3 Optimisation and parameter extraction of circuit models

3.1 Executive summary

Optimised device models are important in the design of electronic devices for specific performance. They help the designer to predict the behaviour of the device in an analogue circuit. However, standard methods of optimisation do not lend themselves to fast computation and may present problems with convergence. The simulated annealing, genetic and structured genetic algorithms are alternative optimisation methods that help to solve the convergence problems by avoiding entrapment in local minima of the solution space. These methods are used for the extraction of the parameters of the equivalent circuit model of the device or to construct its neural network model. The neural network models are black box models that determine device outputs from known inputs after appropriate weighting values have been calculated. Another method of parameter extraction is the semi-analytical procedure that uses an analytic approach together with empirical optimisation methods. Basic expressions and approximations to extract small-signal equivalent circuit parameters are developed so that accurate device models can be obtained.

3.2 Optimisation of device models

Accurate device models are developed to predict behaviour that is in good agreement with experimental observations. The optimisation of the model parameters can be considered to be the curve-fitting of the computed device characteristics to experimental data. Traditional gradient methods are computationally intensive and there are problems with convergence and entrapment in local minima. Optimisation only involves the "trial" of a number of initial solutions to minimise the error.

Combinatorial optimisation overcomes this problem of entrapment in local minima. To set up the device modelling problem as a combinatorial optimisation problem, the limits on each parameter are specified. Parameter values are then discretised so that a large but finite number of solutions is possible. An objective function, also referred to as the *cost function* is given by

$$F(V_{\mathrm{i}}) = \sum_{j=1}^{m} (M_{\mathrm{j}} - M'_{\mathrm{j}})^2, \tag{3.1}$$

where V_i ($i = 1, 2, \ldots, n$; $n =$ number of parameters) = the model parameters to be determined; M_j = a measured characteristic for the jth data point; M_j' = a calculated characteristic which is a function of the parameters V_i for the jth data point; and $m =$ the total number of characteristics to be fitted. Note that $F(V)$ is a least square difference function. The number of possible solutions is extremely large and an exhaustive search for the optimum solution is practically impossible. Hence, it is necessary to employ a heuristic method. Such methods are well-established in CAD tools since they enable the designer to find a feasible solution in a finite period of time. A typical heuristic optimisation process utilises an iterative improvement strategy. At each step of the iteration, the algorithm generates a new solution and tests if it reduces the value of the objective function. If so, it accepts the new solution. If not, another new solution is generated and tested and eventually the global minimum of $F(V)$ in Equation (3.1) should be reached. However, the size of the solution space, as defined by all possible and physically meaningful parameters, is very large and grows exponentially with the number of variables in the model. An exhaustive search for the best solution cannot be performed in a finite period of time.

3.3 Simulated annealing

Most heuristic algorithms search for a solution only in the directions that improve the objective function. This type of heuristic search has a major drawback: it can be easily trapped into the local minima of an objective function. Figure 3.1 demonstrates the problem.

The curve shown in Figure 3.1 is assumed to be the objective function of an iterative improvement process; the circles indicate the costs, i.e. $F(V_i)$ in the least squares objective function given in Equation (3.1) calculated from certain parameter sets. Since

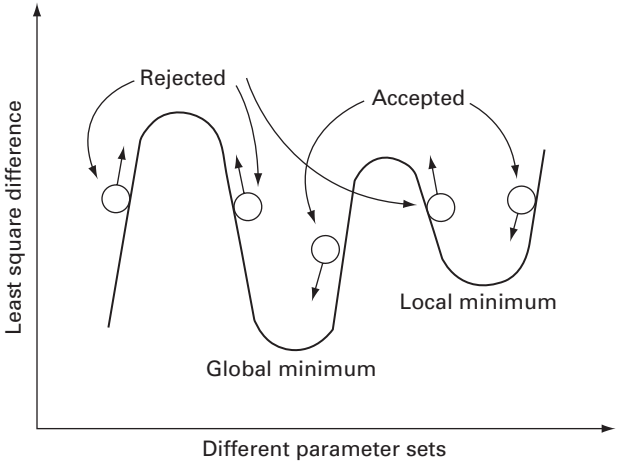

Fig. 3.1 Local minimum trapping in iterative optimisation.

a new set of parameters is generated by introducing small modifications to the model, its corresponding location on the curve is most likely to be somewhere near that of the original configuration. The traditional iterative improvement algorithms only accept parameter sets that have reduced the cost. This criterion of set acceptance implies that the process can only go downhill into a minimum and any uphill movement is forbidden. Depending on the starting point of the search, it is possible that this minimum is only a local minimum of the objective function. This search process generally does not have the capability of climbing over a peak of the curve to reach the global minimum.

The drawback of this process can be overcome by using simulated annealing (SA)[12, 22, 23, 30]. This is a method of finding a near optimal solution for combinatorial optimisation problems. The SA algorithm has the advantage of asymptotically producing the global optimal solution with a probability of one. This is achieved by making the following important modification to conventional heuristic methods. A cost-increasing solution may still be accepted. The probability of acceptance depends on: (1) a parameter called a *pseudo-temperature* T_{k}, which is artificially decreased as the iteration proceeds, and (2) the value of

$$\Delta F(V) = F(V_{\mathrm{k}}) - F(V_{\mathrm{k-i}}) \tag{3.2}$$

where V_{k} and $V_{\mathrm{k}-1}$ are the values of the parameter vector V at iteration steps k and $k-1$. Specifically, a law similar to Boltzmann statistics is used to determine the probability P of accepting a certain cost-increasing solution V_{k} at the kth iteration step. The probability function P is given by

$$P(V_{\mathrm{k}}, T_{\mathrm{k}}) = \exp\left[\frac{-\Delta F(V_{\mathrm{k}})}{T_{\mathrm{k}}}\right]. \tag{3.3}$$

A careful choice of the initial pseudo-temperature T_0 and a rule for decreasing the pseudo-temperature T are necessary to save computation time while being able to escape from the local minima. The temperatures are related by the equation

$$T_{\mathrm{k}} = \alpha T_{\mathrm{k}-1}, \tag{3.4}$$

where α is a constant between 0.8 and 0.95. Its value can be gradually increased from the lowest to the highest value. $T_0 \geq 500$ is a satisfactory choice for device modelling.

At each iteration, new parameter values are generated by first choosing one device parameter V_{i} at random. A user-defined base value $V_{\mathrm{i-base}}$ is multiplied by a random number R, such that $0 \leq R \leq 1$, and a variation $\Delta V_{\mathrm{i}} = R V_{\mathrm{i-base}}$ is introduced into the parameter V_{i}. The new parameter value so obtained is used in the next iteration unless it exceeds prescribed limits, in which case it will be set to the maximum or minimum allowable value.

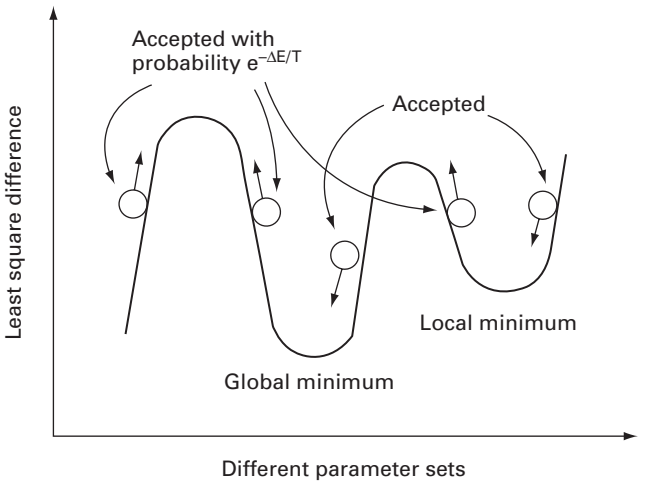

Fig. 3.2 Hill-climbing capability of SA.

A relative stopping criterion is used, since there is no guarantee that the device model can approximate experimental data closely. The optimisation process is stopped when the value of the objective function has remained virtually unchanged for several consecutive iterations (e.g. $\Delta F \leq 0.001$ for 10 consecutive iterations).

Figure 3.2 shows the same objective function as in Figure 3.1 with the hill-climbing capability of SA.

Example: Application of SA to the modelling of a HEMT

Optimisation by SA has been applied to the HEMT [30]. Three test cases are given as examples to demonstrate the optimisation process and evaluate its performance. Parameters are allowed to vary within ±90% of their initial values in test cases A and B. In test case C, the limits are reduced to ±20% for realistic parameters.

Test case A: AC Model of a HEMT

The unilateral power gain U of the HEMT (also known as the MODFET as explained in Chapter 2) was determined by Roblin *et al.* [21]. The values of U are sampled at various frequencies and used as the measured characteristic M_j to be matched to the model. The Mason unilateral power gain (defined in Chapter 5) is expressed in terms of the admittance (Y) parameters as

$$U = \frac{|\,(Y_{21} - Y_{12})\,|^2}{4[\mathrm{Re}(Y_{11})\mathrm{Re}(Y_{22}) - \mathrm{Re}(Y_{12})\mathrm{Re}(Y_{21})]}. \tag{3.5}$$

The Y-parameters are determined by primitive model parameters, i.e. the gate capacitance C_0, the gate length L, the gate width Z, the bias voltage $V = V_{GS} - V_T$ and the parameter k which is given by

$$K = \frac{V_{DS}}{(V_{GS} - V_T)}, \tag{3.6}$$

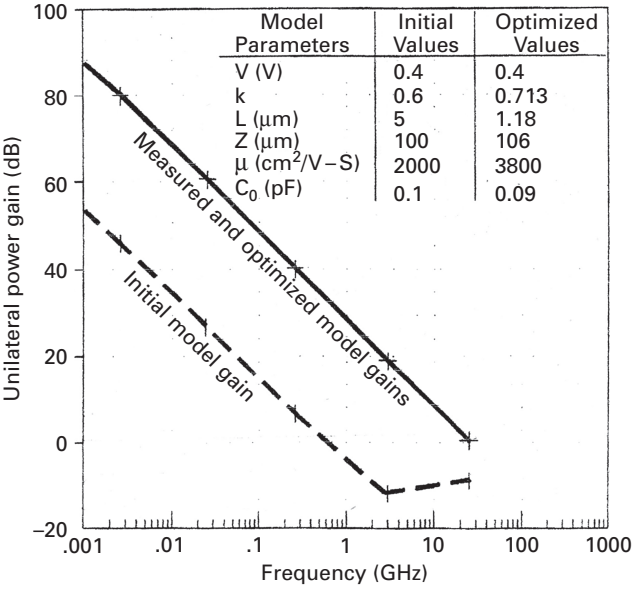

Fig. 3.3 Test case A: unilateral power gain curves and optimisation results [+, measured; solid line, optimised model; dotted line, initial solution] (M-K. Vai, S. Prasad, N. C. Li and F. Kai, *IEEE Transactions on Electron Devices*, Vol. 36, No. 4, pp. 761–762, April 1989. ©1989 IEEE).

where V_{DS} is the drain to source voltage, V_{GS} is the gate to source voltage and V_T is the threshold voltage.

Figure 3.3 compares the measured gain curve with curves generated from an initial model and the final optimised models. The objective function is reduced from an initial value of 4421.83 to 0.161095, virtually a perfect match.

Test case B: Equivalent circuit of a HEMT

The small-signal equivalent circuit of an intrinsic HEMT is used to deduce the unilateral power gain using Equation (3.5). The Y-parameters are expressed in terms of the circuit elements R_i, C_{GS}, R_{DS}, C_{DS}, g_m and τ. The objective function is reduced from 1783.78 to 0.01789 when optimisation is completed. Figure 3.4 shows the agreement between measured and computed values using the optimised model.

Test case C: Device design parameters

In this test, a set of primitive device parameters (gate length, gate width, bias voltage and mobility) are obtained by optimisation such that the device has the highest value of f_{max}, the maximum frequency of oscillation. The analytical expression for f_{max} has been obtained by Roblin *et al.* [21] by setting the unilateral gain equal to 1. The primitive device parameters are related to f_{max} and the objective function is formulated. The initial design gives a value of f_{max} equal to 29.57 GHz. The optimised HEMT has f_{max} equal to 165.81 GHz. Such an optimisation is an aid to the device designer. Table 3.1 gives the optimised values. The initial value of K is 0.6 and the optimised value is 0.95.

Table 3.1 Test case C: optimised design parameters for the highest f_{max}

Model parameters	Initial values	Optimised values
V_T(V)	0.4	0.4
L(μm)	1	0.8
Z(μm)	250	300
Mobility $\mu(\text{cm}^2)/(\text{V-s})$	4000	4800
C_0(pF)	0.1	0.108

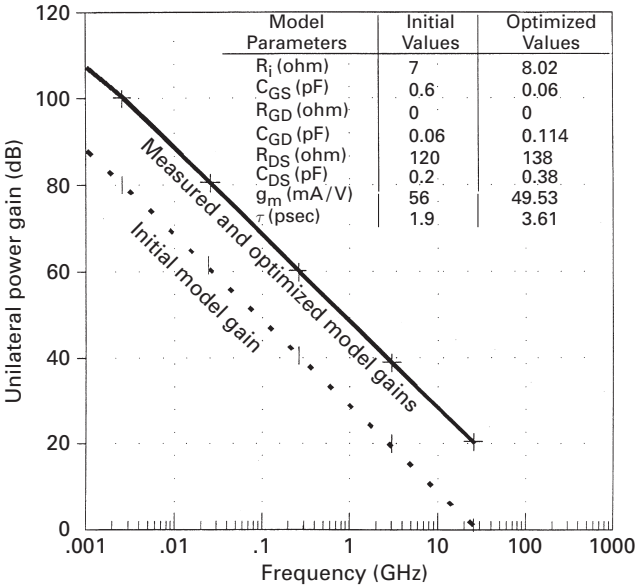

Fig. 3.4 Test case B: unilateral power gain curves and optimisation results [+, measured; solid line, optimised model; dotted line, initial solution] (M-K. Vai, S. Prasad, N. C. Li and F. Kai, *IEEE Transactions on Electron Devices*, Vol. 36, No. 4, pp. 761–762, April 1989 © IEEE).

3.4 Neural networks applied to modelling

The SA algorithm has been applied to device modelling in the previous section. SA avoids the local minimum entrapment problem and has been shown to be preferable to other optimisation methods since it is relatively insensitive to initial conditions. However, it inherits the time-consuming feature of iterative improvement methods. As was shown, the probabilistic hill-climbing operation increases the time taken for the completion of the optimisation. Furthermore, the solution has to slowly "cool down" according to the annealing procedure. Consequently, a large number of intermediate solutions have to be generated and evaluated at each pseudo-temperature which has to be decreased slowly from a large initial value. This disadvantage of the SA algorithm for optimisation can be overcome by using an artificial neural network (ANN).

ANNs are based on the human nervous system which consists of a distribution of neurons to carry messages back and forth to the brain. The ANN has been used successfully in many modelling and optimisation applications in engineering that are particularly useful when several tasks are to be performed in parallel and computation rates are required to be high [31, 37]. Neural networks are particularly attractive because of their speed and accuracy. Hence, neural networks have been developed into an alternative computer-aided approach to model and design devices and circuits. Neural networks represent a robust modelling approach to predict the behaviour of high-speed devices and circuits [29]. In comparison with various statistical methods and curve-fitting approaches for predicting system behaviour, the neural network approach features a learning process which fine tunes neural network parameters to interrelate the variables being modelled. A neural network may be developed to guide the solution generation of an SA optimisation process. This approach utilises the associative capability of a neural network to globally and concurrently evaluate the effect of varying all the parameters. When used in place of a physics-oriented device model, a neural network avoids the need to repeatedly solve the equations that describe the device physics.

Two classes of neural networks are used in modelling:

(i) Multi-layer perceptron neural networks
(ii) Recurrent Hopfield neural networks.

Regardless of the neural network architecture selected for an application, it consists of many processing elements called *neurons*, each connected to many others. Every connection entering a neuron has a weight assigned to it. This weight is used to amplify, attenuate or change the sign of the signal in the incoming connection. The input to the neural network is a vector of the data to be modelled. Each neuron operates on the outputs of other neurons according to its transfer function and delivers an output to other neurons. Often, the transfer function sums the incoming signals to determine the value of the neuron's next output signal. The result is an output vector representing some characteristics associated with the input.

In order to use neural network algorithms, it is necessary to determine an interconnection pattern, the weights and the transfer functions. The creation and training of an appropriate neural network for the problem on hand is difficult and time-consuming. However, an appropriately trained neural network provides fast and efficient solutions that have shown excellent results for different applications such as the modelling of transistor behaviour and microwave circuits as well as microwave impedance matching [28]. A neural network consists of a set of simultaneous non-linear equations that are capable of modelling any continuous function when the appropriate weights are determined. These networks are pictorially represented as neurons (circles) with interconnecting nerves (lines). Neural networks are very flexible tools for device modelling because they can be adapted to model different devices without change of equations. Adapting a physics-based model to a different physical device may involve a radical change of equations. This flexibility gave the impetus to the effort of finding much better ways of applying neural networks to device modelling. However, the problem

of finding the appropriate weights for the network is a highly non-linear one, suggesting the necessity to use stochastic optimising algorithms such as SA and genetic algorithms.

3.4.1 Massively distributed computing networks

A general description of the distributed computing methodology is given by Vai and Prasad [31]. Massively distributed computing networks are a specific form of a non-linear system that maps an input to an output. A distributed computing network can be considered to be an asynchronous array processor with very simple processing elements (i.e. neurons). Figure 3.5 shows a typical processing element, henceforth referred to as a *neuron*, with n inputs $(i_1, \ldots, i_n)$ and one output (Q). An input can be excitatory (indicated by a solid circle) or inhibitory (indicated by a hollow circle) and is assigned a weighting factor W_j. A threshold value T is associated with the neuron.

The function of a neuron can be described by the following equation which combines inputs $i_1, \ldots, i_n$ to form an overall input value I:

$$Q = \sum_{j=1}^{n} W_j \times i_j, \tag{3.7}$$

where W_j is positive for an excitatory input and negative for an inhibitory input. If the overall input value I is above the threshold value T associated with the neuron, the neuron fires and an output of $Q = 1$ is produced. Otherwise, the output remains at $Q = 0$. A neuron is also associated with a time constant (τ) that determines its output response time.

While the operation of a conventional computer is controlled by a series of instructions, called a *programme*, a massively distributed computing network is programmed by wiring up a set of neurons and setting the weights of these interconnections. The function of a distributed computing network can only be determined by considering the network as an integrated entity. No meaningful information can be extracted by examining a neuron isolated from its neighbours.

A distributed computing network is typically implemented by a hardware analogue circuit. Figure 3.6 shows the use of an operational amplifier configured as an integrating adder to carry out the function of a neuron. As shown in Figure 3.6, the input

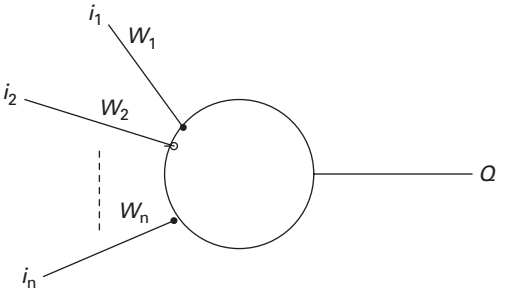

Fig. 3.5 Neuron structure.

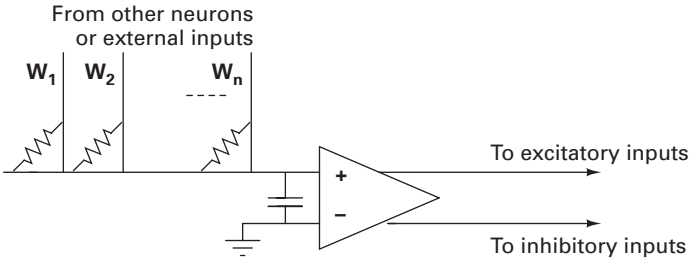

Fig. 3.6 Physical implementation of a neuron (M. Vai and S. Prasad, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 43, No. 5, pp. 1087–1094, May 1995. ©1995 IEEE).

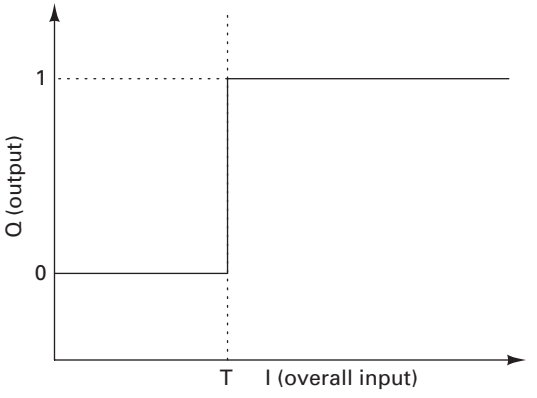

Fig. 3.7 Neuron Transfer Function (M. Vai and S. Prasad, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 43, No. 5, pp. 1087–1094, May 1995. ©1995 IEEE).

weightings of such a neuron can be controlled by choosing appropriate resistance values connecting its inputs to the outputs of other neurons. The time constant (τ) of this neuron is determined by the capacitance connected at the operational amplifier input. The neuron transfer function shown in Figure 3.7 shows the input–output transfer function of a neuron.

3.4.2 Multi-layer perceptron neural networks

The training algorithm called *backpropagation* [16] is used in the application of multilayer perceptron (MLP) neural networks to device modelling. A multi-layer neural network with four layers (one input layer, two hidden layers and one output layer) is shown in Figure 3.8. Referring to the notations in Figure 3.8, $X = (x_1, \ldots, x_i, \ldots, x_m)$ is the input vector; $G = (g_1, \ldots, g_j, \ldots, g_n)$, $H = (h_1, \ldots, h_k, \ldots, h_p)$ and $Y = (y_1, \ldots, y_l, \ldots, y_q)$ are the outputs of the first hidden layer, the second hidden layer and the output layer, respectively; u_{ij} is the weight between the ith neuron and the jth neuron in the first hidden layer; v_{jk} is the weight between the jth neuron in the first hidden layer and the kth neuron in the second hidden layer; and w_{kl} is the weight between the kth neuron in the second hidden layer and the lth neuron in the output layer. Bias

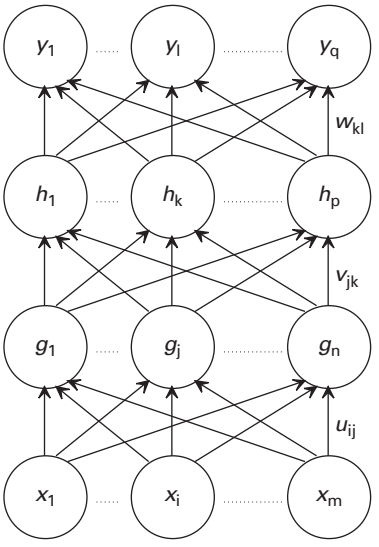

Fig. 3.8 Multi-layer neural network (M. Vai and S Prasad, *Int'l Journal of RF and Microwave CAE*, Vol. 9, No. 3, pp. 187–197, March 1999. © 1999 John Wiley & Sons). Reprinted with permission of John Wiley & Sons, Inc.

terms acting like weights on connections from units whose output is always 1 can also be provided to the neuron. They are not shown in Figure 3.8.

The output of the neural network is computed as

$$y_\ell = \frac{1}{1 + e^{-\gamma_\ell}}, \tag{3.8}$$

where γ_ℓ is the weighted total input to the output neuron ℓ, which is defined as

$$\gamma_\ell = \sum_{k=1}^{p} h_k w_{k\ell}, \tag{3.9}$$

and p is the number of neurons in the second hidden layer. Similarly, the output of the second hidden layer H can be expressed as a function of the output of the first hidden layer G which can in turn be expressed as a function of the input vector X. The backpropagation training algorithm aims to adjust the weights of a MLP neural network in order to minimise the sum-squared error of the network, which is defined as

$$E(n) = \sum_{m=1}^{S} \left\{ \frac{1}{2} \sum_{l=1}^{q} [d_{ml} - y_{ml}(n)]^2 \right\}, \tag{3.10}$$

where n is the epoch number in the training process, S is the number of training data, q is the number of output variables and $d_m = (d_{m1} d_{m2}, \ldots, d_{mq})$ and $y_m = (y_{m1} y_{m2}, \ldots, y_{mq})$ are the mth desired and calculated output vectors, respectively. This is done by continually changing the values of the weights in the direction of steepest descent with respect to the error function E. The iteration process continues until the error function is minimised. The learning is performed by the many presentations of a prescribed set of training examples to the network. One complete presentation of the

entire training set during the learning process is called an *epoch*. The learning process continues on an epoch-by-epoch basis until the weights of the network stabilise and the error function converges to a minimum value.

There are certain problems related to the architecture of an MLP neural network such as the determination of the number of hidden layers and the number of neurons in a hidden layer as well as under-fitting or over-fitting. The backpropagation learning algorithm and its derivatives are sensitive to the number of neurons in hidden layers. In general, a network with too few neurons will fail to model the data (i.e. under-fitting). While the more the number of neurons in hidden layers, the better the network can fit the data; if far too many neurons are used, over-fitting can occur. In the absence of a deterministic approach that can find the number of hidden layers and the number of neurons, a trial and error approach is taken. The hidden layers are adjusted to strike a balance between memorisation and generalisation.

A neural network trained with the relations between device parameters and behaviour can be used in place of conventional device models to speed up the simulation. Once a neural network model is trained, it provides a very fast prediction of results. Figure 3.9 shows a simplified flow chart of the circuit design process. Beginning with an initial solution, a series of solutions is generated. The circuit property of each solution is

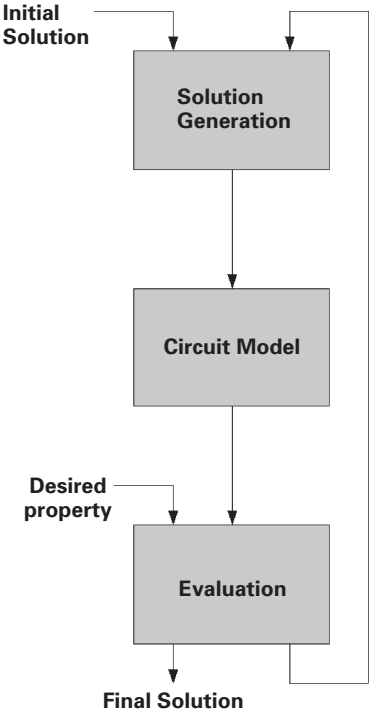

Fig. 3.9 Flow chart for a circuit design process (M. Vai and S. Prasad, *Int'l Journal of RF and Microwave CAE*, Vol. 9, No. 3, pp. 187–197, March 1999. ©1999 John Wiley & Sons). Reprinted with permission of John Wiley & Sons, Inc.

predicted by a circuit model and compared to the desired circuit property. If the solution on hand produces a circuit property close enough to the desired one, the design process is successful and terminated. Otherwise, another solution is generated and the above steps are repeated. The circuit model in Figure 3.9 often includes semiconductor devices which are commonly represented by the equations describing the physics of the particular device or equivalent circuit.

3.4.3 Hopfield recurrent neural networks

Although neural networks are known for their capability of learning the solutions to the problems that they are designed to solve, they also provide a framework for constructing special computing architectures to solve specific problems. The recurrent neural networks described here were proposed by Hopfield and are thus often referred to as *Hopfield networks* [11]. Consider a recurrent neural network of N neurons. If the activation of a neuron is updated according to the equation:

$$V_{\mathrm{i}}(t+1) = \mathrm{sgn}\left(\sum_{j=1}^{N} T_{\mathrm{ij}} V_{\mathrm{j}}(t) + I_{\mathrm{i}}\right), \tag{3.11}$$

where $V_{\mathrm{i}}(t) \in (0, 1)$ is the state of neuron i at moment t, T_{ij} is the weight associated with the link between neurons i and j, I_{i} is the internal threshold parameter of neuron i and $\mathrm{sgn}(x)$ is defined as

$$\mathrm{sgn}(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}. \tag{3.12}$$

It can be shown that an energy function defined as

$$\text{Energy} = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} T_{\mathrm{ij}} V_{\mathrm{i}} V_{\mathrm{j}} - \sum_{i=1}^{N} I_{\mathrm{i}} V_{\mathrm{i}} + K, \tag{3.13}$$

where K is a constant, is minimised.

The significance of a recurrent neural network is its ability to perform associative inference. There is no specific distinction between input and output vectors and a recurrent network perturbed by changing one or more neuron states will evolve into one of its consistent states which are the minima of its energy (Equation 3.13). The recurrent neural network is now applied to implement qualitative models, the function of which is to explore many competing hypotheses in a solution space with constraints. The Neural Network toolbox in MATLAB is available to allow fast and efficient device modelling.

Example: Modelling of HBTs using neural networks

Small-signal models of microwave devices are very useful in circuit design. Neural networks can be trained to learn the non-linear relationship between the small-signal transistor behaviour and the device input bias conditions. Figure 3.10 shows a representation of a neural network small-signal model of an HBT. The inputs of this model are the bias conditions of the device, i.e. the base current I_{b}, the collector–emitter voltage

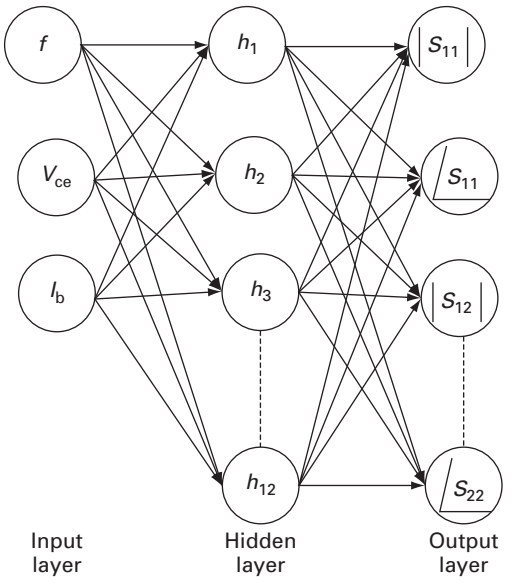

Fig. 3.10 Small-signal neural network model of the HBT.

V_{ce} and the frequency of operation f. The outputs of the model are the S parameters of the device.

As shown in Figure 3.10, the small-signal neural network model is a three-layer model – there are 3 layers in the input layer, 12 neurons in the hidden layer and 8 neurons in the output layer. The 3–12–8 structure of the neural network model is arrived at after various trials as being the best compromise between model accuracy and the time required to train the model. The model is trained with a sample set of measured S-parameter data. Different input bias conditions are then applied to the model and S-parameters derived from the model are compared to measured data. The data used to train the neural network model consist of measured S-parameters at 12 different bias conditions (V_{be} and $I_{c)}$, and a range of frequencies from 10 GHz to 40 GHz. The neural network modelling is tested for AlGaAs HBTs as well as SiGe HBTs. The models are accurate for both the material systems. The method should be applicable for all types of transistors regardless of the material system. The S-parameters for the devices are shown in Figures 3.11 and 3.12 for AlGaAs with bias $V_{ce} = 1.5\,\text{V}$, $I_b = 56\,\mu\text{A}$ and in Figures 3.13 and 3.14 for SiGe with bias $V_{ce} = 1.9\,\text{V}$, $I_b = 53\,\mu\text{A}$.

Figure 3.15 shows a representation of a neural network large-signal model of the device. The inputs of this model are the voltage bias conditions of the device: the collector–emitter voltage V_{ce} and the base–emitter voltage V_{be}. The outputs of the model are the two output currents of the device: the collector current I_c and the base current I_b.

As seen in Figure 3.15, the large-signal neural network model consists of three layers: an input layer consisting of two neurons, a hidden layer consisting of three neurons and an output layer of two neurons. The 2–3–2 structure of the neural network model was arrived at after various trials. It is the best compromise between model accuracy and the time required to train the model. The model is trained with a sample set of measured DC I–V characteristics and Gummel data. Different input voltage bias conditions are

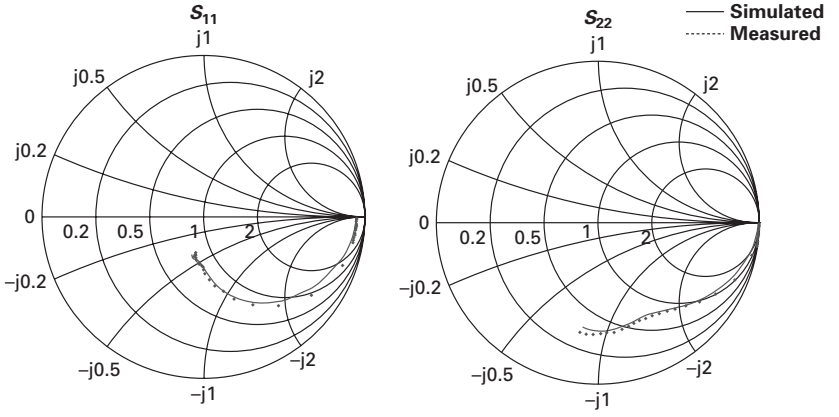

Fig. 3.11 The simulated and measured S_{11} and S_{22} for AlGaAs HBT.

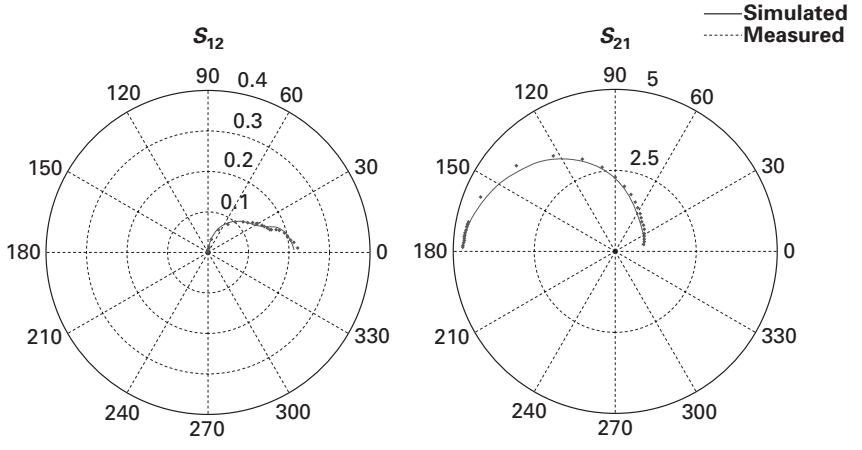

Fig. 3.12 The simulated and measured S_{12} and S_{21} for AlGaAs HBT.

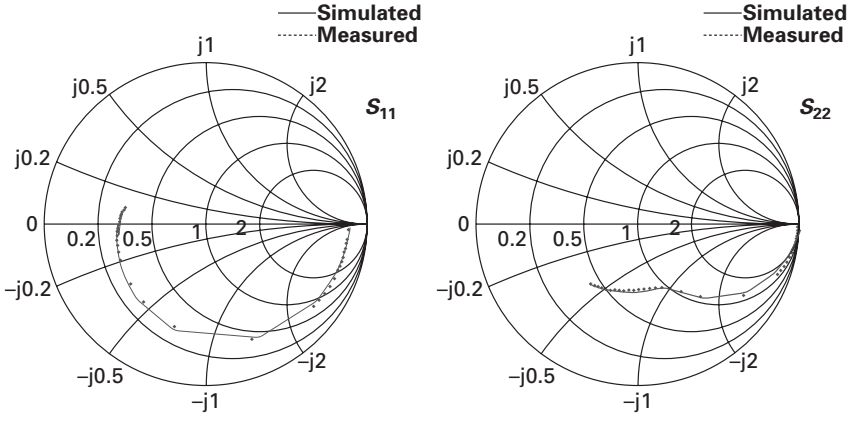

Fig. 3.13 The simulated and measured S_{11} and S_{22} for SiGe HBT.

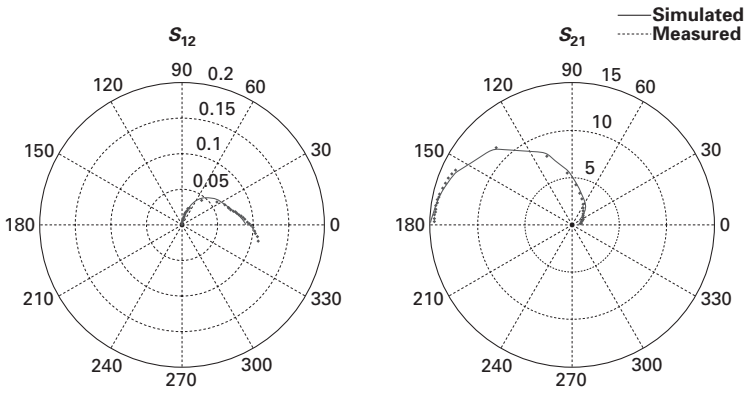

Fig. 3.14 The simulated and measured S_{12} and S_{21} for SiGe HBT.

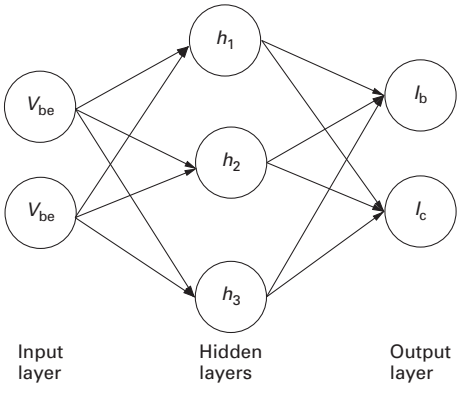

Fig. 3.15 Large-signal neural network model of the HBT.

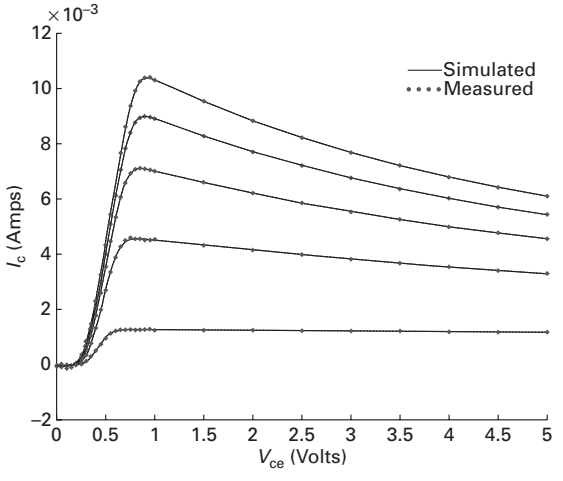

Fig. 3.16 I–V Characteristics of the AlGaAs HBT.

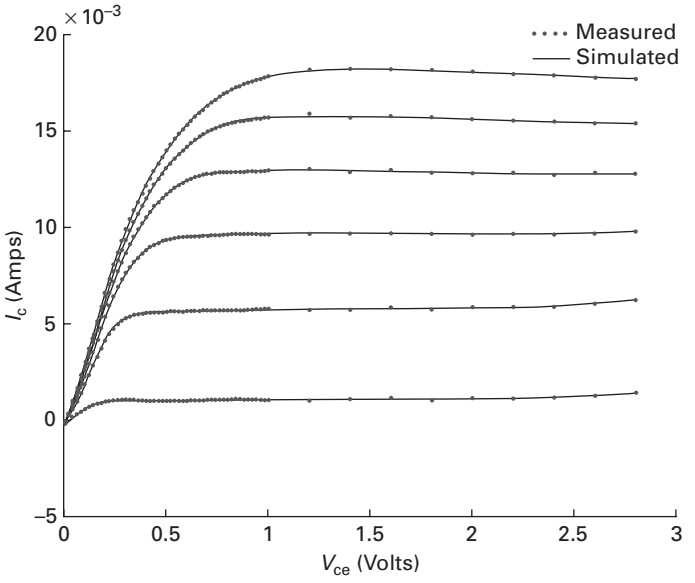

Fig. 3.17 I–V Characteristics of the SiGe HBT.

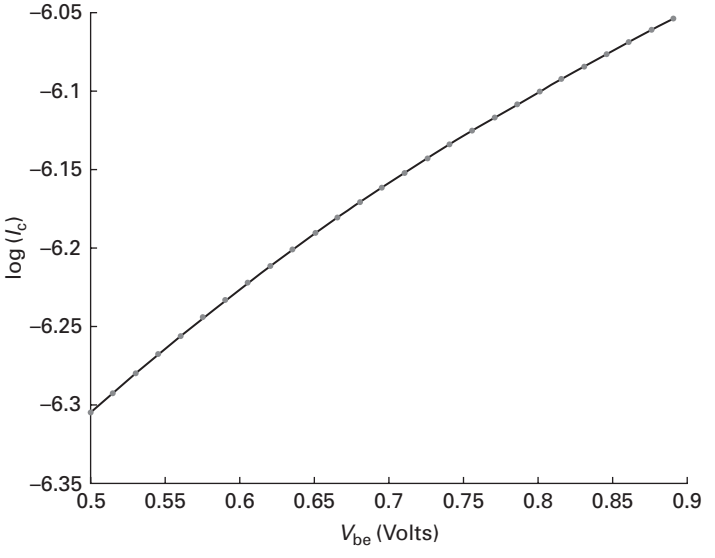

Fig. 3.18 Forward Gummel plot I_c versus V_{be} for AlGaAs HBT.

then applied to the model and output currents derived from the model are compared to measured data. The data used to train the neural network model consist of measured output currents (I_c and I_b) at different voltage bias conditions (V_{be} and V_{ce}). The DC I–V characteristics (measured and simulated) of the devices are shown in Figures 3.16 and 3.17. These figures show the collector current I_c versus the collector–emitter voltage V_{ce} at different base currents I_b.

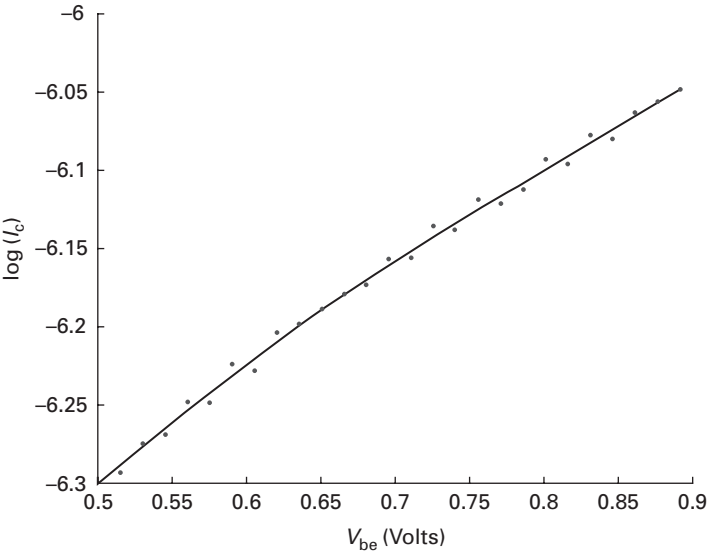

Fig. 3.19 Forward Gummel plot I_b versus V_{be} for AlGaAs HBT.

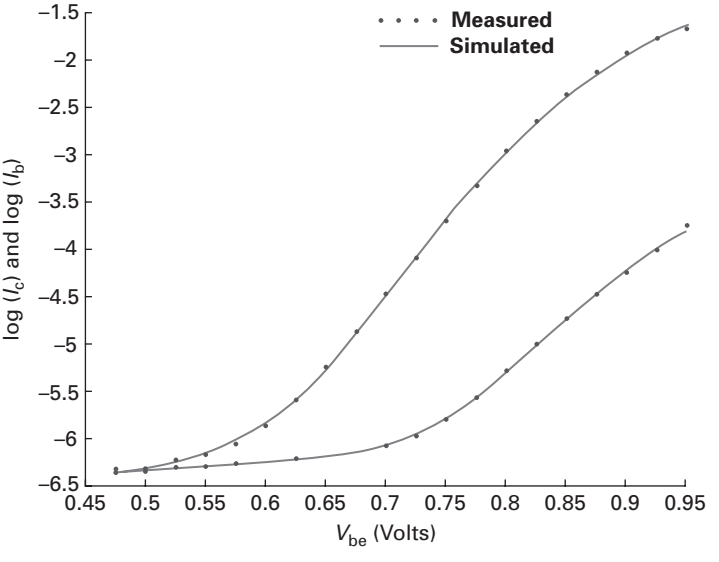

Fig. 3.20 Forward Gummel plot I_c (top curve) and I_b (bottom curve) versus V_{be} for SiGe HBT.

Figures 3.18, 3.19 and 3.20 show the forward Gummel plot of I_c and I_b versus V_{be}.

The success of the application of neural network modelling in the example is evident from the very good correlation between the measured and simulated data.

3.5 Optimisation of neural networks by the genetic algorithm

Like neural networks, genetic algorithms are an optimisation strategy inspired by nature. Based on the Darwinian theory of evolution, these algorithms use the "survival of the fittest" paradigm to find the best solution to a problem [8, 10, 32]. They iteratively evaluate several possible solutions choosing the ones that are the closest fit to the desired solution. The possible solutions are called *chromosomes* and are usually represented as strings of binary numbers called *genes*.

The algorithm begins by randomly generating a number of chromosomes to form a population. Each chromosome is given a rank called a *fitness index* based on its closeness to the desired solutions. The highest ranked chromosomes have a greater chance of being selected for the next stage of the algorithm: the reproduction stage. In this stage, pairs of chromosomes are separated at selected points and their genes are exchanged in a process called *crossover* to form a new generation of chromosomes. Because the most fit chromosomes are likely to be selected for reproduction in every generation, each new population is likely to consist of better solutions. Coupled with the fact that the reproduction process naturally eliminates less fit chromosomes from the population, the population gets pushed towards the desired solution in every new generation. When trying to reach the optimal solution for the problem, a genetic algorithm has to avoid local extrema or pseudo-optimal solutions. To prevent convergence of the algorithm to these local extrema, the genetic algorithm uses a technique called *mutation* which inverts the value of a randomly chosen gene with a given probability. A flow chart for the genetic algorithm is given in Figure 3.21.

The genetic algorithm is a robust optimisation method well-suited to the difficult task of finding the optimum weights for a neural network. However, before it can be used to evolve the weights of a neural network, the following prerequisites must be worked out:

(i) *A suitable problem representation*
A chromosome of a neural network corresponds to the weight matrix for a layer. Substrings of the chromosome are made by concatenating the values of the real-valued weights between each neuron and the neurons in the previous layer. The substrings are then joined together into the chromosome for the layer. In some implementations, the real-valued weights are converted to binary values for enhanced gene exchange during crossover. However, it is sufficient to use strings of real-valued weights in most cases for genetic evolution.

(ii) *A fitness index*
A fitness index measures the closeness of the current output of the neural network to the desired output. It is the least squares norm of the difference.

(iii) *The reproduction strategy*
Reproduction involves two processes: crossover and mutation. During crossover, a pair of chromosomes are separated at a randomly chosen point along their length. The resulting substrings of the pair are switched and then joined to form child chromosomes. This is repeated for every pair of chromosomes in the population

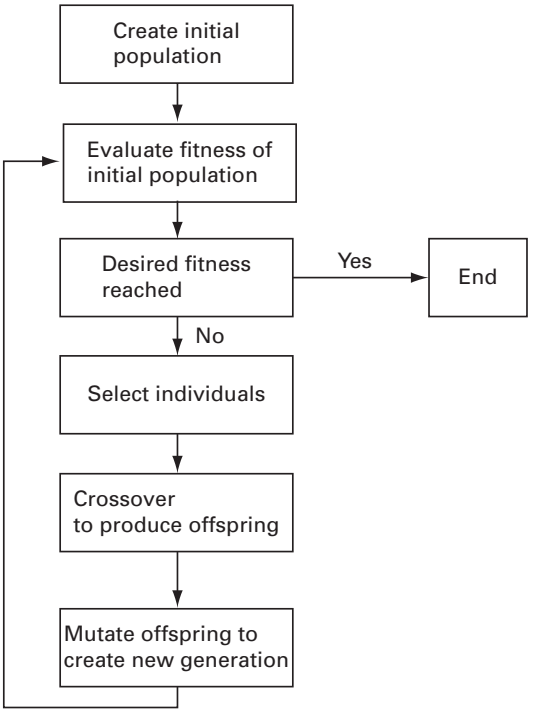

Fig. 3.21 Flow chart for the application of the genetic algorithm.

although some implementations will allow a small percentage of the fittest individuals to pass to the next generation without reproduction. Mutation is applied to a small percentage of the population. In this operation, the value of a gene is changed by adding a small value to it.

(iv) *The termination criterion*
Although, it would seem desirable to train the neural network till the error is as small as possible, this may be neither practical nor ideal. It may take up to several days to train a neural network. In addition, an excessively trained network can fail to produce correct outputs when presented with inputs not included in the training set. This is called *over-fitting*. The appropriate termination criterion for a given application is best determined from experiments. A common way is to terminate the algorithm after a fixed number of iterations.

3.6 Structured genetic algorithm

Neural networks are usually designed by determining the optimum values of weights for a fixed number of neurons. However, it is impossible to know in advance how many neurons are optimal for an application. This difficulty can be avoided by using the structured genetic algorithm (SGA), an algorithm that enables the determination of the optimum number of neurons and weight values simultaneously [4]. SGA uses a

hierarchical representation for the genetic structure in which neurons are a high level layer of binary-valued genes controlling a lower level layer of weight genes. When a high level gene is on (value = 1), the lower level weights genes it controls are activated and used in the computation of the neural network's output. When a high level gene is off, the weights it controls are deactivated. Therefore, SGA is able to evaluate a variable number of neurons in the neural network. Neurons are turned on or off by the processes of mutation or crossover. The equations for SGA applied to a neural network with one hidden layer can be written as follows. Let X be a set of inputs to the networks and U the set of inner layer weights, then the output of a neuron in the middle layer is given by

$$y = \delta f\left(\sum_{i=1}^{m} u_{\mathrm{i}} x_{\mathrm{i}} + \tau\right), \tag{3.14}$$

where $f(x)$ is the sigmoid function:

$$f(x) = \frac{1}{1 + e^{-\mathrm{x}}} \tag{3.15}$$

and δ is one when the neuron is active and zero otherwise.

If the set of the output weights is taken to be V, then the network outputs are given [36] by

$$z_{\mathrm{i}} = \sum_{i=1}^{m} v_{\mathrm{i}} y_{\mathrm{i}}. \tag{3.16}$$

Genetic algorithms are relatively slow algorithms and are best used to determine the weights for just the hidden layers of a neural network. The determined inner weights are then used to calculate the activations of the hidden layer neurons and this reduces the problem to a linear equation, where linear regression can be used to calculate the output layer weights.

The SGA for neural networks is given in Figure 3.22.

Example: Application of SGA to the modelling of a HEMT amplifier

The use of the SGA is illustrated by creating a neural network model of a HEMT class-F power amplifier as shown in Figure 3.23. IMN and OMN denote the input matching network and output matching network. A class-F amplifier is a highly efficient switching amplifier used in mobile commercial and military systems. The design and properties of amplifiers are described in Chapter 5.

Table 3.2 shows the inputs and outputs for one hidden layer neural network model. The upper limit for the number of neurons in the hidden layer was arbitrarily set to 40. Before training, the data were scaled with a linear transform to lie in the range of the sigmoid activation function (0.1–0.9) used for the neurons. At the start of training, the network's weights were initialised to small random values. The input to the SGA is shown in Table 3.3. Using the crossover and mutation operations, the initial weights were optimised until the termination condition was satisfied. The transfer function of the trained neural network and that of the HEMT are plotted in Figure 3.24.

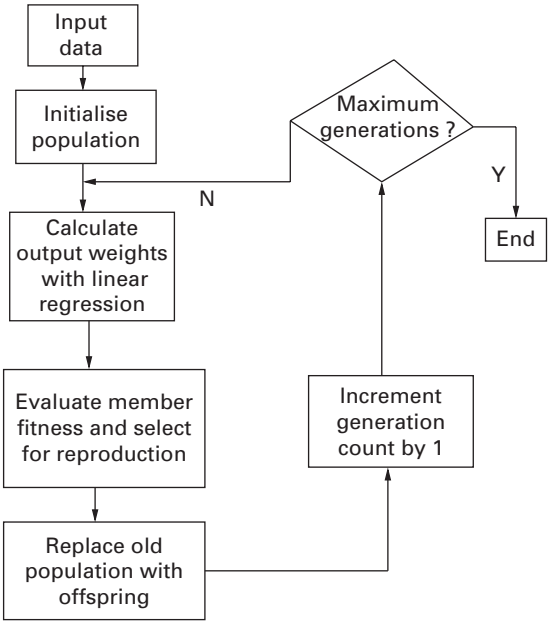

Fig. 3.22 Structured genetic algorithm for neural networks.

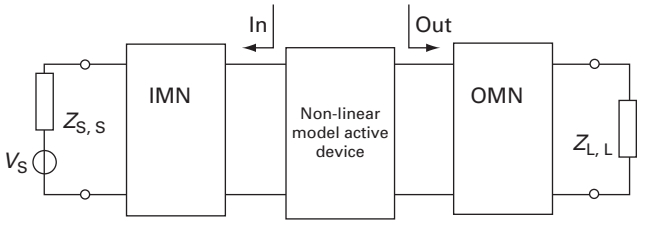

Fig. 3.23 Simplified power amplifier schematic.

Table 3.2 Neural network model for HEMT power amplifier

INPUTS	OUTPUTS
Input power	Output power
Gate voltage	DC gate current
DC drain current	Drain current

Table 3.3 SGA parameters for HEMT power amplifier model

Parameter	Value
Initial population size	200
Number of generations	10000
Number of input neurons	3
Number of output neurons	1
Maximum hidden neurons	40
Number of samples	140
Crossover probability	0.8%
Neuron layer mutation rate	0.0001%
Weights layer mutation rate	0.0001%

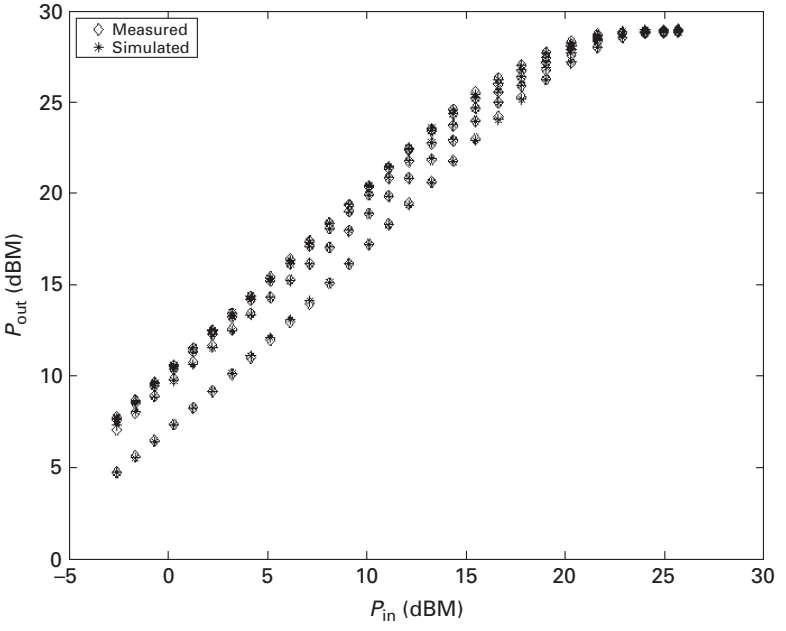

Fig. 3.24 Comparison of measured transfer function to values predicted by the neural network for the HEMT amplifier.

3.7 Semi-analytical device parameter extraction

Accurate, physically meaningful device models are necessary for circuit design, process technology design and optimum device design. The extraction of equivalent circuit parameters has been investigated by researchers for more than a decade [1–3, 5–7, 20, 25–27, 33–35].

A semi-analytical parameter extraction procedure for the HBT equivalent circuit developed by Li and Prasad [15] is presented here as one illustration of parameter extraction. It combines analytical and optimisation approaches. The significance of this procedure is that it is completely general and can be applied to any semiconductor device using the appropriate equivalent circuit.

3.7.1 Theoretical analysis

An AlGaAs common–emitter HBT is used to illustrate this procedure. The small-signal T-model equivalent circuit is shown in Figure 3.25. The box with dashed lines in Figure 3.25 encloses the inner shell without the pad parasitics. All the impedance parameters given below are for the inner shell. There are 16 elements in the equivalent circuit. Only $C_{be}, r_e, \alpha, R_{bc}$ and C_{bc} are considered to be bias-dependent and all the other elements are assumed to be bias-independent. The expressions for the two-port Z-parameters of the inner shell can be simplified in terms of the frequency ranges. The frequency ranges are characterised by $\omega C_{bc} R_{bc} \ll 1$ (low-frequency range), $\omega C_{bc} R_{bi}, \omega C_f R_{bi} \ll 1$ and $\omega C_{bc} R_{bc} \gg 1$ (intermediate frequency range) and $\omega C_{bc} R_{bi} \gg 1$ (high-frequency range). In the measurements on high-speed devices,

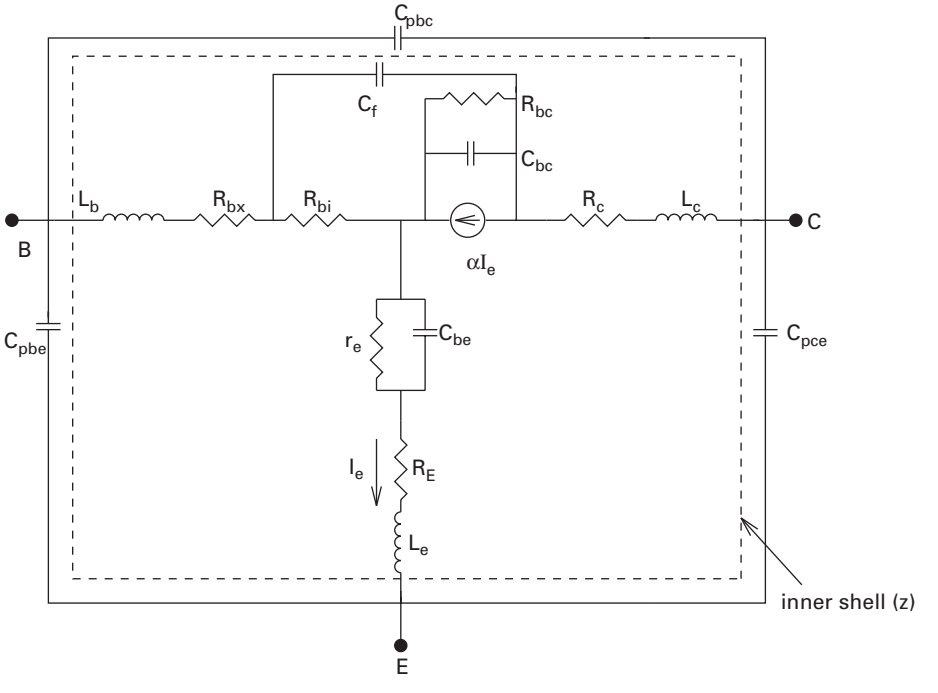

Fig. 3.25 The small-signal equivalent circuit of the AlGaAs/GaAs HBT (B. Li, S. Prasad, L.W. Yang and S. C. Wang, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 46, No. 10, pp. 1427–1435, October 1998. ©1998 IEEE).

most of the frequency data are located in the intermediate frequency range. In the low-frequency range,

$$Z_{11} - Z_{12} = R_{bx} + j\omega L_b + R_{bi} - j\omega R_{bc} R_{bi} C_s \tag{3.17}$$

$$Z_{12} = R_E + j\omega L_e + r_e - j\omega r_e^2 C_{be} + (1-\alpha) \times j\omega R_{bi} C_f (R_{bc} - j\omega R_{bc}^2 C_s) \tag{3.18}$$

$$Z_{22} - Z_{21} = R_C + j\omega L_c + R_{bc} - j\omega R_{bc}^2 C_s, \tag{3.19}$$

Where $C_s = C_{bc} + C_f$.

In the intermediate frequency range, the Z-parameters can be approximated as

$$Z_{11} - Z_{12} = R_{bx} + j\omega L_b + R_{bi} \frac{C_{bc}}{C_s} \tag{3.20}$$

$$Z_{12} = R_E + j\omega L_e + r_e - j\omega r_e^2 C_{be} + (1-\alpha) \times R_{bi} \times \frac{C_f}{C_s} \tag{3.21}$$

$$Z_{22} - Z_{21} = R_C + j\omega L_c + \frac{1}{j\omega C_s} + \frac{1}{\omega^2 R_{bc} C_s} - \frac{R_{bi} C_{bc} C_f}{C_s^2} - j \frac{R_{bi} C_f C_{bc}}{\omega C_s^3 R_{bc}}. \tag{3.22}$$

In the high-frequency range, the simplified relation is

$$Z_{22} - Z_{21} = R_C + j\omega L_c - \frac{1}{\omega^2 R_{bi} C_{bc} C_f} + \frac{1}{\omega^2 R_{bi} C_{bc} C_f} \times \left(\frac{1}{j\omega C_f R_{bi}} + \frac{1}{j\omega C_{bc} R_{bi}} \right) \tag{3.23}$$

$$\alpha = \alpha_0 \left[\frac{1}{1 + \frac{j\omega}{\omega_0}} \right] e^{-j\omega\tau}. \tag{3.24}$$

Example: Extraction of equivalent circuit elements

For the HBT used in this example, the intermediate frequency range is taken to be approximately from 0.5 GHz to 20 GHz, and the high-frequency range should go up to 40 GHz. The condition for the high frequency is relaxed and frequencies over 25 GHz are considered to be in the high-frequency range.

A. Extraction of the parasitic elements

If no test structure is available for extracting the parasitics, it is still possible to extract or estimate the pad capacitances from the HBT under cut-off operation [7, 24]. Under cut-off operation, we have zero V_{BE}, zero I_c, and variable V_{CB}. The HBT equivalent circuit of Figure 3.25 is reduced to the simplified circuit shown in Figure 3.26, provided the influence of the inductances and resistances remains negligible and the conditions

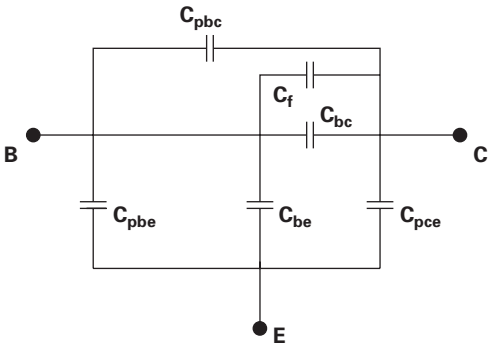

Fig. 3.26 The simplified HBT equivalent circuit under cutoff operations in which both junctions are reverse-biased and the influence of the inductances and resistances remains negligible (B. Li, S. Prasad, L.-W. Yang and S. C. Wang, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 46, No. 10, pp. 1427–1435, October 1998. ©1998 IEEE).

$R_{bc} \gg (1/\omega C_{bc})$ and $R_{bc} \gg (1/\omega C_{be})$ are satisfied. The capacitances in the equivalent circuit can be calculated directly:

$$C_{pbe} + C_{be} = \frac{\mathrm{Im}(Y_{11}) + \mathrm{Im}(Y_{12})}{\omega} \tag{3.25}$$

$$C_{pce} = \frac{\mathrm{Im}(Y_{22}) + \mathrm{Im}(Y_{12})}{\omega} \tag{3.26}$$

$$C_{pbc} + C_f + C_{bc} = -\frac{\mathrm{Im}(Y_{12})}{\omega}. \tag{3.27}$$

In the above equations, the C_{pbe}, C_{pbc}, C_{pce} and C_f are considered to be bias-independent and C_{be} and C_{bc} are bias-dependent elements. The value of C_{pce} can be calculated from Equation (3.26). Figure 3.27 clearly shows that C_{pce} is bias-independent.

C_{be}, the base–emitter junction capacitance, can be described by the equation

$$C_{be} = \frac{C_{jbe0}}{\left[1 + \left(V_{EB}/V_{jb}\right)\right]^{M_{jbe}}}. \tag{3.28}$$

The extraction of C_{pbe} can be carried out by fitting the sum $C_{pbe} + C_{be}$ to Equation (3.28) at different reverse base–emitter voltages or by using the iteration method in which different values of V_{jbe}, M_{jbe} and C_{jbe0} are tried until the plot of $C_{pbe} + C_{be}$ versus $[1 + (V_{EB}/V_{jbe})]^{-M_{jbe}}$ is a straight line shown in Figure 3.28.

Similarly, $(C_f + C_{pbc})$ can be extracted by fitting the sum $(C_f + C_{pbc} + C_{bc})$ to the expression for junction capacitance at different base–collector voltages. However, it must be noted that it is difficult to distinguish between the base–collector coupling capacitance and extrinsic base–collector capacitance [24]. This is due to the fact that the distance between the base probe tip and the collector probe tip in probe stations used for most high-speed measurements is usually longer, and thus the coupling effect between base and collector contacts must be very small; furthermore, the influence of C_{pbc} can be absorbed by the extrinsic base–collector capacitance C_f. Thus, C_{pbc} can

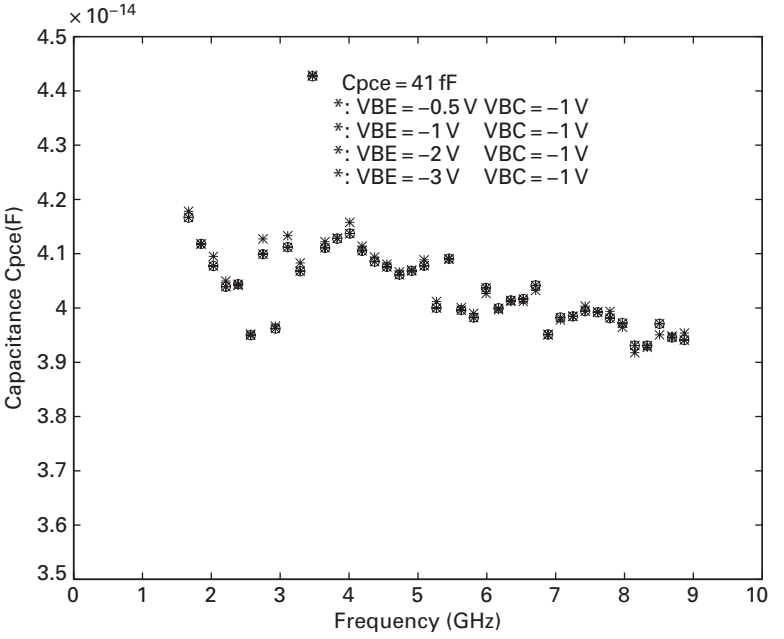

Fig. 3.27 The calculated C_{pce} versus frequency (B. Li, S. Prasad, L.-W. Yang and S. C. Wang, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 46, No. 10, pp. 1427–1435, October 1998. ©1998 IEEE).

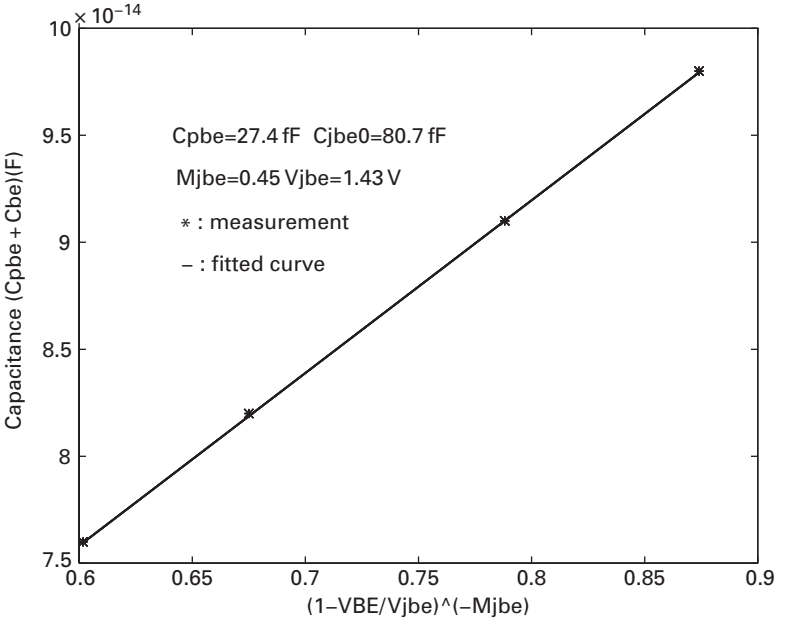

Fig. 3.28 The fitting of the sum $C_{pbe} + C_{be}$ to the expression for junction capacitance (B. Li, S. Prasad, L.-W. Yang and S. C. Wang, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 46, No. 10, pp. 1427–1435, October 1998. ©1998 IEEE).

be chosen to be zero. Such an assumption is also confirmed by the empirical optimisation procedures. The S-parameters measured over the frequency range of interest (here the range chosen was 50 MHz–36 GHz) are first converted to Y-parameters. After de-embedding the effect of the pad capacitances, the Y-parameters of the inner shell are converted to Z-parameters. Most of the elements are extracted from an analysis of the behaviour of the Z-parameters. Certain constraints are obtained to help in conditioning the optimisation procedure and to reduce the uncertainty.

B. The base–collector capacitance C_s

As indicated in Equation (3.66), the following approximation is valid in the intermediate frequency range:

$$\mathrm{Im}(Z_{22} - Z_{21}) = j\omega L_c + \frac{1}{j\omega C_s} - j\frac{R_{bi}C_f C_{bc}}{\omega C_s^3 R_{bc}}. \tag{3.29}$$

At the low end of the intermediate frequency range, the second term is much greater than the other terms on the right side of the equation. C_s can be extracted by the equation:

$$C_s = \frac{1}{\omega \mathrm{Im}(Z_{22} - Z_{21})}. \tag{3.30}$$

The extracted C_s from Equation (3.30) at the bias values for $V_{CE} = \{0.5\,\mathrm{V}, 1.0\,\mathrm{V}, 2.0\,\mathrm{V}, 4.0\,\mathrm{V}, 7.0\,\mathrm{V}\}$ and $I_B = 200\,\mu\mathrm{A}$ is shown in Figure 3.29.

At $V_{CE} = 0.5\,\mathrm{V}$, $1/\omega C_s \simeq 1224\,\Omega$, which is much larger than the other terms in the intermediate frequency range if the values extracted below are used. The base–collector

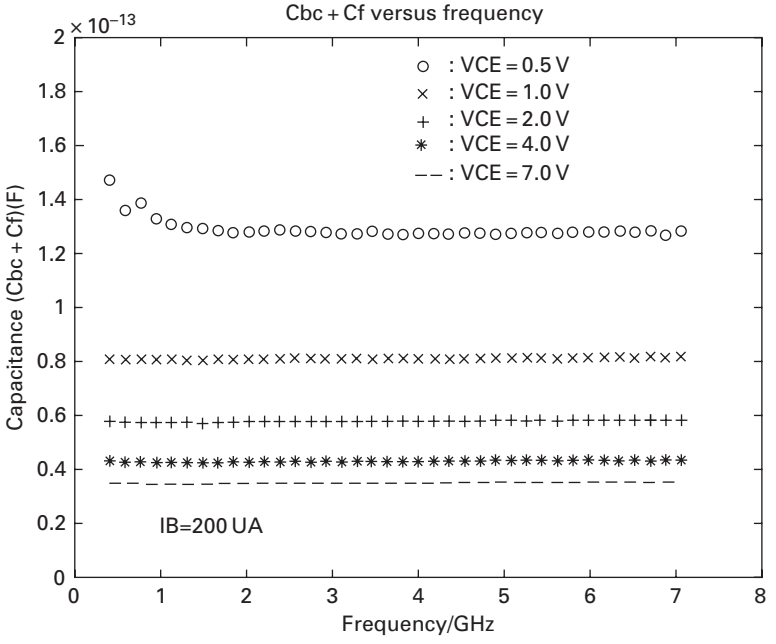

Fig. 3.29 The extracted value C_{bc} at different base collector voltages (B. Li, S. Prasad, L.-W. Yang and S. C. Wang, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 46, No. 10, pp. 1427–1435, October 1998. ©1998 IEEE).

capacitance is observed to decrease as V_{CE} increases. This results from the increased width of the base–collector depletion region due to the increased V_{CE}. The deviation of C_s is less than 5% except when $V_{CE} = 0.5$ V. The base–collector junction is forward-biased at this value of V_{CE} when the intermediate frequency range moves up.

The extrinsic base–collector capacitance is generally a weak function of the base–collector junction voltage. In extreme cases, it can be considered to be independent of the bias variation, or the ratio of the extrinsic capacitance to the total base–collector capacitance is considered to be a constant. Practically, the extrinsic capacitance C_f is the in-between case. For simplicity, the extrinsic capacitance is considered to be fixed and extracted from the values of C_s at the different base–collector voltages. A method similar to that used for the extraction of the value of C_{pbe} and C_{pce} is applied. The value of C_s thus obtained can be compared with what is obtained from the cutoff measurement. The parameters for the base–collector junction capacitance are also extracted from this approach.

C. The collector contact lead inductor L_c

The collector lead conductor L_c can be calculated from Equation (3.29):

$$L_c = \frac{1}{\omega}\left[\mathrm{Im}(Z_{22} - Z_{21}) + \frac{1}{\omega C_s}\right]. \tag{3.31}$$

The third term in Equation (3.29) is assumed to be small enough and is neglected. This straightforward method is not as accurate as expected. The deviation of the extracted value of L_c is large and an accurate value of L_c is difficult to obtain. The reason for this is that the small error resulting from extracting C_s could lead to large errors in L_c. The differentiation of Equation (3.31) yields

$$\begin{aligned}\Delta L_c &= \frac{1}{\omega^2}\left(-\frac{1}{C_s^2}\right)\Delta C_s \\ &= -\frac{1}{\omega^2 C_s} \times \frac{\Delta C_s}{C_s}.\end{aligned} \tag{3.32}$$

L_c is very sensitive to even a 5% error in extracting C_s. The error in estimating L_c resulting from the error in the estimation of C_s is plotted in Figure 3.30. L_c is very sensitive to the error in extracting C_s. However, L_c is less sensitive to the error in estimating C_s if the magnitude of C_s becomes larger. Therefore, a good bias point to extract L_c would be zero bias at which the third term in the equation is negligible and the value of C_s is larger. The L_c extracted at zero bias is shown in Figure 3.31.

D. The base–collector resistance R_{bc}

The real part of $Z_{22} - Z_{21}$ in the middle frequency range is given by

$$\mathrm{Re}(Z_{22} - Z_{21}) = R_c + \frac{1}{\omega^2 R_{bc} C_s} - \frac{R_{bi} C_{bc} C_f}{C_s^2}. \tag{3.33}$$

If the term $1/\omega^2 R_{bc} C_s$ is much larger than the other two terms, R_{bc} can be approximately extracted from the real part of $Z_{22} - Z_{21}$ in the lower middle-frequency range:

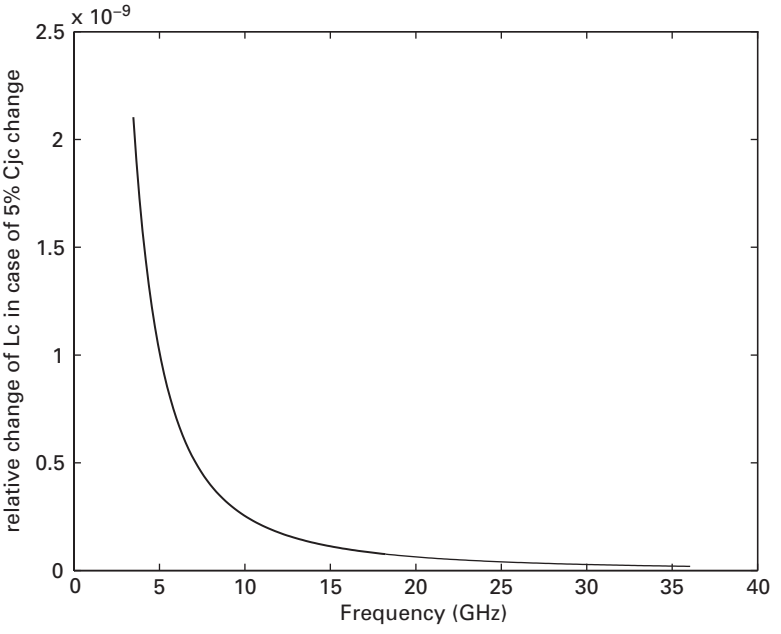

Fig. 3.30 The variation of L_c with a 5% error in the estimation of C_s (B. Li, S. Prasad, L.-W. Yang and S. C. Wang, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 46, No. 10, pp. 1427–1435, October 1998. ©1998 IEEE).

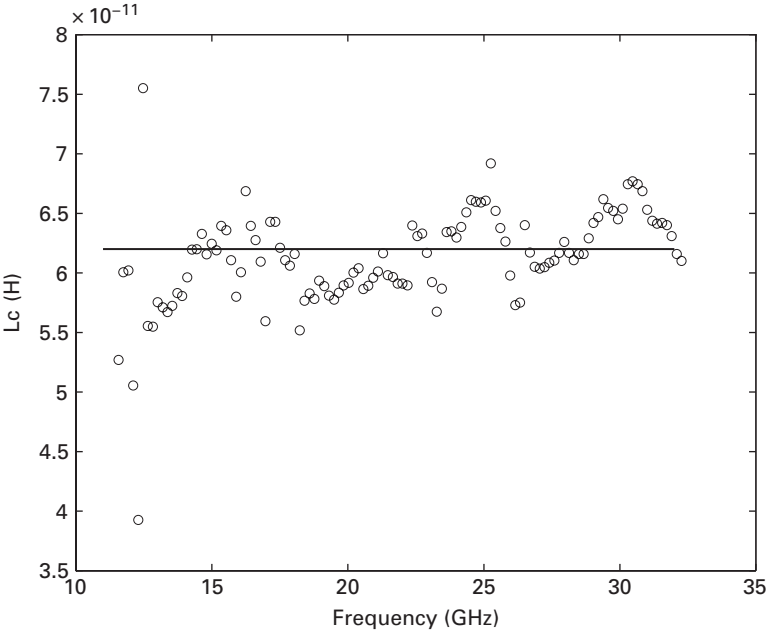

Fig. 3.31 The extracted L_c versus frequency (B. Li, S. Prasad, L.-W. Yang and S. C. Wang, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 46, No. 10, pp. 1427–1435, October 1998. ©1998 IEEE).

$$R_{bc} = \frac{1}{\omega^2 C_s \mathrm{Re}(Z_{22} - Z_{21})} \tag{3.34}$$

The second term in Equation (3.33) is inversely proportional to ω^2; therefore, the magnitude of the second term decreases rapidly. The other two terms cannot be neglected as the frequency increases to a certain point. However, the extracted value of R_{bc} is not significant since the value of R_{bc} is very large and does not affect the frequency response much as long as we are only concerned with forward operation. Figure 3.32 shows the extracted R_{bc} at the bias $I_B = 200\,\mu A$ and $V_{CE} = \{0.5\,V, 1.0\,V, 2.0\,V, 4.0\,V, 7.0\,V\}$. The magnitude of R_{bc} increases as V_{CE} increases.

Figure 3.33 shows the extracted R_{bc} without de-embedding the pad capacitances at the bias $I_B = 200\,\mu A$, $V_{CE} = 1.0\,V$. The magnitude of the calculated R_{bc} is negative beyond 1 GHz. This shows that physically meaningless values may be obtained if no de-embedding procedure is carried out.

E. The collector extrinsic resistance R_c

The R_c could be extracted by plotting $\mathrm{Re}(Z_{22} - Z_{21})$ versus $1/\omega^2$ in the high frequency range. The y-axis intercept is the value of R_c. The requirement for the high-frequency range is difficult to be achieved and the conditions for the requirement are relaxed. C_s is bias-dependent and the larger value of C_s could be obtained from S-parameters at zero bias. R_c should be extracted from zero bias by this method since the R_c is more significant in Equation (3.23) at zero bias.

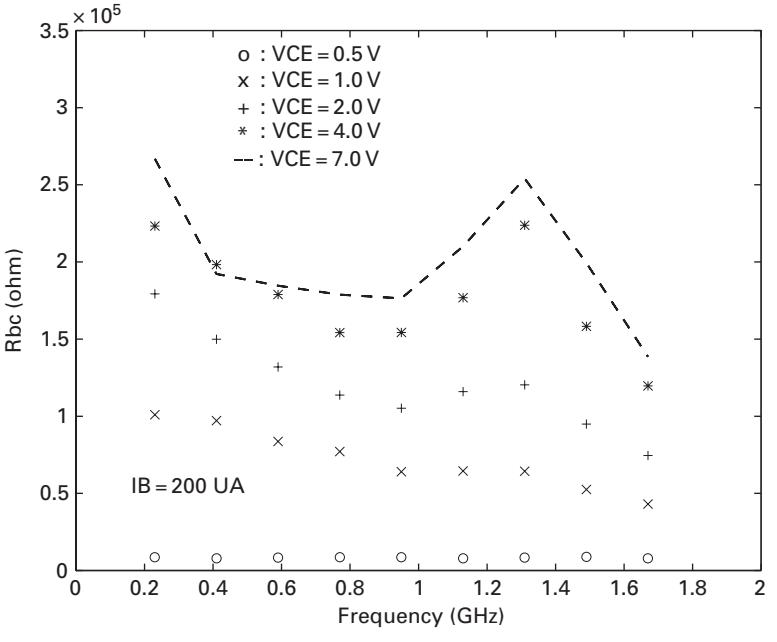

Fig. 3.32 The extracted R_{bc} at different base–collector voltages (B. Li, S. Prasad, L.-W. Yang and S. C. Wang, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 46, No. 10, pp. 1427–1435, October 1998. ©1998 IEEE).

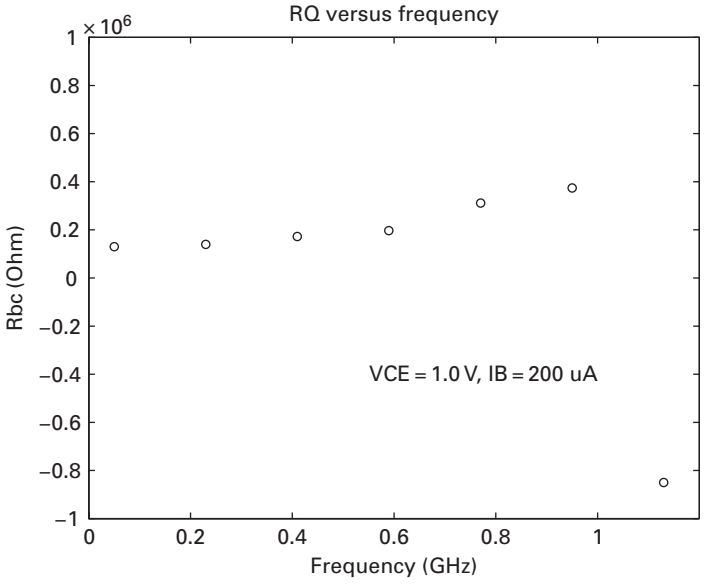

Fig. 3.33 The extracted R_{bc} at the bias $I_B = 200\,\mu A$, $V_{CE} = 1.0\,V$ without de-embedding the pad capacitances (B. Li, S. Prasad, L.-W. Yang and S. C. Wang, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 46, No. 10, pp. 1427–1435, October 1998. ©1998 IEEE).

F. The base contact lead inductor L_b

Applying the first-order approximation, L_b can be easily extracted from the imaginary part of $Z_{11} - Z_{12}$ in the middle-frequency range. That is

$$L_b = \mathrm{Im}(Z_{11} - Z_{12})/\omega. \tag{3.35}$$

The extracted L_b at different biases is shown in Figure 3.34 without the de-embedding procedure.

The dependence of the value of L_b on the bias V_{CE} is attributable to the pad capacitance. After the de-embedding procedure is carried out, the extracted L_b is shown in Figure 3.35.

The magnitude variation of L_b at the different biases is very small and almost negligible, so L_b can be considered to be independent of bias.

G. The intrinsic and extrinsic base resistances

In principle, the sum of the intrinsic and extrinsic base resistances, $R_{bx} + R_{bi}$, can be extracted from the low-frequency data, and the extrinsic base resistance, R_{bx}, can be extracted from the high-frequency data if the equivalent circuit shown in Figure 3.25 describes the frequency response of the HBT accurately. However, most of the frequency data are located in the intermediate frequency range. The requirement for the high frequency condition is difficult to be satisfied, and the data at extremely low frequencies are not available. The constraints on the base resistances thus can be obtained from the real part of $Z_{11} - Z_{12}$ in the intermediate frequency range:

$$\mathrm{Im}(Z_{11} - Z_{12}) = R_{bx} + R_{bi}\frac{C_{bc}}{C_s}. \tag{3.36}$$

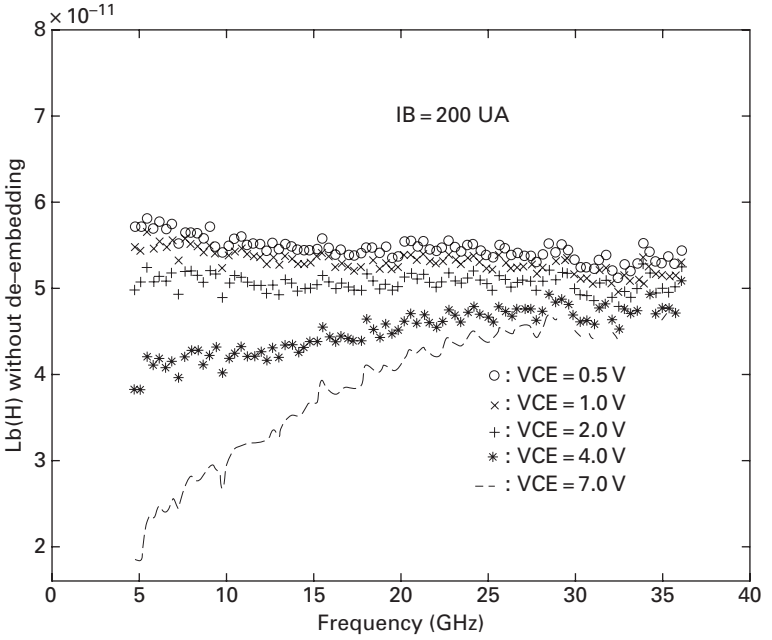

Fig. 3.34 The L_b versus frequency, in which the pad capacitances have not been de-embedded (B. Li, S. Prasad, L.-W. Yang and S. C. Wang, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 46, No. 10, pp. 1427–1435, October 1998. ©1998 IEEE).

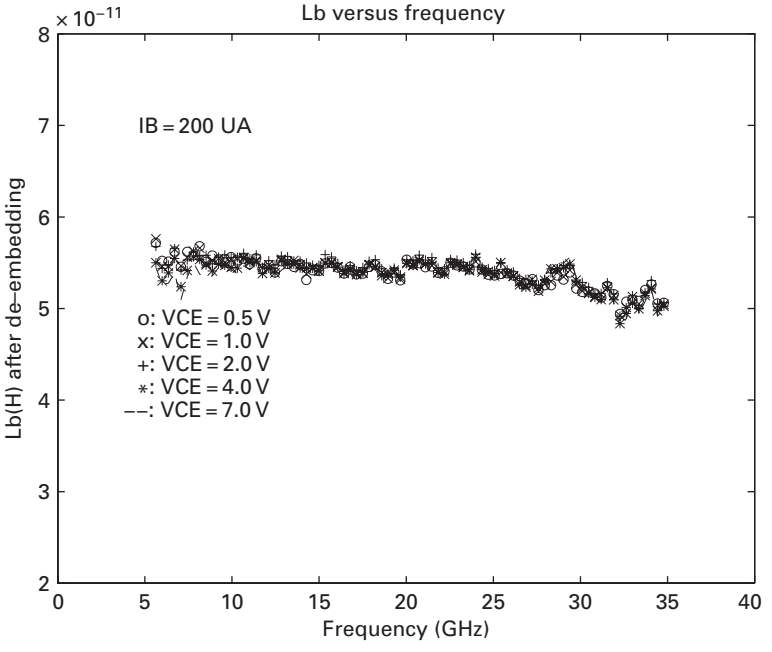

Fig. 3.35 The L_b versus frequency, in which the pad capacitance effect has been de-embedded (B. Li, S. Prasad, L.-W. Yang and S. C. Wang, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 46, No. 10, pp. 1427–1435, October 1998. ©1998 IEEE).

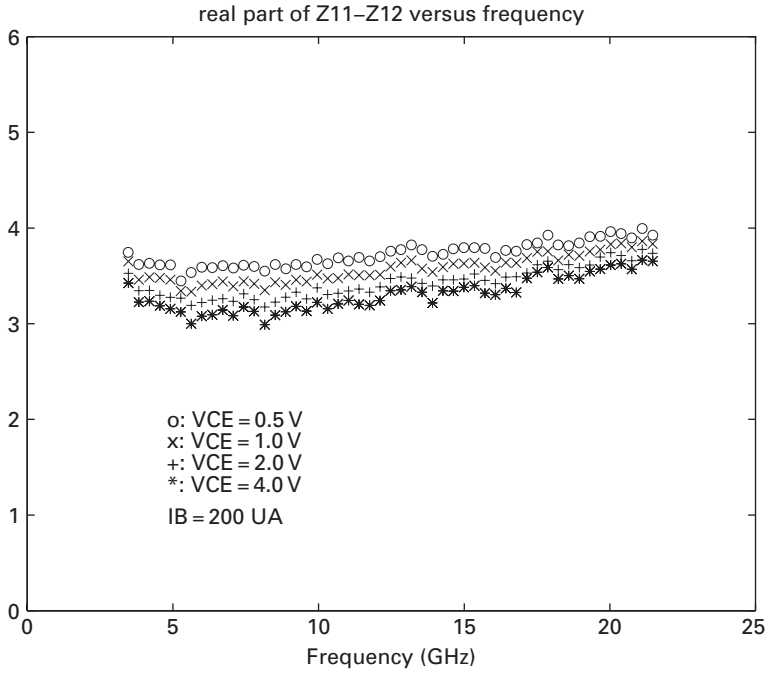

Fig. 3.36 The values of $R_{bx} + R_{bi}(C_{bc}/C_s)$ after de-embedding the pad capacitances (B. Li, S. Prasad, L.-W. Yang and S. C. Wang, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 46, No. 10, pp. 1427–1435, October 1998. ©1998 IEEE).

The pad capacitances have a significant effect on the value obtained for $R_{bx} + R_{bi}(C_{bc}/C_s)$. The maximum variations of $R_{bx} + R_{bi}(C_{bc}/C_s)$ before and after de-embedding the pad capacitances are 5 and 0.5 respectively. The result after removing the pad parasitics is shown in Figure 3.36.

This value is used to constrain the optimisation procedure in order to obtain accurate values of R_{bx} and R_{bi}. The variation of $R_{bx} + R_{bi}(C_{bc}/C_s)$ is due to the change of C_{bc} with the base–collector voltage V_{CE}. C_{bc} decreases as V_{CE} increases. This causes the ratio $C_{bc}/(C_{bc} + C_f)$ to decrease and hence the magnitude of $R_{bx} + R_{bi}(C_{bc}/C_s)$ decreases.

H. The emitter resistance R_E and base–emitter resistance r_e

$R_E + r_e$ can be obtained from the real part of Z_{12} in the intermediate frequency range. With the high collector current where the neutral base recombination is the dominant recombination, $R_E + r_e$ can be expressed as

$$R_E + r_e = R_E + \frac{n_f\, kT}{q\, I_E}. \tag{3.37}$$

The real part of Z_{12} in the intermediate frequency range is the sum of $r_e + R_E$. The plot of $r_e + R_E$ versus $1/I_E$ would give the values of R_E, r_e and ideality factor n_f.

I. The emitter lead inductor and base–emitter capacitance

$L_e - C_{be}r_e^2$ can be obtained from the imaginary part of Z_{12} in the low middle frequency range. In the case of high collector currents, the fraction of the depletion capacitance

in the base–emitter capacitance C_{be} is small and C_{be} can be approximated to be proportional to I_E, and we also have $r_e \propto 1/I_E$. Therefore, the y intercept of the plot of $L_e - r_e^2 C_{be}$ versus $1/I_E$ gives the value of L_e. The value of $L_e - C_{be} r_e^2$ at $f = 5\,\text{GHz}$ is used for this purpose. $L_e - C_{be} r_e^2$ is plotted versus $1/I_e$ and shown in Figure 3.37.

Based on the values of L_e and r_e obtained previously, the value of C_{be} can be easily calculated. The value of C_{be} obtained in this way only serves to give the initial value of C_{be}. An accurate value of C_{be} is obtained from the optimisation procedure. It is noted that the magnitude of C_{be} is not sensitive to the optimisation procedure. This was also reported in [24] where the value of C_{be} is calculated from f_α (where f_α is the transport factor α cutoff frequency and $f_\alpha = 1/r_e C_{be}$). An accurate value of C_{be} is extremely difficult to obtain since changing the value of C_{be} does not change the error of optimisation much over the bias ranges for this example.

J. The transport factor α

The transport factor α can be calculated directly by

$$\alpha = \frac{Z_{21} - Z_{12}}{Z_{22} - Z_{21} - R_c - j\omega L_c}. \tag{3.38}$$

Assuming a single-pole approximation, one can write

$$\alpha = \frac{\alpha_0}{1 + \dfrac{j\omega}{\omega_\alpha}} e^{-j\omega\tau}, \tag{3.39}$$

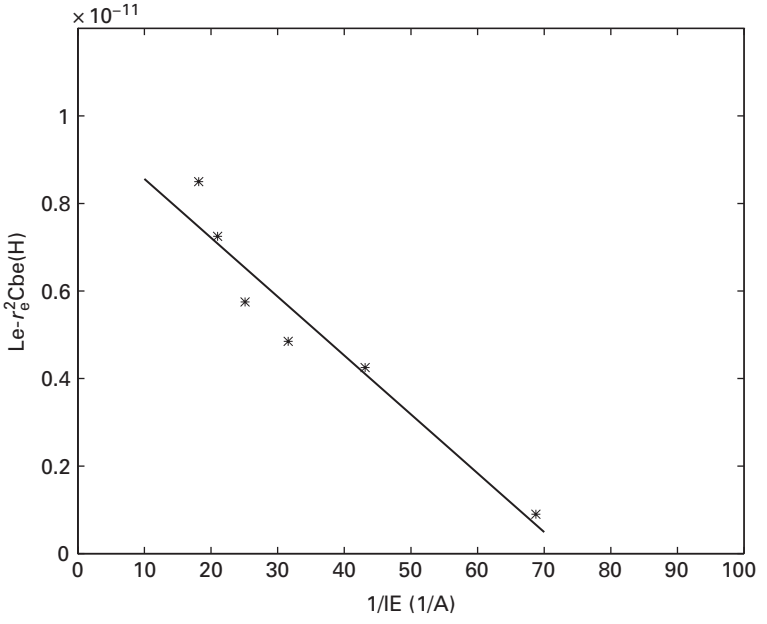

Fig. 3.37 $L_e - C_{be} r_e^2$ versus the inverse emitter current. $L_e = 9.99\,\text{pH}$ (B. Li, S. Prasad, L.-W. Yang and S. C. Wang, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 46, No. 10, pp. 1427–1435, October 1998. ©1998 IEEE).

where ω_α can be expressed as

$$\omega_\alpha = \frac{1}{C_{be} r_e} = \frac{1}{\tau_b} \tag{3.40}$$

where τ_b is the base transit time and is related to physical parameters by $W_{BC}^2/2D_n$ for npn HBTs. The magnitude of $\alpha(\omega)$ at $I_B = 200\,\mu A$ with different collector–emitter voltages is shown in Figure 3.38 together with the fitted curve of the magnitude of α.

α_0 is obtained by taking the value of $|\alpha|$ at low frequency and ω_α (and therefore the base transit time $\tau_b = 1/\omega_\alpha$) can be calculated directly at each frequency using

$$\tau_b = \frac{\sqrt{\alpha_0^2 - |\alpha(\omega)|^2}}{\omega|\alpha(\omega)|}. \tag{3.41}$$

The calculated τ_b at $I_B = 200\,\mu A$ with different collector–emitter voltages is shown in Figure 3.39. Since the base is heavily doped, the base width modulation effect in the HBT is negligible and therefore τ_b should be a weak function of I_B and V_{CE}. The dependence of τ_b on the base current and collector–emitter voltage is not completely clear. One possible reason for the α dependence on V_{CE} is the self-heating effect in the HBT. The diffusion coefficient $D_n = (kT/q)\mu_n$ is a function of the temperature in which $\mu_n \propto T^{-s}$. An often quoted value of s is 2.3 (for intrinsic GaAs). It is noted that the D_n decreases when the dissipated power in the HBTs increases. Therefore, the τ_b increases with larger V_{CE} values. The emitter–collector phase delay time can be calculated by

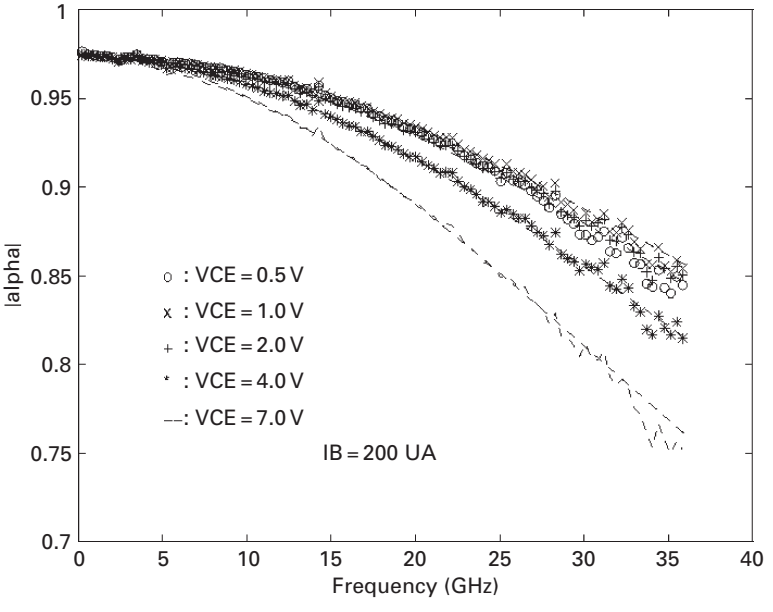

Fig. 3.38 $|\alpha|$ versus frequency at different collector emitter voltages (B. Li, S. Prasad, L.-W. Yang and S. C. Wang, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 46, No. 10, pp. 1427–1435, October 1998. ©1998 IEEE).

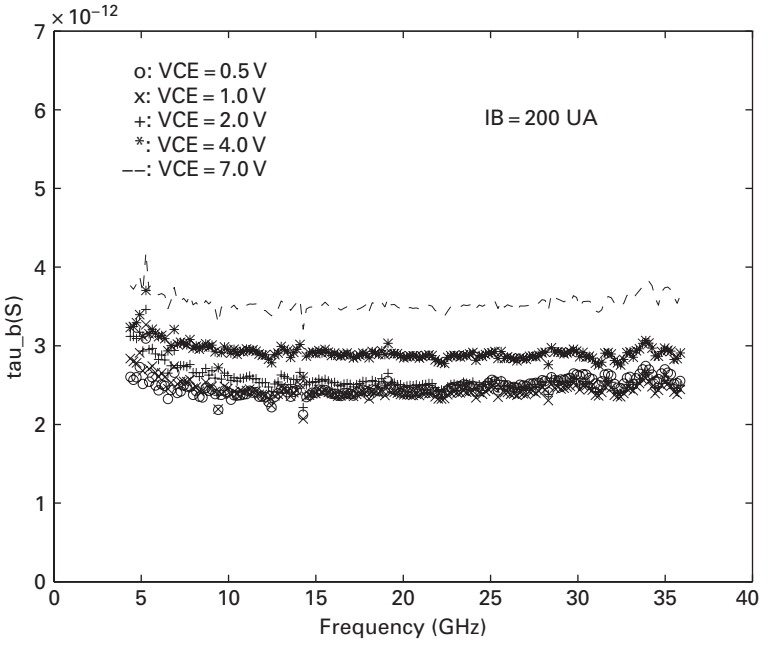

Fig. 3.39 τ_b versus frequency at different collector–emitter voltages (B. Li, S. Prasad, L.-W. Yang and S. C. Wang, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 46, No. 10, pp. 1427–1435, October 1998. ©1998 IEEE).

$$\tau = \frac{1}{\omega}\left[-\angle\alpha(\omega) - \tan^{-1}\left(\frac{\omega}{\omega_\alpha}\right)\right], \tag{3.42}$$

where $\tau = (m/1.2)\tau_b + \tau_c$ and $m \simeq 0.22$. The calculated τ versus frequency at $I_B = 200\,\mu A$ with different values of V_{CE} is shown in Figure 3.40.

When the collector–emitter voltage increases, the collector transit time $\tau_c = W_{BC}/2v_{sat}$ increases due to the larger base–collector space region. Therefore, the emitter–collector delay time increases as expected with the larger collector–emitter voltage. This might be explained by the self-heating effect in the HBT. The emitter–collector delay time is a monotonously increasing function of τ_b and τ_c. As the power dissipated in the HBT increases, the temperature of the intrinsic part of the HBT increases. That causes the diffusion coefficient D_n to decrease and results in the larger base transit time τ_b, and hence the larger emitter–collector delay time.

3.7.2 Results of the parameter extraction

The values of the bias-independent elements are given in Table 3.4.

All of the bias-independent elements are extracted from the procedure described above except for L_e, R_{bi} and R_{bx}. Accurate values of L_e, R_{bx} and R_{bi} are obtained from the empirical optimisation procedure. Let $a = R_{bx} + R_{bi}(C_{bc}/C_s)$ and $\gamma = C_{bc}/C_s$. The initial values of R_{bi} and R_{bx} are estimated from the variation of a. We have $R_{bx} = \Delta a/\Delta\gamma$. The calculated values of R_{bx} and R_{bi} are listed in Table 3.4. Instead of defining just the absolute error and just the relative error, the mixed relative and absolute

Table 3.4 The bias-independent parameters

Parameters	Values (analytical)	Values (optimised)
C_{pbc}(fF)	0	0
C_{pbe}(fF)	27.4	27.4
C_{pce}(fF)	41	42
C_f(fF)	16.5	16.5
L_b(pH)	55	55
L_e(pH)	9.9	5.46
L_c(pH)	61	61
R_{bx}(Ω)	1.38	1.42
R_{bi}(Ω)	2.3	4.049
R_E(Ω)	1.832	1.832
R_c(Ω)	4.99	4.99
Error	2.2 %	0.43 %

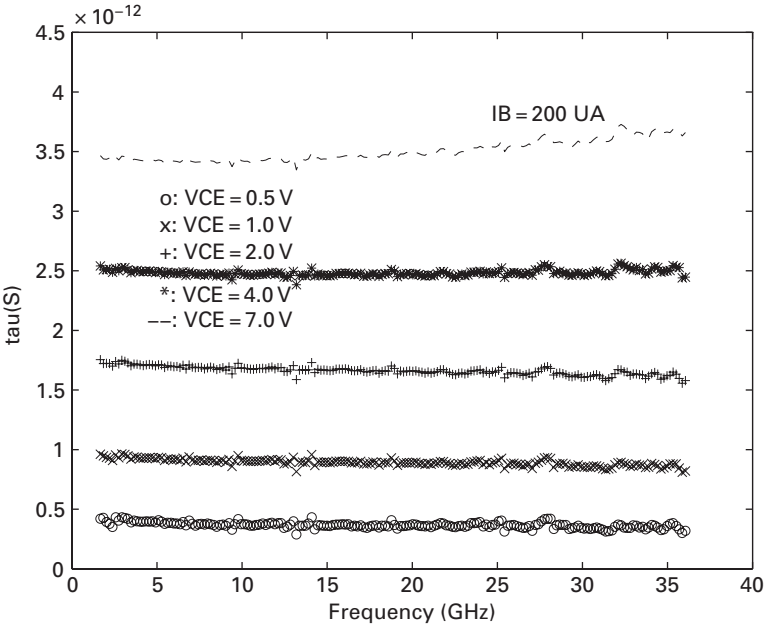

Fig. 3.40 τ versus frequency at different collector–emitter voltage (B. Li, S. Prasad, L.-W. Yang and S. C. Wang, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 46, No. 10, pp. 1427–1435, October 1998. ©1998 IEEE).

errors are used to obtain the best fit between the measurement and the simulation. The optimisation error is defined by

$$E = 1/4N \sum_{i=1}^{N} \left[\left(\frac{S_{11}^{mea} - S_{11}^{sim}}{S_{11}^{mea}} \right)^2 + \left(\frac{S_{12}^{mea} - S_{12}^{sim}}{S_{12}^{mea}} \right)^2 + \left(\frac{S_{21}^{mea} - S_{21}^{sim}}{S_{21}^{mea}} \right)^2 + \left(\frac{S_{22}^{mea} - S_{22}^{sim}}{S_{22}^{mea}} \right)^2 \right], \quad (3.43)$$

Table 3.5 The bias-dependent parameters

V_{CE}(V)	0.50000	1.0000	2.0000	4.0000	7.0000
I_C(mA)	6.0	6.2	6.1	5.9	5.6
C_{bc} (fF)	109.1	64.75	41.50	26.8	18.70
$r_e(\Omega)$	4.168	4.0432	4.218	4.518	5.018
$R_{bc}(k\Omega)$	8.6	70	121	180	200
α_0	0.9741	0.9751	0.9751	0.9740	0.9740
C_{be}^{A}(fF)	344	344	344	344	344
f_α^{A}(GHz)	63	66	66	55	45
τ^{A} (ps)	0.42	0.95	1.7	2.52	3.42
C_{be}^{O} (fF)	289	238	239	268	293
f_α^{O}(GHz)	58.2	58.9	52.3	47.10	39.9
τ^{O}(ps)	0.158	0.69	1.09	2.08	3.14
Error^{A}	0.98%	0.84%	0.86%	0.62%	0.9%
Error^{O}	0.50%	0.69%	0.43%	0.41%	0.56%

where N is the number of the frequency points. The errors between the measured and simulated S-parameters are also listed in Table 3.4.

The optimisation is carried out at the bias $I_B = 200\,\mu A$ and $V_{CE} = 2.0V$. The error between the measurement and the simulation at the bias $I_B = 200\,\mu A$ and $V_{CE} = 2.0\,V$ before the optimisation is already 2.2%. The bias-dependent parameters $C_{be}, r_e, C_{bc}, R_{bc}, \alpha_0, f_\alpha$ and τ at constant base current $I_B = 200\,\mu A$ and $V_{CE} = \{0.5\,V, 1.0\,V, 2.0\,V, 4.0\,V, 7.0\,V\}$ are given in Table 3.5. Superscript A represents the results from optimisation and superscript O represents the results from the direct analysis. Once the values of the bias-independent elements are known, all the bias-dependent values can be easily calculated and no further optimisation is needed. It is seen in Table 3.5 that, by using the directly calculated values of the bias-dependent elements, the error between simulation and measurement is very small. All the errors are less than 1%. Optimisations are also used. Only the three elements C_{be}, f_α and τ are optimised. The errors after optimisation are given in the Table 3.5. The variation of r_e is dependent on the collector current and the self-heating effect. As explained previously, accurate values of C_{be} are very difficult to obtain. The variation of C_{be} may result from the numerical techniques. The bias-dependent parameters $C_{be}, r_e, C_{bc}, R_{bc}, \alpha_0, f_\alpha$ and τ at constant collector–emitter voltage $V_{CE} = 2.0\,V$ and $I_B = \{200\,\mu A, 400\,\mu A, 600\,\mu A, 800\,\mu A, 1000\,\mu A\}$ are given in Table 3.6. The errors in using the analytical approach and in using the optimisation procedure based on the initial values obtained from the analytical approach are both given. It is seen that the errors, using the analytical approach, become higher if the collector currents increase. This is because the self-heating effect becomes more significant when the collector currents increase. However, the bias-independent elements are forced to be fixed in all the extraction procedures and they are practically functions of the device temperature. The thermal effect is absorbed by the bias-dependent elements after optimisation. Thus the errors become smaller. As expected, C_{be} increases with increased collector

Table 3.6 The bias-dependent parameters

$I_B(\mu A)$	200.000	400.000	600.000	800.000	1000.000
$I_C(mA)$	6.1	14.1	22.6	30.19	38.8
$C_{bc}(fF)$	41.50	37.4	32.7	28.75	27.15
$r_e(\Omega)$	4.218	1.798	1.133	0.8482	0.6912
$R_{bc}(k\Omega)$	200	200	200	200	200
α_0	0.9751	0.9785	0.9796	0.9790	0.9790
$C_{be}^{A}(fF)$	344	863	1275	1704	2190
$f_\alpha^{A}(GHz)$	66	72	94	114	102
$\tau^{A}(ps)$	1.7	1.6	1.52	1.73	1.2
$C_{be}^{O}(fF)$	0.239	0.682	1.070	1.537	1.984
$f_\alpha^{O}(GHz)$	52.3	64.4	79.1	88.9	81.2
$\tau^{O}(ps)$	1.09	1.02	0.98	0.68	0.45
$Error^{A}$	0.86%	2.6%	3.2%	5.2%	7.6%
$Error^{O}$	0.43%	0.62%	1.1%	2.1%	2.9%

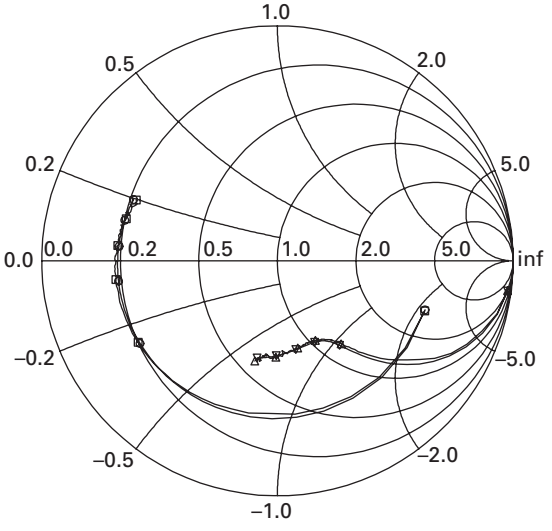

Fig. 3.41 The simulated and measured S_{11} and S_{22}. ○: measured S_{11}; □: simulated S_{11}; ∇: measured S_{22}; △: simulated S_{22} (B. Li, S. Prasad, L.-W. Yang and S. C. Wang, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 46, No. 10, pp. 1427–1435, October 1998. ©1998 IEEE).

currents. It is observed from Table 3.6 that C_{bc} decreases with the increased collector currents. One possible explanation is the self-heating effect and the modification of the base–collector space charge region by the injected carriers [24]. The effect of self-heating is not discussed here. The simulated and measured S-parameters at the bias $I_B = 200\,\mu A$ and $V_{CE} = 2.0\,V$ are shown in Figures 3.41–3.43. The excellent fit between the measured and modelled data shows that this procedure may be used successfully in device parameter extraction.

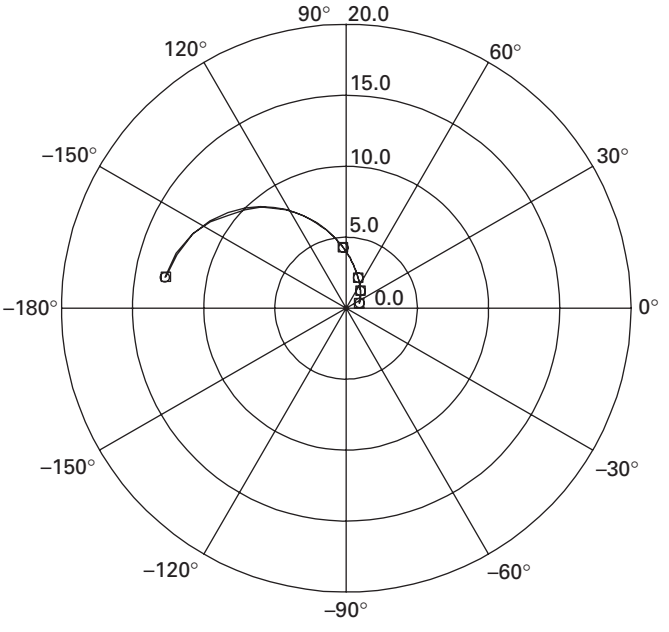

Fig. 3.42 The simulated and measured S_{21}. ○: measured S_{21}; □: simulated S_{21} (B. Li, S. Prasad, L.-W. Yang and S. C. Wang, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 46, No. 10, pp. 1427–1435, October 1998. ©1998 IEEE).

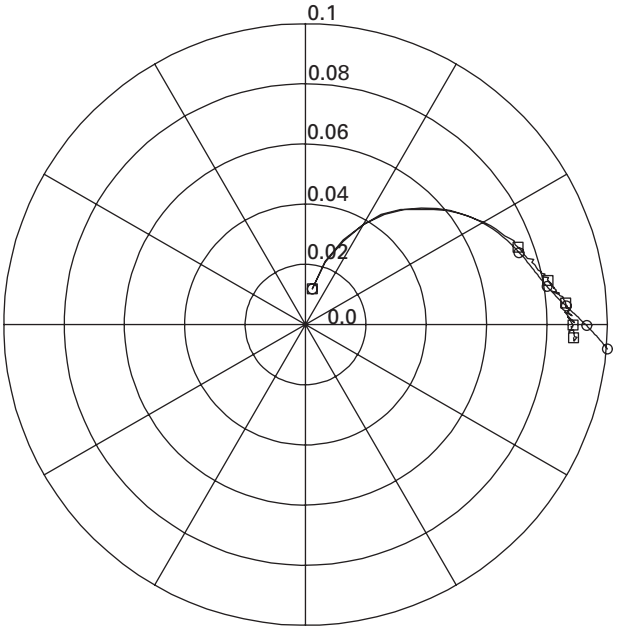

Fig. 3.43 The simulated and measured S_{12}. ○: measured S_{12}; □: simulated S_{12} (B. Li, S. Prasad, L.-W. Yang and S. C. Wang, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 46, No. 10, pp. 1427–1435, October 1998. ©1998 IEEE).

In summary, the pad capacitances are extracted from the HBTs under cutoff operation. Most of the elements are obtained from the analysis of the behaviour of the Z-parameters. The values of uncertain elements are obtained from the optimisation at a specific bias. The initial values of these uncertain elements are also obtained from the analytical approach.

3.8 Basic expressions for small-signal parameter extraction

Small-signal equivalent circuit parameter extraction has been much addressed by several researchers [3, 6, 19, 20, 25, 26, 33, 34]. The basic expressions and approximations for the Z-parameters in different frequency ranges and under different bias conditions are treated in this section [14].

3.8.1 Theoretical approximations of Z-parameters for the HBT

The simplified T equivalent circuit of the HBT after de-embedding pad capacitances is shown in Figure 3.44. The T equivalent circuit is used here since it is more physically meaningful than the π equivalent circuit. The latter circuit is the mathematical representation of transistor operation. It is also easier to infer the large-signal model from the

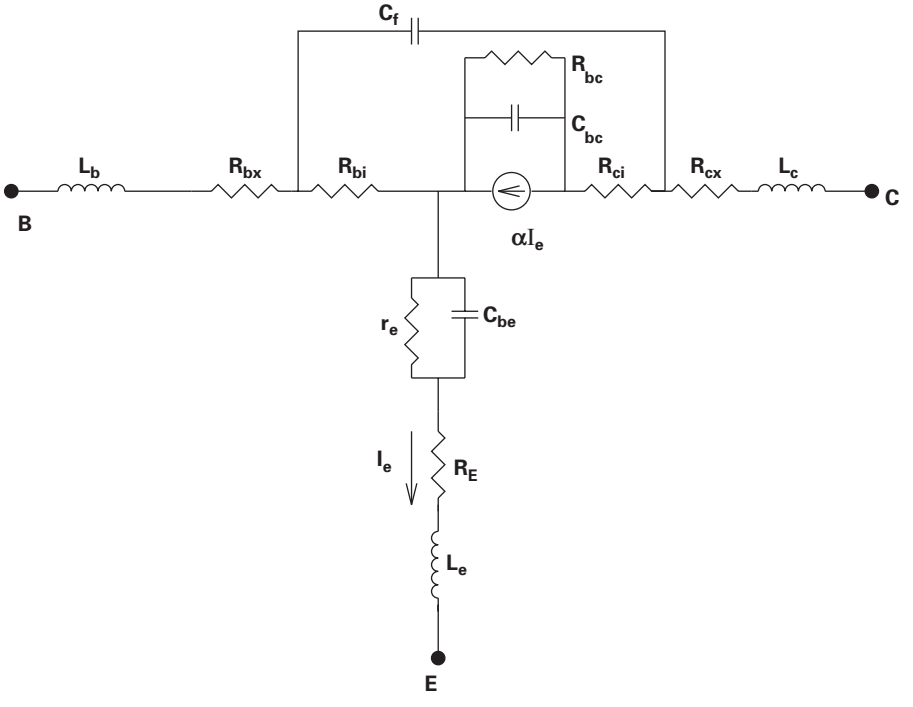

Fig. 3.44 Simplified T-type equivalent circuit, in which pad parasitics have been de-embedded (B. Li and S. Prasad, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 47, No. 5, pp. 534–539, May 1999. ©1999 IEEE).

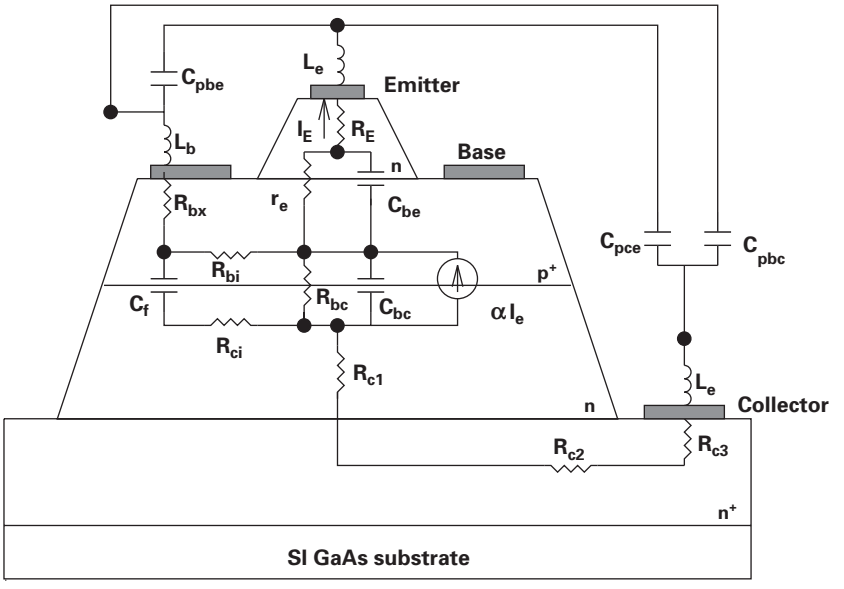

Fig. 3.45 Physical significance of the elements in the small-signal equivalent circuit (B. Li and S. Prasad, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 47, No. 5, pp. 534–539, May 1999. ©1999 IEEE).

bias-dependent element parameters by the T equivalent circuit since there are straightforward physical equations for the bias-dependent intrinsic elements. Figure 3.45 depicts the physical significance of each of the circuit elements in Figure 3.44 and also includes the extrinsic parasitic capacitances C_{pbe}, C_{pbc} and C_{pce}. The active portion of the HBT was modelled using C_{be}, C_{bc}, r_{e}, αI_{e}, R_{bc} and C_{f}. R_{E} is the extrinsic emitter resistance which consists of the contact resistance and emitter region resistance. The extrinsic collector resistance was divided into three parts: R_{c1}, R_{c2} and R_{c3}, which are, respectively, the resistance due to the n-collector, the n^+ access region, and the collector contact. The intrinsic collector resistance is represented by R_{ci} which characterises the distribution effect of the base–collector junction at the collector side. R_{c1}, R_{c2} and R_{c3} are lumped together as R_{cx} in Figure 3.44. Similarly, the extrinsic base resistance consists of a contact resistance R_{b1} and an access resistance R_{b2}. R_{b1} and R_{b2} are lumped together as R_{bx} in Figure 3.44. R_{bi} is the intrinsic base resistance. Finally, the distribution effect of the base–collector junction is modelled by the elements R_{bi}, R_{bx}, C_{f}, C_{bc} and R_{bc}. C_{pbe}, C_{pbc} and C_{pce} modelled the coupling between the base–emitter, the base–collector and the collector–emitter interconnection layers. L_{e}, L_{b} and L_{c} are the contact leads of the emitter, the base and the collector respectively:

$$Z_{\mathrm{BC}} = \frac{R_{\mathrm{bc}}}{1 + j\omega R_{\mathrm{bc}} C_{\mathrm{bc}}} \tag{3.44}$$

$$Z_{\mathrm{E}} = R_{\mathrm{E}} + j\omega L_{\mathrm{e}} + \frac{r_{\mathrm{e}}}{1 + j\omega r_{\mathrm{e}} C_{\mathrm{be}}} \tag{3.45}$$

$$Z_F = \frac{1}{j\omega C_f} \tag{3.46}$$

$$Z_C = R_c + j\omega L_c \tag{3.47}$$

$$Z_B = R_{bx} + j\omega L_b \tag{3.48}$$

$$\alpha = \alpha_0 \frac{1}{1 + \dfrac{j\omega}{\omega_\alpha}} e^{-j\omega\tau}. \tag{3.49}$$

The two-port network Z-parameters are given as follows (the derivation for two-port Z-parameters of this equivalent circuit is straightforward using basic circuit theory and the details are not given here):

$$\begin{aligned}
Z_{11} &= Z_B + Z_E + \frac{R_{bi}[Z_F + R_{ci} + (1-\alpha)Z_{BC}]}{\triangle} \\
Z_{12} &= Z_E + R_{bi} \times \frac{R_{ci} + (1-\alpha)Z_{BC}}{\triangle} \\
Z_{21} &= Z_E + \frac{R_{ci}R_{bi} + (1-\alpha)Z_{BC}R_{bi} - \alpha Z_F Z_{BC}}{\triangle} \\
Z_{22} &= Z_C + Z_E + \frac{[(1-\alpha)Z_{BC} + R_{ci}](Z_F + R_{bi})}{\triangle},
\end{aligned} \tag{3.50}$$

where $\triangle = R_{bi} + R_{ci} + Z_F + Z_{BC}$.

Let $R_{ci} = 0$ (in most cases, it is hard to distinguish between R_{ci} and R_{cx}, and R_{ci} and R_{cx} are lumped together as R_c), the following equations are obtained after some simple calculations:

$$Z_{11} - Z_{12} = Z_B + \frac{Z_F R_{bi}}{Z_{BC} + Z_F + R_{bi}} \tag{3.51}$$

$$Z_{12} = Z_E + \frac{(1-\alpha)Z_{BC}R_{bi}}{Z_{BC} + Z_F + R_{bi}} \tag{3.52}$$

$$Z_{12} - Z_{21} = \frac{\alpha Z_F Z_{BC}}{Z_{BC} + Z_F + R_{bi}} \tag{3.53}$$

$$Z_{22} - Z_{21} = Z_C + \frac{Z_F Z_{BC}}{Z_{BC} + Z_F + R_{bi}}. \tag{3.54}$$

The most important term is

$$\Re = \frac{Z_F Z_{BC}}{Z_{BC} + Z_F + R_{bi}}. \tag{3.55}$$

This term $\Re$ is a complicated function of frequency and all the right sides of the Equations (3.51)–(3.54) can be expressed as the sum of a simple function of frequency and this term (except for a scaling constant).

Within the extreme low frequency range in which $\omega C_{bc} R_{bc} \ll 1$,

$$\begin{aligned}\Re &\simeq \frac{Z_F Z_{BC}}{Z_{BC} + Z_F} \\ &\simeq \frac{R_{bc}}{j\omega(C_f + C_{bc})R_{bc} + 1} \\ &\simeq R_{bc} - j\omega R_{bc}^2 C_s,\end{aligned} \tag{3.56}$$

where we have used the assumption $R_{bc} \gg R_{bi}$ and let $C_s = C_f + C_{bc}$, since this is almost always the case in the forward or saturation applications. A typical extreme low frequency can be calculated by using the typical element values. In modern process technology in which the base resistance and base–collector capacitance have been greatly reduced, the typical values $C_{bc} \simeq 5 \times 10^{-14}$ F and $R_{bc} \simeq 5 \times 10^4\ \Omega$, therefore the frequency $f \simeq 0.1/(2\pi R_{bc} C_{bc}) = 6.8$ MHz. For microwave applications, this typical frequency is very low. Substituting the expression in Equation (3.56) into Equations (3.51), (3.52) and (3.54), the equations reduce to

$$Z_{11} - Z_{12} = R_{bx} + j\omega L_b + R_{bi} - j\omega R_{bc} R_{bi} C_s \tag{3.57}$$

$$\begin{aligned}Z_{12} = {} & R_E + j\omega L_e + r_e - j\omega r_e^2 C_{be} \\ & + (1 - \alpha) \times j\omega R_{bi} C_f (R_{bc} - j\omega R_{bc}^2 C_s)\end{aligned} \tag{3.58}$$

$$Z_{22} - Z_{21} = R_C + j\omega L_c + R_{bc} - j\omega R_{bc}^2 C_s. \tag{3.59}$$

The base–emitter resistance is small in forward bias and therefore, Z_E is approximated as $R_E + j\omega L_e + r_e - j\omega r_e^2 C_{be}$ in this analysis. Z_{12} can be simplified by discarding the higher-order terms:

$$Z_{12} = R_E + j\omega L_e + r_e - j\omega r_e^2 C_{be} + j(1 - \alpha)\omega R_{bi} R_{bc} C_s. \tag{3.60}$$

At low frequency, $\alpha \to \alpha_{DC} \to 1$. Hence, Z_{12} is further simplified as

$$Z_{12} = R_E + j\omega L_e + r_e - j\omega r_e^2 C_{be}. \tag{3.61}$$

The intermediate frequency range is characterised by $\omega C_{bc} R_{bi}, \omega C_f R_{bi} \ll 1$, but $\omega C_{bc} R_{bc} \gg 1$. The approximation for $\Re$ is given as

$$\begin{aligned}\Re &= \frac{Z_F Z_{BC}}{Z_F + Z_{BC} + R_{bi}} \\ &\simeq \frac{Z_F Z_{BC}}{Z_F + Z_{BC}} \times \left(1 - \frac{R_{bi}}{Z_F + Z_{BC}}\right) \\ &\simeq \left(\frac{1}{j\omega C_s} + \frac{1}{\omega^2 C_s^2 R_{bc}}\right)\left(1 - \frac{j\omega C_f C_{bc} R_{bi}}{C_s}\right) \\ &\simeq \frac{1}{j\omega C_s} + \frac{1}{\omega^2 C_s^2 R_{bc}} - \frac{R_{bi} C_{bc} C_f}{C_s^2} - \frac{j R_{bi} C_f C_{bc}}{\omega C_s^3 R_{bc}} \\ &\simeq \frac{1}{j\omega C_s} + \frac{1}{\omega^2 C_s^2 R_{bc}} - \frac{R_{bi} C_{bc} C_f}{C_s^2},\end{aligned} \tag{3.62, 3.63}$$

where higher-order terms have been discarded. A typical intermediate frequency is calculated by using the typical element values. Assume $R_{bi} = 10\,\Omega$, the typical value of frequency is $0.1/(2\pi R_{bi}C_{bc}) \simeq 0.64\,\text{GHz}$. The value $\omega C_{bc}R_{bc} \simeq 10$ appropriately justifies the assumption. Substituting the approximation in Equation (3.63) into Equations (3.51), (3.52) and (3.54), we obtain

$$Z_{11} - Z_{12} = R_{bx} + j\omega L_b + R_{bi}\frac{C_{bc}}{C_s} - \frac{jR_{bi}C_f}{\omega R_{bc}C_s^2} \tag{3.64}$$

$$Z_{12} = R_E + j\omega L_e + r_e - j\omega r_e^2 C_{be} + (1-\alpha)R_{bi}\left(\frac{C_f}{C_s}\right) \tag{3.65}$$

$$Z_{22} - Z_{21} = R_C + j\omega L_c + \frac{1}{j\omega C_s} + \frac{1}{\omega^2 R_{bc}C_s} - \frac{R_{bi}C_{bc}C_f}{C_s^2}. \tag{3.66}$$

As indicated in the above expressions, initially $1/\omega^2(C_f + C_{bc})^2 R_{bc}$ is much larger than $R_{bi}C_{bc}C_f/C_s^2$ and R_c. $R_e(Z_{22} - Z_{21})$, where R_e denotes the real part can hence be used to extract R_{bc}. With increase of frequency, the second term decreases rapidly and $\text{Re}(Z_{22} - Z_{21})$ is possible to be negative as described in [20], where the intrinsic base resistance R_{bi} is much higher in InP-based HBTs than in GaAs-based HBTs. After the higher-order terms are discarded, Equations (3.64) and (3.65) are reduced to

$$Z_{11} - Z_{12} = R_{bx} + j\omega L_b + R_{bi}\frac{C_{bc}}{C_s} \tag{3.67}$$

$$Z_{12} = R_E + j\omega L_e + r_e - j\omega r_e^2 C_{be} \tag{3.68}$$

The fifth term in Equation (3.65) consists of only a very small fraction of the real part of Z_{12}; therefore, the approximation is generally reasonable. However, for InP-based HBTs it may account for up to 18% of the real part of Z_{12} due to the large base resistance [20]. Self-consistent iterations can be used to correct the approximation errors [20].

The very high frequency range is defined by $\omega C_{bc}R_{bi}, \omega C_f R_{bi} \gg 1$

$$\begin{aligned}\Re &\simeq \frac{Z_F Z_{BC}}{R_{bi}} \times \frac{1}{1 + \dfrac{Z_F + Z_{BC}}{R_{bi}}} \\ &\simeq -\frac{1}{\omega^2 C_{bc}C_f R_{bi}} \times \left(1 - \frac{Z_F + Z_{BC}}{R_{bi}}\right) \\ &\simeq -\frac{1}{\omega^2 R_{bi}C_{bc}C_f} + \frac{1}{\omega^2 R_{bi}C_{bc}C_f}\left(\frac{1}{j\omega C_f R_{bi}} + \frac{1}{j\omega C_{bc}R_{bi}}\right).\end{aligned} \tag{3.69}$$

The typical very high frequency is calculated by the values given above: $f = 10/(2\pi R_{bi}C_{bi}) = 63.8\,\text{GHz}$. Since the very high frequency is very hard to reach, we normally relax the high frequency requirement to above 20 GHz. Substituting Equation (3.69) into Equations (3.51), (3.52) and (3.54), we obtain

$$Z_{11} - Z_{12} = R_{bx} + j\omega L_b + \frac{1}{j\omega C_f} + \frac{1}{\omega^2 R_{bi}C_{bc}C_f}\left(\frac{C_{bc}}{C_f + 1}\right) \tag{3.70}$$

$$Z_{12} = R_{\mathrm{E}} + j\omega L_{\mathrm{e}} + r_{\mathrm{e}} - j\omega r_{\mathrm{e}}^2 C_{\mathrm{be}} + \frac{(1-\alpha)}{j\omega C_{\mathrm{bc}}} + \frac{(1-\alpha)}{\omega^2 R_{\mathrm{bi}} C_{\mathrm{bc}} C_{\mathrm{f}}}\left(\frac{C_{\mathrm{f}}}{C_{\mathrm{bc}}} + 1\right) \tag{3.71}$$

$$Z_{22} - Z_{21} = R_{\mathrm{C}} + j\omega L_{\mathrm{c}} - \frac{1}{\omega^2 R_{\mathrm{bi}} C_{\mathrm{bc}} C_{\mathrm{f}}} + \frac{1}{\omega^2 R_{\mathrm{bi}} C_{\mathrm{bc}} C_{\mathrm{f}}}\left(\frac{1}{j\omega C_{\mathrm{f}} R_{\mathrm{bi}}} + \frac{1}{j\omega C_{\mathrm{bc}} R_{\mathrm{bi}}}\right). \tag{3.72}$$

We divide the frequency range into three subdivisions. The boundaries of these three ranges are not strict and highly dependent on the applied biases, process technology and device design. The magnitudes of R_{bc}, R_{bi}, C_{F}, C_{bc} and r_{e} affect the definitions of the frequency range. The network analyser is used to measure S-parameters and then the S-parameters are converted to Z-parameters. The lowest frequency of the network analyser is about several tenths of MHz. Therefore, S-parameter data at the low frequency (satisfied by the condition $\omega R_{\mathrm{bc}} C_{\mathrm{s}} \ll 1$) are not of much use for analysis except in the reverse or deep saturation region.

For the analytical approach to parameter extraction, we clarify that the intermediate frequency range is below 5 GHz and the high frequency range is greater than 10 GHz. The method of clarifying the frequency range before extracting the element parameters is arbitrary; however, the justification needs to be made during the procedure of parameter extraction.

3.8.2 Z-Parameters at zero bias or 'cold' biases

At zero bias ($V_{\mathrm{BE}} = 0\,\mathrm{V}$ and $V_{\mathrm{CE}} = 0\,\mathrm{V}$) or 'cold' bias ($V_{\mathrm{BE}} = 0\,\mathrm{V}$ and $I_{\mathrm{B}} = 0\,\mathrm{A}$; 'cold' is adopted from the MESFET modelling techniques), the current gain becomes zero and the device behaves like a passive component. The base–emitter resistance r_{e} is also much larger. The magnitude of this resistance is normally greater than $10^4\ \Omega$, which is within the same order of magnitude as the collector–base resistance. Therefore, the equivalent circuit becomes much simpler and so are the Z-parameters. With $\alpha = 0$, we have

$$Z_{11} - Z_{12} = Z_{\mathrm{B}} + \frac{Z_{\mathrm{F}} R_{\mathrm{bi}}}{Z_{\mathrm{BC}} + Z_{\mathrm{F}} + R_{\mathrm{bi}}} \tag{3.73}$$

$$Z_{12} = Z_{21} = Z_{\mathrm{E}} + \frac{Z_{\mathrm{BC}} R_{\mathrm{bi}}}{Z_{\mathrm{BC}} + Z_{\mathrm{F}} + R_{\mathrm{bi}}} \tag{3.74}$$

$$Z_{22} - Z_{21} = Z_{\mathrm{C}} + \frac{Z_{\mathrm{F}} Z_{\mathrm{BC}}}{Z_{\mathrm{BC}} + Z_{\mathrm{F}} + R_{\mathrm{bi}}}. \tag{3.75}$$

The approximation of the Z-parameters is almost the same as before except for Z_{12} and Z_{21}. The approximation of Z_{E} in this case will be rather different. The Z_{E} at the low frequency range in which $\omega r_{\mathrm{e}} C_{\mathrm{be0}} \ll 1$ is satisfied is given by

$$Z_{\mathrm{E}} \simeq R_{\mathrm{E}} + j\omega L_{\mathrm{e}} + r_{\mathrm{e}} - j\omega r_{\mathrm{e}}^2 C_{\mathrm{be0}}. \tag{3.76}$$

Substituting into Equation (3.74), Z_{12} can be written as

$$Z_{12} \simeq R_{\mathrm{E}} + j\omega L_{\mathrm{e}} + r_{\mathrm{e}} - j\omega r_{\mathrm{e}}^2 C_{\mathrm{be0}} + j\omega R_{\mathrm{bi}} C_{\mathrm{f}}(R_{\mathrm{bc}} - j\omega R_{\mathrm{bc}}^2 C_{\mathrm{s}}). \tag{3.77}$$

For simplicity, higher-order terms can be discarded in the above equation without affecting the accuracy.

When the frequency increases to the point at which $\omega r_e C_{be0} \gg 1$, Z_E is approximated as

$$Z_E = R_E + j\omega L_e + \frac{1}{j\omega C_{be0}} + \frac{1}{\omega^2 r_e C_{be}^2}. \tag{3.78}$$

Actually, the assumption of $\omega r_e C_{be0}$, $\omega R_{bc} C_{bc0} \gg 1$ and $\omega C_{be0} R_{bi}$, $\omega C_{bc0} R_{bi} \ll 1$ is easily satisfied in the frequency range of interest; therefore,

$$Z_{11} - Z_{12} = R_{bx} + j\omega L_b + \frac{C_{bc}}{C_s} R_{bi} \tag{3.79}$$

$$Z_{12} = R_E + j\omega L_e + \frac{1}{j\omega C_{be0}} + \frac{1}{\omega^2 r_e C_{be}^2} + \frac{C_f}{C_s} R_{bi} \tag{3.80}$$

$$Z_{22} - Z_{21} = R_C + j\omega L_c + \frac{1}{j\omega C_s} + \frac{1}{\omega^2 R_{bc} C_s^2}. \tag{3.81}$$

In the above equations, the higher-order terms have been discarded for simplicity. If the frequency increases further and $1/(\omega C_{be0})$, $1/(\omega C_{bc0}) \ll R_{bi}$, then Z_{12} is given by

$$Z_{12} = R_E + j\omega L_e + \frac{1}{j\omega C_{be0}} + \frac{1}{\omega^2 r_e C_{be}^2} + \frac{1}{j\omega C_{bc}} + \frac{1}{\omega^2 R_{bi} C_{bc}} \left(\frac{1}{C_{bc}} + \frac{1}{C_f} \right). \tag{3.82}$$

In the case of both junctions being reverse-biased, the approximation of the Z-parameters might be rather similar to this case at zero bias.

3.8.3 Parameter extraction

The model parameter extractions are based on the assumption that all the extrinsic and parasitic elements are bias-independent and only elements in the intrinsic part of the device vary with bias. This also means that the extrinsic and parasitic elements can be extracted by using multiple bias information. Some element values are sensitive to the bias point at which the extraction procedure is carried out.

Since the conditions at the intermediate frequency are easily satisfied in the frequency measurement range, we should make the most use of the Z-parameters in the intermediate frequency range. The extraction procedure is shown in Figures 3.46–3.49. The information for extracting the element parameters is overwhelming; hence different schemes can be developed. The extraction procedure shown here is just one of them. The superscripts used are explained as follows: R represents the real part of the corresponding Z-parameters; I the imaginary part; F the forward bias; 0 the zero bias; L the low frequency range; M the intermediate frequency range; and H the high frequency range.

From $Z_{11} - Z_{12}$, we extract L_b by plotting $\mathrm{Im}(Z_{11} - Z_{12})/\omega$ versus frequency in high frequency range. $R_{bx} + R_{bi}$ can be obtained from $\mathrm{Re}(Z_{11} - Z_{12})$ at the extreme low frequency. R_{bx} might be obtained from $\mathrm{Re}(Z_{11} - Z_{12})$ at the high frequency. The difference will be R_{bi}.

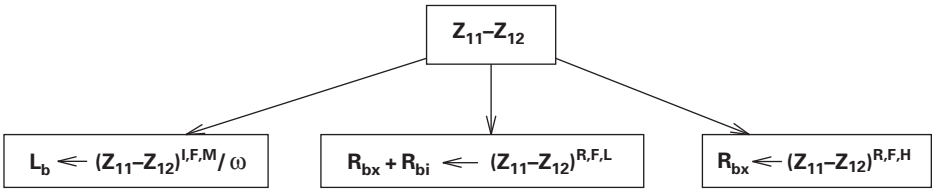

Fig. 3.46 The element values extracted from $Z_{11} - Z_{12}$ (B. Li and S. Prasad, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 47, No. 5, pp. 534–539, May 1999. ©1999 IEEE).

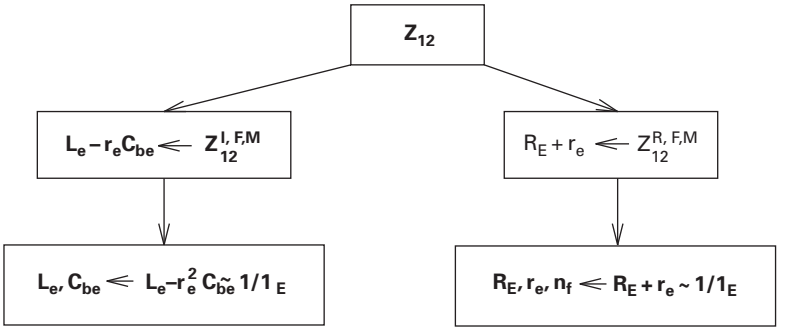

Fig. 3.47 The element values extracted from Z_{12} (B. Li and S. Prasad, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 47, No. 5, pp. 534–539, May 1999. ©1999 IEEE).

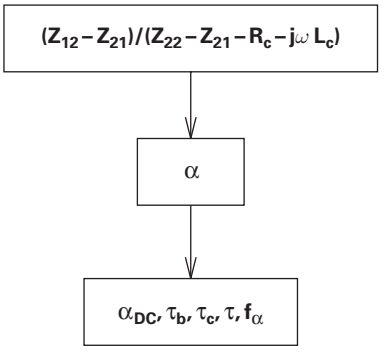

Fig. 3.48 The element values extracted from $(Z_{12} - Z_{21})/(Z_{22} - Z_{21} - R_c - j\omega L_c)$ (B. Li and S. Prasad, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 47, No. 5, pp. 534–539, May 1999. ©1999 IEEE).

Figure 3.47 shows how to extract R_E, r_e, L_e and C_{be}. $R_E + r_e$ is easily obtained from the real part of Z_{12} in the low or intermediate frequency range and we have

$$R_E + r_e = R_E + \frac{n_f kT}{q I_E}. \tag{3.83}$$

Therefore, we plot the curve $R_E + r_e$ versus $1/I_E$. R_E, r_e and n_f should be easily extracted. In the case of high collector currents, the fraction of the depletion capacitance

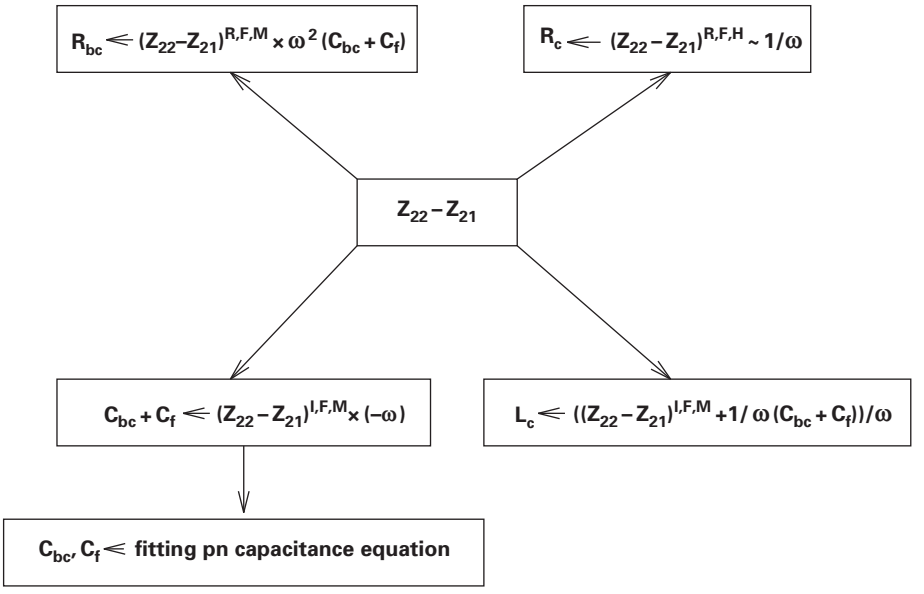

Fig. 3.49 The element values extracted from $Z_{22} - Z_{21}$ (B. Li and S. Prasad, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 47, No. 5, pp. 534–539, May 1999. ©1999 IEEE).

in the base–emitter capacitance C_{be} is small and C_{be} can be approximated to be proportional to I_E, and we also have $r_e \propto 1/I_E$. Therefore, the y intercept of plot of $L_e - r_e^2 C_{be}$ versus $1/I_E$ gives the value of L_e. Once L_e is known, C_{be} can be easily calculated.

The extraction from $Z_{22} - Z_{21}$ is somewhat more complex. R_{bc} is extracted from the expression $\mathrm{Re}(Z_{22} - Z_{21})\omega^2 C_s$ at the intermediate frequency or from the real part of $Z_{22} - Z_{21}$ at the low frequency. R_c is the vertical axis intercept of the plot of $\mathrm{Re}(Z_{22} - Z_{21})$ versus $1/\omega^2$ in the high frequency range. $C_f + C_{bc}$ is extracted from $(Z_{22} - Z_{21}) \times (-\omega)$ at the intermediate frequency. There are two different assumptions about C_f. One assumes that C_f is bias-independent. Therefore, we can obtain C_f and C_{bc} by fitting $C_f + C_{bc}$ into the following bias-dependent equation:

$$C_f + C_{bc} = C_f + \frac{C_{jbc0}}{\left(1 - \dfrac{V_{BC}}{V_{jbc}}\right)^{M_{jbc}}}. \tag{3.84}$$

On the other hand, C_f might be considered to vary with the bias rather than be fixed as in the Gummel–Poon model. The fraction of C_f in the base–collector capacitance is constant. For this case, given R_{bx}, R_{bi} and C_{bc}/C_s can be obtained from the real part of $Z_{11} - Z_{12}$ at the intermediate frequency or calculated from a knowledge of the geometry. We can calculate C_{bc} and C_f once the fraction is given.

L_c is calculated by the expression $\mathrm{Im}[(Z_{22} - Z_{21}) + 1/(\omega(C_{bc} + C_f))]/\omega$ in the high frequency range.

α can be extracted easily once we know R_c and L_c. This is shown in Figure 3.47. The extraction of parameters related to α has been reported by Pehlke and Pavlidis [19].

3.8.4 The approximation at $R_{bi} = 0$

This is the case in which the distributed effect of the base–collector region is not necessarily taken into account. The equivalent circuit is reduced and so are the Z-parameters. Let $R_{bi} = 0$ in Equations (3.51)–(3.54), we have

$$Z_{11} - Z_{12} = R_b + j\omega L_b \tag{3.85}$$

$$Z_{12} = R_E + j\omega L_e + r_e - j\omega r_e^2 C_{be} \tag{3.86}$$

$$Z_{12} - Z_{21} = \alpha \times Z_{BC} \tag{3.87}$$

$$Z_{22} - Z_{21} = R_C + j\omega L_c + Z_{BC}. \tag{3.88}$$

The element parameter extraction becomes trivial in this particular situation.

In the extreme low frequency range,

$$Z_{22} - Z_{21} = R_C + j\omega L_c + R_{bc} - jR_{bc}^2 C_{bc}. \tag{3.89}$$

In the intermediate frequency range,

$$Z_{22} - Z_{21} = R_C + j\omega L_c + \frac{1}{j\omega C_{bc}} + \frac{1}{\omega^2 R_{bc} C_{bc}^2}. \tag{3.90}$$

R_b and L_b are given by Equation (3.85). Equation (3.86) gives R_E, r_e, L_e and C_{be}. Equation (3.88) gives R_c, R_{bc}, C_{bc} and L_c. The extraction of α is the same as described above.

The T small-signal equivalent circuit for HBTs after de-embedding pad capacitances has been used to derive the Z-parameter expressions. The simplification of the Z-parameters based on the range of measured frequencies was developed in order to provide guidelines for directly extracting element parameters. The information from the measured S-parameters is normally overwhelming and different schemes of extraction procedure could be developed from these approximations. One fully analytical extraction procedure has been provided here. Z-parameters at multiple biases are utilised to extract the small-signal parameters. These results can also be used for the parameter extractions of BJTs.

3.9 Small-signal model of the collector-up (inverted) HBT

The device used in this example is a $5 \times 10\,\mu m^2$ InGaAs/InAlAs/InP inverted HBT with $f_T = 23$ GHz and $f_{max} = 20$ GHz. The small-signal equivalent circuit of the device [17, 18] is shown in Figure 3.50.

L_b, L_c and L_e and C_{pbe}, C_{pbc} and C_{pce} are parasitic inductances and capacitances respectively, and R_b, R_c and R_e are extrinsic resistances. The active portion of the HBT is modelled by intrinsic elements C_e, r_e, C_{jc}, α and R_{jc}, where $\alpha = \alpha_F/[1 + j(f/f_\alpha)]e^{-j\omega\tau}$. α_F is the dc value of the transport factor, τ is the transit time of collector current and f_α is the α 3 dB frequency. RF-measurements indicate that the HBT under zero bias ($I_b = 0$ A, $V_{CE} = 0$ V) can be represented by a passive network. Therefore, in

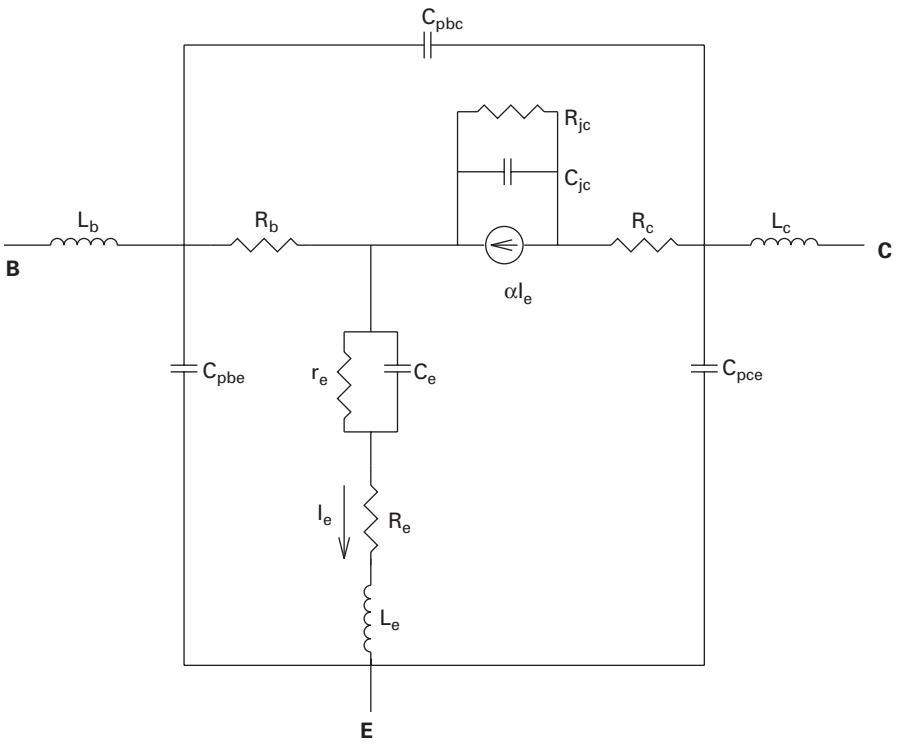

Fig. 3.50 Small-signal equivalent circuit of the inverted HBT (B. Li and S. Prasad, *IEEE Transactions on Microwave Theory and Techniques*, Vol. 45, No. 7, pp. 1135–1137, July 1997. ©1997 IEEE).

this case, the transport factor α is negligible. Moreover, the dynamic resistance of the base–collector p–n junction and the base–emitter p–n junction is assumed to be very large. The uncertainty of numerical optimisation in zero bias can be reduced. The initial values that are assumed in optimisation are calculated from the physical and geometrical parameters. The parasitic elements, C_{pbe}, C_{pbc}, C_{pce}, L_{b}, L_{e} and L_{c}, are obtained from zero bias numerical optimisation and assumed to be invariant with bias [18]. Their values are listed in Table 1.

The other elements under non-zero bias are extracted by the following analytical approach:

- Convert the S-parameters to Z-parameters and remove the parasitic series elements L_{b} and L_{c};
- Convert the Z-parameters to Y-parameters and remove the parasitic shunt elements C_{pbe}, C_{pce} and C_{pbc};
- Convert the Y-parameters to h-parameters.

The elements of the equivalent circuit, excluding the parasitic effects, are easily extracted using the procedure described in [19].

The elements R_{b}, L_{e}, R_{e} and R_{c} are basically constant over the entire frequency range of interest and do not show significant variation with bias. Therefore,

these elements can be considered to be fixed. The bias-dependent elements are C_e, C_{jc}, R_{jc}, α_F, τ, f_α and r_e. The consideration of the bias variation of these elements is sufficient for accurate small-signal modelling [17, 18].

3.10 Problems

(1) The scattering parameters of the HEMT in the common source configuration are given. The Y-parameters are related to the equivalent circuit parameters as follows:

$$Y_{11} + Y_{12} = j\omega C_{GS} \tag{3.91}$$

$$-Y_{12} = j\omega C_{GD} \tag{3.92}$$

$$Y_{22} + Y_{12} = g_D + j\omega C_{DS} \tag{3.93}$$

$$Y_{21} - Y_{12} = \frac{g_m}{1 + j\left(\frac{\omega}{\omega_0}\right)\tau_1}, \tag{3.94}$$

where

$$\tau_1 = \omega_0 \frac{-C_{DS}}{g_D} \tag{3.95}$$

$$\omega_0 = \frac{\mu_n (V_{GS} - V_T)}{L^2}. \tag{3.96}$$

The scattering parameters were measured at $V_{GS} - V_T = 0.4\,\text{V}$. The mobility $\mu_n = 4400\,\text{cm}^2/\text{V-s}$ and $L = 1\,\mu\text{m}$. $V_{DS} = 0.5\,\text{V}$. $g_m = 0.049\,\text{S}$, $g_D = 0.014\,\text{S}$.

Use the SA algorithm to obtain optimised values of the equivalent circuit after converting the scattering parameters to admittance parameters [14] as shown in Table 3.7.

(2) What are the advantages of combinatorial optimisation methods over analytical methods of device optimisation? What are the disadvantages?

(3) A researcher wants to optimise the parameters of a small-signal equivalent circuit for the HBT using the genetic algorithm. The parameters are shown in Table 3.8. Assume that there is a procedure for measuring the S-parameters of the HBT for each variation of the parameters of the circuit shown in the table.
   (a) Write down the steps of a genetic algorithm for the optimisation.
   (b) Write a fitness function for the genetic algorithm.
   (c) Modify this genetic algorithm for optimising the small-signal parameters of a FET. Compare the difficulty of doing this to the difficulty of modifying an analytic method for the same procedure.

(4) A student wants to create a neural network model for a power amplifier with the parameters shown in Table 3.9. He/she decides to use the SGA to simultaneously optimise both the weight values and the number of neurons in the neural network.

Table 3.7 HEMT scattering parameters

Frequency (GHz)	S_{11}	S_{12}	S_{21}	S_{22}
5	$0.55\angle -158$	$0.05\angle 1.80$	$0.53\angle 14.5$	$0.83\angle 176$
10	$0.75\angle -166$	$0.06\angle 180$	$0.63\angle 18.5$	$0.82\angle 173$
20	$0.8\angle 179$	$0.07\angle 26$	$0.68\angle 28.5$	$0.82\angle 167$
30	$0.826\angle 169$	$0.07\angle 35$	$0.079\angle 33.5$	$0.83\angle 161$
40	$0.79\angle 161$	$0.009\angle 33$	$0.09\angle 34.5$	$0.81\angle 155$

Table 3.8 Circuit parameters

Component	Definition
L_b(pH)	Base inductance
L_c(pH)	Collector inductance
L_e(pH)	Emitter inductance
C_c(fF)	Collector capacitance
C_e(pF)	Emitter capacitance
$R_b(\Omega)$	Base resistance
$R_c(\Omega)$	Collector resistance
τ(ps)	Transit time
β_o	Current gain

Table 3.9 Neural network inputs and outputs

Inputs	Outputs
Gate length	Output power
Gate voltage	DC gate current
DC drain current	Drain current

(a) Draw a diagram of a chromosome that can be used by the SGA.
(b) Is there any other way to simultaneously optimise both the weight values and the number of neurons in the network?

(5) The scattering parameters for an HBT are given in Table 3.10.

Use the semi-analytical parameter extraction method to determine the element values in the equivalent circuit and then obtain the scattering parameters to compare with the given measured parameters.

(6) Assume that the base–collector extrinsic capacitance could be lumped into the intrinsic C_{bc} and develop the strategy for the small-signal model parameter extraction procedure.

Table 3.10 HBT S-parameters

Frequency (GHz)	$\|S_{11}\|$	$\angle S_{11}$	$\|S_{21}\|$	$\angle S_{21}$	$\|S_{12}\|$	$\angle S_{12}$	$\|S_{22}\|$	$\angle S_{22}$
0.45	0.955	−4.3	4.651	174.83	0.008	85.1	0.997	−2.78
0.5	0.954	−4.83	4.648	174.25	0.009	84.2	0.996	−3.1
0.6	0.953	−5.78	4.638	173.19	0.01	84.11	0.996	−3.73
0.7	0.952	−6.71	4.631	172.13	0.012	83.44	0.995	−4.32
0.8	0.951	−7.66	4.622	171.05	0.013	82.7	0.994	−4.93
0.9	0.95	−8.68	4.614	170.05	0.015	82.79	0.993	−5.51
1	0.948	−9.52	4.601	168.94	0.016	82.61	0.992	−6.12
2	0.92	−18.95	4.457	158.29	0.034	75.73	0.974	−12.12
4	0.833	−35.63	4.005	139.02	0.062	66.77	0.921	−22.78
6	0.726	−49.36	3.482	122.93	0.086	55.57	0.853	−31.24
8	0.628	−60.18	3.001	109.46	0.102	48.19	0.797	−38.1
10	0.544	−68.72	2.593	98.35	0.115	42.08	0.75	−43.67
12	0.476	−74.48	2.264	89.05	0.123	37	0.716	−48.35
14	0.43	−79.5	2.009	81.24	0.132	34.11	0.694	−52.23
16	0.39	−84.14	1.805	73.81	0.141	30.94	0.68	−56.36
18	0.352	−87.56	1.633	66.91	0.149	27.43	0.669	−60.25
20	0.326	−88.9	1.489	61.32	0.153	25.01	0.663	−63.83
22	0.306	−89.49	1.382	55.44	0.158	22.24	0.661	−67.38
24	0.299	−90.98	1.293	50.96	0.163	20.41	0.662	−70.79
26	0.292	−94.44	1.227	46.22	0.167	19.91	0.672	−73.97
28	0.283	−95.78	1.173	41.1	0.176	18.33	0.68	−77.75
30	0.272	−98.22	1.127	35.42	0.184	16.84	0.69	−81.95
32	0.2569	−99.7	1.073	29.92	0.194	14.21	0.688	−85.81
34	0.246	−102.13	1.024	25.25	0.201	11.68	0.692	−89.2
36	0.238	−100.66	0.974	21.13	0.206	8.92	0.697	−92.91
38	0.232	−101.63	0.963	16.41	0.213	6.63	0.699	−96.4
40	0.223	−101.26	0.918	11.77	0.22	3.11	0.694	−99.63

(7) What are the advantages of the inverted (also known as the *emitter-down* or *collector-up*) HBT? If it is promising in some power application, why is it not popular yet? Compare the emitter-up HBT and the emitter-down HBT in terms of the device parameters?

(8) Use the small-signal equivalent circuit of the collector-up HBT (Figure 3.50) to determine the basic expressions for the circuit parameters.

(9) The base–collector capacitance C_f is much smaller in the inverted HBT. Does that make the model parameter extraction easy?

(10) In compact device modelling, layout pads for measurement are not the part of the device which needs to be removed (De-embedding procedure) from the measurement data. In industry, open/short structures are used.
   (a) Design your own open/short structure. How do you think the pad size, ground layout, probe type or frequency affect your design?
   (b) Draw the equivalent circuit for the open structure.
   (c) Draw the equivalent circuit for the short structure.
   (d) The S-parameters for the open/short structures are given. What information can you extract from them?

(11) An engineer is developing the diode model at a particular bias. Outline the procedure he/she is going to develop. Use the given S-parameter file. If there is inductance in the structure that he/she forgets to remove, how will that affect the final result? If the resistor and inductor are ignored, do you get the right result? Explain your answer.

(12) There is a set of S-parameter files which were taken for the diode. Use the files to extract the diode capacitance model parameters, assuming that parasitic elements have been de-embedded out.

(13) Given the HBT S-parameters measured at the cold condition, use what you learned to extract the model parameters.

(14) Given the HBT S-parameters measured at the constant V_{ce} condition, use what you learned to extract the model parameters.

(15) Given the HBT S-parameters measured at constant I_b condition, use what you learned to extract the model parameters.

(16) The intrinsic part of the small-signal equivalent circuit of the MESFET which is similar to the hybrid pi model of the bipolar transistor is given in Figure 3.51. Outline the extraction procedure to extract the model parameters assuming negligible g_{ds}. Now the full equivalent circuit of the MESFET is given in Figure 3.52. Assume $L_g = L_d = 25\,\text{pH}$, $L_s = 10\,\text{pH}$, $R_s = R_d = 10\,\Omega$, $R_g = 1\,\Omega$,

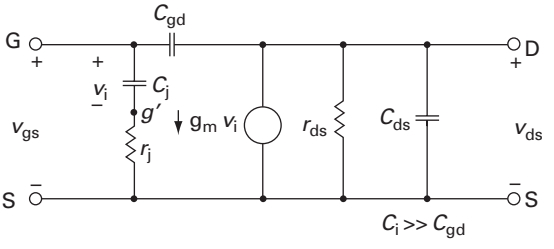

Fig. 3.51 Intrinsic MESFET circuit.

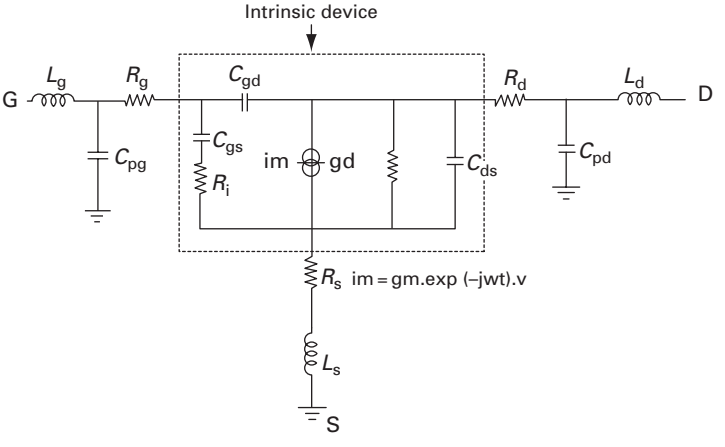

Fig. 3.52 MESFET equivalent circuit.

$C_{pg} = C_{pd} = 25$ fF and $g_{ds} = 0$ S. Extract the other model parameters based on the given set of S-parameters.

Note that the necessary data files are available on the Web labelled according to the problem numbers.

References

[1] Bousnina S., Falt C., Mandeville P., Kouki A. B., Ghannouchi F. M. (2002). An accurate on-wafer deembedding technique with application to HBT devices characterizaion. *IEEE Trans. Microw. Theory Tech. MTT-50*, 2 (February), 420–424.

[2] Bousnina S., Mandeville P., Kouki A. B., Surridge R., Ghannouchi F. M. (2002). Direct parameter-extraction method for HBT small signal model. *IEEE Trans. Microw. Theory Tech. MTT-50*, 2 (February), 529–536.

[3] Costa D., Liu W., Jr. J. S. H. (1991). Direct extraction of the AlGaAs/GaAs heterojunction bipolar transistor small-signal equivalent circuit. *IEEE Trans. Electron Devices ED-38*, 9 (September), 2018–2024.

[4] Dasgupta D., McGregor D. (1992). SGA: a structured genetic algorithm. Tech. Rep. IKBS-8-92, University of Strathclyde.

[5] Degachi L., Ghannouchi F. M. (2006). Systematic and rigorous extraction method of HBT small-signal model parameters. *IEEE Trans. Microw. Theory Tech. MTT-54*, 2 (February), 682–688.

[6] Dvorak M. W., Bolognesi C. (2003). On the accuracy of direct extraction of the heterojunction-bipolar-transistor equivalent-circuit model parameters C_π, C_{BC} and R_E. *IEEE Trans. Microw. Theory Tech. MTT-51*, 6 (June), 1640–1649.

[7] Gobert Y., Tasker P. J., Bachem K. H. (1997). A physical, yet simple, small-signal equivalent circuit for the heterojunction bipolar transistor. *IEEE Trans. Microw. Theory Tech. MTT-45*, 1 (January), 149–153.

[8] Goldberg D. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley.

[14] Gonzalez G. (1997). *Microwave Transistor Amplifiers, Analysis and Design*. Prentice Hall.

[10] Holland J. H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press.

[11] Hopfield J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Nat. Acad. Sci. 81*, 10 (May), 3088–3092.

[12] Kirkpatrick S., Gelatt C. D. Jr, Vecchi M. P. (1983). Optimization by simulated annealing. *Science 220*, 4598 (May), 671–680.

[13] Li B., Prasad S. (1997). Harmonic and two-tone intermodulation distortion analyses of the inverted InGaAs/InAlAs/InP HBT. *IEEE Trans. Microw. Theory Tech. MTT-45*, 7 (July), 1135–1137.

[14] Li B., Prasad S. (1999). Basic expressions and approximations in small-signal parameter extraction for HBT's. *IEEE Trans. Microw. Theory Tech. MTT-47*, 5 (May), 534–539.

[15] Li B., Prasad S., Yang L.-W., Wang S. C. (1998). A semianalytical parameter-extraction procedure for HBT equivalent circuit. *IEEE Trans. Microw. Theory Tech. MTT-46*, 10 (October), 1427–1435.

[16] Lippmann R. (1987). An introduction to computing with neural nets. *IEEE ASSP Mag. 4*, 5 (April), 4–22.

[17] Meskoob B., Prasad S., Vai M., Fonstad C., Vlcek J. C., Sato H. (1992). A small-signal equivalent circuit for the collector-up InGaAs/InAlAs/InP heterojunction bipolar transistor. *IEEE Trans. Electron Devices ED-39*, 11 (November), 2629–2632.

[18] Meskoob B., Prasad S., Vai M., Vlcek J. C., Sato H., Fonstad C. G. (1992). Bias-dependence of the intrinsic element values of InGaAs/InAlAs/InP inverted heterojunction bipolar transistor. *IEEE Trans. Microw. Theory Tech. MTT-40*, 5 (May), 1012–1014.

[19] Pehlke D. R., Pavlidis D. (1992). Evaluation of the factors determining HBT high frequency performance by direct analysis of S-parameter data. *IEEE Trans. Microw. Theory Tech. MTT-40*, 12 (December), 2367–2373.

[20] Rios J. M. M., Lunardi L. M., Chandrasekhar S., Miyamoto Y. (1997). A self-consistent method for complete small-signal extraction of InP-based heterojunction bipolar transistors (HBTs). *IEEE Trans. Microw. Theory Tech. MTT-45*, 1 (January), 39–45.

[21] Roblin P., Kang S., Ketterson A., Morkoc H. (1987). Analysis of MODFET microwave characteristics. *IEEE Trans. Electron Devices ED-34*, 9 (September), 1919–1927.

[22] Romeo F., Sangiovanni-Vincentelli A. (1987). Probabilistic hill climbing algorithms: properties and applications. In *1985 Chapel Hill Conference on VLSI*, H. Fuchs, ed. Computer Science Press, 671–680.

[23] Rutenbar R. A. (1989). Simulated annealing algorithm: an overview. *IEEE Circ. Dev. Mag. 5*, 1 (January), 19–26.

[24] Samelis A., Pavlidis D. (1997). DC to high-frequency HBT-model parameter evaluation using impedance block conditioned optimization. *IEEE Trans. Microw. Theory Tech. MTT-45*, 6 (June), 886–897.

[25] Schaper U., Holzapfl B. (1995). Analytical parameter extraction of the HBT equivalent circuit with T-like topology from measured S-parameters. *IEEE Trans. Microw. Theory Tech. MTT-40*, 3 (March), 493–498.

[26] Sheinman B., Wasige E., Rudolph M., Sidorov V., Cohen S., Ritter D. (2002). A peeling algorithm for extraction of the HBT small-signal equivalent circuit. *IEEE Trans. Microw. Theory Tech. MTT-50*, 12 (December), 2804–2810.

[27] Spiegel S. J., Ritter D., Hamm R. A., Feygenson A., Smith P. R. (1995). Extraction of the InP/GaInAs heterojunction bipolar transistor small-signal equivalent circuit. *IEEE Trans. Electron Devices ED-42*, 6 (June), 1059–1064.

[28] Vai M., Prasad S. (1993). Automatic impedance matching with a neural network. *IEEE Microw. Guided Wave Lett. MGWL-3*, 10 (October), 353–354.

[29] Vai M., Prasad S. (1999). Neural networks in microwave circuit design – beyond black-box models. *Int. J. RF and Microwave CAE 9*, 3 (May), 187–197.

[30] Vai M., Prasad S., Li N., Kai F. (1989). Modeling of microwave semiconductor devices using simulated annealing optimization. *IEEE Trans. Electron Devices ED-36*, 4 (April), 761–762.

[31] Vai M., Prasad S. (1995). Microwave circuit analysis and design by a massively distributed computing network. *IEEE Trans. Microw. Theory Tech. MTT-43*, 5 (May), 1087–1094.

[32] vanRooij A., Jain L. C., Johnson R. P. (1996). *Neural Network Training Using Genetic Algorithms*. World Scientific.

[33] Wei C. J., Huang J. C. M. (1995). Direct extraction of equivalent circuit parameters for heterojunction bipolar transistors. *IEEE Trans. Microw. Theory Tech. MTT-43*, 9 (September), 2035–2039.

[34] Willén B., Rohner M., Schnyder I., Jäckel H. (2002). Improved automatic parameter extraction of InP-HBT small signal equivalent circuits. *IEEE Trans. Microw. Theory Tech. MTT-50*, 2 (February), 580–583.

[35] Yang T. -R., Tsai J. M. -L., Ho C. -L., Hu R. (2007). SiGe HBTs small-signal pi modeling. *IEEE Trans. Microw. Theory Tech. MTT-55*, 7 (July), 1417–1424.

[36] Zaabab A. H., Zhang Q., Nakhla M. (1995). A neural network modeling approach to circuit optimization and statistical design. *IEEE Trans. Microw. Theory Tech. MTT-43*, 6 (June), 1349–1358.

[37] Zhang Q. -J., Gupta K., Devabhaktuni V. K. (2003). Artificial neural networks for RF and microwave design–from theory to practice. *IEEE Trans. Microw. Theory Tech. MTT-51*, 4 (April), 1339–1350.