

1997

Expertise in Context

Human and Machine

Edited by

Paul J. Feltovich

Kenneth M. Ford

&

Robert R. Hoffman

AAAI Press / The MIT Press

Menlo Park, California / Cambridge, Massachusetts / London, England

Abduction and Abstraction in Diagnosis: A Schema-based Account

Carl R. Stern & George F. Luger

Introduction

The activity of constructing explanations is strongly goal-dependent. This dependency has recently been emphasized in the “content theory of explanation” proposed by Leake (1992). Leake observes that the information that a good explanation must provide is closely tied to the reasons for constructing the explanation. Leake fleshes out his analysis with a taxonomy of general explanatory goals and an analysis of the requirements imposed by each type of goal.

Our study of expert performance in the area of semiconductor component failure analysis supports Leake’s account of goal-dependency. We find that the patterns of diagnostic explanation produced by failure analysts are closely correlated with the need to support different kinds of remedial practices (i.e., different ways of addressing the reliability concerns raised by component failures). Explanations of component failures exhibit a distinct range of forms corresponding to the different causal dimensions addressed by remediation, for example, the component design, the component manufacturing process, the surrounding circuit design, and the stresses (electrical, mechanical, thermal) originating from an external environment. We have frequently observed that failure analysts working in different settings (e.g., for a component manufacturer vs. a circuit assembly manufacturer) tend to emphasize different causal dimensions.

Despite the fact that component failures typically result from an interaction of factors, diagnostic explanations usually focus on only one causal di-

mension, treating the others as incidental. The diagnostician's selection of a diagnostic hypothesis provides a context for interpreting evidence, for selectively emphasizing or ignoring certain kinds of data, and for constructing causal theories about the sequence of events resulting in the failure. Within the framework of this general hypothesis, the diagnostician's application of causal knowledge is controlled by the goal of producing a detailed causal explanation of a certain form.

Our model of diagnosis is based on the observation and analysis of expert performance in the area of semiconductor component failure analysis. We have worked with five failure analysts over nearly half a decade in the process of constructing a failure analysis expert system.¹ The weakness of our original rule-based expert system in capturing the diagnostic problem solving behavior of human experts motivated the development of a second architecture. In this architecture, explanation patterns are encoded in *schemas*. A schema specifies a general pattern of causation as a causal sequence in which each step of the sequence is characterized by causal processes of a certain type. Using this schema representation, we have developed a schema-based abduction algorithm that implements an important modification of the usual abductive chaining algorithm (Levesque 1989). In schema-based abduction, search for causal processes to explain unexplained conditions is restricted to the class of causal processes specified by the schema at the current step.

We now present some observations regarding the patterns of investigation and hypothesis formation followed by human experts in semiconductor component failure analysis; this is followed by a general discussion of certain related cognitive issues. In the next two sections we give a specification of our architecture for diagnosis. We then present an extended example of failure analysis using this architecture.

Expertise in Context: Component Level Failure Analysis

Semiconductor component failure analysis offers an important example of expertise in context (Luger and Stern 1992). The failure analyst is presented with an initial set of signs, for example, the abnormal behavior of a diode after *burn in*,² and is required to organize an investigation based on an interpretation of those signs. The analyst begins the analysis by gathering information about the history and vulnerabilities of the device as well as the particular circumstances of the current failure. The initial visual and electrical examinations are conducted against the background of this information. Based on the initial examination, the analyst adopts a prioritized list of hypotheses—the failure mechanisms which could account for the abnormal device behavior. Data gathering then proceeds, focused by the active hypothesis set.

As new information is acquired, some hypotheses are dropped while oth-

ers are modified. In the light of hypothesis revision, observations that once were considered relevant are pushed aside, and new observations become critical. Eventually the investigation stabilizes on a sufficiently well-established hypothesis and the focus changes. The goal becomes one of establishing certain details in the causal scenario which are relevant to the task of fixing the problem or preventing it in the future. The final outcome of this process, if it is successful, is an explanation of the device malfunction suitably focused and precise enough to support corrective action.

An essential element of the investigative process is the use of hypotheses to organize search. The problem solver hypothesizes conditions which are not directly in evidence, conditions which might account for the device's anomalous behavior. Initially these conditions may be specified in a very general or abstract way. Evidence gathering is then directed towards confirming or disconfirming these hypotheses as well as elaborating the hypotheses in more detail.

Although the initial stage of the investigation involves a parallel investigation of competing hypotheses, each hypothesis can be seen to define a particular investigative context. These contexts are distinguished by the scope of relevant data, the set of patterns used for reasoning about the evidence, and the set of methods appropriate for correcting the problem. For example, environmentally-induced failures are investigated and remediated differently from manufacturing defects, which are in turn handled differently from failures resulting from wearout mechanisms. It is therefore important to constrain the type of failure mechanism involved as soon as possible in order to narrow the scope of the investigation.

Semiconductor failure analysis involves the initial adoption of candidate hypotheses, that is, conditions not in evidence, to explain the failure. This pattern of reasoning was characterized by the philosopher Peirce as a peculiar form called *abduction*, to be distinguished from the more familiar *deduction* and *induction* (Peirce 1958). It has been studied recently by workers in the AI research community (e.g., Levesque 1989; Charniak and Shimony 1990; Pearl 1987), as well as in the cognitive science community (e.g., Fel'tovich *et al.* 1984; Kuipers and Kassirer 1984).

Our study of semiconductor failure analysis has led us to examine abductive problem solving more closely. In an effort to understand better the inner logic of diagnostic investigations, we have analyzed the structure of abductive hypotheses in semiconductor failure analysis and examined the way in which these hypotheses organize investigations.

The abductive hypotheses used in semiconductor failure analysis are called *failure mechanisms*. Failure mechanism represent abstract patterns of causation, codifying the accumulated experience of experts both in understanding and responding to recurring patterns of failures over time. During initial hypothesis formation, failure mechanisms are treated as simple associations be-

tween sets of symptoms and types of causes. However, as the investigation of individual hypotheses progress, *deeper* knowledge from the domain of semiconductor physics is brought to bear. To understand the *logic* underlying the application of this deeper domain knowledge, we believe it is necessary to view failure mechanisms as complex structures representing key elements of the causal chains which produce failures. Viewed in this way, these abstract causal patterns help us to understand the specific sequence of data gathering steps and interpretive reasoning by which human problem solvers pursue the investigation of hypotheses.

We call the representation of these recurring patterns of causation *schemas* because of their role in organizing and interpreting the diagnostician's experience. A schema is defined as a cognitive structure which guides the application of concepts, in this case causal laws, to experience. Schema-based pattern recognition involves the interpreter's use of schemas to actively construct perceptual or conceptual patterns which fit the data. This notion is distinguished from simple pattern matching, where the interpreter selects one of a predefined set of stored patterns based on criteria such as identity or closeness to the data. The term schema is thus used in a sense similar to that first proposed by Kant (1781/1964) and later developed by Bartlett (1932), Newell and Simon (1972), and Piaget (1970).

Causal Associations and the Heuristics of Diagnosis

In the area of semiconductor failure analysis, as in many other diagnostic domains, knowledge of first principles is insufficient for proficiency in diagnosis. In addition to a knowledge of semiconductor physics, engineers require an extensive period of training and experience before they become competent failure analysts. One reason for this is that there can be a large gulf between observed symptoms on the one hand and the laws of semiconductor physics on the other. Computationally speaking, the search for explanations from first principles involves too large a search space.

The gulf between first principles and observed symptoms is mediated by the recognition of recurring causal patterns or scenarios. The diagnostician searches for indications of these causal patterns in the preliminary data. The semiconductor component failure analyst learns to recognize and reason about a set of potential *failure mechanisms*. These represent the commonly occurring patterns of causation to which experts attribute component failures. For transistors and diodes, the experts we interviewed recognize between 40 and 60 different failure mechanisms.

Failure analysts associate failure *mechanisms* with failure *modes*. A failure mode is a general class of behavior under which a set of observable symptoms has been subsumed. For transistors, failure modes include: short, open, resistive, reverse bias leakage, low gain, intermittent, etc. Examples of associations

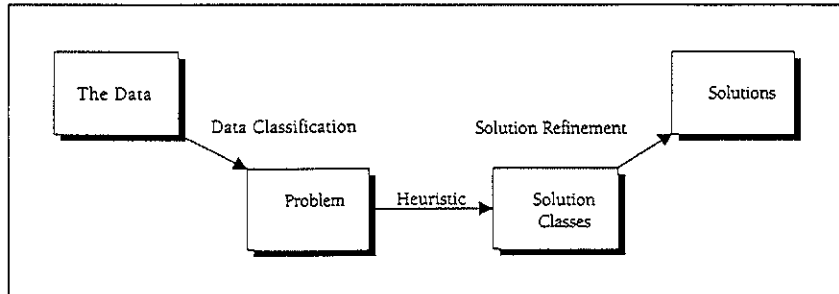


Figure 1. Clancey's heuristic classification architecture. The essential steps are data classification, heuristic matching, and solution refinement.

between mechanisms and modes are: contamination causes reverse bias leakage; particles cause intermittent shorts; faulty die attach causes high series resistance; electrical overstress causes opens; faulty wire bonds cause opens.

It is a mistake to construe such associations as deterministic relations between cause and effect. Contamination does not always produce leakage; particles do not always produce shorts. It would be more appropriate to view these as rough statistical correlations between types of effects and types of causes, that is, between failure modes and failure mechanisms. The rules in our expert system estimate the likelihood of failure mechanisms based on the failure mode along with contextual factors such as device structure and history. The important point, however, is that failure modes denote general types of failures and failure mechanisms denote general patterns of causation. To attempt to redefine these associations in a way that renders them more deterministically causal would undermine their heuristic function in the formation of hypotheses. It is precisely the generality of these associations that allows them to provide a useful decomposition of the global solution space during the early stages of inquiry.

This pattern of problem solving follows closely that described by Clancey (1985) in his analysis of *heuristic classification* architectures. Clancey discovered that a large class of expert systems employ a similar method of problem solving. This architecture is illustrated in Figure 1. The method is based on identifying a finite set of problem classes and solution classes. Problem data are first analyzed and identified with a problem class. Then a method of heuristic classification or matching is used to map the problem class into a solution class. Clancey used the term *heuristic classification*, as opposed to *simple classification*, to describe the process of associating elements from distinct classification hierarchies. Finally, a solution refinement method is used to generate and validate a concrete solution from the solution class. Clancey recognized that this order of steps was not necessarily sequential; the stages

of the problem classification, heuristic matching, and solution refinement are often interleaved. He did, however, propose this as a knowledge level specification of the logical structure of problem solving in an important class of expert systems.

Our experience indicates that semiconductor component failure analysts use symptom classification and heuristic matching in generating an hypothesis set. What remains to be described is how these hypotheses are investigated and elaborated into viable causal explanations. This corresponds to the solution refinement stage described by Clancey. We believe we have discovered an interesting and important characterization of solution refinement with respect to semiconductor failure analysis. We have observed that causal hypotheses function as schemas for the construction of causal explanations from the domain laws and the facts of the case.

Failure Mechanisms and Explanation Schemas

When expert failure analysts are asked to explain their actions and reasoning, much of their discussion is phrased in terms of the failure mechanisms whose presence they are trying to establish or eliminate. Experts, however, do not usually articulate the content and structure of these failure mechanisms unless asked. Nonetheless, this content and structure is part of the understanding implicit in their practice, and it is useful to ask them to articulate it. When they are pushed, what experts often describe is a set of stereotyped failure scenarios. These are patterns of causation consisting of events or device states connected by transitions, where the transitions are law-governed processes or mechanisms.

It is important for the knowledge engineer to determine the structure corresponding to each failure mechanism in order to understand why particular test procedures or measurements are performed and how test results are interpreted. Simply put, a failure mechanism represents a story pattern explaining why a failure occurred. A story conforming to this pattern has events or states of a specific type linked by processes or mechanisms of a certain type. The failure analyst attempts to match events in the current situation to those in the story. Once those events are known he or she can then verify that the processes or mechanisms linking events or states conform to the constraints imposed by the causal first principles of the domain. (We have created a set of schema graphs corresponding to the causal mechanisms which experts use in transistor failure analysis, and describe several of them later in this chapter.)

Situated Versus Context Free Knowledge

The scientific laws used in diagnosis, for example, the laws of semiconductor physics, represent a context free form of knowledge. The generality of these

laws can be seen from the fact that the same laws can be applied to a wide variety of different situations or circumstances. The heuristic associations between failure mechanisms and failure modes represent the other extreme. These correlations vary both in content and strength depending on a variety of circumstances, including device structure, failure history, and failure analysis goals. The process of identifying failure mechanisms is thus *semiotic* in the sense that it involves interpreting signs within a pragmatic context.

In forming an hypothesis regarding the cause of leakage,³ for example, an expert will take into account whether a transistor is NPN or PNP⁴ because contamination induced inversion⁵ is much more strongly correlated with leakage in PNP transistors. He may also take into account whether the manufacturer has had a history of problems with contamination, and whether other devices from the same lot show signs of contamination. Similarly, experts take into account at what stage in its life cycle a device failed, because, for example, wearout mechanisms such as metal migration and whisker formation become more likely when a device has seen extended testing or use.

We also found that analysts with different goals and resources generally produce explanations with a different focus and structure. For example, failure analyses conducted by engineers at a manufacturing facility employed a larger set of causal mechanisms relating to process control in manufacturing than commercial failure analyses conducted on behalf of end users. Moreover, even when the same manufacturing defects were described, the explanation of defects in the manufacturer's failure analysis reports were focused on the details required to determine necessary changes in the manufacturing process, whereas the customer's failure analysis reports were focused on the details required to *detect* those defects in the lot acceptance process.

It is useful to contrast the schema-based approach to diagnosis with that of model-based diagnosis (Davis and Hamscher 1992). Both methods use reasoning based on first principles to identify and refine causal explanations. Both methods need to employ knowledge of device structure in order to apply causal knowledge. However, in model-based reasoning, the device structure is formalized in advance into a context free description. As Davis and Hamscher acknowledge, the construction of useful and appropriate models is, to a great extent, a black art. A model is necessarily an abstraction: it captures only certain aspects of device structure, while omitting others. The trick in creating a model is to choose a suitable abstraction, one that does not abstract out any elements of device structure required to account for a malfunction in some future situation.

In our approach the device structure is formalized in the context of an hypothesis regarding the type of causality responsible for a failure. This means that only those aspects of device structure are examined that are relevant to the hypothesized failure mechanism. Thus the hypothesis provides a context, determining what aspects of structure need to be formalized.

A Computational Architecture for Schema-Based Diagnosis

The architecture we propose uses schemas to investigate causal mechanisms and construct explanations. Explanation schemas represent an organized body of knowledge related to a causal pattern. Each schema describes a causal mechanism which is capable of producing a determinate range of effects. The mechanism is characterized by a set of events or states and their causal connections. This defines the attributes that must be specified in order to instantiate the causal mechanism vis-à-vis the current situation. The schema is also associated with a body of "compiled knowledge" used by human problem solvers to test for the presence of the mechanism and to propose corrective action. Finally the schema *graph* gives a concise description of the pattern of causal connections between events or states, serving as a template for constructing explanations from the causal domain theory and the facts of the case.

Schema-based diagnosis involves five steps: 1) generation of an hypothesis set, 2) hypothesis pruning, 3) hypothesis instantiation, 4) explanation construction and validation, and 5) explanation repair. In the following subsections we describe each of these steps.

Generation of an Hypothesis Set

We use a heuristic classification approach to identify a set of candidate mechanisms that will account for the observed fault/malfunction. Initial observations regarding the symptom or malfunction must be elaborated by further observations or measurements. Additional evidence gathering in conjunction with data abstraction is used to locate the malfunction within a classification hierarchy of problem types. Problem types are then matched against solution types, that is, causal mechanisms which can produce the observed problems. Each mechanism corresponds to an abstract pattern of causation. A mechanism is composite in the sense that it comprises a causal chain, distinguished by the constituent event types and causal processes. The identification of a mechanism activates a schema for reasoning about that type of mechanism.

Hypothesis Pruning

The hypothesis set, represented by the set of activated schemas, is tested and pruned. As mentioned above, each schema is associated with a set of tests or observations designed to confirm or disconfirm the presence of that mechanism. Some tests provide specific support for individual mechanisms while others provide general criteria for discriminating between classes of causal mechanisms. General tests discriminate between classes by producing additional data which some mechanisms can "explain" and others cannot. The test procedures are collectively assembled and correlated in order to select the least

cost test for pruning the hypothesis set. Testing and planning steps are alternated until the number of hypotheses on the “discriminant” cannot be further reduced. The remaining mechanisms are ranked in order of likelihood.

Hypothesis Instantiation

The most likely remaining hypothesis is selected for expansion. The corresponding schema graph is applied to the current situation. Nodes in the schema graph, representing the events in the causal scenario, are correlated with observed or hypothesized events in the current situation. Facts or data from the current situation are used to determine event attributes and properties in the schema graph. Attributes or properties required by the schema but currently unknown may trigger further testing or observation.

Explanation Construction and Validation

The schema graph is used as a template to construct an explanation from the causal domain theory and the known facts of the case. The causal links in the uninstantiated schema graph represent causal relations at a very general and abstract level. Consider, for example, one of the simplest schema graphs, that for an electrical overstress⁶ induced open. This is illustrated in Figure 2. In this schema graph, two key causal links are “excessive current causes temperature elevation” and “excessive temperature causes melting.”

Once the nodes of the graph are bound to a specific set of events, these causal links need to be reconstructed from the causal domain theory at a more concrete and detailed level. The application of the causal domain theory starts from the observed symptoms or malfunction (the bottom of the schema graph) and proceeds upwards from effect to cause through the instantiated nodes of the graph. This procedure propagates constraints, inferring characteristics of the cause from those of the effect. This procedure can serve to confirm or disconfirm the explanation.

In the current example, if we are using electrical overstress to explain a melted bond wire, we can determine from the composition of the bond wire material the minimum temperature required to melt it. We thus specify “excessive temperature” with an exact number. We can then infer from the thickness and resistance of the bond wire material an exact range of current over time which would be required to cause the melting. If, for example, the device could not have seen that level of current, then an inconsistency is detected which disconfirms the hypothesis of simple electrical overstress.

Explanation Repair

The detection of an inconsistency between the inferred and actual properties of an object in the schema graph constitutes a potential disconfirmation of

cost test for pruning the hypothesis set. Testing and planning steps are alternated until the number of hypotheses on the “discriminant” cannot be further reduced. The remaining mechanisms are ranked in order of likelihood.

Hypothesis Instantiation

The most likely remaining hypothesis is selected for expansion. The corresponding schema graph is applied to the current situation. Nodes in the schema graph, representing the events in the causal scenario, are correlated with observed or hypothesized events in the current situation. Facts or data from the current situation are used to determine event attributes and properties in the schema graph. Attributes or properties required by the schema but currently unknown may trigger further testing or observation.

Explanation Construction and Validation

The schema graph is used as a template to construct an explanation from the causal domain theory and the known facts of the case. The causal links in the uninstantiated schema graph represent causal relations at a very general and abstract level. Consider, for example, one of the simplest schema graphs, that for an electrical overstress⁶ induced open. This is illustrated in Figure 2. In this schema graph, two key causal links are “excessive current causes temperature elevation” and “excessive temperature causes melting.”

Once the nodes of the graph are bound to a specific set of events, these causal links need to be reconstructed from the causal domain theory at a more concrete and detailed level. The application of the causal domain theory starts from the observed symptoms or malfunction (the bottom of the schema graph) and proceeds upwards from effect to cause through the instantiated nodes of the graph. This procedure propagates constraints, inferring characteristics of the cause from those of the effect. This procedure can serve to confirm or disconfirm the explanation.

In the current example, if we are using electrical overstress to explain a melted bond wire, we can determine from the composition of the bond wire material the minimum temperature required to melt it. We thus specify “excessive temperature” with an exact number. We can then infer from the thickness and resistance of the bond wire material an exact range of current over time which would be required to cause the melting. If, for example, the device could not have seen that level of current, then an inconsistency is detected which disconfirms the hypothesis of simple electrical overstress.

Explanation Repair

The detection of an inconsistency between the inferred and actual properties of an object in the schema graph constitutes a potential disconfirmation of

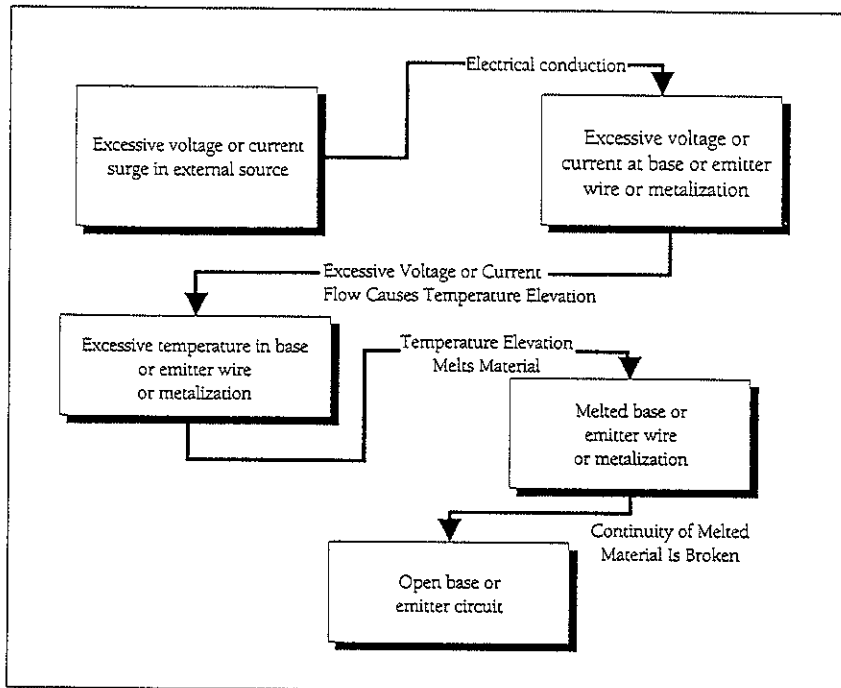


Figure 2. A linear schema graph for electrical-overstress-induced-open. Excessive voltage or current causes localized temperature elevation, resulting in melted wire or metallization.

the explanation. At this point, two courses of action are possible. Either the current hypothesis is discarded or we attempt to repair it. If the hypothesis is discarded, the procedure returns to step 3 and begins to instantiate the next hypothesis on the hypothesis list.

If, on the other hand, no other hypotheses remain, or no other hypothesis has a similar weight of evidence supporting it, we may choose to repair the current explanation. In the melted bond wire example, we may conjecture that the bond wire was pinched or otherwise thinned at the location where it melted, thus reducing the amount of current required to melt it. To accomplish such an explanation repair, we need to locate the source of the inconsistent constraint, the causal link where it originated, and modify one of the conditions on which the constraint depended.

The problem of explanation repair is one of the most difficult faced by our method. We explore two general approaches to the problem: 1) to encode the repair strategies employed by human experts in the area and 2) to use abductive reasoning to reconstruct from scratch the schema subgraph where the inconsistency was detected.

Schema-Based Abduction

At the heart of the procedure described above is a process of reasoning which we call *schema-based abduction*. This refers to the method by which the causal links specified by the schema graph are reconstructed from the causal domain theory in the context of a set of situation-specific bindings. We next characterize this mode of reasoning more precisely, relating it to the conventional notion of abduction.

Abduction, as it is ordinarily understood by the logic community in AI, is a mode of inference which generates candidate explanations for an otherwise unexplained set of observations O . Abductive inference allows us to *assume* facts not directly in evidence. More formally, an hypothesis H is a minimal abductive explanation for observation set O if:

- i O is not entailed by the current background knowledge K
- ii $H \cup K$ entails O
- iii No proper subset of H has property ii
- iv H is consistent with K

Abductive reasoners have been implemented that generate the complete set of minimal abductive explanations using a relatively straightforward backchaining approach. They rely on the "inference rule":

$$\text{abduce}(B, \alpha \rightarrow B) = \alpha$$

where α is a conjunction of literals. Backchaining proceeds by taking each $a_i \in \alpha$, $a_i \notin K$, as a new abductive subgoal. Such algorithms work best over sets of propositional Horn clauses. Typically these systems take as input a set A of abducible propositions, that is, propositions which the abductive reasoner is allowed to include in the hypothesis H . Even for a propositional Horn clause language, however, this task has been shown to be NP-hard.

The schema-based abduction which we employ represents a highly constrained form of abductive reasoning. In schema-based abduction the objects related as cause and effect are already given in abstract form in the schema graph. Abductive reasoning consists not in inferring causes from effects but in inferring *properties* of the cause from the *properties* of the effect. More precisely, for any given link in the schema graph, the set A of abducibles is restricted to properties of the objects or events playing the role of cause in that particular link. The effect of abduction is thus a kind of constraint propagation, where instance bindings representing known properties in the effect are used to create new instance bindings in the cause. Backchaining search in schema-based abduction is thus considerably simplified. Its flow is determined by the preset pattern of links in the schema graph.

There are many cases where backward constraint propagation requires a method other than simple abductive inference. Consider the case discussed

earlier in which electrical overstress is the hypothesis used to explain a melted bond wire. We can infer from the composition of the bond wire and its thickness the temperature required to melt it. From this we can infer a minimum range of current over time required to produce that temperature. The propagation of constraints in this case requires an equation solving and equality substitution capability. We thus identify the term *schema-based abduction* with a broadened notion of abduction that includes methods for backward constraint propagation such as equation solving and equality substitution.

Schema-Based Diagnosis: An Extended Example

We next analyze a typical situation from the semiconductor failure analysis domain in terms of schema-based problem solving. A bipolar transistor is brought in with a complaint of "low gain." A series of standard electrical measurements confirms low gain at low current (low gain in low current hFE) as well as discovering a second abnormal electrical characteristic: high collector base⁷ leakage (high ICBO).

The first phase of reasoning, corresponding to the data abstraction phase in heuristic classification, involves the firing of a classification rule. High ICBO and low hFE together are classified as a collector base leakage problem:

$$\text{result(ICBO, high)} \wedge \text{result(hFE_lo, low gain)} \Rightarrow \text{problem(C-B, leakage)}$$

This reduction can be readily reconstructed from the domain theory. The explanation is that collector base leakage reduces gain by lowering the effective base drive. Note that reasoning so far relies solely on representations of internal device structure and function. This type of reasoning is well handled by the model-based paradigm.

In the next phase of reasoning, we seek the cause of high collector base leakage. Heuristic classification rules identify three types of causal mechanisms which can explain collector base leakage:

$$\text{problem(C-B, leakage)} \Rightarrow \text{hypothesis(electrical overstress)}$$

$$\text{problem(C-B, leakage)} \Rightarrow \text{hypothesis(contamination)}$$

$$\text{problem(C-B, leakage)} \Rightarrow \text{hypothesis(mask misalignment/overetch)}$$

These rules all fire, causing the activation of three schemas. The data structure for a typical schema is described in Figure 3.

Each of the schemas is located within a schema hierarchy, reflecting the hierarchical structure of causal abstractions used to organize the search for explanations. An example of a schema hierarchy for *bridging faults* is given in Figure 4.

We use the available evidence to activate the most specific hypotheses pos-

Name	electrical overstress induced leakage
Attribute	type of overstress (voltage/current) location of overstress (emitter/base) intensity (pulse/power) source type
Indications	test equipment or application clustering linear junction characteristic leakage is stable over temperature
Test Procedures	delid: visual examination; look for orange peel or burn delid: SEM; look for pitting or tunneling delid: deprocessing; look for channels, damage to Si crystalline structure
Micro-theory	overstress induced leakage
Schema Graph	EOS-Leakage
Subschemas	ESD induced leakage oscillation overstress induced leakage pulse power overstress induced leakage

Figure 3. The data structure used for electrical-overstress-induced-leakage. The schema includes not only needed values but also "compiled knowledge" relating to indications and testing.

sible, that is, the deepest hypothesis in the schema hierarchy. Suppose, for example, that our transistor is a PNP device. In the schema hierarchy for bridging faults, we first activated the *contamination* schema among others. Using additional evidence, we then specialize the hypothesis of *contamination*. The most common form of contamination is Na^+ . Na^+ contamination in PNP transistors typically produces *inversion*, resulting in a characteristic signature of high ICBO leakage. Because our transistor is PNP and because of the presence of ICBO leakage, we thus specialize the hypothesis of contamination using the following rule:

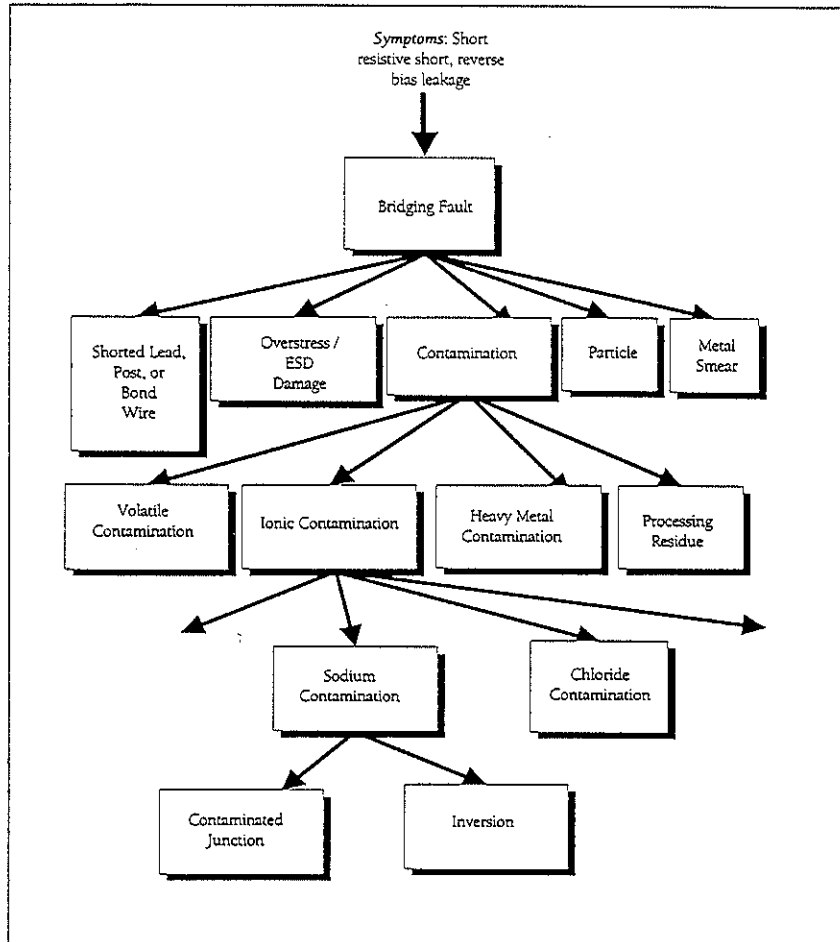


Figure 4. The schema hierarchy for bridging faults. Diagnosis seeks to identify a hypothesis as deep in the hierarchy as possible.

hypothesis(contamination) \wedge device(polarity, PNP) \wedge problem(C-B, leakage) \wedge NOT(problem(E-B, leakage)) \Rightarrow hypothesis(inversion)

Similarly, we might fire a rule activating ESD⁸ as a subhypothesis of electrical overstress based on the fact that the degradation is localized and that small signal devices are particularly sensitive to ESD damage:

hypothesis(overstress) \wedge problem(C-B, leakage) \wedge NOT(problem(E-B, leakage)) \wedge device(power, small signal) \Rightarrow hypothesis(ESD)

After the hypothesis set is generated, the *hypothesis pruning* stage begins.

Each schema provides a standard set of procedures for gathering evidence related to that particular causal mechanism. Electrical overstress, for example, is investigated by gathering details about the device's history and its possible exposure to an overstress or ESD environment. Similarly, mask misalignment is generally a wafer level problem; it can thus be investigated by determining if other chips from the same wafer are similarly degraded.

The internal verification procedures for overstress and mask misalignment/over-etch both involve cutting off the package lid and examining the internal structure of the device. Since these procedures are potentially destructive of evidence, we put them off as long as possible. The internal verification procedure for contamination, on the other hand, is usually non-destructive. This procedure depends on the temperature sensitivity of contamination. Since many contaminants are dispersed by elevated temperatures, the device is baked and then electrically retested to see if its electrical characteristics have improved. Let us suppose that they have. This fact then increases the probability of contamination, without completely eliminating the other hypotheses.

Up until now we have pruned and ordered the hypothesis list using non-destructive procedures. A nondestructive approach was necessary in order to allow for the possibility of backtracking. Beyond this point we engage in procedures which involve irreversible changes to the device and thus potential destruction of evidence. From this point on, we organize testing based on a careful consideration of potential gains versus potential costs, including destruction of evidence. To focus evidence gathering and precisely define the goals of each procedure, we continue the investigation in the context of constructing a detailed explanation based on the most likely hypothesis.

Explanation construction proceeds by instantiating the schema graph for the hypothesized mechanism. We illustrate again the notion of a schema graph by presenting the schema graph for *Inversion Induced Leakage* in Figure 5.

In constructing an explanation, the instantiated schema graph is traversed backwards or abductively, reconstructing the causal links using the causal domain theory. This serves two main purposes: to test the viability of the explanation by determining consistency with known facts, and to serve as a source of potential tests by fleshing out required conditions or assumptions. Returning to our example, the hypothesis of *inversion* is elaborated by proceeding backwards from the node *High ICBO* (at the lower right hand corner of the *Inversion Schema* graph). According to the graph, high ICBO is caused by an exposed junction along the edge of the chip. To reconstruct this causal link from the domain theory, we must make use of the fact that the chip surface along the edge is rough because of the way the chip is split off or sawed from the wafer. It is this roughness, the absence of a regular crystalline structure, that produces low level leakage when there is an exposed junction. Reconstructing this causal link from the domain theory thus fleshes out a hid-

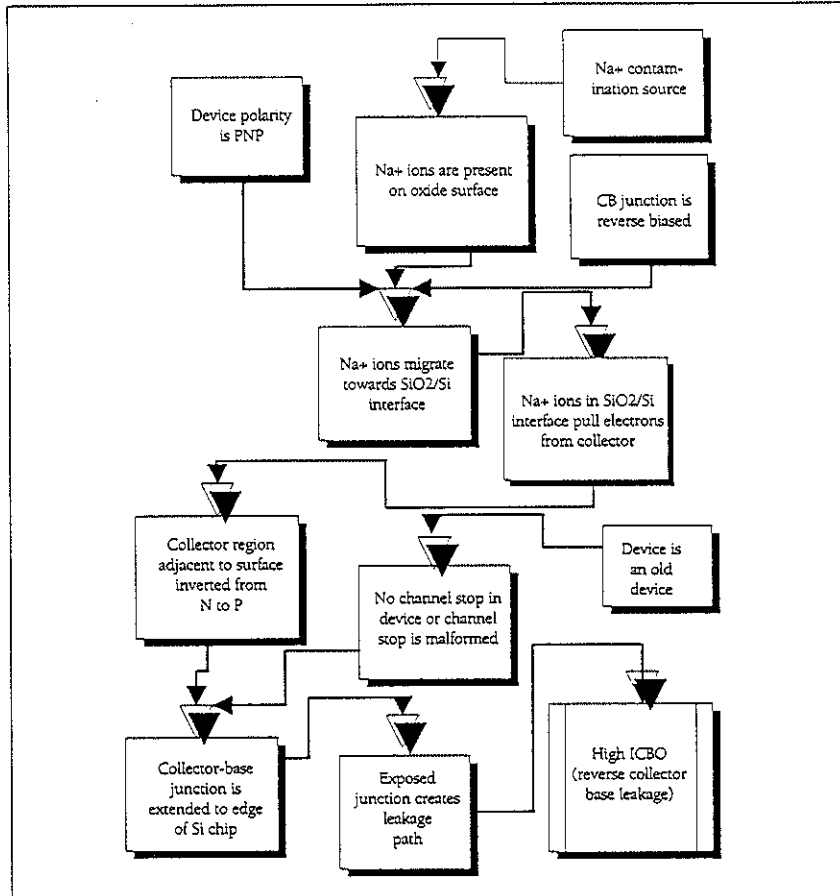


Figure 5. The schema graph for inversion-induced-leakage. Sodium ions produce an inverted region, extending junction out to the edge of the chip. Leakage occurs at the chip's edge.

den assumption, an assumption which might be violated if, for instance, a new method for separating chips from the wafer were invented.

Let us suppose that we have abductively regressed to the condition *Collector-base junction extended to edge of chip* (lower left-hand corner of the schema graph). The two conditions required to produce this result are 1) the collector base region is inverted and 2) no channel stop is present to prevent the extension of the inverted region to the chip's edge. Inversion has been such a common cause of PNP transistor problems that modern PNP transistors are almost always built with a channel stop to prevent the inverted region from

reaching the edge of the chip. The presence of a properly functioning channel stop in the transistor would thus be inconsistent with an explanation based on inversion. The reconstruction of this causal link thus requires that we establish either 1) the absence of a channel stop or 2) a defect in the channel stop.

Suppose we determine, after a low power internal visual examination, that a channel stop is present. This determination then focuses the examination of the die on the search for a defect in the channel stop. The discovery of such a defect would significantly increase the force of evidence behind an inversion-based explanation. It would also specialize the explanation, adding an important twist: the cause of the leakage problem is Na⁺ contamination in conjunction with a faulty channel stop. The process of explanation construction thus results in an explanation richer in detail and more useful than the abstract explanatory hypothesis from which we started.

Conclusion

We propose an architecture for integrating heuristic problem solving and causal reasoning. This approach uses heuristic matching at a high level of abstraction to frame initial hypotheses. These are then refined to form fully articulated explanations. The schema-based architecture we propose involves a dynamic process of interrogation, explanation generation, and hypothesis evaluation. This in effect supports search through alternative interpretation spaces constrained by the fit of hypotheses to data.

In addition, the schema-based architecture provides a means for modeling the practical dimension of explanation construction. We have observed in our work with human failure analysts that the structure and focus of explanations generally reflect the practical need to support specific remedial practices. The use of a schema-based architecture to control explanation construction allows us to model this practical dimension, generating explanations which are consistent with the evidence and the applicable laws of semiconductor physics but which at the same time embody a structure developed over time to support certain types of corrective action.

Notes

1. This expert system, called DSFAX, was developed for Sandia National Laboratories from 1988 to 1994.
2. *Burn in* is a test procedure in which a device is powered up over an extended period of time under carefully controlled conditions in order to identify potential defects.
3. Leakage denotes the existence of a small channel current path between device locations or across a junction which is reverse biased.

4. NPN and PNP stand for the two possible polarity structures of bipolar transistors.
5. Inversion is a phenomenon in which positively charged ions attract negative ions from an adjacent P region, effectively creating a thin N region across which a leakage current can flow.
6. Electrical overstress is an excessive current or voltage applied to the device.
7. The two junctions of a bipolar transistor are the collector base (CB) and emitter base (EB).
8. ESD, or ElectroStatic Discharge, is a short duration high voltage current resulting from a static buildup, usually from a human source.

References

- Bartlett, F. (1932). *Remembering*. London: Cambridge University Press.
- Charniak, E. & Shimony, S. (1990). Probabilistic semantics for cost based abduction. *Proceedings of the Eighth National Conference on Artificial Intelligence* (pp. 106-111). Menlo Park, CA: AAAI/MIT Press.
- Clancey, W.E. (1985). Heuristic classification. *Artificial Intelligence*, 27, 289-350.
- Davis, R. & Hamscher, W. (1988). Model based reasoning: Troubleshooting. In W. Hamscher, L. Console, & J. deKleer (Eds.), (1992), *Readings in Model Based Diagnosis*. San Mateo, CA: Morgan Kaufman.
- Doyle, J. (1983). Methodological simplicity in expert system construction: The case of judgments and reasoned assumptions. In I. Shafer & J. Pearl (Eds.), (1990), *Uncertain Reasoning* (pp. 689-693). San Mateo, CA: Morgan Kaufman.
- Eco, U. (1976). *A Theory of Semiotics*. Bloomington: Indiana University Press.
- Feltovich, P.J., Johnson, P.E., Moller, J.H., & Swanson, D.B. (1984). LCS: The role and development of medical knowledge in diagnostic expertise. In W.J. Clancey & E.H. Shortliffe (Eds.), *Readings in Artificial Intelligence in Medicine: The First Decade*. Reading, MA: Addison Wesley.
- Gallanti, M., Roncato, M., Stefanini, A., & Torielli, G. (1989). A diagnostic algorithm based on models at different levels of abstraction. *Proceeding of the Eleventh International Joint Conference on Artificial Intelligence* (pp. 1350-1355). San Mateo, CA: Morgan Kaufman.
- Kant, I. (1781/1964). *Critique of Pure Reason*. N.K. Smith (Trans.). New York: St. Martin's Press.
- Kass, A., Leake, D., & Owens, C. (1986). SWALE: A program that explains. In *Explanation Patterns: Understanding Mechanically and Creatively* (pp. 232-254). Hillsdale, NJ: Lawrence Erlbaum.
- Kuipers, B. & Kassirer, J. (1984). Causal reasoning in medicine: analysis of a protocol. *Cognitive Science*, 8, 363-385.
- Leake, D. (1992). *Evaluating Explanations: A Content Theory*. Hillsdale, NJ: Lawrence Erlbaum.
- Levesque, H.J. (1989). A knowledge level account of abduction. *Proceeding of the Eleventh International Joint Conference on Artificial Intelligence* (pp. 1061-1067). San Mateo, CA: Morgan Kaufman.
- Luger, G.F. & Stern, C. (1992). Expert Systems and the abductive circle. In R. Jorna & B. von Heusden (Eds.), *Cognition and Semiotics*. Berlin: Walter de Gruyter.

- Newell, A. & Simon, H. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice Hall.
- Ng, H.T. & Mooney, R.J. (1990). On the role of coherence in abductive explanation. *Proceedings of the Eighth National Conference on Artificial Intelligence* (pp. 337-342). Menlo Park, CA: AAAI/MIT Press.
- Patil, R. (1981). Causal representation of patient illness for electrolyte and acid-base diagnosis. In *MIT/LCS/TR-267*. Cambridge: MIT Press.
- Pearl, J. (1987). On evidential reasoning in a hierarchy of hypotheses. *Artificial Intelligence*, 28, (1986) 9-15. [Reprinted in I. Shafer & J. Pearl (Eds.), (1990), *Uncertain Reasoning* (pp. 449-451). San Mateo, CA: Morgan Kaufman.]
- Peirce, C.S. (1958). *Collected Papers: 1931-1958*. Cambridge MA: Harvard University Press.
- Peng, Y. & Reggia, J.A. (1990). *Abductive Inference Methods for Diagnostic Problem-Solving*. Berlin: Springer Verlag.
- Piaget, J. (1970). *Structuralism*. New York: Basic Books.
- Poole, D. (1989). Normality and faults in logic-based diagnosis. *Proceeding of the Eleventh International Joint Conference on Artificial Intelligence* (pp. 1304-1310). San Mateo, CA: Morgan Kaufman.
- Reggia, J., Nau, D.S., & Wang, P.Y. (1983). Diagnostic expert systems based on a set covering model. *International Journal of Man-Machine Studies*, 19(5), 437-460.
- Selman, B. & Levesque, H.J. (1990). Abductive and default reasoning: A computational core. *Proceedings of the Eighth National Conference on Artificial Intelligence* (pp. 343-348). Menlo Park, CA: AAAI/MIT Press.
- Torasso, P. & Console, L. (1989). *Diagnostic Problem Solving: Combining Heuristic, Approximate, and Causal Reasoning*. London: North Oxford Academic.