Received: 30 December 2019 | Revised: 11 March 2020 | Accepted: 28 April 2020

DOI: 10.1002/jnm.2762

WILEY

SPECIAL ISSUE PAPER

Statistical modeling of frequency responses using linear Bayesian vector fitting

Simon De Ridder | **Dirk Deschrijver** | **Domenico Spina** | **Dries Vande Ginste** | **Tom Dhaene**

IDLab, Department of Information Technology, Ghent University – IMEC, Ghent, Belgium

Correspondence
Simon De Ridder, IDLab, Department of Information Technology, Ghent University – IMEC, Ghent, Belgium.
Email: simon.deridder@edpnet.be

Abstract

This article presents a Bayesian extension of the vector fitting (VF) procedure for rational approximation of frequency-domain responses. The proposed method treats the linear part of VF in a Bayesian way, while propagating distributions through the nonlinear part by sampling. As such, it is capable of providing data-driven uncertainty information along with the rational fit. The Bayesian VF technique is applied to two realistic design examples, a double folded stub filter and a RAM memory channel, demonstrating its validity and highlighting three potential applications of this novel framework.

KEYWORDS

adaptive sampling, bayesian inference, macromodeling

1 | INTRODUCTION

Black-box macromodeling is a popular tool to approximate the dynamic behavior of complex systems in terms of low complexity models. The most accomplished approaches to black-box macromodeling are the Löwner matrix method,[1,2] and the vector fitting (VF) algorithm.[3-6] By virtue of its robustness and modeling power, several macromodeling techniques based on the VF algorithm have been developed in recent years, in order to characterize the behavior of distributed elements.[7-9]

As any measurement of an electronic device is affected by a modicum of noise, and due to discrete sampling of the frequency response, macromodels are subject to model uncertainty. Most macromodeling techniques, including VF, however, consist of a deterministic interpolation of the response, based on those samples. These models display no measure of model uncertainty with regard to the value of their interpolation. A statistical treatment of these methods would be an efficient means to assess this model uncertainty.

In this article, such a statistical treatment is presented, based on Bayesian linear regression combined with sampling-based propagation of model uncertainty. Two insightful examples demonstrate that the novel Bayesian version of VF gives sharp bounds of the model uncertainty. With the aid of these examples, it is shown that such uncertainty information can be used for various purposes, such as characterizing device responses in the presence of noisy or missing data, to verify functionality and compliance, or for adaptive sampling. Note that in Reference 10, the idea of using Bayesian VF for adaptive sampling was introduced first, but only for a simple one-port example of an antenna. In this work, the linear Bayesian vector fitting (LB-VF) framework is fully developed and tested on two multiport systems (a microwave filter and a RAM memory channel) for various applications.

The remainder of this article is structured as follows. In section 2, the VF algorithm is briefly explained. Then, the proposed Bayesian extension is detailed in section 3. In section 4, the LB-VF framework is applied to two realistic examples. Finally, in section 5, the article is concluded.

2 | VECTOR FITTING

The vector fitting (VF) algorithm is a well-known technique to build a rational macromodel that approximates a transfer function.[3-6] This transfer function or (for multiport S-parameters) transfer matrix $\overline{\overline{S}}(s)$ is approximated by a function $\overline{\overline{F}}(s)$ as a sum of partial fractions:

$$\overline{\overline{S}}(s) \approx \overline{\overline{F}}(s) = \sum_{i=1}^{N} \frac{\overline{\overline{R}}_i}{s - a_i} + \overline{\overline{D}} + s\overline{\overline{E}}. \tag{1}$$

Here, the a_i and $\overline{\overline{R}}_i$ are the poles and the corresponding residue matrices, respectively. These poles and residue matrices can either be real or constitute complex conjugate pairs. Furthermore, using a suitable pole-flipping scheme,[3] stability can be enforced by keeping the real part of the poles negative. The Laplace variable is denoted by s. $\overline{\overline{D}}$ and $\overline{\overline{E}}$ are optional constant and linear terms, describing the asymptotic behavior of the transfer function.

2.1 | Sanathanan-Koerner iteration

An established way to obtain such a fit is by iteratively relocating a set of starting poles until convergence, using Sanathanan-Koerner (SK) iterations.[11] This technique fits a function $\overline{\overline{F}}(s)$ to data samples of $\overline{\overline{S}}(s)$ by first choosing a set of 'starting' poles $\{a_i^0\}$ and introducing a numerator function (matrix) $\overline{\overline{p}}(s)$ and a denominator function $q(s)$ such that:

$$\overline{\overline{S}}(s) \approx \overline{\overline{F}}(s) = \frac{\overline{\overline{p}}(s)}{q(s)} = \frac{\sum_{i=1}^{N} \frac{\overline{\overline{r}}_i}{s - a_i^0} + \overline{\overline{d}} + s\overline{\overline{e}}}{\sum_{i=1}^{N} \frac{\widetilde{r}_i}{s - a_i^0} + \widetilde{d}}. \tag{2}$$

In the original VF implementation,[3] $\widetilde{d} = 1$, while for relaxed VF $\widetilde{d}$ is a free variable, but an extra equation is added to avoid trivial solutions (see[4] for more details).

The initial poles can then be relocated by solving the linear least squares problem $q(s)\overline{\overline{S}}(s) \approx \overline{\overline{p}}(s)$ for $\{\overline{\overline{r}}_i\}$, $\{\widetilde{r}_i\}$, $\overline{\overline{d}}$, $\overline{\overline{e}}$ and $\widetilde{d}$. If samples of $\overline{\overline{S}}(s)$, denoted as $\left\{\overline{\overline{S}}_j\right\}$, are known at N_s frequency points $\{s_j\}$, $j = 1, ..., N_s$, this can be summarized in the complex-valued matrix equation:

$$\overline{\overline{A}}\,\overline{x} = \overline{b}. \tag{3}$$

For example, for nonrelaxed VF and a scalar $S(s)$, this results in:

$$\overline{\overline{A}} = \begin{bmatrix} \frac{1}{s_1 - a_1^0} & \cdots & \frac{1}{s_1 - a_N^0} & 1 & s_1 & \frac{-S_1}{s_1 - a_1^0} & \cdots & \frac{-S_1}{s_1 - a_N^0} \\ \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ \frac{1}{s_{N_s} - a_1^0} & \cdots & \frac{1}{s_{N_s} - a_N^0} & 1 & s_{N_s} & \frac{-S_{N_s}}{s_{N_s} - a_1^0} & \cdots & \frac{-S_{N_s}}{s_{N_s} - a_N^0} \end{bmatrix},$$
$$\overline{x} = [r_1 \ \ldots \ r_N \ d \ e \ \widetilde{r_1} \ \ldots \ \widetilde{r_N}]^{\mathrm{T}},$$
$$\overline{b} = [S_1 \ \ldots \ S_{N_s}]^{\mathrm{T}}. \tag{4}$$

In the case of relaxed VF, $\widetilde{d}$ is included in $\overline{x}$ and an additional condition

$$\Re\left(\sum_{k=1}^{N_s}\left(\sum_{i=1}^{N}\frac{\widetilde{r}_i}{s_k-a_i^0}+\widetilde{d}\right)\right)=N_s \tag{5}$$

is added to avoid a zero solution.[4]

In the case of a complex conjugate pair r_i and $r_{i+1}=r_i^*$, r_i+r_{i+1} and $\Im(r_i-r_{i+1})$ are considered instead. Likewise, $\overline{A_i}+\overline{A_{i+1}}$ and $\Im\left(\overline{A_i}-\overline{A_{i+1}}\right)$ replace $\overline{A_i}$ and $\overline{A_{i+1}}$ as the corresponding columns of $\overline{\overline{A}}$. When $\overline{\overline{S}}$ itself is complex, the system

$$\overline{\overline{A'}}\,\overline{x}=\overline{b'} \tag{6}$$

with

$$\overline{\overline{A'}}=\begin{bmatrix}\Re\left(\overline{\overline{A}}\right)\\ \Im\left(\overline{\overline{A}}\right)\end{bmatrix}\quad\text{and}\quad\overline{b'}=\begin{bmatrix}\Re\left(\overline{b}\right)\\ \Im\left(\overline{b}\right)\end{bmatrix} \tag{7}$$

is solved instead. In this way, the linear system (6) is always real-valued.

When dealing with an n by m matrix-variate $\overline{\overline{S}}(s)$, it is vectorized and for each independent element l a matrix $\overline{\overline{A}}_l$ and vector $\overline{b}_l$ of the same form as $\overline{\overline{A'}}$ and $\overline{b'}$ are constructed. Each such $\overline{\overline{A}}_l$ is QR-decomposed as $\overline{\overline{A}}_l=\overline{\overline{Q_lP_l}}$, and the lower right parts of each $\overline{\overline{P}}_l$ are vertically stacked: $\overline{\overline{A''}}=\begin{bmatrix}\overline{\overline{P}}_{1,LR}^T & \dots & \overline{\overline{P}}_{nm,LR}^T\end{bmatrix}^T$. Likewise, a stacked vector $\overline{b''}$ is constructed from the lower part of each $\overline{\overline{Q}}_l$ and $\overline{b}_l$: $\overline{b''}=\begin{bmatrix}\overline{b}_1^T\overline{\overline{Q}}_{i,L} & \dots & \overline{b}_{nm}^T\overline{\overline{Q}}_{nm,L}\end{bmatrix}^T$. The coefficients of $q(s)$ in $\overline{x}$, denoted $\overline{x}_L$, are then computed by solving

$$\overline{\overline{A''}}\,\overline{x}_L=\overline{b''}, \tag{8}$$

yielding a common form for $q(s)$. For more details, see Reference 5.

Using the initial poles $\{a_i^0\}$ and the newly obtained $\{\widetilde{r}_i\}$ and $\widetilde{d}$, the zeros of $q(s)$ are then calculated. As both $q(s)$ and $\overline{\overline{p}}(s)$ share the same poles $\{a_i^0\}$, these poles cancel out, and the zeros of $q(s)$ are the relocated poles $\{a_i\}$ of $\overline{\overline{F}}(s)$. These are found by solving an eigenvalue problem, based on the minimal LTI state-space realization of $q(s)$[12]:

$$\{a_i\}=\mathrm{eig}\left(\overline{\overline{A}}_q-\overline{B}_qD_q^{-1}\overline{C}_q\right), \tag{9}$$

where $\overline{\overline{A}}_q$ is a matrix with the starting poles $\{a_i^0\}$ as diagonal elements, $\overline{B}_q$ is a column vector of ones, $\overline{C}_q$ is a row vector containing the $\{\widetilde{r}_i\}$, and D_q equals $\widetilde{d}$. Thus, a relocated set of poles is obtained. In order to enforce stability, poles whose real part is positive are flipped to the left half of the complex plane.[3] The procedure is then iterated by replacing the initial starting poles with the relocated poles.

2.2 | Residue identification

Once the poles are relocated to their final position, identifying the residues $\overline{\overline{R}}_i$ comes down to solving a linear system for each element in (1):

$$\overline{\overline{AX}}=\overline{\overline{B}}, \tag{10}$$

where

$$\begin{aligned}
\overline{\overline{A}} &= \begin{bmatrix} \frac{1}{s_1-a_1} & \cdots & \frac{1}{s_1-a_N} & 1 & s_1 \\ \vdots & & \vdots & \vdots & \vdots \\ \frac{1}{s_{N_s}-a_1} & \cdots & \frac{1}{s_{N_s}-a_N} & 1 & s_{N_s} \end{bmatrix}, \\
\overline{\overline{X}} &= \begin{bmatrix} R_{1,1} & \cdots & R_{1,nm} \\ \vdots & & \vdots \\ R_{N,1} & \cdots & R_{N,nm} \\ D_1 & \cdots & D_{nm} \\ E_1 & \cdots & E_{nm} \end{bmatrix}, \\
\overline{\overline{B}} &= \begin{bmatrix} S_{1,1} & \cdots & S_{1,nm} \\ \vdots & & \vdots \\ S_{N_s,1} & \cdots & S_{N_s,nm} \end{bmatrix}.
\end{aligned} \tag{11}$$

Similarly as for (4), both the sum of, and (the imaginary part of) the difference between complex conjugate pairs of residues are considered instead of those residues themselves. When $\overline{\overline{S}}$ itself is complex, the real and imaginary parts of $\overline{\overline{A}}$ and $\overline{\overline{B}}$ are vertically stacked. As such, the system (10) is ensured to be real-valued.

3 | LINEAR BAYESIAN VECTOR FITTING

3.1 | Sampling $\{\widetilde{\boldsymbol{r}_i}\}$ and $\widetilde{\boldsymbol{d}}$

Revisiting (8) with the final, converged VF poles as the starting poles $\{a_i^0\}$, $\overline{b''}$ is modeled by a Gaussian distribution:

$$\overline{b''} \sim N\left(\overline{b''}|\overline{\overline{A''}}\overline{x}, \sigma^2\overline{\overline{I}}\right) \tag{12}$$

This allows one to treat the linear system (8) in a Bayesian manner when the samples $\{S_j\}$ are subject to uncertainty, solving for the distribution of $\overline{x}$, rather than for the optimal value. A conjugate prior for $\overline{x}$ and σ^2 is the following:

$$\mathrm{P}\left(\overline{x}, \sigma^2\right) \sim N\left(\overline{x}|\overline{x_0}, \sigma^2\overline{\overline{\Lambda_0}}^{-1}\right)\mathrm{IG}\left(\sigma^2|\alpha_0, \beta_0\right), \tag{13}$$

which is a σ^2-dependent Gaussian distribution for $\overline{x}$, and an inverse-gamma distribution for σ^2. The parameters $\overline{x_0}$, $\overline{\overline{\Lambda_0}}$, α_0 and β_0 define this prior. They can either be set by prior knowledge, or chosen to represent the least informative distribution of the form (13). In the latter case, used in this article as well, the uninformative (Jeffrey's) prior is set to

$$\mathrm{P}\left(\overline{x}, \sigma^2\right) \sim \left(\sigma^2\right)^{-1}. \tag{14}$$

Using Bayes' theorem, some calculations (see Appendix A) now yield a posterior distribution of $\overline{x}$ having the same form:

$$\mathrm{P}\left(\overline{x}, \sigma^2|\overline{b''}\right) \sim N\left(\overline{x}|\overline{x_f}, \sigma^2\overline{\overline{\Lambda_f}}^{-1}\right)\mathrm{IG}\left(\sigma^2|\alpha_f, \beta_f\right). \tag{15}$$

The parameters $\overline{x_f}$, $\overline{\overline{\Lambda_f}}$, α_f and β_f can be obtained using the update rules:

$$\overline{x_f} = \overline{\overline{\Lambda_f}}^{-1}\left(\overline{\overline{\Lambda_0}}\,\overline{x_0} + \overline{\overline{A''}}^{\mathrm{T}}\overline{b''}\right) \tag{16}$$

$$\overline{\overline{\Lambda_f}} = \overline{\overline{\Lambda_0}} + \overline{\overline{A''}}^{\mathrm{T}}\overline{\overline{A}} \tag{17}$$

$$\alpha_f = \alpha_0 + \frac{N_b}{2} \tag{18}$$

$$\beta_f = \beta_0 + \frac{\overline{b''}^{\mathrm{T}}\overline{b''} - \overline{x_f}^{\mathrm{T}}\overline{\overline{\Lambda_f}}\,\overline{x_f} + \overline{x_0}^{\mathrm{T}}\overline{\overline{\Lambda_0}}\,\overline{x_0}}{2}, \tag{19}$$

where N_b is the length of $\overline{b''}$. It is easily verified that, when using (14) as a prior, the results in (15)-(19) still hold, and they are the same as if one were to choose $\overline{x_0} = \overline{0}$, $\overline{\overline{\Lambda_0}} = \overline{\overline{0}}$ and $\alpha_0 = \beta_0 = 0$.

Since σ^2 is still unknown, it can and should be marginalized out, in order to derive a marginal posterior for $\overline{x}$ (see Appendix B):

$$\mathrm{P}\left(\overline{x}|\overline{b''}\right) = \int \mathrm{P}\left(\overline{x}, \sigma^2|\overline{b''}\right) d\left(\sigma^2\right) = t_{2\alpha_f}\left(\overline{x}|\overline{x_f}, \left(\frac{\alpha_f}{\beta_f}\overline{\overline{\Lambda_f}}\right)^{-1}\right). \tag{20}$$

This is a multivariate *t*-distribution.

The marginal likelihood $\mathrm{P}\left(\overline{b''}\right)$ can be readily derived by integrating the product of the likelihood (12) and the prior (13) (see Appendix C):

$$\mathrm{P}\left(\overline{b''}\right) = \iint \mathrm{P}\left(\overline{b''}|\overline{x}, \sigma^2\right)\mathrm{P}\left(\overline{x}, \sigma^2\right) d\overline{x} d\left(\sigma^2\right) = (2\pi)^{-\frac{N_b}{2}} \sqrt{\frac{\left|\overline{\overline{\Lambda_0}}\right|}{\left|\overline{\overline{\Lambda_f}}\right|}} \frac{\beta_0^{\alpha_0}}{\beta_f^{\alpha_f}} \frac{\Gamma\left(\alpha_f\right)}{\Gamma(\alpha_0)} \tag{21}$$

The only dependency on $\overline{b''}$ in this expression stems from the posterior parameters (16)-(19). This quantity is useful because it is a *measure* of how well the model fits the data, for any value of its parameters. Hence, (21) can be used to evaluate the quality of the starting poles, or their number.

3.2 | Pole distribution

Since the nonlinear step in (9) precludes an analytic propagation of the posterior distribution of $\{\widetilde{r_i}\}$ and $\widetilde{d}$ in (20) to the relocated poles $\{a_i\}$, the distribution of the latter must be approximated. This is done by drawing samples from the posterior distribution in (20) and solving (9) for each of these samples. The real part of any unstable sampled pole is inverted to ensure stability.

3.3 | Sampling the residues

Similarly to the system described in section 3.1, the solution to (10) can also be treated in a Bayesian way. In this case, the stochastic variability is modeled by a matrix normal distribution:

$$\overline{\overline{B}} \sim \mathcal{MN}_{N_s,\mathrm{nm}}\left(\overline{\overline{B}}|\overline{\overline{AX}}, \overline{\overline{I}}, \overline{\overline{\Sigma}}\right). \tag{22}$$

The conjugate prior is now of the form:

$$\mathrm{P}\left(\overline{\overline{X}},\overline{\overline{\Sigma}}\right)\sim\mathcal{MN}_{N,nm}\left(\overline{\overline{X}}|dle|\overline{\overline{X_0}},\overline{\overline{\Lambda_0}}^{-1},\overline{\overline{\Sigma}}\right)\mathcal{W}^{-1}\left(\overline{\overline{\Sigma}}|dle|\overline{\overline{V_0}},\nu_0\right),\tag{23}$$

which is a $\overline{\overline{\Sigma}}$-dependent matrix normal distribution times an inverse-Wishart distribution for $\overline{\overline{\Sigma}}$ itself. After very similar calculations to the ones in section 3.1, the posterior is shown to become:

$$\mathrm{P}\left(\overline{\overline{X}},\overline{\overline{\Sigma}}|\overline{\overline{B}}\right)\sim\mathcal{MN}_{N,\mathrm{nm}}\left(\overline{\overline{X}}|\overline{\overline{X_f}},\overline{\overline{\Lambda_f}}^{-1},\overline{\overline{\Sigma}}\right)W^{-1}\left(\overline{\overline{\Sigma}}|\overline{\overline{V_f}},\nu_f\right),\tag{24}$$

where $\overline{\overline{X}}$ is an $N\times nm$ matrix, and where the parameters can be calculated as:

$$\overline{\overline{X_f}}=\overline{\overline{\Lambda_f}}^{-1}\left(\overline{\overline{\Lambda_0}}\,\overline{\overline{X_0}}+\overline{\overline{A}}^{\mathrm{T}}\overline{\overline{B}}\right)\tag{25}$$

$$\overline{\overline{\Lambda_f}}=\left(\overline{\overline{\Lambda_0}}+\overline{\overline{A}}^{\mathrm{T}}\overline{\overline{A}}\right)\tag{26}$$

$$\overline{\overline{V_f}}=\overline{\overline{V_0}}+\overline{\overline{B}}^{\mathrm{T}}\overline{\overline{B}}-\overline{\overline{X_f}}^{\mathrm{T}}\overline{\overline{\Lambda_f}}\,\overline{\overline{X_f}}+\overline{\overline{X_0}}^{\mathrm{T}}\overline{\overline{\Lambda_0}}\,\overline{\overline{X_0}}\tag{27}$$

$$\nu_f=\nu_0+N_B,\tag{28}$$

with N_B the number of rows in $\overline{\overline{B}}$.

In this case, the uninformative prior $\mathrm{P}\left(\overline{\overline{X}},\overline{\overline{\Sigma}}\right)\propto\left|\overline{\overline{\Sigma}}\right|^{-nm}$ yields a posterior of the same form where $\overline{\overline{X_0}}=\overline{\overline{\Lambda_0}}=\overline{\overline{V}}=\overline{\overline{0}}$ and $\nu_0=nm-1$ (or $\nu_0=0$ for the Jeffrey's prior $\mathrm{P}\left(\overline{\overline{X}},\overline{\overline{\Sigma}}\right)\propto\left|\overline{\overline{\Sigma}}\right|^{-\frac{nm+1}{2}}$).

The unknown parameter $\overline{\overline{\Sigma}}$ can again be marginalized out to obtain:

$$\begin{aligned}\mathrm{P}\left(\overline{\overline{X}}|\overline{\overline{B}}\right)&=\int\mathrm{P}\left(\overline{\overline{X}},\overline{\overline{\Sigma}}|\overline{\overline{B}}\right)d\left(\overline{\overline{\Sigma}}\right)\\&=\mathrm{T}_{N_x,\mathrm{nm}}\left(\overline{\overline{X}}|\nu_f-\mathrm{nm}+1,\overline{\overline{X_f}},\overline{\overline{\Lambda_f}}^{-1},\overline{\overline{V_f}}\right),\end{aligned}\tag{29}$$

where N_x is the number of rows in $\overline{\overline{X}}$. This is the matrix-variate t-distribution.

3.4 | Sampling pole-residue models

VF models (1) can now be sampled for a given transfer function using the scheme outlined in Figure 1. First, according to section 3.1, $\{\widetilde{r}_i\}$ and $\widetilde{d}$ are sampled N_p times from their posterior distribution (20). Then, for each such sample, a set of new poles are calculated by finding the zeros of (this sample of) $q(s)$ (section 3.2). Next, for each pole set, N_r residue sets are sampled as described in section 3.3. Each pole-residue set now forms a different rational model that is drawn from the space of probable models given the data and the prior information.

4 | APPLICATION EXAMPLES AND NUMERICAL RESULTS

4.1 | Double folded stub filter

The statistical framework introduced above is now applied to a double folded stub filter, shown in Figure 2.[13] This design is a standard example in Keysight's Advanced Design System (ADS).[14] As a realistic design example, it is used in the following to highlight some potential applications of the novel LB-VF framework.

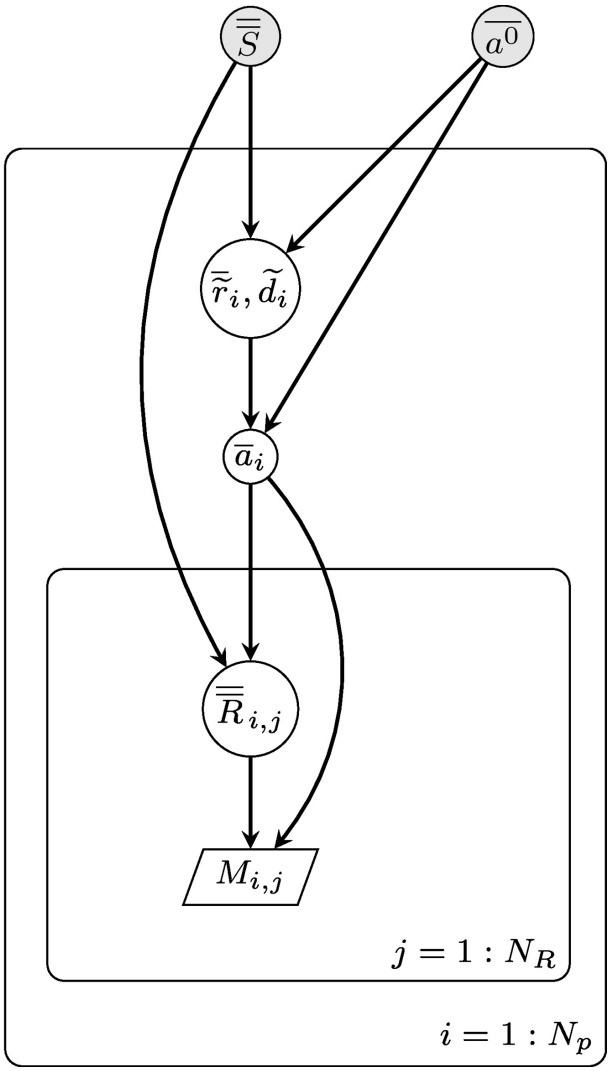


FIGURE 1 Linear Bayesian vector fitting framework: Graph of the sampling of VF models ($M_{i,j}$), where N_p pole sets are sampled, and N_R residue matrix sets are sampled for each of these pole sets

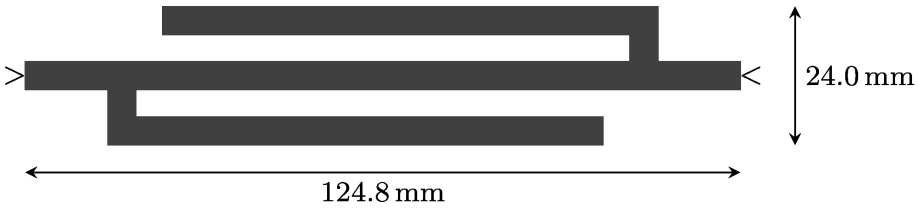


FIGURE 2 Geometry of the double folded stub filter[13]

4.1.1 | Uncertainty arising from additive noise

As mentioned before, an amount (even if minimal) of noise is always present in measurement data. It can be very insightful for a designer to know in what range the actual transfer function could fall, given the noisy measurement data. To this end, the frequency response—specifically, the *S*-parameter matrix—of the stub was simulated using ADS in 101 points over a frequency range from 1 to 30 GHz. In order to emulate noise (eg, stemming from measurement errors), uncorrelated Gaussian noise with zero mean and a standard deviation (SD) of 0.01 was added on top of the simulated *S*-parameters. Note that, because σ^2 was marginalized out of (20) [and likewise $\overline{\overline{\Sigma}}$ out of (29)], and because uninformative priors were used, no knowledge of the magnitude of the noise is needed when applying LB-VF.

After a few iterations with regular VF to relocate the 15 starting poles, LB-VF was applied to the data, starting from those relocated poles. An uninformative prior was used to sample $N_p = 500$ pole sets, and for each pole set, $N_r = 20$ residue sets were sampled, for a total of 10 000 models.

The sampled poles are shown in Figure 3, while Figures 4 and 5 show the S_{11} and S_{21} parameters, respectively. As can be seen in these figures, the 99.73% (3σ) confidence bound encompasses most of the original data. It ought to be noted that these bounds are overconfident, as they are still conditioned on the locations of the (relocated) starting poles, their number, and the rational form of the model. If the magnitude of the noise in the original data increases, the posterior distributions (20) and (29) widen, the spread of the samples increases, and the confidence bounds widen.

4.1.2 | Effect of missing data

To demonstrate the behavior of LB-VF when part of the data is missing, the data from 6.8 to 9.7 GHz for the double folded stub was discarded before relocating the initial poles or applying the LB-VF model.

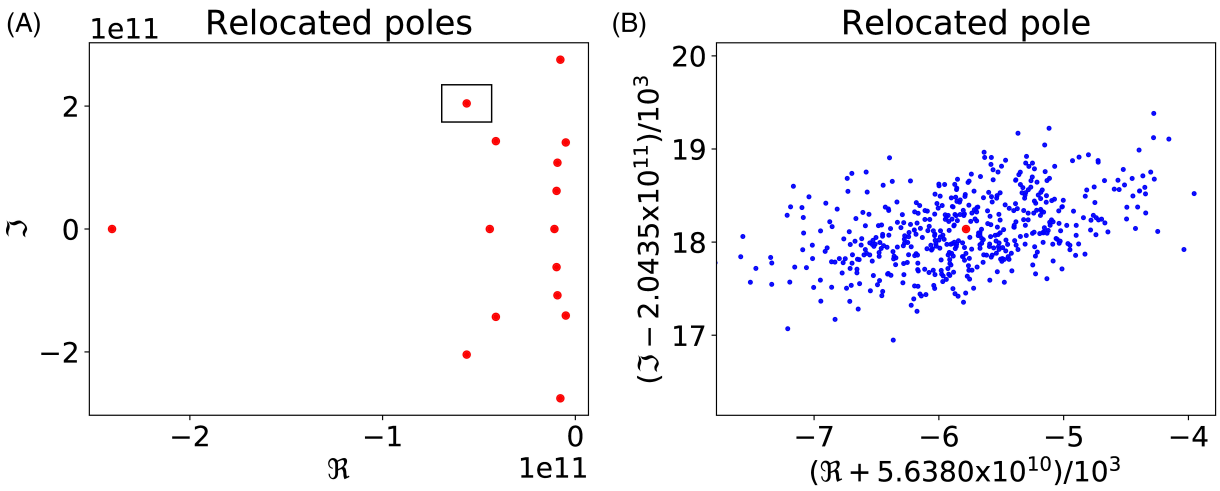


FIGURE 3 The sampled distribution of the poles as a result of linear Bayesian vector fitting applied to the S_{11} parameter of the double folded stub filter. The 500 samples for each pole are shown as blue dots, while the poles obtained with regular VF are plotted in red. The locations of all relocated poles is shown in (A), while a closer look at the samples around on of the poles (indicated by the box) is shown in (B)

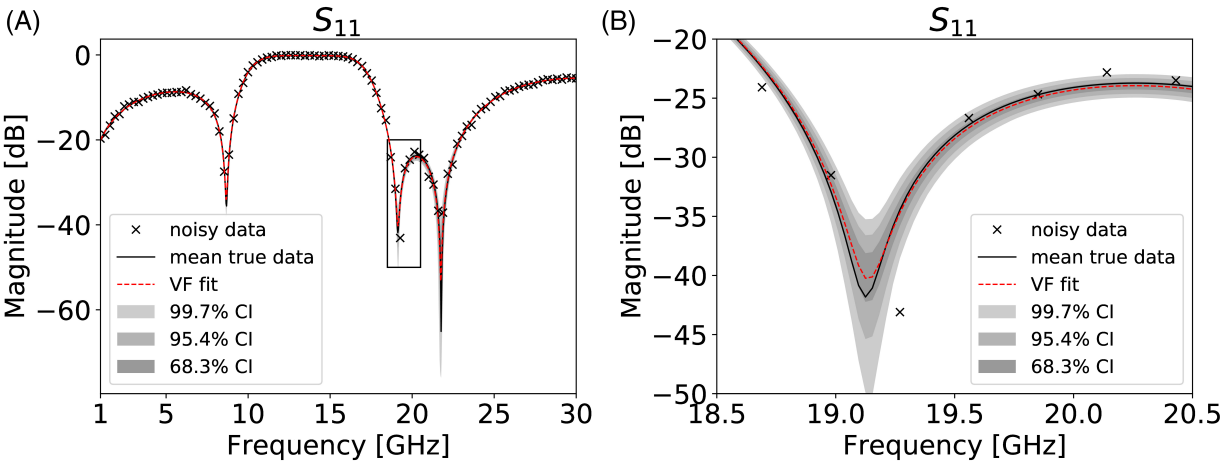


FIGURE 4 The result of linear Bayesian vector fitting applied to the S_{11} parameter of the double folded stub filter. The 10 000 rational models were used to construct confidence intervals of 68.27, 95.45, and 99.73% (the 1-, 2-, and 3-σ bounds of a Gaussian distribution). These confidence levels are shown in dark gray, gray, and light gray, respectively. The mean linear Bayesian vector fitting model, in this case corresponding to the regular VF fit, is shown as a dashed red line. Simulations of the nominal design (without noise) at 1001 frequency points, produced using ADS,[14] are shown as a black line. The noisy data used to fit the models is shown as black crosses. The entire frequency range (1-30 GHz) is shown in (A), while the (18.5-20.5 GHz) range is shown in (B)

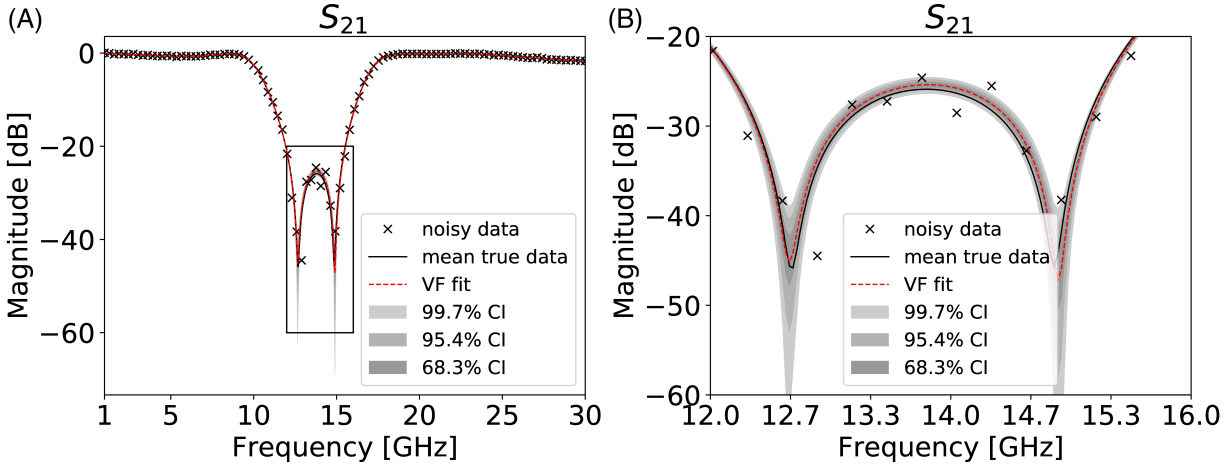


FIGURE 5 The result of linear Bayesian vector fitting applied to the S_{21} parameter of the double folded stub filter. Colors are as in Figure 4

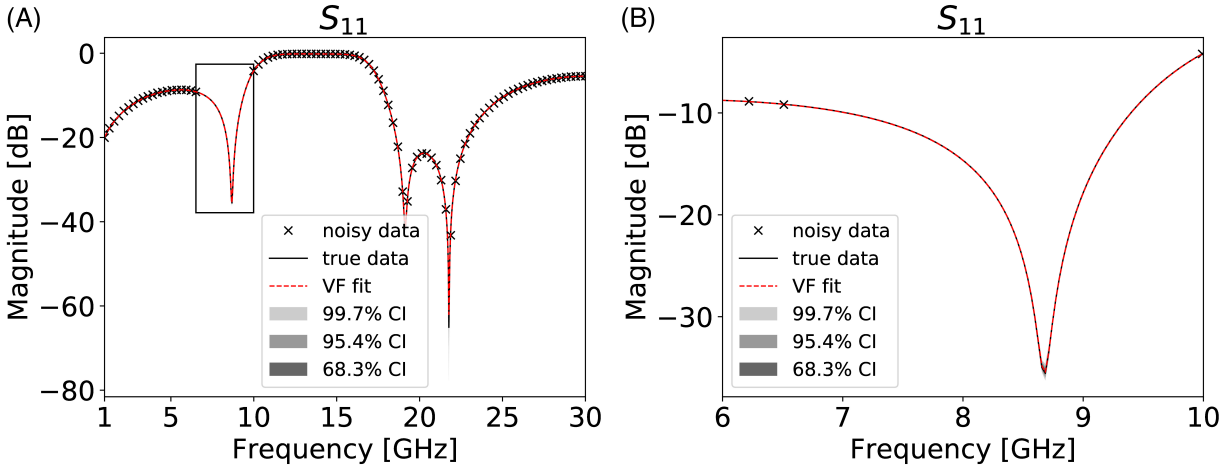


FIGURE 6 The result of linear Bayesian vector fitting applied to the S_{11} parameter of the double folded stub filter, with the data from 6.8 to 9.7 GHz removed, and noise with SD 10^{-3}. Colors as in Figure 4

With part of the data removed, and with low amount of additive Gaussian noise (a SD of 10^{-3}), the confidence bounds are very narrow, as shown in Figure 6. With a higher amount of noise (a SD of 10^{-2}), the confidence bounds increase as well, as illustrated in Figure 7. These figures show some overconfidence, due to the conditioning on the starting poles. Nevertheless, the confidence bounds can indicate when a design is not robust enough to remain above or below a certain threshold, given the uncertainty. As the classical VF fit (here coinciding with the mean of the LB-VF model) is quasi the same in both cases, such information is not present in a deterministic setting.

4.1.3 | Adaptive frequency sampling

If simulating a device at a certain frequency is computationally expensive, a full sweep over the frequency range of interest becomes cumbersome or even impossible given time constraints. Therefore, it is advisable to fully characterize the device with as few simulations as possible. A successful way to accomplish this is by simulating one frequency sample at a time, and build a rational model in the process. This is known as adaptive frequency sampling (AFS).[15-18]

A distinct advantage of having a measure of model uncertainty, such as LB-VF provides, is that it allows an efficient sequential sampling of the device responses in function of frequency. An LB-VF model's intrinsic uncertainty removes the necessity to rely on differences between models of different orders, as heuristic AFS algorithms[15-18]

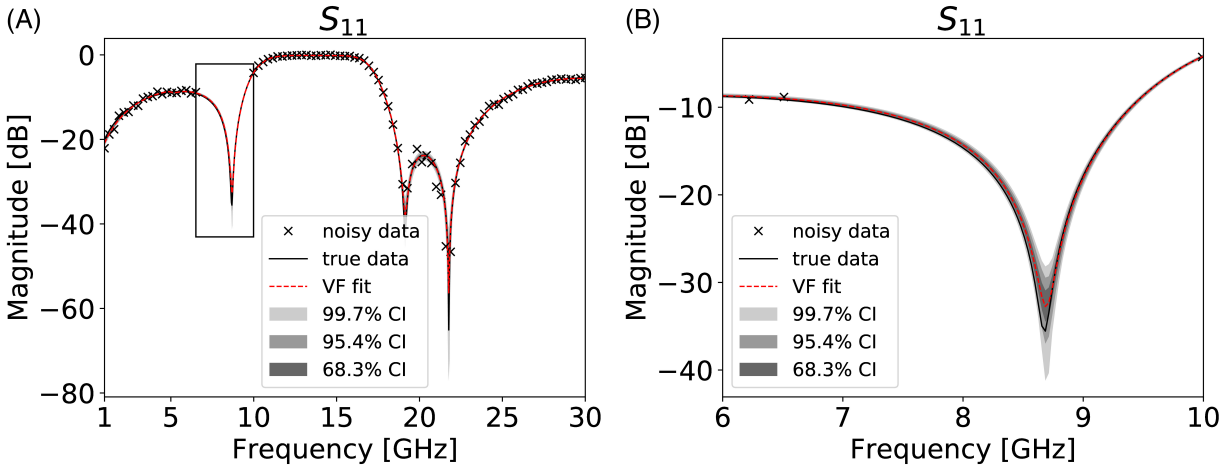


FIGURE 7 The result of linear Bayesian vector fitting applied to the S_{11} parameter of the double folded stub filter, with the data from 6.8 to 9.7 GHz removed, and noise with SD 10^{-2}. Colors are as in Figure 6

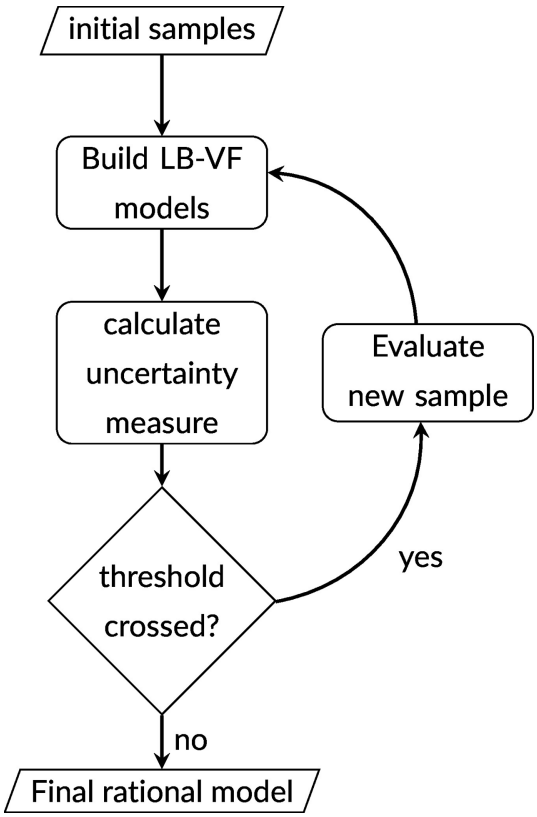


FIGURE 8 Flowchart of the proposed AFS strategy

do. Nevertheless, it is advantageous to consider several LB-VF models, in order to mitigate the overconfidence stemming from conditioning on the starting poles.

The uncertainty calculated from each of these models by grace of the samples drawn from them, is combined by weighting them by their normalized marginal likelihoods with respect to the pole relocation system (21). These weights are used to obtain a weighted SD of all samples. In order to increase the spread of sampled points, a small Gaussian-shaped penalty, with an amplitude of half the maximum weighted SD, and a SD of 10% of the distance between points, is subtracted around the known frequency points. This also adds an additional exploration focus, which can help to identify previously undetected resonances.

The AFS scheme used is summarized in Figure 8. After four initial frequency points are evaluated, all of the LB-VF models with highest order N are built, sampled from, and used to construct the uncertainty. The frequency point with

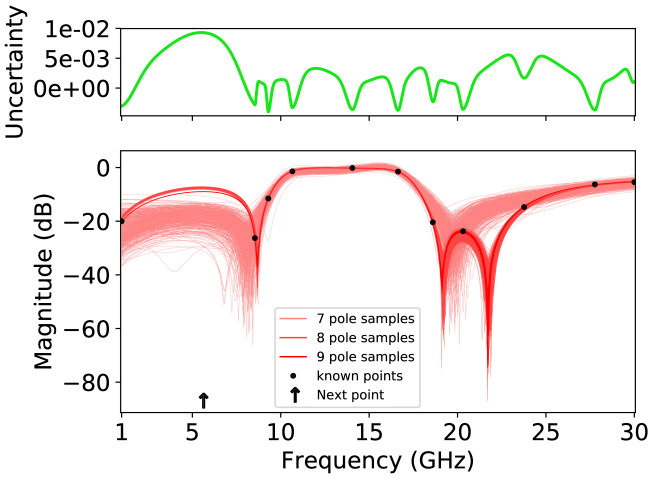


FIGURE 9 The eighth iteration in the AFS algorithm, for the S_{11} parameter of the double folded stub filter. In the lower plot, the known data points are represented as black dots. Samples drawn from linear Bayesian vector fitting models of different orders (500 each) are plotted in various shades of red, proportional to their log-likelihood. The upper plot shows the overall uncertainty measure in green. An arrow is also shown in the lower plot, at the frequency where this uncertainty is highest, and thus where the next evaluation will be done

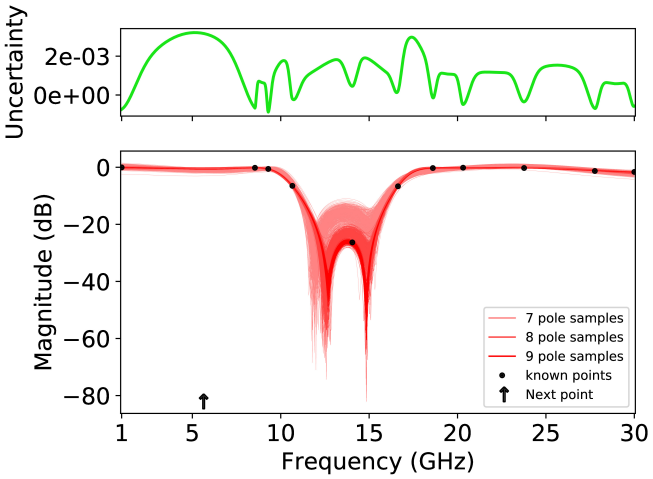


FIGURE 10 As Figure 9, but for the S_{21} parameter of the double folded stub filter. Note that the selection of the next sample is motivated by the higher uncertainty for the S_{11} parameter

highest uncertainty is determined and a new sample at this frequency is computed. For matrix-variate $\overline{\overline{S}}$, priority is given to the diagonal elements' uncertainty, and only the element with the highest uncertainty determines the next evaluation point. After this, new models are built, and so on. Using LB-VF models of different orders helps mitigate the conditioning on the number of poles, because of the weighted average being taken. In Reference 10, the 10 highest order models were used, to maximally utilize this effect. Naturally, taking more model orders into account increases the computational cost. Because of this, and for the sake of efficiency, only up to three of the highest order LB-VF models are built in this work, as a compromise. When the uncertainty no longer exceeds an a priori set threshold over the entire frequency range, convergence is assumed, and the model with the highest marginal likelihood is chosen as a suitable macromodel for the device response.

We apply the proposed AFS scheme using the measure of uncertainty described above to the example of the double folded stub filter. Note that no noise is added in this case. Figures 9 and 10 show the samples drawn from the LB-VF models and how they lead the selection of the next frequency point. After a total of 17 evaluations, the mean of the best fit [with $N = 15$ in (1)] achieves a total root mean squared error (RMSE) of −83.3 dB with respect to 1001 linearly spaced simulated samples. For comparison, a standard VF fit based on 17 uniformly spaced samples achieves an RMSE of −53.6 dB.

4.2 | RDRAM memory channel

The proposed framework is now applied to a Directed Rambus DRAM (RDRAM) memory channel.[19,20] The 4-port data in this example was measured using a vector network analyzer (from 50 MHz to 2.5 GHz), and shows significant coupling and reflection, and a large delay.

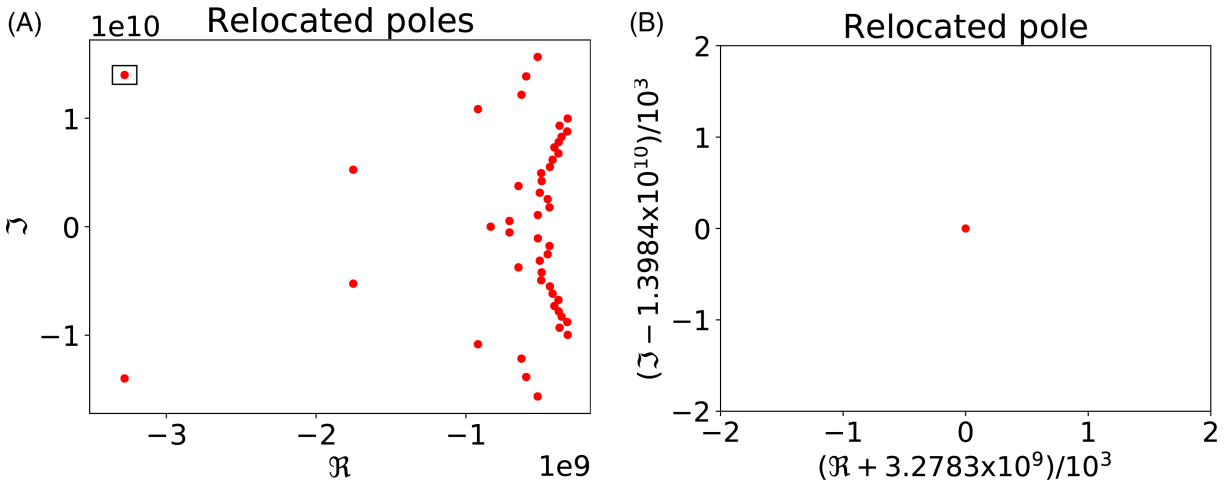


FIGURE 11 The sampled distribution of the poles as a result of linear Bayesian vector fitting applied to the S_{11} parameter of the RDRAM memory channel. The 500 samples for each pole are shown as blue dots, while the poles obtained with regular VF are plotted in red

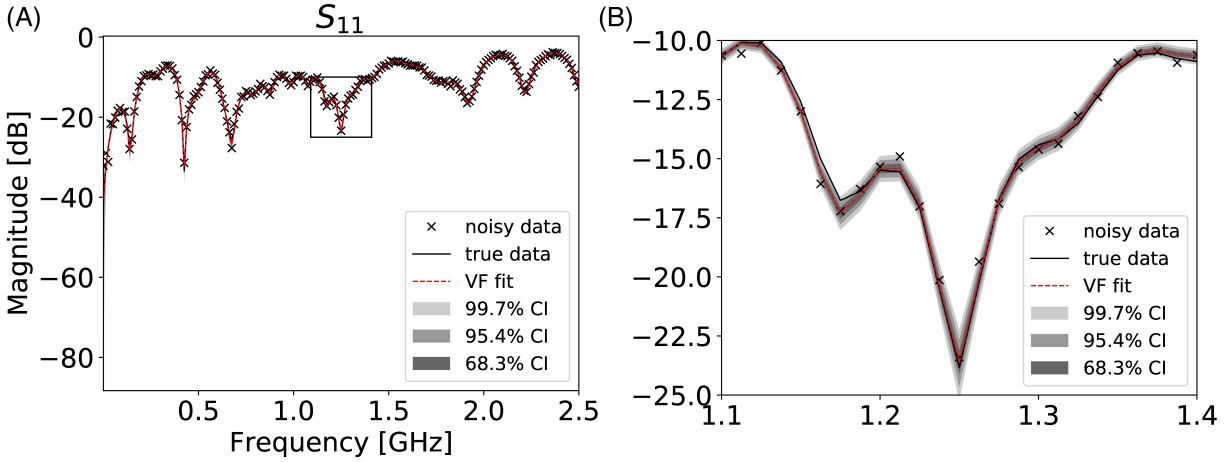


FIGURE 12 The result of linear Bayesian vector fitting applied to the S_{11} parameter of the RDRAM memory channel. The 10 000 rational models were again used to construct confidence intervals of 68.27, 95.45, and 99.73% (the 1-, 2-, and 3-σ bounds of a Gaussian distribution). Colors are as in Figure 4

4.2.1 | Uncertainty arising from additive noise

In this example, 51 out of the 201 measured points, in the frequency range of 0 to 2.5 GHz, were uniformly selected. Gaussian noise with zero mean and a SD of 0.01 was again added to the original data points. As for the previous example, an uninformative prior was used to sample $N_p = 500$ pole sets, and for each pole set, $N_r = 20$ residue sets were sampled, for a total of 10 000 models.

The application of the proposed LB-VF framework leads to the results shown in Figures 11-13, obtained for a model with 47 poles. Again, the mean of the LB-VF model, which coincides with a traditional VF fit, differs somewhat from the true data (the 201 measured points), but the latter remains within at least the predicted 99.73% confidence bound.

4.2.2 | Effect of missing data

As an additional example to verify the performance of the proposed modeling framework, the data from 1.50 to 1.75 GHz of the RDRAM memory channel was omitted before relocating the initial poles and applying the LB-VF model. Gaussian noise (with a SD of 10^{-2}) was also added. The benefit of a stochastic approach over a deterministic one

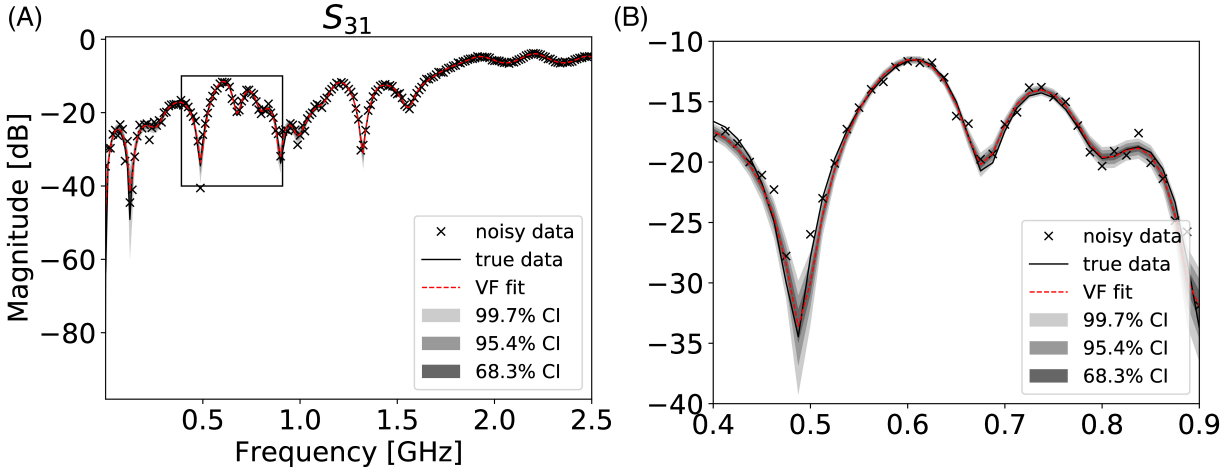
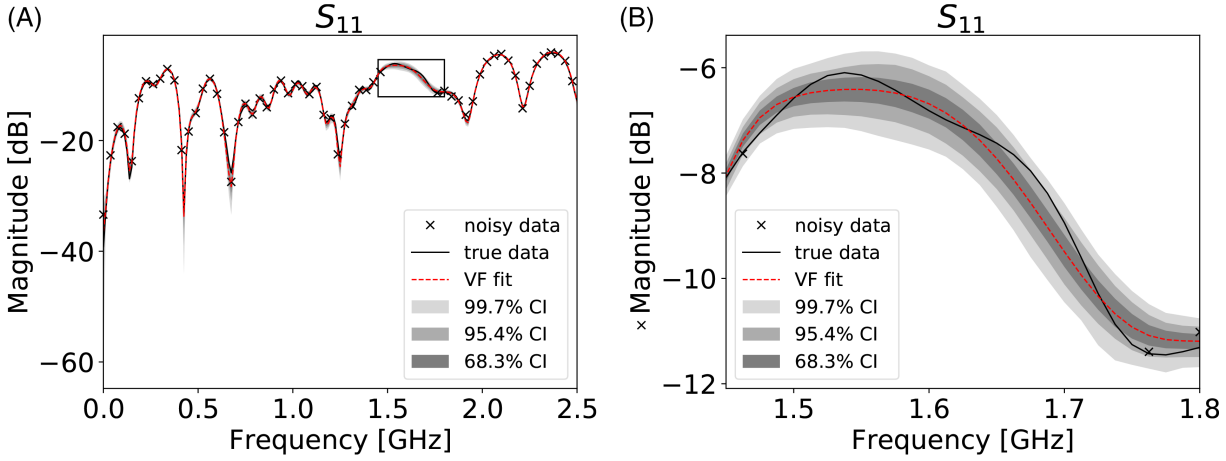


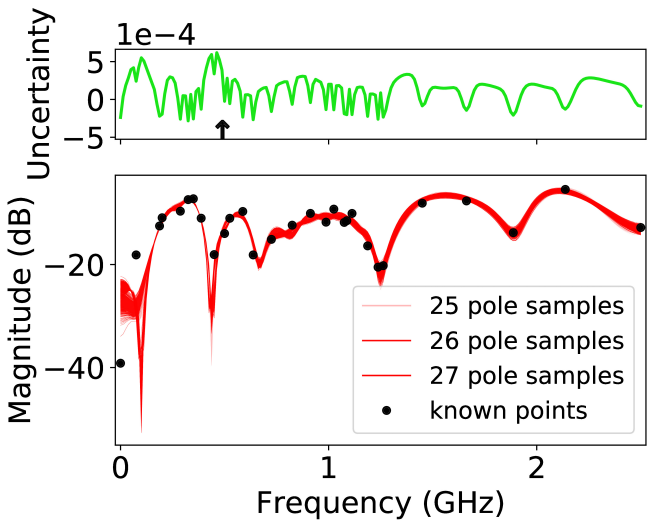
FIGURE 13 The result of linear Bayesian vector fitting applied to the S_{31} parameter of the RDRAM memory channel. Colors are as in Figure 4



(a) Entire frequency range. **(b)** Zoom of (a).

FIGURE 14 The result of linear Bayesian vector fitting applied to the S_{11} parameter of the RDRAM memory channel, with the data from 1.50 to 1.75 GHz removed, and noise with SD 10^{-3}. Colors are as in Figure 4

FIGURE 15 The 26th iteration in the adaptive frequency sampling algorithm, for the S_{11} parameter of the RDRAM memory channel. In the lower plot, the known data points are represented as black dots. Samples drawn from linear Bayesian vector fitting models of different orders (500 each) are plotted in various shades of red, proportional to their log-likelihood. The upper plot shows the overall uncertainty measure in green. An arrow is also shown in the lower plot, at the frequency where this uncertainty is highest, and thus where the next evaluation will be done



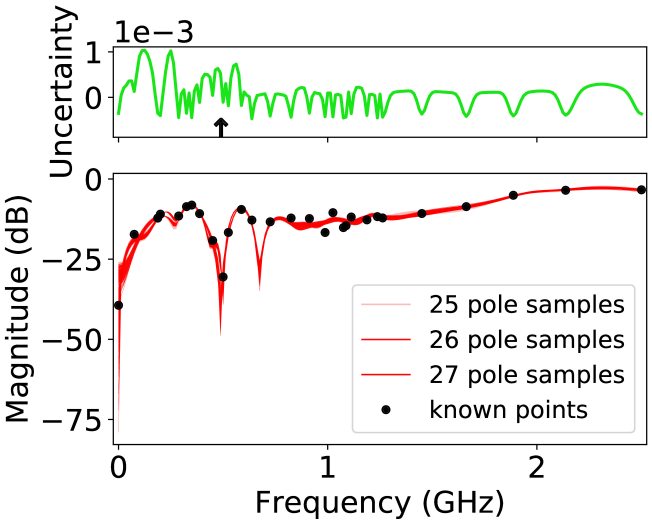


FIGURE 16 As Figure 15, but for S_{22}

is clear for this example as well. The obtained confidence bounds encompass most of the true function in the region where data is missing, as shown in Figure 14.

4.2.3 | Adaptive frequency sampling

The AFS algorithm outlined in section 4.1.3 can also be applied to the RDRAM memory channel example. Again, no noise (apart from that present in the data) is added in this case. An example of the samples drawn and used in the AFS algorithm is shown in Figures 15 and 16. Because the available measured data is a discrete set, the closest known point to the point of maximum uncertainty in each AFS step is added to the set of known points. After 66 evaluations, the mean of the best fit ($N = 63$) attains a total root mean squared error (RMSE) of −59.8 dB with respect to all measured data (201 points) available. For comparison, a standard VF fit based on 66 uniformly spaced samples achieves an RMSE of −43.3556 dB.

5 | CONCLUSIONS

In this article, a framework was introduced that expands the traditional VF macromodeling algorithm by extending it with stochastic information in a Bayesian manner. This is done by solving VF's linear systems using Bayesian linear regression, and sampling to propagate uncertainty through the nonlinear part of the VF algorithm.

The framework is able to provide quantitative information concerning the VF model uncertainty. As such, even without prior knowledge about the nature of the uncertainty in the data, it can provide insight into how confident a VF model is at any given frequency.

The applicability and potential of this framework is showcased by applying it to two realistic designs in three real-world applications: quantifying model uncertainty in the case of noisy observations, displaying model confidence in the case of missing data, and constructing a rational model with a minimal amount of evaluations using AFS. In each of these examples, the appositeness of the LB-VF framework is demonstrated.

ORCID

Simon De Ridder https://orcid.org/0000-0003-0493-0160
Domenico Spina https://orcid.org/0000-0003-2379-5259

REFERENCES

1. Lefteriu S, Antoulas AC. A new approach to modeling multiport systems from frequency-domain data. *IEEE Trans Comput-Aid Des Integr Circ Syst*. 2010;29(1):14-27. https://doi.org/10.1109/TCAD.2009.2034500.
2. Kabir M, Khazaka R. Macromodeling of distributed networks from frequency-domain data using the Loewner matrix approach. *IEEE Trans Microwave Theory Tech*. 2012;60(12):3927-3938.

3. Gustavsen B, Semlyen A. Rational approximation of frequency domain responses by vector fitting. *IEEE Trans Power Deliv*. 1999;14(3):1052-1061. https://doi.org/10.1109/61.772353.
4. Gustavsen B. Improving the pole relocating properties of vector fitting. *IEEE Trans Power Deliv*. 2006;21(3):1587-1592. https://doi.org/10.1109/TPWRD.2005.860281.
5. Deschrijver D, Mrozowski M, Dhaene T, De Zutter D. Macromodeling of multiport systems using a fast implementation of the vector fitting method. *IEEE Microwave Wirel Compon Lett*. 2008;18(6):383-385. https://doi.org/10.1109/LMWC.2008.922585.
6. Grivet-Talocia S, Gustavsen B. *Passive Macromodeling: Theory and Applications*. Hoboken, New Jersey: Wiley; 2015:239.
7. Spina D, Ferranti F, Antonini G, Dhaene T, Knockaert L, Vande GD. Time-domain Green's function-based parametric sensitivity analysis of multiconductor transmission lines. *IEEE Trans Compon Packag Manuf Technol*. 2012;2(9):1510-1517. https://doi.org/10.1109/TCPMT.2012.2186570.
8. Sahouli M, Dounavis A. An instrumental-variable QR decomposition vector-fitting method for modeling multiport networks characterized by Noisy frequency data. *IEEE Microwave Wirel Compon Lett*. 2016;26(9):645-647. https://doi.org/10.1109/LMWC.2016.2598141.
9. Grivet-Talocia S, Fevola E. Compact parameterized black-box modeling via Fourier-rational approximations. *IEEE Trans Electromagn Compat*. 2017;59(4):1133-1142. https://doi.org/10.1109/TEMC.2017.2649100.
10. De Ridder S, Deschrijver D, Spina D, Dhaene T, Vande GD. Adaptive frequency sampling using linear Bayesian vector fitting. *Electron Lett*. 2018;55:74-76. https://doi.org/10.1049/el.2018.6668.
11. Sanathanan C, Koerner J. Transfer function synthesis as a ratio of two complex polynomials. *IEEE Trans Autom Control*. 1963;8(1):56-58. https://doi.org/10.1109/TAC.1963.1105517.
12. Deschrijver D, Haegeman B, Dhaene T. Orthonormal vector fitting: a robust macromodeling tool for rational approximation of frequency domain responses. *IEEE Trans Adv Packag*. 2007;30(2):216-225.
13. Bakr MH, Bandler JW, Biernacki RM, Chen SH, Madsen K. A trust region aggressive space mapping algorithm for EM optimization. *IEEE Trans Microwave Theory Tech*. 1998;46(12):2412-2425. https://doi.org/10.1109/22.739229.
14. Keysight EEsof EDA. Advanced Design System. 2018. 471.update1.0.
15. Antonini G, Deschrijver D, Dhaene T. Broadband rational macromodeling based on the adaptive frequency sampling algorithm and the partial element equivalent circuit method. *IEEE Trans Electromagn Compat*. 2008;50(1):128-137. https://doi.org/10.1109/TEMC.2007.913225.
16. Dhaene T, Ureel J, Faché N, De Zutter D. Adaptive frequency sampling algorithm for fast and accurate S-parameter modeling of general planar structures. 1995;3:1427-1430.
17. Deschrijver D, Dhaene T. Passivity-based sample selection and adaptive vector fitting algorithm for pole-residue modeling of sparse frequency-domain data. 2004: 68–73
18. Mutonkole N, de Villiers DIL. Adaptive frequency sampling for radiation patterns and S-parameters of antennas. 2017: 3195–3199
19. Beyene WT, Yuan X, Cheng N, Shi H. Design, modeling, and hardware correlation of a 3.2Gb/s/pair memory channel. *IEEE Trans Adv Packag*. 2004;27(1):34-44. https://doi.org/10.1109/TADVP.2004.825463.
20. Stevens N, Deschrijver D, Dhaene T. Fast automatic order estimation of rational macromodels for signal integrity analysis. 2007: 89–92.

How to cite this article: De Ridder S, Deschrijver D, Spina D, Vande Ginste D, Dhaene T. Statistical modeling of frequency responses using linear Bayesian vector fitting. *Int J Numer Model El*. 2020;33:e2762. https://doi.org/10.1002/jnm.2762

APPENDIX A: DERIVATION OF POSTERIOR DISTRIBUTION FOR THE $\sigma(s)$-RESIDUE SYSTEM (8)

Using Bayes' theorem, (12) and (13) can be multiplied to obtain the form of the posterior distribution (with N_b the length of $\overline{b''}$ and N_x the length of $\overline{x}$):

$$\begin{aligned}
&\mathrm{P}\left(\overline{x},\sigma^2|\overline{b''}\right) \propto \mathrm{P}\left(\overline{b''}|\overline{x},\sigma^2\right)\mathrm{P}(\overline{x},\sigma^2)\\
&\propto N\left(\overline{b''}|\overline{\overline{A''}}\overline{x},\sigma^2\overline{\overline{I}}\right)N\left(\overline{x}|\overline{x_0},\sigma^2\overline{\overline{\Lambda_0}}^{-1}\right)\mathrm{IG}(\sigma^2|\alpha_0,\beta_0)\\
&\propto (\sigma^2)^{-\frac{N_b}{2}}\exp\left(\frac{-1}{2\sigma^2}\left(\overline{b''}-\overline{\overline{A''}}\overline{x}\right)^{\mathrm{T}}\left(\overline{b''}-\overline{\overline{A''}}\overline{x}\right)\right)\\
&\times(\sigma^2)^{-\frac{N_x}{2}}\exp\left(\frac{-1}{2\sigma^2}(\overline{x}-\overline{x_0})^{\mathrm{T}}\overline{\overline{\Lambda_0}}(\overline{x}-\overline{x_0})\right)(\sigma^2)^{-\alpha_0-1}\exp\left(-\frac{\beta_0}{\sigma^2}\right).
\end{aligned} \tag{A.1}$$

With some algebraic manipulation one can show that:

$$\left(\overline{\overline{b''}}-\overline{\overline{A''}}\overline{\overline{x}}\right)^{\mathrm{T}}\left(\overline{\overline{b''}}-\overline{\overline{A''}}\overline{\overline{x}}\right)=\left(\overline{\overline{b''}}-\overline{\overline{A''}}\overline{\overline{\hat{x}}}\right)^{\mathrm{T}}\left(\overline{\overline{b''}}-\overline{\overline{A''}}\overline{\overline{\hat{x}}}\right)+\left(\overline{\overline{x}}-\overline{\overline{\hat{x}}}\right)^{\mathrm{T}}\overline{\overline{A''}}^{\mathrm{T}}\overline{\overline{A''}}\left(\overline{\overline{x}}-\overline{\overline{\hat{x}}}\right), \tag{A.2}$$

where

$$\overline{\overline{\hat{x}}}=\left(\overline{\overline{A''}}^{\mathrm{T}}\overline{\overline{A''}}\right)^{-1}\overline{\overline{A''}}^{\mathrm{T}}\overline{\overline{b''}}. \tag{A.3}$$

This is easily verified by substituting (A.3) into the right hand side of (A.2). The second term in the right hand side of (A.2) can now be combined with the argument of the second exponential of (A.1) to obtain:

$$\begin{aligned}&\left(\overline{\overline{x}}-\overline{\overline{\hat{x}}}\right)^{\mathrm{T}}\overline{\overline{A''}}^{\mathrm{T}}\overline{\overline{A''}}\left(\overline{\overline{x}}-\overline{\overline{\hat{x}}}\right)+\left(\overline{\overline{x}}-\overline{\overline{x_0}}\right)^{\mathrm{T}}\overline{\overline{\Lambda_0}}\left(\overline{\overline{x}}-\overline{\overline{x_0}}\right)\\&=\left(\overline{\overline{x}}-\overline{\overline{x_f}}\right)^{\mathrm{T}}\overline{\overline{\Lambda_f}}\left(\overline{\overline{x}}-\overline{\overline{x_f}}\right)-\overline{\overline{x_f}}^{\mathrm{T}}\overline{\overline{\Lambda_f}}\overline{\overline{x_f}}+\overline{\overline{x_0}}^{\mathrm{T}}\overline{\overline{\Lambda_0}}\overline{\overline{x_0}}+\overline{\overline{\hat{x}}}^{\mathrm{T}}\overline{\overline{A''}}^{\mathrm{T}}\overline{\overline{A''}}\overline{\overline{\hat{x}}},\end{aligned} \tag{A.4}$$

with

$$\overline{\overline{\Lambda_f}}=\overline{\overline{\Lambda_0}}+\overline{\overline{A''}}^{\mathrm{T}}\overline{\overline{A''}} \tag{A.5}$$

$$\overline{\overline{x_f}}=\overline{\overline{\Lambda_f}}^{-1}\left(\overline{\overline{\Lambda_0}}\overline{\overline{x_0}}+\overline{\overline{A''}}^{\mathrm{T}}\overline{\overline{b''}}\right). \tag{A.6}$$

After some more calculation, we can also write:

$$\left(\overline{\overline{b''}}-\overline{\overline{A''}}\overline{\overline{\hat{x}}}\right)^{\mathrm{T}}\left(\overline{\overline{b''}}-\overline{\overline{A''}}\overline{\overline{\hat{x}}}\right)+\overline{\overline{\hat{x}}}^{\mathrm{T}}\overline{\overline{A''}}^{\mathrm{T}}\overline{\overline{A''}}\overline{\overline{\hat{x}}}=\overline{\overline{b''}}^{\mathrm{T}}\overline{\overline{b''}} \tag{A.7}$$

All this then leads to the posterior distribution:

$$\mathrm{P}\left(\overline{\overline{x}},\sigma^2|\overline{\overline{b''}}\right)\propto\left(\sigma^2\right)^{-\frac{N_x}{2}}\exp\left(\frac{-1}{2\sigma^2}\left(\overline{\overline{x}}-\overline{\overline{x_f}}\right)^{\mathrm{T}}\overline{\overline{\Lambda_f}}\left(\overline{\overline{x}}-\overline{\overline{x_f}}\right)\right)\left(\sigma^2\right)^{-\alpha_f-1}\exp\left(-\frac{\beta_f}{\sigma^2}\right), \tag{A.8}$$

in which

$$\alpha_f=\alpha_0+\frac{N_b}{2} \tag{A.9}$$

$$\beta_f=\beta_0+\frac{-\overline{\overline{x_f}}^{\mathrm{T}}\overline{\overline{\Lambda_f}}\overline{\overline{x_f}}+\overline{\overline{x_0}}^{\mathrm{T}}\overline{\overline{\Lambda_0}}\overline{\overline{x_0}}+\overline{\overline{b''}}^{\mathrm{T}}\overline{\overline{b''}}}{2}. \tag{A.10}$$

This posterior is proportional to (15).

APPENDIX B: DERIVATION OF THE MARGINAL POSTERIOR DISTRIBUTION $\mathrm{P}\left(\overline{\overline{\boldsymbol{x}}}|\overline{\overline{\boldsymbol{b}''}}\right)$

Using this result, the marginal posterior distribution can be derived by integrating the posterior with respect to σ^2 as follows:

$$
\begin{aligned}
&\mathrm{P}\left(\overline{x}|\overline{b''}\right)=\int_0^{+\infty}\mathrm{P}\left(\overline{x},\sigma^2|\overline{b''}\right)d(\sigma^2)\\
&=\int_0^{+\infty}(2\pi\sigma^2)^{-\frac{N_x}{2}}\left|\overline{\overline{\Lambda_f}}\right|^{\frac{1}{2}}\exp\left(\frac{-1}{2\sigma^2}\left(\overline{x}-\overline{x_f}\right)^{\mathrm{T}}\overline{\overline{\Lambda_f}}\left(\overline{x}-\overline{x_f}\right)\right)\\
&\quad\times\frac{\beta_f^{\alpha_f}}{\Gamma\left(\alpha_f\right)}\left(\sigma^2\right)^{-\alpha_f-1}\exp\left(\frac{-\beta_f}{\sigma^2}\right)d\left(\sigma^2\right)\\
&=\frac{(2\pi)^{-\frac{N_x}{2}}\beta_f^{\alpha_f}}{\Gamma\left(\alpha_f\right)\left|\overline{\overline{\Lambda_f}}\right|^{-\frac{1}{2}}}\frac{\Gamma\left(\frac{N_x}{2}+\alpha_f\right)}{\left[\frac{\left(\overline{x}-\overline{x_f}\right)^{\mathrm{T}}\overline{\overline{\Lambda_f}}\left(\overline{x}-\overline{x_f}\right)}{2}+\beta_f\right]^{\frac{N_x}{2}+\alpha_f}}\\
&\times\int_0^{+\infty}\mathrm{IG}\left(\sigma^2|\frac{N_x}{2}+\alpha_f,\frac{\left(\overline{x}-\overline{x_f}\right)^{\mathrm{T}}\overline{\overline{\Lambda_f}}\left(\overline{x}-\overline{x_f}\right)}{2}+\beta_f\right)d(\sigma^2)\\
&=\frac{\beta_f^{\alpha_f}\Gamma\left(\frac{N_x}{2}+\alpha_f\right)}{\Gamma\left(\alpha_f\right)(2\pi)^{\frac{N_x}{2}}\left|\overline{\overline{\Lambda_f}}\right|^{-\frac{1}{2}}}\left[\frac{\left(\overline{x}-\overline{x_f}\right)^{\mathrm{T}}\overline{\overline{\Lambda_f}}\left(\overline{x}-\overline{x_f}\right)}{2}+\beta_f\right]^{-\frac{N_x+2\alpha_f}{2}}\\
&=t_{2\alpha_f}\left(\overline{x}|\overline{x_f},\left(\frac{\alpha_f}{\beta_f}\overline{\overline{\Lambda_f}}\right)^{-1}\right)
\end{aligned}
\tag{B.1}
$$

APPENDIX C: DERIVATION OF THE MARGINAL LIKELIHOOD $\mathrm{P}\left(\overline{\boldsymbol{b}''}\right)$

As shown in (21), the marginal likelihood can be obtained by integrating the product of the likelihood (12) and the prior (13):

$$
\begin{aligned}
&\mathrm{P}\left(\overline{b''}\right)=\int\int(2\pi\sigma^2)^{-\frac{N_b}{2}}\exp\left(\frac{-1}{2\sigma^2}\left(\overline{b''}-\overline{\overline{A''}}\overline{x}\right)^{\mathrm{T}}\left(\overline{b''}-\overline{\overline{A''}}\overline{x}\right)\right)\\
&\times(2\pi\sigma^2)^{-\frac{N_x}{2}}\left|\overline{\overline{\Lambda_0}}\right|^{\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2}(\overline{x}-\overline{x_0})^{\mathrm{T}}\overline{\overline{\Lambda_0}}(\overline{x}-\overline{x_0})\right)\\
&\times\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}\left(\sigma^2\right)^{-\alpha_0-1}\exp\left(-\frac{\beta_0}{\sigma^2}\right)d\overline{x}d\sigma^2\\
&=\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}(2\pi)^{-\frac{N_b+N_x}{2}}\left|\overline{\overline{\Lambda_0}}\right|^{\frac{1}{2}}\int[\left(\sigma^2\right)^{-\frac{N_b+N_x+2\alpha_0}{2}-1}\exp\left(\frac{-\beta_f}{2\sigma^2}\right)\\
&\times\int\exp\left(\frac{-1}{2\sigma^2}\left(\overline{x}-\overline{x_f}\right)^{\mathrm{T}}\overline{\overline{\Lambda_f}}\left(\overline{x}-\overline{x_f}\right)\right)d\overline{x}]d\sigma^2\\
&=\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}(2\pi)^{-\frac{N_b}{2}}\left|\overline{\overline{\Lambda_0}}\right|^{\frac{1}{2}}\left|\overline{\overline{\Lambda_f}}\right|^{\frac{-1}{2}}\int\left(\sigma^2\right)^{-\frac{N_b+2\alpha_0}{2}-1}\exp\left(\frac{-\beta_f}{2\sigma^2}\right)d\sigma^2\\
&=(2\pi)^{-\frac{N_b}{2}}\sqrt{\frac{\left|\overline{\overline{\Lambda_0}}\right|}{\left|\overline{\overline{\Lambda_f}}\right|}}\frac{\beta_0^{\alpha_0}}{\beta_f^{\alpha_f}}\frac{\Gamma\left(\alpha_f\right)}{\Gamma(\alpha_0)}
\end{aligned}
\tag{C.1}
$$