Extraction of Piecewise-Linear Analog Circuit Models from Trained Neural Networks using Hidden Neuron Clustering

Simona Doboli
Computer Science Department
Hofstra University
Hempstead, NY, 11549
Email: cscszd@hofstra.edu

Gaurav Gothoskar and Alex Doboli
VLSI Systems Design Laboratory
Electrical and Computer Engineering Department
SUNY Stony Brook, Stony Brook, NY, 11794-2350
Email: {gaurav, adoboli}@ece.sunysb.edu

Abstract

This paper presents a new technique for automatically creating analog circuit models. The method extracts - from trained neural networks - piecewise linear models expressing the linear dependencies between circuit performances and design parameters. The paper illustrates the technique for an OTA circuit for which models for gain and bandwidth were automatically generated. The extracted models have a simple form that accurately fits the sampled points and the behavior of the trained neural networks. These models are useful for fast simulation of systems with non-linear behavior and performances.

1. Introduction

The need for mixed analog-digital designs is predicted to dramatically increase over the next years [3]. There is a lack of systematic design methods and efficient general-purpose synthesis tools for analog circuits [3]. As a result, analog designs continue to seize a considerable portion of the total design time for mixed-signal systems [3]. This paper presents a new analog circuit modeling method that can be efficiently used for both circuit simulation and synthesis.

Analog circuit models (macromodels) express mathematical relationships between *significant* electrical and geometrical parameters of a circuit (like device sizes, layout parasitics, signal frequencies, noise etc) and *specific* performance attributes (such as gain, bandwidth, power consumption, slew rate etc) [3]. Circuit models are very important for speeding up the convergence of simulation-based circuit synthesis tools. Most of the time, models are used to quickly find the performance attributes of the explored designs. Periodically, exhaustive circuit simulations are performed to correct the inaccuracies introduced by the models. It has been shown that performance estimation through a combined circuit model evaluation and circuit simulation offers good accuracy levels while significantly reducing synthesis time.

This paper presents a new technique for extracting piecewise linear models from trained neural networks. A model is a set of linear dependencies between circuit performances and design parameters. As experiments show, the produced piecewise linear models have a simple form that accurately fits the sampled points. Also, piecewise linear models are a promising method for approximating nonlinear behavior and performances with small errors [4]. Our work proposes a method to systematically create piecewise linear models used for simulation [4].

The model extraction method begins by training a neural network - using backpropagation algorithm - to a desired accuracy. Next, a pruning method is applied to eliminate the neurons and weights with insignificant contributions to the accuracy of the network. Then, the nonlinear activation function of each hidden neuron is approximated with a piecewise linear function [5] with a variable number of segments. The number of segments, its limits and the linear approximation on each segment are automatically determined by a clustering algorithm. Finally, the linear functions of the hidden neurons are composed to generate the piecewise linear functions of the model output. The regions were each linear output model is active are found by iteratively solving a linear system of inequalities and adjusting its limits. Previously [1], the activation function of hidden neurons was approximated with three linear segments for all neurons. Here, we introduce an adaptive linearization method - the clustering algorithm - that significantly improves the accuracy of the approximation.

2. Extraction method

The neural network considered here is a three layer feed-forward network with N input neurons, H hidden neu-

1530-1591/03 $17.00 © 2003 IEEE

rons and O output neurons. The activation function of the hidden neurons is the sigmoidal: $\phi(x) = \frac{1}{1+exp(-\lambda x)}$, with $0 < \lambda \leq 1$. The weighted sum input into a hidden neuron and into an output neuron are respectively: $h_j = \sum_{i=1}^{N} w_{ji} x_i, \quad h_k = \sum_{j=1}^{H} w_{kj} y_j$, where x_i is the output of input neuron i. The output of hidden neuron j is: $y_j = \phi(h_j)$, while the output neurons are linear: $y_k = h_k$.

The steps of the model extraction method are:

1. *Training and pruning the neural network.* A neural network is first trained and then pruned. In this process, the number of hidden neurons is chosen such that small training and testing errors are obtained.
2. *Linearization of the activation function of the hidden neurons.* The values of the activation function of each hidden neuron are first clustered into a number of sets, each corresponding to a linear segment. The clustering algorithm determines: the number of clusters, its limits and the equation of the linear segment for each cluster - the slope and intercept of the linear segment that passes through the limits of a cluster.
3. *Extraction of the linear models.* This step consists in finding the non-empty regions in the input space where combinations of linear regions in the hidden neurons are valid. For each such a region the linear model of the network output is expressed in terms of the network weights.

2.1. Clustering algorithm

The activation values of the hidden neurons are computed using all input data points. The clustering algorithm is a modified agglomerative technique [2]. A linear segment passing through each pair of consecutive output values is defined by its slope and intercept. Then the distance between two segments is computed as the cosine of their angle.

Initially, the number of clusters is equal to the number of segments between consecutive output points. Then, iteratively, the closest pair of clusters is merged until the following stopping criteria goes up: $J(t) = \frac{N_c(t)}{(N_i-1)} + \frac{1}{N_i}\sum_{n=1}^{N_i} |y_{lj}(x_n) - y_j(x_n)|$, where $N_c(t)$ is the number of clusters at step t, N_i is the initial number of points, $y_{lj}(x_n)$ is the linear output for input point x_n, $y_j(x_n)$ is the original sigmoidal output. The first term penalizes a large number of clusters, while the second term penalizes a large linearization error. Initially, the linearization error is zero and the penalty for a large number of clusters is one: $J(0) = 1$. As clusters are merged, the first term goes down, while the second term goes up. The linearization error dominates the sum, and at one point the values of $J(t)$ go up. This is when merging stops, and the final number of clusters is equal to the number of linear regions for hidden neuron j.

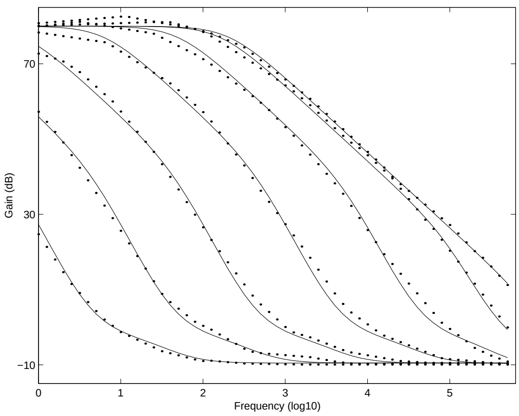

Figure 1. Gain frequency response.

3. Results

The extraction method is applied to the amplitude frequency response of an analog transconductance amplifier (OTA) at different parasitic levels. The data is obtained using SPICE simulations of the analog circuit. There are two inputs - frequency and parasitics - and one output - the gain.

A three layer neural network with $I = 2$ input and $H = 7$ hidden neurons is trained. Pruning eliminates eight weights from the initial number of 21 weights. Two of the hidden neurons have constant outputs given by the bias weight. The rest of the hidden neurons are clustered into 6, 9, 3 and 8 clusters respectively. Finally, 136 linear models were obtained. The result of the piecewise linear extraction method is presented in Figure 1. The dotted plots represent the piecewise linear model output, while the lines represent the true values. Each curve is for a different parasitics value. It can be seen that the piecewise linear approximation is very accurate. Similar results were obtained for modeling an operational amplifier circuit, and a high-frequency OTA as well.

References

[1] S. Doboli, et al. Piecewise Linear Modeling of Analog Circuits based on Model Extraction from Trained Neural Networks. In Proc. of *BMAS*, 2002.

[2] R.O. Duda and P.E. Hart. *Pattern classification and scene analysis*, New York: Wiley, 1973.

[3] G. Gielen, R. Rutenbar, “Computer Aided Design of Analog and Mixed-signal Integrated Circuits”, *Proc. of IEEE*, vol 88, No 12, pp. 1825-1852, 2000.

[4] D. Leenaerts, W. van Bokhoven, “Piecewise Linear Modeling and Analysis”, *Kluwer*, 1998.

[5] R. Setiono, et al. Extraction of rules from artificial neural networks for nonlinear regression. *IEEE Trans. Neural Networks*, 13(3):564–577, 2002.