On the use of causal feature selection in the context of machine-learning indirect test

M. J. Barragan*, G. Leger†, F. Cilici*‡, E. Lauga-Larroze‡, S. Bourdel‡, S. Mir*
*Univ. Grenoble Alpes, CNRS, Grenoble INP[1], TIMA F-38000, Grenoble, France
†Instituto de Microlectrónica de Sevilla, CSIC-Universidad de Sevilla, 41092, Seville, Spain
‡Univ. Grenoble Alpes, RFICLab 38000, Grenoble, France

***Abstract*—The test of analog, mixed-signal and RF (AMS-RF) circuits is still considered as a matter of human creativity, and although many attempts have been made towards their automation, no accepted and complete solution is yet available. Indeed, capturing the design knowledge of an experienced analog designer is one of the key challenges faced by the Electronic Design Automation (EDA) community. In this paper we explore the use of causal inference tools in the context of AMS-RF design and test with the goal of defining a methodology for uncovering the root causes of performance variation in these systems. We believe that such an analysis can be a promising first step for future EDA algorithms for AMS-RF systems.**

I. Introduction

Nowadays, functional test continues to be the standard for AMS-RF production test. However, this central position has been recently challenged. The wide variety of AMS-RF circuits and huge number of analog specifications make functional test a lengthy and expensive task. The cost of assuring quality is the second reason: key economic sectors –such as automotive– are pushing to improve quality and reliability, to the point where functional test becomes insufficient. Machine learning indirect test [1] is a promising strategy for overcoming the issues of traditional functional test. The basic idea is to replace costly specification measurements by a set of simpler measurements, often called signatures or features. Machine learning regression algorithms are then used to map features to specifications. This has many interesting advantages. Among others, the simple measurements are designed to be cheaper than their functional counterparts and can be also tailored to yield a higher defect coverage. However, machine learning indirect test is not free of shortcomings. One of the key issues is the lack of a methodology for proposing a robust set of features to extract a meaningful regression model. It is often a matter of creativity based on a precise knowledge of the Device Under Test (DUT). Former works in this line [2]–[8] have essentially focused on maximizing the predicting power of the feature set by performing feature selection on a large set of candidates. For that purpose, they resorted to different methods ranging from filtering to wrapper or a combination of both. The main advantage of filtering methods is that they do not use the training algorithm and are usually more computationally efficient than wrapper approaches which perform a combinatorial search with some form of optimizer.

[1]Institute of Engineering Univ. Grenoble Alpes

These methods rely on implicit correlations. Indeed, in a purely observational scenario, uncovering causal mechanisms is unnecessary for making good predictions [9]. However, this is not satisfying from a robustness perspective since the information on the underlying mechanisms that link the features to the metric of interest is lost. These mechanisms are the causal relations that reflect the rules of physics governing the circuit behavior. Let us consider the following thought experiment. The environmental temperature and the gain of an amplifier can appear to be correlated. Obviously the temperature is the cause of the change in gain and not the contrary. However, in a purely observational scenario, we could devise a simple test that predicts the gain of the amplifier by measuring the number of glasses of water that the test engineer drinks, with the rationale that temperature could cause the test engineer to drink more or less water. In this particular environment, for this particular test engineer, this could even lead to good results but, clearly, by no means this is a robust test.

In this work we explore novel causal inference algorithms in the context of feature selection problems in order to enhance the robustness of the predictions. In this line, Section II presents the basic concepts behind causality inference algorithms. Then, Section III illustrates the application of the technique to a case study –a mm-wave Power Amplifier (PA) designed in a BiCMOS 55 nm technology. Section IV summarizes the main contributions of this paper and discusses the potential future applications of the proposed technique.

II. Theoretical basis

A. Basic concepts in causality inference

From an engineering point of view the concept of causality is intuitively linked to modifying the evolution of a system: if we modify a system in a certain way, then the system will react accordingly, following a cause-effect evolution. Mathematically, causality relations between random variables describing the state of a system can be conveniently represented using causal Bayesian networks [10]. We will illustrate the concept of causal Bayesian networks with a simple example. Let us consider the directed acyclic graph (DAG) in Fig. 1. For reference, we recall that in a DAG a node A is denoted as the parent of node B (conversely, B is the child of A) if there is a direct edge from A to B, while A is the ancestor of B (B is the descendant of A) if there is a direct path from A to

978-3-9819263-2-3/DATE19/©2019 EDAA

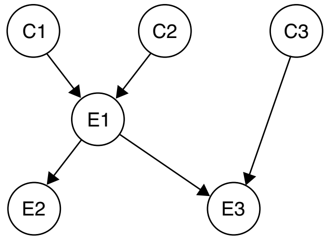

Fig. 1. Example of causal Bayesian network.

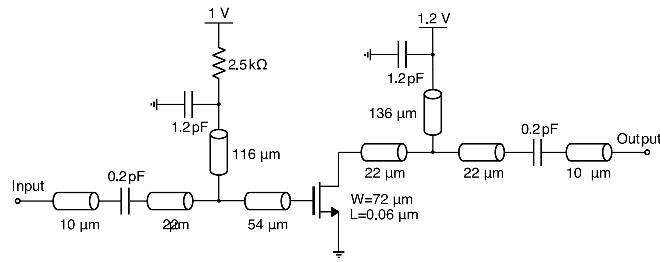

Fig. 2. Transistor-level schematic of the Power Amplifier case study

B. For instance, in Fig. 1, $C1$ is a parent of $E1$, and it is also an ancestor of $E2$ and $E3$.

A Bayesian network, also denoted as Belief Network or Probabilistic Network [11], is the triplet of: a) a set of random variables, b) the joint probability distribution of these variables and c) a DAG whose nodes correspond to each of the random variables. Such a triplet is a Bayesian network if it complies with the Markov condition, that is, if any node A in the network is statistically independent of all non-descendants of A, given the parents of A. For instance, in the example in Fig. 1 the Markov condition requires that $C3$ is statistically independent from $E1$, since $C3$ is not a descendant of $E1$.

A causal Bayesian network is a Bayesian network with the additional condition that each edge in the network represents a cause-effect relationship. Using Pearl's notation [10], if there is an edge from node A to node B and we set node A to a desired value (denoted as $do(A)$), then A is a cause of B (B is the effect of A) if the conditional probability $P(B|do(A))$ is different than the raw probability $P(B)$, that is, if manipulating the value of A the resulting distribution of B is different compared to the normal evolution of the system.

It is important to notice the difference between $P(B|do(A))$ and $P(B|A)$. If $P(B|A) \neq P(B)$ we can only conclude that A and B are correlated when the system evolves according to its governing joint probability density function. On the other hand, the condition $P(B|do(A)) \neq P(B)$ has a more profound meaning, since it implies that any change imposed on A will have an effect in the distribution of B. Intuitively, this is the mathematical difference between correlation and causation.

B. Markov blanket and causal feature selection

The Markov blanket associated to a given node in a Bayesian network is by definition the set of nodes shielding this node from the influence of the rest of the nodes in the network [10]. That is, the evolution of a given node in a Bayesian network can be explained by the state of the nodes in its Markov blanket, independently of the state of the rest of the nodes in the network. In other words, if the state of the Markov blanket of a variable is known, the other nodes will not give any extra information about the state of this variable. This does not imply that we can determine exactly the state of the target variable but it does guarantee that it is the best that we can do with the variables that we have.

If the network is a causal Bayesian network, it can be proved that any node has a unique Markov blanket composed by its direct parents (its causes), direct children (its effects) and its spouses (other direct causes of its direct effects). Going back to our example in Fig. 1, for instance, the Markov blanket of node $C1$ is composed by nodes $C2$ (its spouse) and $E1$ (its child). Conceptually, if we think of these variables as binary events to simplify the discussion, if we want to know if event $C1$ happened, we can just observe the variables in its Markov blanket. Thus, if $E1$ happened and $C2$ did not, it is clear that $C1$ did happen. However, if $E1$ and $C2$ both happened, the state of $C1$ cannot be determined. The information in the Markov blanket is not complete, but no other variable in the network will give us any additional information about $C1$. These properties of the Markov blanket have a direct application in the field of feature selection. In a machine learning scenario, variables in the Markov blanket are likely to be the most robust candidate features for adequately capturing the behavior of the regression target.

In this line, causal discovery machine learning algorithms [12], [13] have been proposed for inferring causal relationships from sets of observational data. Generally, these algorithms start by identifying the shape of the DAG around the target variable and then they test conditional independence and dependence hypothesis between the identified variables to determine the Markov blanket of the target variable. Multiple implementations have been published in the last few years that optimize it for applications with large sets of variables with high connectivity. These algorithms differ in the way that the variables are ranked and in the way that the conditional independence and dependence hypothesis are checked. Readers are referred to [12], [13] for a detailed discussion of causal discovery machine learning algorithms.

III. Application example

The selected case study is a single stage 60 GHz PA designed in a 55 nm BiCMOS technology. Fig. 2 shows the transistor-level schematic of the PA. The PA input and output nodes are matched to a 50 Ω impedance at 60 GHz using microstrip line stubs. Despite the apparent simplicity of the circuit, it is difficult to identify *a priori* the dominant sources of performance degradation of the PA, since it is affected by variations of both the active and passive components. At the same time, the circuit is simple enough to facilitate the interpretation of the results. This makes it an interesting case study for the proposed causal inference methodology.

In this paper we propose a proof-of-concept experiment aimed at analyzing the root-causes of performance variation in the presence of process variation. The proposed experiment is carried out at electrical simulation level. A set of 2000 instances of the PA were generated using the Monte Carlo models in the Process Design Kit (PDK) of the technology. In our simulation environment, we have direct access to the complete set of possible root-causes of performance variation, that is, the complete set of Monte Carlo process parameters in the PDK models. The selected 55 nm BiCMOS technology has 154 Monte Carlo process parameters. As an illustrative example of application, we propose to use machine learning causality discovery to explore this high-dimensionality space and find the Markov blanket of one of the specifications of the considered case study, namely the gain of the PA at 60 GHz. For this experiment, we use the Incremental Association-Based Markov Blanket (IAMB) algorithm included in the Causal Explorer Toolbox [12] in Matlab on the 2000 generated Monte Carlo instances. It is important to notice that, in this particular experiment, the features (i.e. the process parameters) can only be independent causes of the performance degradation. This is not a realistic test scenario, but it opens the possibility to validate the causal features selection with classical feature selection algorithms for regressing the gain of the PA. Indeed, the features that are not degradation causes should neither be selected by causal discovery nor by performance-oriented feature selection. Conversely, any feature that gives unique information should be selected by both. In that particular case of known independent causal features, we could as well consider a simple sensitivity analysis instead of conventional feature selection.

Table I shows the Monte Carlo process parameters in the Markov blanket associated to the gain of the PA, obtained with the IAMB algorithm and ranked by causal relevance. The Markov blanket is composed of only 22 parameters from the total 154 candidates. For confidentiality reasons, we cannot disclose the actual names of the model parameters in the PDK. For convenience, model parameters corresponding to a given device have been renamed as “device_i”, where “device” is the device name and “i” is a blind index.

In a DAG representation, these 22 parameters are parents of the target specification and they causally explain the complete variation of the gain of the PA under process variations. To further validate this claim, Fig. 3 shows a scatterplot of the actual versus predicted gain specification for a regression model trained with the parameters in the Markov blanket. The model is a perceptron neural network trained using 1700 instances of the PA, while 300 instances were used as independent validation set. The generalization error in the validation set is of only 0.02 dB, which again validates the relevance of the causal features

For reference, Table I shows also the 22 most relevant parameters as selected by the hybrid feature selection algorithm in [8], ranked by relevance. As expected, in this proof-of-concept example the obtained sets of parameters are almost identical (only the two less relevant parameters differ and the ranking order is similar) for both causal and non-causal selection algorithms which further validates the results of the proposed causal analysis.

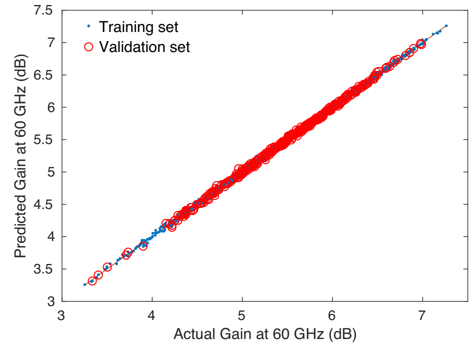

Fig. 3. Prediction of PA Gain at 60 GHz using the set of variables in the Markov blanket of the Gain.

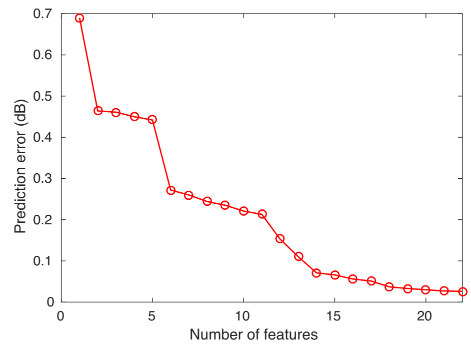

Fig. 4. Prediction error for the PA Gain at 60 GHz versus the number of features used for training. Features are selected from the Markov blanket following their ranking order.

Although the results seem equivalent, it should be remarked that the causal discovery algorithm is a supervised filtering algorithm, while the hybrid feature selection uses a wrapper. In other words, hybrid feature selection relies on training regression models and optimizing the prediction error, which is computationally expensive. Causal discovery, being a filtering method, does not need model training, which saves significant processing resources and opens the door to high-dimensionality problems. Even more, it can be shown that this filtering algorithm actually finds all significant features for the prediction of the PA gain. This is a particularly important result since correlation-based filters such as the ones in [5], [7] usually detect only first-order contributions. In this line, Fig. 4 shows the evolution of the prediction error of the gain as a function of the number of features in the Markov blanket used for the training (selected by ranking order), for devices in the validation set. It is clear to see that the set of 22 selected features actually explain the observed variation of the PA gain, as the prediction error reaches a plateau at 0.02 dB. Actually, adding any more parameters to this set does not improve the prediction error, which is a strong indication of the completeness of these 22 features for explaining the PA gain variation.

Additionally, given the topological simplicity of the case study, it is possible to interpret the obtained results from an electrical point of view. As it is shown in Table I, the parameters in the Markov blanket of the gain are related to the impedance of MOM capacitors and microstrip lines, ohmic losses, and the operation point of the NMOS transistor.

TABLE I
MARKOV BLANKET INFERENCE VS HYBRID FEATURE SELECTION ON THE SPACE OF MONTE CARLO MODEL PARAMETERS

Markov blanket of PA gain (electrical effect in PDK models)	Feature selection for PA gain regression (electrical effect in PDK models)
CMOM_1 (MOM capacitor ohmic losses)	Microstrip_1 (Microstrip line impedance)
Microstrip_1 (Microstrip line impedance)	NMOS_1 (NMOS gate resistance)
Microstrip_2 (Microstrip line ohmic losses)	Microstrip_2 (Microstrip line ohmic losses)
CMOM_2 (MOM capacitor impedance)	CMOM_1 (MOM capacitor ohmic losses)
Microstrip_3 (Microstrip line impedance)	Microstrip_3 (Microstrip line impedance)
NMOS_1 (NMOS gate resistance)	NMOS_5 (NMOS parasitic capacitances)
Microstrip_4 (Microstrip line impedance)	Microstrip_6 (Microstrip line impedance)
Microstrip_5 (Microstrip line impedance)	NMOS_2 (NMOS gate resistance)
NMOS_2 (NMOS gate resistance)	Microstrip_4 (Microstrip line impedance)
NMOS_3 (NMOS gate resistance)	NMOS_4 (NMOS parasitic capacitances)
NMOS_4 (NMOS parasitic capacitances)	Microstrip_5 (Microstrip line impedance)
NMOS_5 (NMOS parasitic capacitances)	Microstrip_7 (Microstrip line impedance)
Microstrip_6 (Microstrip line impedance)	CMOM_2 (MOM capacitor impedance)
Microstrip_7 (Microstrip line impedance)	NMOS_3 (NMOS gate resistance)
Microstrip_8 (Microstrip line impedance)	NMOS_7 (NMOS parasitic capacitances)
NMOS_6 (NMOS gate oxide)	Microstrip_8 (Microstrip line impedance)
NMOS_7 (NMOS parasitic capacitances)	NMOS_6 (NMOS gate oxide)
CMOM_3 (MOM capacitor impedance)	Microstrip_9 (Microstrip line impedance)
NMOS_8 (NMOS output resistance)	NMOS_9 (NMOS threshold voltage)
NMOS_9 (NMOS threshold voltage)	CMOM_3 (MOM capacitor impedance)
Microstrip_9 (Microstrip line impedance)	Microstrip_10 (Microstrip line impedance)
NMOS_9 (NMOS transconductance)	CMOM_4 (MOM capacitor impedance)

Notice that these causal relationships, that basically reflect a trade-off between load adaptation, electrical losses and transistor amplification, are in the know-how of an expert mm-wave designer. However, in this case they were obtained automatically by the causal discovery algorithm, without the need of previous electrical knowledge of the PA.

IV. CONCLUSION AND FUTURE WORK

This paper explores the use of machine learning causal discovery algorithms in the context of analog test. The basic concepts behind advanced causality inference techniques have been laid out, and an application example has been presented using a 60 GHz PA in a 55 nm BiCMOS technology. Preliminary results show that machine learning causal discovery algorithms may open the door to interesting advances in the fields of analog design and test.

Firstly, it is clear that causality inference may improve current machine learning indirect test frameworks by guiding the design of test features towards the Markov blanket of the target specification. Moreover, the resulting causal features can potentially improve defect detection and diagnosis. In our future work, we intend to advance this research line towards the design of robust defect filters based on causal features. Such defect filters should be more robust to unknown defects than current defect filters based on observational features.

From a designer point of view, the use of causal discovery algorithms may also be interesting. The extraction of the Markov blanket of the set of specifications of a given circuit is a valuable diagnosis tool. This, for instance, could help designing dedicated tuning knobs in complex systems that take advantage of the unveiled cause-effect relationships.

ACKNOWLEDGMENT

This work has been partially funded by the Spanish Government project TEC2015-68448-R (co-funded by FEDER program) and by the CNRS PICS project IndieTEST PICS07703.

REFERENCES

[1] P. N. Variyam and A. Chatterjee, "Enhancing test effectiveness for analog circuits using synthesized measurements," in *VLSI Test Symposium*, 1998, pp. 132–137.

[2] S. S. Akbay and A. Chatterjee, "Feature extraction based built-in alternate test of RF components using a noise reference," in *VLSI Test Symposium*, 2004, pp. 273–278.

[3] H. Ayari *et al.*, "Smart selection of indirect parameters for DC-based alternate RF IC testing," *VLSI Test Symposium*, pp. 19–24, 2012.

[4] H.-G. Stratigopoulos *et al.*, "RF Specification Test Compaction Using Learning Machines," *Trans. on VLSI Systems*, vol. 18, no. 6, pp. 998–1002, June 2010.

[5] M. Barragan and G. Leger, "Efficient selection of signatures for analog/RF alternate test," in *Test Symposium (ETS), 2013 18th IEEE European*, May 2013, pp. 1–6.

[6] M. J. Barragan and G. Leger, "A procedure for alternate test feature design and selection," *Design & Test*, vol. 32, no. 1, pp. 18–25, 2015.

[7] S. Larguech *et al.*, "Efficiency evaluation of analog/rf alternate test: Comparative study of indirect measurement selection strategies," *Microelectronics Journal*, vol. 46, no. 11, pp. 1091–1102, 2015.

[8] G. Leger and M. J. Barragan, "Brownian distance correlation-directed search: A fast feature selection technique for alternate test," *Integration, the VLSI Journal*, vol. 55, pp. 401–414, 2016.

[9] I. Guyon *et al.*, *Computational Methods of Feature Selection*. Chapman and Hall, CRC Press, 2007, ch. Causal feature selection.

[10] J. Pearl, *Causality*. Cambridge University Press, 2009.

[11] R. E. Neapolitan, *Probabilistic reasoning in expert systems: Theory and algorithms*. John Wiley and Sons, 1990.

[12] C. F. Aliferis *et al.*, "Causal Explorer: A Causal Probabilistic Network Learning Toolkit for Biomedical Discovery." in *Int. Conf. on Mathematics and Engineering Techniques in Medicine and Biological Sciences, METMBS*, vol. 3, 2003, pp. 371–376.

[13] ——, "Local causal and markov blanket induction for causal discovery and feature selection for classification part i: Algorithms and empirical evaluation," *The Journal of Machine Learning Research*, vol. 11, pp. 171–234, 2010.