Symbolic Regression for Derivation of an Accurate Analytical Formulation Using "Big Data": An Application Example

Peyman Mahouti [1], Filiz Güneş [1], Mehmet A. Belen [2], and Salih Demirel [1]

[1] Department of Electronic and Communication Engineering
Yıldız Technical University, Istanbul, Turkey

[2] Department of Electric and Electronic Engineering
Artvin Çoruh University, Istanbul, Turkey
pmahouti@yildiz.edu.tr, gunes@yildiz.edu.tr, mehmetalibelen@artvin.edu.tr, salihd@yildiz.edu.tr

Abstract ─ With emerging of the Big Data era, sample datasets are becoming increasingly large. One of the recently proposed algorithms for Big Data applications is Symbolic Regression (SR). SR is a type of regression analysis that performs a search within mathematical expression domain to generate an analytical expression that fits large size dataset. SR is capable of finding intrinsic relationships within the dataset to obtain an accurate model. Herein, for the first time in literature, SR is applied to derivate a full-wave simulation based analytical expression for the characteristic impedance Z_0 of microstrip lines using Big Data obtained from an 3D-EM simulator, in terms of only its real parameters which are substrate dielectric constant ε, height h and strip width w within 1-10 GHz band. The obtained expression is compared with the targeted simulation data together with the other analytical counterpart expressions of Z_0 for different types of error function. It can be concluded that SR is a suitable algorithm for obtaining accurate analytical expressions where the size of the available data is large and the interrelations within the data are highly complex, to be used in Electromagnetic analysis and designs.

Index Terms ─ Big Data application, characteristic impedance, microstrip line, Symbolic Regression.

I. INTRODUCTION

The emerging of the big data era poses a challenge to traditional machine learning algorithms. In many research area, big data is a collection of datasets so large and complex that they become difficult to process using available database management tools or traditional data processing applications. Big data is usually composed of datasets with sizes beyond ability of the commonly used software tools, which are unable to capture, curate, manage, or process such data within a reasonable elapsed time [1-2]. Thus, a challenging problem is being faced in terms of hardware capabilities for the traditional algorithms where the training time for the algorithm becomes very lengthy which would significantly decreases the efficiency of the system design optimization process. To compensate the requirements of today's data analytics, many methods and tools have been lately established. Various systems have been developed mainly by the industry to support Big Data analysis, including Google's Map Reduce, Yahoo's PNUTS, Microsoft's SCOPE, and Walmart Labs Muppet. Each of these tools possess a unique algorithm for classification, clustering, and collaborative filtering, dimension reduction, miscellaneous approximations. Recently, a proposal has been put forward for a new machine learning tool based on Symbolic Regression (SR): Eureqa for Big Data applications [3-10]. Eureqa is a breakthrough technology that uncovers the intrinsic relationships hidden within complex data. Traditional machine learning techniques like neural networks and regression machines are capable tools for prediction, but become impractical when "solving the problem" involves understanding how you arrive at the answer. Eureqa uses a machine learning technique called Symbolic Regression to unravel the intrinsic relationships in data and describe them in simple mathematical equations. By using Symbolic Regression (SR), Eureqa can create incredibly accurate predictions that are easily explained and shared with others.

SR is a type of regression analysis that performs a search within mathematical expressions domain to create a model that fits to a given dataset. Commonly in the starting point of a search, there is not a particular model, not unless one is given by the user. Instead, initial expressions are formed by randomly combining mathematical blocks such as mathematical operators, analytic functions, constants, and variables. New equations are then formed by recombining previous equations, using genetic programming. Since SR does not require a specific model, it is not affected by users or unknown gaps in problem domain. The software uses

Submitted On: December 30, 2016
Accepted On: March 19, 2017

1054-4887 © ACES

evolutionary search to determine mathematical equations that describe sets of data in their simplest form. This class of algorithms are based on Darwinian Theory of evolution and one of its main attributes is that there is no calculated single solution, but a class of possible solutions at once. This class of possible and acceptable solutions is called “population”. Members of this population are called “individuals” and mathematically said, they represent possible solution, i.e., solution which can be realized in real world application. Main aim of evolutionary algorithms is to find the best solution of all during evolutionary process. Evolutionary algorithms differ among themselves in many points of view like individual representation (binary, decimal) or offspring creation (standard crossover, arithmetic operations, vector operations, etc.).

In this work, the aim is briefly to use Symbolic Regression SR in processing Big Data to obtain a full - wave Simulation-based analytical expression for the characteristic impedance Z_0 of a microstrip line in terms of only its real parameters to the contrary of the counterpart analytical expressions which require additional calculations for an effective dielectric constant ε_{eff} for different w/h ratios. In fact, there have been many different proposed equation sets for the calculation of Z_0. The most commonly used expressions are proposed by Pozar [11] and Balanis [12]. However, in all of these expressions, Z_0 is calculated using an effective dielectric permittivity ε_{eff} through the helping relationships obtained with the empirical methods.

To obtain “Big Data”, an efficient and fast 3D simulation tool is needed for gathering the required data for training and validation in the modelling process. Sonnet's suites [13] of high-frequency electromagnetic (EM) software are aimed at today's demanding design challenges involving predominantly planar (3D planar) circuits and antennas such as microstrip, stripline, coplanar waveguide, multi/single layer PCB and combinations with vias, vertical metal sheets (z-directed strips), and any number of layers of metal traces embedded in stratified dielectric material. By using Sonnet, it is possible to obtain a high number of samples for modelling the Z_0 of a given substrate (h, ε_r) for different values of the lines’ widths in a very short period of time. After obtaining the required data from Sonnet's suites, the data were inputted into the SR in the “Eureqa” environment for modelling characteristic impedance of microstrip lines.

In the study case section, all the data obtained by using Sonnet will be applied to the expressions given by [13-14] and SR of the Eureqa, respectively. Then, their performances are compared according to four commonly used error metrics, which are Mean Absolute Error (MAE), Relative Mean Absolute Error (RMAE), Maximum Error within the whole test data (MXE) and Root Mean Squared Error (RMSE). Furthermore, comparison between the approaches is also presented as figures for different values for h, ε_r and w within 1 GHz-10 GHz bandwidth. The flowchart of the methodology for obtaining a full-wave simulation based analytical expression for Z_0 is presented in Fig. 1. The paper is organized as follows: In Section 2, a brief explanation about Z_0 and the current mostly used analytical expressions in literature are presented. The Section 3 presents a general overview of the electromagnetic simulations and the obtained data. In Section 4, a brief explanation about working mechanism of SR and some example of its usage in science and engineering fields is presented alongside of the methodology that is implemented to forming the full - wave Simulation-based analytical expression for Z_0. Section 5 gives results of the study case. Finally, the paper ends with the conclusions in the Section 6.

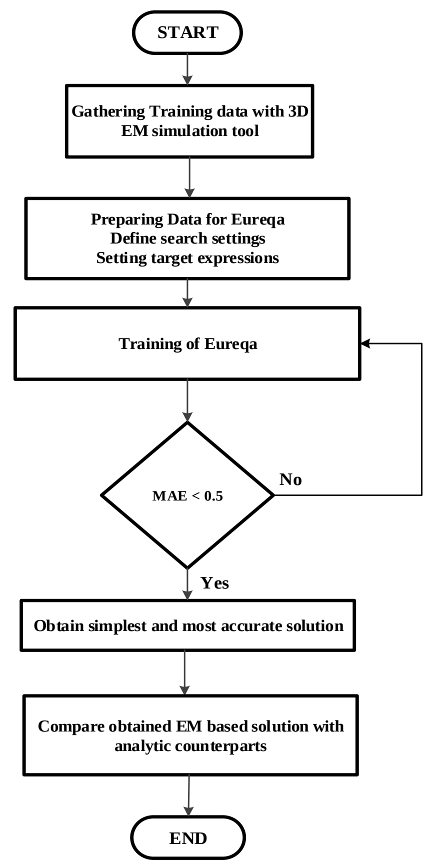


Fig. 1. Flowchart of the proposed methodology.

II. ALTERNATIVE FORMULATION OF THE CHARACTERISTIC IMPEDANCE OF MICROSTRIP LINES

Characteristic impedance is an important design parameter in microstrip-based microwave circuit designs such as filter, antenna, power divider, coupler, oscillator, and amplifier applications. Also, it is essential in the design of interdigital capacitance and stub inductance.

There are many analytical expressions for the calculation of microstrip transmission lines of Z_0. The well-known equations which are credited to Pozar [11] and Balanis [12] are the most commonly used expressions. These expressions require pre-calculation of ε_{eff} and some sub-equations, where the main equation of Z_0 changes depending on the ratio of transmission lines widths (w) and substrate's height (h). In [11], the solution only depends on the pre-calculation of ε_{eff} and the condition of w/h ratio, the effect of the frequency is not taken into account, either. On the other hand, in addition to the pre-calculation of ε_{eff} and the condition of w/h ratio, the effect of the transmission line thickness T and the frequency were taken into the account in [12]. In our study, the expressions in [11] and [12] are presented in 4 different cases for benchmarking purposes.

Case 1

The analytical expression in [11] is given in Eqs. (1-3), where the expression depends on the ratio of w/h and requires the pre-calculation of Eq. (1). For the sake of simplicity, Eqs. (1-3) will hereafter be named Case 1:

$$\varepsilon_{eff} = \frac{\varepsilon_r+1}{2}+\frac{\varepsilon_r-1}{2}\left(\left(1+12\frac{h}{w}\right)^{-\frac{1}{2}}\right). \tag{1}$$

For $w/h \le 1$,

$$Z_0 = \frac{60}{\sqrt{\varepsilon_{eff}}}\ln\left(\frac{8h}{w}+\frac{w}{4h}\right). \tag{2}$$

For $w/h \ge 1$,

$$Z_0 = \frac{120\pi}{\sqrt{\varepsilon_{eff}}\left(\mathrm{w/h}+1.393+0.667\ln\left(\mathrm{w/h}+1.444\right)\right)}. \tag{3}$$

Case 2

In [12], a more detailed expression set is provided for analytical calculation of Z_0. Firstly, all the expressions are solved for f=10 GHz conditions, then the frequency-dependent expression is calculated by using previous values as sub-expressions. In Eqs. (4-5), the effect of the transmission line thickness is considered in the calculation of the effective width value of microstrip transmission line for f=10 GHz. Hereafter, Eqs. (4-12) will be named Case 2:

for $w/h \le 1$,

$$\frac{w_{eff}(0)_{f=0}}{h} = \frac{w}{h}+\frac{1.25}{\pi}\frac{T}{h}\left(1+\ln\left(\frac{2h}{T}\right)\right). \tag{4}$$

For $w/h \ge 1$,

$$\frac{w_{eff}(0)_{f=0}}{h} = \frac{w}{h}+\frac{1.25}{\pi}\frac{T}{h}\left(1+\ln\left(\frac{4\pi w}{T}\right)\right). \tag{5}$$

After calculating the effective width value of the transmission line, Z_0 (f=0) can be obtained depending on the ratio of $w_{eff}(0)/h$ as follows:

For $\frac{w_{eff}(0)}{h} \le 1$,

$$\varepsilon_{eff}(0) = \frac{\varepsilon_r+1}{2}+\frac{\varepsilon_r-1}{2}\left(\left(1+12\frac{h}{w_{eff}(0)}\right)^{-\frac{1}{2}}+0.04\left(1-\frac{w_{eff}(0)}{h}\right)^2\right), \tag{6}$$

$$Z_0(0)_{f=0} = \frac{60}{\sqrt{\varepsilon_{eff}(0)}}\ln\left(\frac{8h}{w_{eff}(0)}+\frac{w_{eff}(0)}{4h}\right), \tag{7}$$

For $\frac{w_{eff}(0)}{h} \ge 1$,

$$\varepsilon_{eff}(0) = \frac{\varepsilon_r+1}{2}+\frac{\varepsilon_r-1}{2}\left(1+12\frac{h}{w_{eff}(0)}\right)^{-\frac{1}{2}}, \tag{8}$$

$$Z_0(0) = \frac{\dfrac{120\pi}{\sqrt{\varepsilon_{eff}(0)}}}{\dfrac{w_{eff}(0)}{h}+1.393+0.677\ln\left(\dfrac{w_{eff}(0)}{h}+1.444\right)}. \tag{9}$$

Finally, the effect of the frequency is taken into the account in Eq. (12) by using sub-equations given in Eqs. (10-11):

$$\varepsilon_{eff}(f) = \varepsilon_r - \frac{\varepsilon_r-\varepsilon_{eff}(0)}{1+\dfrac{\varepsilon_{eff}(0)}{\varepsilon_r}\left(\dfrac{f}{f_t}\right)^2}, \tag{10}$$

$$f_t = \frac{Z_0(0)}{2\mu_0 h}, \tag{11}$$

$$Z_0(f) = Z_0(0)\sqrt{\frac{\varepsilon_{eff}(0)}{\varepsilon_{eff}(f)}}. \tag{12}$$

Case 3

In examples given in [12], the T/h ratio is commonly considered equal to zero. In our study, we also ignored the effect of T/h ratio in Case 3 for the sake of better benchmarking results. Thus, Eqs. (4-5) are simplified as follows:

for $w/h \le 1$ and T/h=0,

$$\frac{w_{eff}(0)}{h} = \frac{w_{eff}(f=0)}{h} = \frac{w}{h}. \tag{13}$$

For $w/h \ge 1$ and T/h=0,

$$\frac{w_{eff}(0)}{h} = \frac{w_{eff}(f=0)}{h} = \frac{w}{h}. \tag{14}$$

In our studies, we have observed that when an EM-based simulation result is taken as the reference point, Case 2 can be improved by considering T/h ratio equal to zero. The results can be clearly seen in Figs. 4-6 and Tables 4 and 5.

Case 4

As it can be seen in Figs. 4-6, commonly, the result of Sonnet for Z_0 slightly increases with the frequency. However, in Cases 2 and 3, $Z_0(f)$ is inversely proportional with the frequency. Thus, by ignoring the frequency effect in Eq. (12), Case 4 can be expressed as follows where its performance is better than Cases 2 and 3:

$$Z_0(f) = Z_0(0) \quad for \quad \sqrt{\frac{\varepsilon_{eff}(0)}{\varepsilon_{eff}(f)}} = 1, \ \frac{T}{h} = 1. \qquad (15)$$

In Eqs. (1-15) the analytical expressions of two well-known references had been given in 4 different cases. Although these have admissible results, they require:

- Pre-calculation of ε_{eff} ;
- Condition of *w/h* ratio;
- Additional sub-equations.

In the next sections, by using Sonnet and Eureqa, a new EM based analytical expression for calculation of Z_0 is presented, which is independent from the ratio of *w/h* and does not require pre-calculation of ε_{eff} as well as being more simple and accurate than the counterpart analytical expressions given in Cases 1-4.

III. ELECTROMAGNETIC SIMULATIONS WITH SONNET'S SUITS

Sonnet's suites of high frequency EM software are aimed at today's demanding design challenges involving predominantly planar (3D planar) circuits [13]. The software requires a physical description of your circuit and employs a rigorous Method-of-Moments EM analysis based on Maxwell's equations that includes all parasitic, cross-coupling, enclosure and package resonance effects. Thus, it is a very effective tool for calculating the Z_0 of a microstrip lines.

In our study, a series of simulations are performed in Sonnet's suites with a range of parameters for the combination of each microstrip parameters such as substrate height *h*, dielectric permittivity ε_r, widths of the microstrip line *w* and frequency *f*. The values given in Tables 1 and 2 were simulated with Sonnet to obtain a high accuracy rate in the training process. Training data is given in Table 1 while Table 2 contains the test data including the most commonly used substrates (FR4, Rogers, Duroid). Tables 3-6 presents the variations of characteristic impedance Z_0 according to the microstrip parameters of width *w*, operation frequency *f*, substrate height *h* and the dielectric constant ε_r. As it can be observed from these tables, the most effective parameters can be ordered as: *w*, *h*, ε_r while the effect of operation frequency *f* can be ignored.

Some typical examples of the training data are given in Table 7. Due to the training features of Eureqa normalized with 50 ohm are used. Also, in the study case, all the data in Tables 1 and 2 were used for performance benchmarking of the equations given in Cases 1-4 and the expression obtained by Eureqa. The results of the benchmarking are given in Tables 8 and 9.

Table 1: Training data

Parameters	Value	Sample Number
h (mm)	0.3, 0.5, 0.7, 1, 1.5, 2, 2.5	7
ε_r	1, 2, 3, …., 6	6
w (mm)	0.2, 0.25, 0.3,…, 1, 1.2, 1.4, 1.6, 1.8, 2, 2.2, 2.5, 2.7, 3, 3.3, 3.5, 3.7, 4, 4.3, 4.5, 5	33
Thickness (µm)	35	1
f (GHz)	1, 2, 3, …., 10	10
Total samples	7x7x33x1x10=13860	

Table 2: Test data

Parameters	Duroid 5880	Duroid 6006	FR4	Roger 3003	Roger 3006
h (mm)	0.254 0.784 3.175	0.254 1.27 2.5	1.6	0.25 0.75 1.52	0.25 0.64 1.27
ε_r	2.2	6.15	4.6	3	6.15
w (mm)	0.2, 0.25, 0.3,…, 11.2, 1.4, 1.6, 1.8, 2, 2.2, 2.5, 2.7, 3, 3.3, 3.5, 3.7, 4, 4.3, 4.5, 5				
Thickness T (µm)	35				
f (GHz)	1, 2, 3, 4, 5, 6, 7, 8, 9, 10				
Sample number	990	990	330	990	990
Total samples	4290				

Table 3: Variation of Z0 due to the effect of microstrip width w for (h=1.5 mm, εr=3, f=5 GHz)

h (mm)	ε_r	*w* (mm)	*f* (GHz)	Z_0/50 (Sonnet)
1.5	3	0.2	5	3.417
1.5	3	0.3	5	3.058
1.5	3	0.5	5	2.614
1.5	3	1	5	2.024
1.5	3	1.6	5	1.639
1.5	3	3.5	5	1.058
1.5	3	5	5	0.834

Table 4: Variation of Z0 due to the frequency f for (h=1.5 mm, εr=3, w=1 mm)

h (mm)	ε_r	*w* (mm)	*f* (GHz)	Z_0/50 (Sonnet)
1.5	3	1	1	2.015
1.5	3	1	3	2.019
1.5	3	1	5	2.024
1.5	3	1	7	2.029
1.5	3	1	10	2.035

Table 5: Variation of Z0 due to the effect of substrate height h for (εr=3, w=1 mm, f=5 GHz)

h (mm)	ε_r	*w* (mm)	*f* (GHz)	$Z_0/50$ (Sonnet)
0.3	3	1	5	0.836
0.5	3	1	5	1.164
0.7	3	1	5	1.410
1	3	1	5	1.691
1.5	3	1	5	2.024
2	3	1	5	2.263
2.5	3	1	5	2.446

Table 6: Variation of Z0 due to the effect of substrate dielectric constant εr for (h=1.5 mm, w=1 mm, f=5 GHz)

h (mm)	ε_r	*w* (mm)	*f* (GHz)	$Z_0/50$ (Sonnet)
1.5	1	1	5	2.940
1.5	2	1	5	2.359
1.5	3	1	5	2.024
1.5	4	1	5	1.801
1.5	5	1	5	1.637
1.5	6	1	5	1.511

Table 7: Example of the data input into Eureqa

ID	*f* (GHz)	*h* (mm)	ε_r	*w* (mm)	Z_0 (Sonnet)	$Z_0/50$
1	1	0.3	1	0.4	111.729	2.235
2	1	1.5	5	1.4	70.272	1.405
3	1	2.5	6.15	3	56.5	1.130
4	2	1	5	3.5	32.120	0.642
5	3	3.175	2.2	2	112.75	2.255
6	4	2.5	6	5	43.321	0.866
7	6	0.75	3	1	73.2	1.464
8	7	0.5	2	0.75	81.149	1.623

IV. OBTAINING THE EM-SIMULATION BASED ANALYTICAL EXPRESSION WITH SYMBOLIC REGRESSION

In this section, firstly, a brief introduction about SR and its usage in various applications in science and engineering are provided. After that, the implementation of the SR's training process for obtaining the EM-based analytical expression of Z_0 is presented.

A. Symbolic Regression with Eureqa

As it has been mentioned before, Eureqa [14] is a breakthrough technology that uncovers the intrinsic relationships hidden within complex data using SR. As referenced in [15], Schmidt and Lipson developed a computing system that could observe a phenomenon in nature and then automatically identify various laws of nature and invariant equations that it obeyed. For example, they had the system to observe a double pendulum swinging chaotically using motion-tracking. Without any knowledge of physics or geometry, the system identified the exact energy conservation and momentum relations that governed its dynamics. In [16], Ceperic, Bako and Baric presented the use of Symbolic Regression for the black-box modelling of non-linear dynamic behaviour in the AC/DC rectifiers of radio-frequency identification (RFID) circuits. Also, there are many recent publications in the literature which used Eureqa's SR algorithm in order to solve or obtain solutions for various problems in astronomy, biology, chemistry, computer science, physics and many other fields [15-17].

Symbolic Regression is based on evolutionary algorithms. This type of algorithms are inspired from Darwinian theory of evolution. In evolutionary algorithms there is no calculated single solution. Instead, there is a class of possible solutions called as "population" at once. Individuals of this population represents the possible solution of real world problem. Main aim of these type of algorithms is to find the optimal solution during evolutionary process.

SR is a process in which the given dataset fitted by suitable mathematical formulas. This process is commonly used when some data of unknown process are obtained. For the first time, the idea of how to solve various problems using SR by means of evolutionary algorithms (EAs), come from Koza who used genetic algorithm (GA) in so called genetic programming (GP) [15-16]. GP is basically SR which is done by evolutionary algorithms. Main principle of GP is such that expressions are represented in chromosomes like syntactic trees. The syntactic tree form of Eq. (16) is given in Fig. 2 for a better understanding. Based on GA principles, new individuals (Childs) whose representations are in fact new expressions which are evaluated by fitness are created either by random generators or by exchanging parent's parts by crossover operators, or mutation. This process can be simply seen in Fig. 3, where Eqs. (16-17) are taken as parents and Eqs. (18-20) are the children of these parents through cross over or mutation operations:

$$P_1 = f(a,b,c,d,e) = \cos(ab) + \sqrt{c-d}, \tag{16}$$

$$P_2 = f(a,b,c,d,e) = a^2 + \frac{c}{b-d} - 12e, \tag{17}$$

$$C_1 = \cos(ab) - 12e, \tag{18}$$

$$C_2 = a^2 + \frac{c}{b-d} + \sqrt{c-d}, \tag{19}$$

$$C_3 = \sin\left(\frac{a}{b}\right) + \sqrt{c-d}. \tag{20}$$

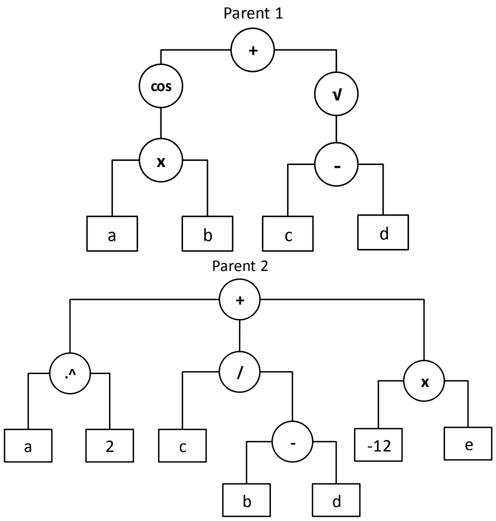


Fig. 2. Examples for Syntactic Tree.

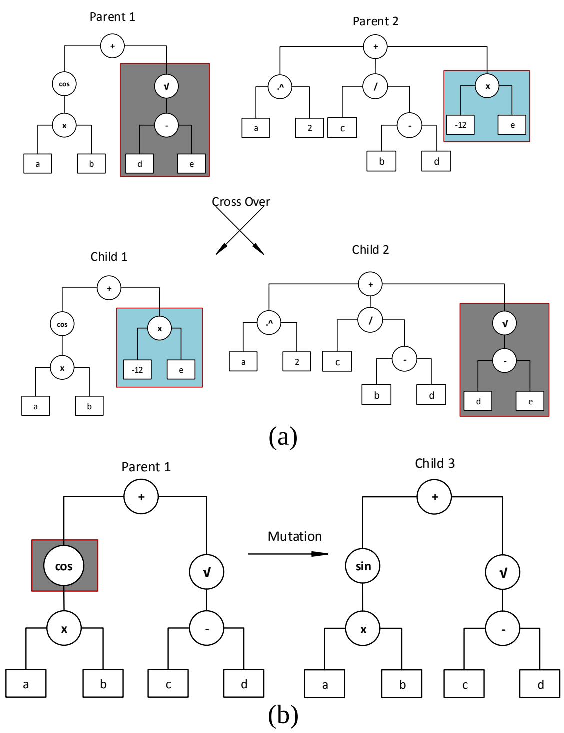


Fig. 3. An example of (a) crossover and (b) mutation in syntactic tree form.

B. Implementation process in Eureqa for Z_0

As mentioned before, SR is a process in which the given dataset is fitted by suitable mathematical formulas. For this mean, firstly the data in Tables 1 and 2 were input into Eureqa as the training data with the following commands to start the search settings for creating a fitted mathematical /analytical expression of Z_0:

- Shuffle and split data points equally for training and test purposes;
  - Use all basic formula blocks;
  - Use all trigonometry blocks;
  - Use all exponential blocks;
  - Use all squashing blocks;
  - Use mean absolute error as error metric (default).
  - Target expression is expressed as follows:

$$Z_0 = f\left(h, \varepsilon_r, w, f\right). \quad (21)$$

Unfortunately, Eureqa did not provide an appropriate expression as it was expected. By checking the results of Sonnet given in Table 4, one can observed that the effect of frequency can be ignored as compared with the effects of *h*, ε_r and *w*. Thus, frequency is discarded from the target expression in Eq. (21) and some changes are made in Eureqa's searching commands as follows:

- Treat all data equally both for training and test;
- Use all basic formula blocks;
- Use all trigonometry blocks;
- Use all exponential blocks;
- Use only Gaussian from squashing blocks;
- Use Mean Absolute Error MAE as error metric (default).
- In order to increase the complexity of the predicted expression user can define the target expression as the sum of two or more functions in case of need. Here Z0 is defined as the sum of the two ingredients as follows as target expression:

$$Z_0 = f_0\left(h, \varepsilon_r, w\right) + f_1\left(h, \varepsilon_r, w\right). \quad (22)$$

Thus, by using SR algorithm Eureqa builds up an expression consisting of summation of the two different functions as given in Eq. (22), which is expected to give more accurate results than a single function. The training stage of Eureqa was performed by the Intel Core i5 CPU, 3.3 GHz Processor, 4 GB RAM. After approximately 10 hours of training, the MAE value dropped by less than 0.5, which was chosen as the target error value, therefore, we stopped the training process. Eureqa sorts the candidate's expression according to their complexity and fitness value of expressions. Eureqa's sorting algorithm suggests that the expression given in Eq. (23) is the optimal expressions:

$$Z_0 = 3.5 - 0.725\varepsilon_r + \frac{112.5}{w/h + 2.47} + \frac{300}{1 + 0.217\varepsilon_r + (\varepsilon_r + 0.57)(w/h) + 0.87e^{-\left(0.22 + \frac{0.75}{\varepsilon_r + w} + \frac{0.09}{w/h}\right)^2}}. \quad (23)$$

V. CASE STUDY

For the case study purposes, in Table 9, performance benchmarking of each analytic expression are presented for the data examples given in Table 7. Also, approximately 20.000 samples given in Tables 1 and 2 were taken as testing instances for each of the Cases 1-4 and the solution in Eq. (12). Performance results given in Table 5 were calculated by using the following error metrics:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|T_i - P_i\right|, \tag{24}$$

$$RMAE = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|T_i - P_i\right|}{\left|T_i\right|}, \tag{25}$$

$$MXE = \max\left(\left|T_i - P_i\right|\right), \tag{26}$$

$$RMSE = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\left(T_i - P_i\right)^2}, \tag{27}$$

where, *T*: Target, *P*: Predicted, *N*: total sample.

Table 8: Z_0 results for samples in Table 3

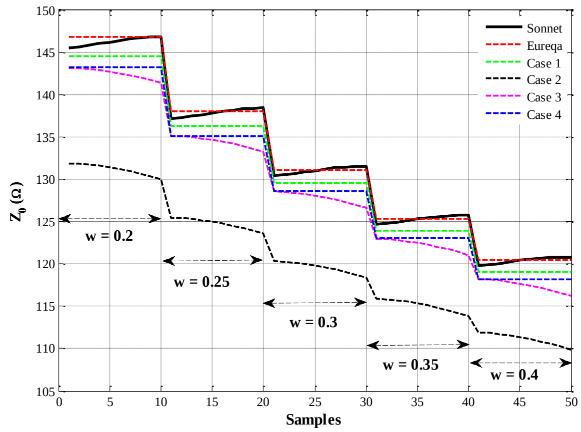
ID	Z_0	Eureqa	Case 1	Case 2	Case 3	Case 4
1	111.7	112.3	110.6	103.9	110.6	110.6
2	70.2	70.2	70.4	67.6	69.3	69.3
3	56.5	56.5	55.7	54.8	55.6	55.7
4	32.1	32.1	31.8	31.3	31.7	31.8
5	112.7	115.6	116.3	113.8	115.7	116.2
6	43.3	42.8	42.3	40.5	40.9	42.3
7	73.2	73.1	72.6	70.1	72.4	72.6
8	81.1	80.9	80.8	77.4	80.7	80.8

Table 9: Performance benchmarking of Cases 1-4 and Eureqa

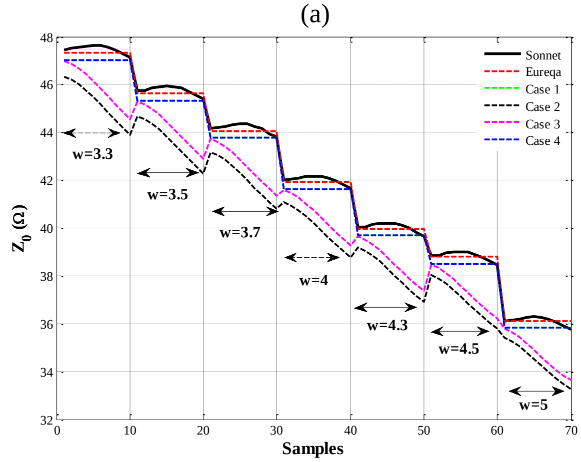
	MAE	RMAE	MXE	RMSE
Eureqa	0.4942	0.0078	7.1105	0.0061
Case 1	1.0197	0.0117	8.3851	0.0116
Case 2	4.2904	0.0513	24.292	0.0404
Case 3	1.6788	0.0214	14.642	0.0178
Case 4	1.1455	0.0128	9.9481	0.0127

Furthermore, in Figs. 4-6, benchmarking results of the 4 cases and SR applied to the 3 different substrates are given as compared to targeted values obtained with Sonnet within 1 GHz-10 GHz bandwidth for 11 different w values. Here the test data given in the Table 2 is used. Figure 4 depicts all the Z_0 results for FR4, while Fig. 5 gives the results for Duroid 6006 and Fig. 6 shows the results for Roger 3003. In order to depict all the results of the different w values for the different cases within 1-10 GHz bandwidth in a single plane, all the results are compressed in x-axis for each value of w. On the x-axis, each sample in the w-band is taken with 1 GHz intervals using Sonnet suit.

As it can be seen from the figures, the values of Z_0 obtained by Sonnet slightly increase with the frequency on the contrary to Case 3 and 4. As a consequence, the proposed EM-based analytical expression for Z_0 is better than the other expressions. Also, it should be noted that the proposed expression is not only better in performance benchmarking, but it is also, simpler than the other expressions because it does not require pre-calculation of ε_{eff} as well as it does not require different expressions for the ratio of the *w/h*.



(a)



(b)

Fig. 4. Z_0 benchmarking results of FR4: (a) *w*=[0.2 0.25 0.3 0.35 0.4], and (b) *w*=[3.3 3.5 3.7 4 4.3 4.5 5], *h*=0.64 mm.

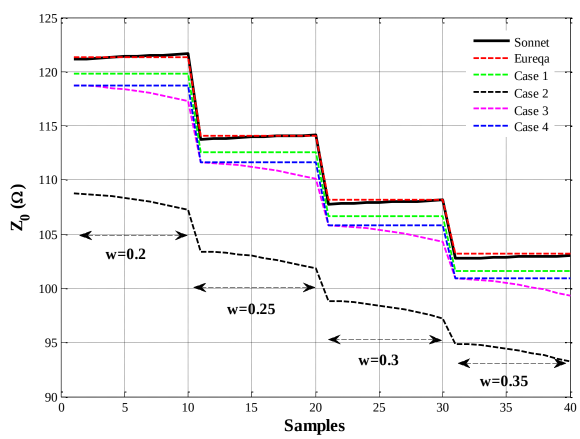


Fig. 5. Z_0 benchmarking results of Duroid 6006 (h=1.52 mm) *w*=[0.2 0.25 0.3 0.35].

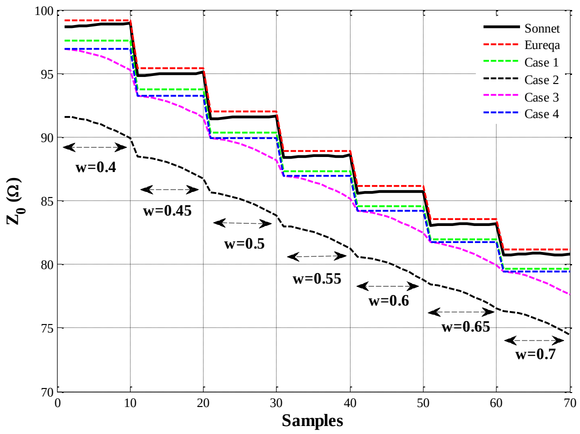


Fig. 6. Z_0 benchmarking results of Roger 3003 (h=0.64 mm) *w*=[0.4 0.45 0.5 0.55 0.6 0.65 0.7].

VI. CONCLUSION

In this study, by using 3D EM simulation and Symbolic Regression tools, an EM-based analytical expression for the characteristic impedance Z_0 of a microstrip transmission line is obtained and compared with the other analytical expressions in literature for both simulation and measurement results. The obtained expression is simpler and does not require additional calculation of ε_{eff} and does not require to use different expressions with respect to the ratio of the height to the width of the microstrip transmission line.

Furthermore, the proposed expression can easily be deployed on similar microstrip-based microwave circuit designs such as filter, amplifier, and antenna designs to acquire more accurate results, where the precise value of Z_0 is an important intrinsic design parameter.

Also, all the data obtained by Sonnet in Tables 1 and 2 are shared with detailed descriptions about the data set in [18] for all the readers who are interested in this field and would like to make contributions to this subject.

In future studies, the analytical expressions for more complex design parameters such as gain, far field, near field, scattering parameters, extraction of LC parameters for interdigital capacitance and stub inductance will be studied.

ACKNOWLEDGMENT

We would like to express our special appreciation and gratitude to Sonnet Software, Inc. [13] and Nutonian, Inc. [14] for the researcher's licenses they provided for our use. With their kind support, we have managed to offer this study to the interest of readers.

REFERENCES

[1] C. Doulkeridis and K. Norvag, "A survey of large-scale analytical query processing in Map Reduce," *VLDB Journal*, vol. 23, no. 3, pp. 355-380, 2014.

[2] Big Data: A Parallel Particle Swarm Optimization-Back-Propagation Neural Network Algorithm Based on Map Reduce.

[3] M. F. Korns, "Extremely accurate symbolic regression for large feature problems," Genetic Programming Theory and Practice XII, Genetic and Evolutionary Computation, pp. 109-131, 2015.

[4] N. Allgaier and R. McDevitt, *Reverse Engineering the Brain with Eureqa*. UVM, 2011.

[5] I. Icke and C. Joshua, "Improving genetic programming based symbolic regression using deterministic machine learning," *Bongard. Congress on Evolutionary Computation (CEC), 2013 IEEE*, pp. 1763-1770, 2013.

[6] E. Magalhaes, et al., "Experimental-based subsystem models for simulation of heterogeneous optical networks," *J. Microw. Optoelectron. Electromagn. Appl*. [online], vol. 13, no. 2, [cited 2016-10-04], pp. 197-213, 2014.

[7] F. Renner, A. Schwab, R.-P. Kapsch, Ch. Makowski, and D. Jannek, "An approach to an accurate determination of the energy spectrum of high-energy electron beams using magnetic spectrometry," *JINST*, vol. 9, p. P03004, 2014.

[8] G. D. Rodriguez, I. Velásquez, D. Cachi, and D. Inga, "Surrogate modeling with genetic programming applied to satellite communication and ground stations," *Aerospace Conference, 2012 IEEE*, Big Sky, MT, pp. 1-8, 2012. DOI: 10.1109/AERO.2012.6187326

[9] S. T. Goh, S. A. Zekavat, and O. Abdelkhalik, "Space-based wireless solar power transfer via a network of LEO satellites: Doppler effect analysis," *Aerospace Conference, 2012 IEEE*, Big Sky, MT, pp. 1-7. 2012. DOI: 10.1109/AERO.2012.6187081

[10] H. Kharbanda, M. Krishnan, and R. H. Campbell, "Synergy: A middleware for energy conservation in mobile devices," *2012 IEEE International Conference on Cluster Computing*, Beijing, pp. 54-62, 2012. DOI: 10.1109/CLUSTER.2012.64

[11] D. M. Pozar, *Microwave Engineering*. 4th edition, Willey, 2011. ISBN 978-0-470-63155-3.

[12] C. A. Balanis, *Advanced Engineering Electromagnetics*. 2nd edition, John Wiley & Sons, 2012.

[13] www.sonnetsoftware.com

[14] www.nutonian.com

[15] M. Schmidt and H. Lipson, "Distilling free-form natural laws from experimental data," *Science*, vol. 324, no. 5923, 2009, pp. 81-85, 2009.

[16] V. Ceperic, N. Bako, A. Baric, "A symbolic regression-based modelling strategy of AC/DC rectifiers for RFID applications," *Expert Systems with Applications*, vol. 41, iss. 16, pp. 7061-7067, 2014.

[17] M. M. Hasan, A. Sharma, F. Johnson, G. Mariethoz, A. Seed, "Correcting bias in radar Z-R relationships due to uncertainty in point rain gauge

networks," *Journal of Hydrology*, 2014. DOI: http://dx.doi.org/10.1016/j.jhydrol.2014.09.060
[18] www.neuralnetworkdb.com

Peyman Mahouti received his Ph.D. degree in Electronics and Communication Engineering from the Yıldız Technical University in 2016. He has been currently working as a Teaching Assistant in the Yıldız Technical University Department of Electronic and Communication Engineering. The main research areas are optimization of microwave circuits, broadband matching circuits, and device modeling, computer-aided circuit design, and microwave amplifiers.

Filiz Güneş received her M.Sc. degree in Electronics and Communication Engineering from the Istanbul Technical University. She attained her Ph.D. degree in Communication Engineering from the Bradford University in 1979. She is currently a Full Professor in Yıldız Technical University. Her current research interests are in the areas of multivariable network theory, device modeling, computer-aided microwave circuit design, monolithic microwave integrated circuits, and antenna designs.

Mehmet Ali Belen received his Ph.D. degree in Electronics and Communication Engineering from Yıldız Technical University. His current research interests are in the areas of multivariable network theory, device modeling, computer aided microwave circuit design, monolithic microwave integrated circuits, and antenna arrays. He has been currently working as an Assistance Professor in the Artvin Çoruh University Department of Electric and Electronic Engineering.

Salih Demirel has received M.Sc. and Ph.D. degrees in Electronics and Communication Engineering from Yıldız Technical University, Istanbul, Turkey in 2006 and 2009, respectively. He has been currently working as an Associated Professor in the same department. His current research interests are among of microwave circuits especially optimization of microwave circuits, broadband matching circuits, device modeling, computer-aided circuit design, and microwave amplifiers.