes, P.S., *Introduction to Logic* (Van Nostrand, New York, 1957).
32. Waldhauer, F.D., *Feedback* (Wiley, New York, 1982).
33. Wiener, N., *The Human Use of Human Beings: Cybernetics and Society* (Houghton Mifflin, Boston, 1950).
34. Williams, B.C., Qualitative analysis of MOS circuits, *Artificial Intelligence* **24** (1984) this volume.
35. Mayr, O., *The Origins of Feedback Control* (MIT Press, Cambridge, MA, 1970).