International Journal of Innovative Computing, Information and Control
Volume **8**, Number **10(A)**, October **2012**
 ISSN 1349-4198
pp. 6761–6775

ON FIXED-ORDER FILTER DESIGN FOR UNCERTAIN CASCADED 2-1 SIGMA-DELTA MODULATORS: A DILATED PLANT APPROACH

Chun-Chen Lin[1], Yung-Shan Chou[1] and Venkataramanan Balakrishnan[2]

[1]Department of Electrical Engineering
Tamkang University
No. 151, Yingzhuan Rd., Tanshui Dist., New Taipei City 25137, Taiwan
894350015@s94.tku.edu.tw

[2]School of Electrical and Computer Engineering
Purdue University
West Lafayette, IN 47907, USA

Received May 2011; revised November 2011

Abstract. *This paper presents a new method to design the digital filters for correcting uncertain 2-1 cascaded sigma-delta ($\Sigma\Delta$) modulators. The main contribution of this paper consists of two parts. First, we develop a new filter design method, based on H_∞ loop shaping technique, to deal with a certain weighted matching condition with polytopic uncertainties in parameters. The feature of the proposed method is to show the filter order can be independent of the weighting function and determined beforehand. Therefore, in contrast to the conventional H_∞ loop shaping design method, lower-order filters can be obtained by using the proposed method. The second contribution is the application of the proposed method. For uncertain cascaded $\Sigma\Delta$ modulators, a low-order filter with the same order of the nominal filter is designed, which can efficiently reduce the H_∞ norm of the noise transfer function in the signal frequency band. Consequently, the signal-to-noise ratio (SNR) performance is improved. We compare the proposed method with other existing designs and establish its efficacy.*
Keywords: Uncertainty, Cascaded sigma-delta modulator, Reduced-order filter, H_∞ loop shaping, Linear matrix inequality

1. **Introduction.** Sigma-delta ($\Sigma\Delta$) modulators [1] are important devices which have found widespread application in high-speed analog-to-digital (A/D) conversion for modern digital signal processing. In order to achieve higher signal-to-noise-ratio (SNR) performance of a $\Sigma\Delta$ modulator, a common way is to increase the oversampling ratio (OSR) and modulator order. For a realizable OSR, increasing the modulator order is a good way to achieve fine signal resolution, because this directly leads to high-order noise shaping. However, stability is a serious problem for single-stage $\Sigma\Delta$ modulators of order higher than two. Therefore, multi-stage (or cascaded) architecture becomes a good alternative to cope with the stability problem. For achieving fine SNR, cascaded $\Sigma\Delta$ modulators rely heavily on accurate matching of analog stages with digital filters to prevent leakage of quantization noise. However, in fact, perfect matching is impossible due to the limited accuracy of the implementation technologies. Techniques that do not explicitly account for this often yield designs that perform poorly [1].

Thus, in recent years, considerable effort has been devoted to the study of robust matching filters under the framework of "model matching" [2-4]. The basic idea is to recast the filter design problem as a problem of minimizing the worst-case value, over all possible uncertainties, of a measure of a certain model mismatch (we will present details in Section 2). The specific mismatch measure that has been most often used is the H_∞

norm which measures the peak value of the mismatch over all frequencies. To achieve higher-order noise shaping, a weighting function was introduced in [2]; the uncertainties were of the "polytopic" type, and the problem became that of the minimization of the H_∞ norm of the weighted matching error over a linearized polytopic model. While the introduction of weighting functions is useful in shaping the noise transfer function (NTF) so as to increase the SNR, it also increases the order of the filter, which in turn leads to increased complexity of circuit implementation. To alleviate this problem, two fixed-order designs have been proposed [3,4]. In [3], the central polynomial linear matrix inequality (LMI) method has been employed. In particular, the order of the resulting infinite impulse response (IIR) filter is independent of that of the introduced weighting function. Moreover, the filter order can be chosen to be any positive integer. In contrast to the mathematical approach in [2,3], a design based more on engineering insight was presented in [4]. A low-frequency linearized model of a 2-1 cascaded modulator was derived and, again a fixed-order (but finite impulse response (FIR)) filter design was presented based on a formal optimization method.

In this paper, we revisit the weighted H_∞ norm minimization formulation in [2], and directly address the issue of high filter order. Our main contribution is to develop a new reduced-order filter design procedure that yields filters whose order is equal to the plant order, independent of the weighting function. We show that this approach yields filters whose performance compares favorably with those presented in [3,4]. The rest of this work is organized as follows. In Section 2, an uncertain cascaded 2-1 $\Sigma\Delta$ modulator is briefly described and the problem formulation is presented. In Section 3, the proposed fixed-order IIR filter design is provided. Section 4 shows the simulation results with comparison to some of the existing works. Section 5 is the conclusion.

2. **Problem Formulation.** A block-diagram of a cascaded 2-1 $\Sigma\Delta$ modulator is depicted in Figure 1 where H_i $(i = 1, 2, 3)$ and F are, respectively, the integrators and the digital filter. U denotes the input signal. E_j $(j = 1, 2)$ represent the quantization noises. Y is the output signal. z^{-1} is one-step delay. Accordingly, the signal transfer function (*STF*) and the noise transfer function (*NTF*) of the cascaded 2-1 modulator are defined as the transfer functions from the signals U, E_1, and E_2 to the output signal Y, respectively.

It follows that

$$\begin{aligned} Y(z) = {} & STF_1(z) \times U(z) + (NTF_1(z) - F(z) \times STF_2(z))\, E_1(z) \\ & - F(z) \times NTF_2(z) \times E_2(z) \end{aligned} \tag{1}$$

where

$$\begin{aligned} STF_1(z) &= \frac{H_1(z)\, H_2(z)}{1 + z^{-1}H_2(z) + z^{-1}H_1(z)\, H_2(z)}, \\ NTF_1(z) &= \frac{1}{1 + z^{-1}H_2(z) + z^{-1}H_1(z)\, H_2(z)}, \\ STF_2(z) &= \frac{H_3(z)}{1 + z^{-1}H_3(z)}, \\ NTF_2(z) &= \frac{1}{1 + z^{-1}H_3(z)}. \end{aligned} \tag{2}$$

Considering the case of ideal components, in order to eliminate the effects on the output Y due to the 1st-stage quantization error E_1, which is the most significant one among the noises E_1 and E_2, the matching filter F is chosen as

$$F(z) = NTF_1(z)/STF_2(z) \tag{3}$$

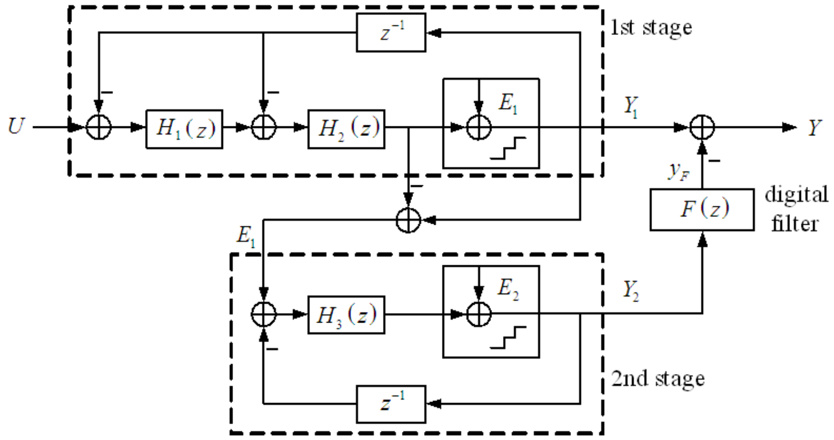

FIGURE 1. A cascaded 2-1 ΣΔ modulator with 1-bit quantizers

to meet the matching condition (4)

$$NTF_1(z) - F(z) \times STF_2(z) = 0. \tag{4}$$

Then (1) reduces to

$$Y(z) = STF_1(z) \times U(z) - F(z) \times NTF_2(z) \times E_2(z) \tag{5}$$

and only leaves the input signal U and the second-stage quantization noise E_2. It is readily checked that a third-order noise shaping, i.e., $F(z) \times STF_2(z)$, is obtained.

With compared to low-order case, it is known that the third-order noise shaping can shape the spectrum of the quantisation noise to place more of its energy outside the signal band. However, perfect cancellation of E_1 is not possible in practice owing to non-ideal analog components. Recall that two common sources of analog imperfections in a ΣΔ modulator are finite amplifier gain and mismatch in capacitor values. These non-idealities affect the locations of singularities (poles and zeros) of the transfer functions, NTF_1 and STF_2, and then result in an imperfective cancellation on (4). Consequently, it results in low-order noise shaping and degrades SNR performance to some degree.

From [2-4], these non-idealities are modeled as parametric uncertainties in the gains and poles of the integrators. Therefore, a non-ideal integrator can be modeled as

$$H_i(z) \cong \frac{1-\delta_a}{1-(1-\delta_b)z^{-1}}, \quad i = 1, 2, 3 \tag{6}$$

where $\delta_a \in [0, 1)$ and $\delta_b \in [0, 1)$ are the parameter deviations from the nominal values. Based on (6), an important issue arises that is how large values of δ_a and δ_b will have great effect on the system performance?

From [5,6], one can find that the modulator performance degrades seriously when the values of δ_a and δ_b are larger than 0.001. This implies that the matching filter $F(3)$ is not suitable to be used for achieving a fine noise shaping when the uncertainties in the gain and poles of the integrator do not lie in the ranges $0 \le \delta_a \le 0.001$, $0 \le \delta_b \le 0.001$. In fact, if the modulators are not implemented with careful layout, larger unit capacitor area, and enough finite DC gain, the values of δ_a and δ_b would be larger than 0.001. Therefore, the goal of this work is to redesign the filter for achieving fine noise shaping when the uncertainties are larger than 0.001.

For the later development, we assume the integrators $H_2(z)$ and $H_3(z)$ are ideal, and thus the uncertain $NTF_1(z)$ becomes

$$NTF_1(z) = \frac{1 - (2 - \delta_b) z^{-1} + (1 - \delta_b) z^{-2}}{1 - (-\delta_b + \delta_a) z^{-1}} \tag{7}$$

and $STF_2(z) = 1$. Our work is to minimize the effect of quantization noise $E_1(z)$ on the output $Y(z)$ in the signal band. This can be formulated as a weighted H_∞ norm minimization problem

$$\min_{F(z)} \| W(z) (NTF_1(z) - F(z)) \|_\infty \tag{8}$$

where $W(z)$ is a weighting function that is employed to shape the noise transfer function $(NTF_1(z) - F(z))$ for $E_1(z)$. We assume that the transfer functions $W(z)$, $F(z)$, and $NTF_1(z)$ have the following state-space realizations:

$$W(z) := C_W (zI - A_W)^{-1} B_W + D_W \tag{9}$$

$$F(z) := C_F (zI - A_F)^{-1} B_F + D_F \tag{10}$$

$$NTF_1(z) := C_1 (zI - A_1)^{-1} B_1 + D_1 \tag{11}$$

where

$$A_1 = \begin{pmatrix} -\delta_b + \delta_a & 1 \\ 0 & 0 \end{pmatrix}, \quad B_1 = \begin{pmatrix} -2 + \delta_a \\ 1 - \delta_b \end{pmatrix}, \quad C_1 = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad D_1 = 1 \tag{12}$$

and

$$\begin{aligned} &A_F \in R^{nf \times nf}, \quad A_W \in R^{nw \times nw}, \quad A_1 \in R^{np \times np}, \\ &B_1 \in R^{np \times 1}, \quad C_1 \in R^{1 \times np}, \quad D_1 \in R^{1 \times 1}. \end{aligned} \tag{13}$$

In order to take the uncertainties δ_a and δ_b into account in problem (8), we assume the uncertain matrices (A_1, B_1, C_1, D_1) of $NTF_1(z)$ belong to the following uncertainty polytope [7]:

$$\Omega = \left\{ (A_1, B_1, C_1, D_1) \,|\, (A_1, B_1, C_1, D_1) = \sum_{i=1}^{m} \alpha_i \left(A_1^{(i)}, B_1^{(i)}, C_1^{(i)}, D_1^{(i)} \right), \; \alpha_i \geq 0, \sum_{i=1}^{m} \alpha_i = 1 \right\}. \tag{14}$$

Here giving the values of α_i, $i = 1, \cdots, m$, with $\alpha_i \geq 0$ and $\alpha_1 + \alpha_2 + \cdots + \alpha_m = 1$ produces an element of Ω.

Accordingly, it is known that the conventional H_∞ loop-shaping technique can be employed to design the matching filter $F(z)$ for (8), where its order is equal to the sum of plant order and the order of the weighting function, i.e., $nf = np + nw$. Hence, the hardware realization of the robust matching filter is more complex than that of the conventional filter (3). Therefore, we present a new filter design method to alleviate the problem in the following section.

3. **Main Results.** In this section, we present a new method to design the fixed-order filter for the weighted H_∞ norm minimization problem. To produce the filters of desired order nf, we introduce the delay elements of appropriate order to adjust the size of the relevant transfer functions.

For proceeding, based on (13), two parameters M_d and N_d are defined in Table 1. In Table 1, one can find the values of M_d and N_d for four cases, where $round(x)$ denotes rounding up, i.e., $round(x)$ is the smallest positive integer that satisfies $round(x) \geq x$.

Table 1. Selection of M_d and N_d

	M_d	N_d
Case 1. $np \geq nf$ and $nw \geq nf$	$M_d = round\,(np/nf)$	$N_d = round\,(nw/nf)$
Case 2. $np \geq nf$ and $nw < nf$	$M_d = round\,(np/nf)$	$N_d = 1$
Case 3. $np < nf$ and $nw \geq nf$	$M_d = 1$	$N_d = round\,(nw/nf)$
Case 4. $np < nf$ and $nw < nf$	$M_d = 1$	$N_d = 1$

Next, two transfer functions are considered, i.e.,

$$Z_p(z) = \left(z^{-1}\right)^{mp}, \quad Z_w(z) = \left(z^{-1}\right)^{mw}, \tag{15}$$

where

$$mp = M_d \times nf - np \text{ and } mw = N_d \times nf - nw. \tag{16}$$

Accordingly, a dilated plant and a dilated weighting function are considered as

$$NTF_{1d}(z) := NTF_1(z) \times Z_p(z), \tag{17}$$

and

$$W_d(z) := W(z) \times Z_w(z), \tag{18}$$

respectively. In this case, the state-space realization of $Z_p(z)$ and $W_d(z)$ are defined as

$$Z_p(z) := C_d\,(zI - A_d)^{-1}\,B_d + D_d \tag{19}$$

and

$$W_d(z) := \bar{C}_W\left(zI - \bar{A}_W\right)^{-1}\bar{B}_W + \bar{D}_W, \tag{20}$$

respectively.

From the property of delay element, it follows that $NTF_{1d}(z) - F(z) \times Z_p(z)$ and $W_d(z)$ give

$$\begin{aligned}\|NTF_{1d}(z) - F(z) \times Z_p(z)\|_\infty &= \|(NTF_1(z) - F(z)) \times Z_p(z)\|_\infty \\ &= \|NTF_1(z) - F(z)\|_\infty\end{aligned}$$

and

$$\|W_d(z)\|_\infty = \|W(z) \times Z_w(z)\|_\infty = \|W(z)\|_\infty,$$

respectively. More explicitly, delay element is regarded as a solution to increase the order of $\left[\begin{array}{cc} NTF_1(z) & 1 \end{array}\right]^T$ and $W(z)$ and the bode magnitude of the dilated transfer functions $\left[\begin{array}{cc} NTF_1(z) & 1 \end{array}\right]^T \times Z_p(z)$ and $W_d(z)$ is not to be altered. In what follows, the weighted matching error $W(z)\,(NTF_1(z) - F(z))$ is replaced by

$$W_d(z)\,(NTF_{1d}(z) - F(z) \times Z_p(z)). \tag{21}$$

Then, it is readily verified the state-space realization of (21) denoted by $T(z) := C_T\,(zI - A_T)^{-1}\,B_T + D_T$ is given by

$$T(z) := \sum_{i=1}^{m} \alpha_i \left(\begin{array}{c|c} A_T^{(i)} & B_T^{(i)} \\ \hline C_T^{(i)} & D_T^{(i)} \end{array}\right) = \sum_{i=1}^{m} \alpha_i \left(\begin{array}{cc|c} A_M^{(i)} & B_M^{(i)}\bar{C}_W & B_M^{(i)}\bar{D}_W \\ 0 & \bar{A}_W & \bar{B}_W \\ \hline C_M^{(i)} & D_M^{(i)}\bar{C}_W & D_M^{(i)}\bar{D}_W \end{array}\right) \tag{22}$$

where

$$\begin{aligned} A_M^{(i)} &= \begin{bmatrix} \bar{A}_1^{(i)} & 0 \\ B_F\bar{C}^{(i)} & A_F \end{bmatrix}, \quad B_M^{(i)} = \begin{bmatrix} \bar{B}_1^{(i)} \\ B_F\bar{D}^{(i)} \end{bmatrix}, \\ C_M^{(i)} &= \begin{bmatrix} \bar{C}_1^{(i)} - D_F\bar{C}^{(i)} & -C_F \end{bmatrix}, \quad D_M^{(i)} = \begin{bmatrix} \bar{D}_1^{(i)} - D_F\bar{D}^{(i)} \end{bmatrix} \end{aligned} \tag{23}$$

and

$$\bar{A}_1^{(i)} = \begin{bmatrix} A_d & 0 \\ B_1^{(i)} C_d & A_1^{(i)} \end{bmatrix}, \quad \bar{B}_1^{(i)} = \begin{bmatrix} B_d \\ B_1^{(i)} D_d \end{bmatrix},$$
$$\bar{C}^{(i)} = \begin{bmatrix} C_d & 0 \end{bmatrix}, \quad \bar{C}_1^{(i)} = \begin{bmatrix} D_1^{(i)} C_d & C_1^{(i)} \end{bmatrix},$$
$$\bar{D}^{(i)} = D_d, \quad \bar{D}_1^{(i)} = D_1^{(i)} D_d$$

and

$$A_F \in R^{nf \times nf}, \quad \bar{A}_W \in R^{(nw+mw)\times(nw+mw)},$$
$$\bar{A}_1^{(i)} \in R^{(np+mp)\times(np+mp)}, \quad \bar{B}_1^{(i)} \in R^{(np+mp)\times 1},$$
$$\bar{C}_1^{(i)} \in R^{1\times(np+mp)}, \quad \bar{D}_1 \in R^{1\times 1}.$$

This recasts the design problem (8) to that of finding a filter of form (10) via the solution of the following optimization problem:

$$\min_{A_F, B_F, C_F, D_F} \gamma \tag{24}$$

subject to $||C_T (zI - A_T)^{-1} B_T + D_T||_\infty < \gamma$.

To solve the optimization problem described above, we first provide the following lemma.

Lemma 3.1. [2] *Suppose that the filter matrices (A_F, B_F, C_F, D_F) are known. For all (A_1, B_1, C_1, D_1) belonging to Ω, the condition $||C_T (zI - A_T)^{-1} B_T + D_T||_\infty < \gamma$ holds if there exist a matrix G and matrices $P^{(i)} = P^{(i)T}$ $(i = 1, \ldots, m)$ satisfying*

$$\begin{bmatrix} G + G^T - P^{(i)} & 0 & GA_T^{(i)} & GB_T^{(i)} \\ 0 & I & C_T^{(i)} & D_T^{(i)} \\ A_T^{(i)T} G^T & C_T^{(i)T} & P^{(i)} & 0 \\ B_T^{(i)T} G^T & D_T^{(i)T} & 0 & \gamma^2 I \end{bmatrix} > 0, \quad i = 1, \ldots, m. \tag{25}$$

In Lemma 3.1, we suppose that the filter matrices (A_F, B_F, C_F, D_F) are known. The goal of this paper is to design the filter such that (24) is minimized. Therefore, the filter matrices are variables to be determined in this work. We now present Theorem 3.1 that states that there exist the filter matrices if certain matrix inequality constraints (26) are satisfied.

Theorem 3.1. *For all m vertices*

$$\left(\bar{A}_1^{(i)}, \bar{B}_1^{(i)}, \bar{C}_1^{(i)}, \bar{D}_1^{(i)}\right) \quad (i = 1, \ldots, m),$$

there exists a suboptimal filter (10) of order nf to problem (24) if optimization problem (26) is feasible for $i = 1, \ldots, m$.

$$\min_{\chi, \phi, P_{11}^{(i)} = (P_{11}^{(i)})^T, P_{g12}^{(i)}, P_{13}^{(i)}, P_{g22}^{(i)} = (P_{g22}^{(i)})^T, P_{g23}^{(i)}, P_{33}^{(i)} = (P_{33}^{(i)})^T, G_{11}, G_{13}, G_{31}, G_{33}, M, N, S, Q_A, Q_B, Q_C, Q_D \ (i=1,\ldots,m)} \gamma \tag{26}$$

subject to

$$\left[\begin{array}{cccccccc} G_{11}+G_{11}^T-P_{11}^{(i)} & \chi M^T+S^T-P_{g12}^{(i)} & G_{13}+G_{31}^T-P_{13}^{(i)} & 0 & G_{11}\bar{A}_1^{(i)}+\chi Q_B\bar{C}^{(i)} & \chi Q_A & G_{11}\bar{B}_1^{(i)}\bar{C}_W+\chi Q_B\bar{D}^{(i)}\bar{C}_W+G_{13}\bar{A}_W & G_{11}\bar{B}_1^{(i)}\bar{D}_W+\chi Q_B\bar{D}^{(i)}\bar{D}_W+G_{13}\bar{B}_W \\ * & M^T+M-P_{g22}^{(i)} & N+M\phi^T-P_{g23}^{(i)} & 0 & S\bar{A}_1^{(i)}+Q_B\bar{C}^{(i)} & Q_A & S\bar{B}_1^{(i)}\bar{C}_W+Q_B\bar{D}^{(i)}\bar{C}_W+N\bar{A}_W & S\bar{B}_1^{(i)}\bar{D}_W+Q_B\bar{D}^{(i)}\bar{D}_W+N\bar{B}_W \\ * & * & G_{33}+G_{33}^T-P_{33}^{(i)} & 0 & G_{31}\bar{A}_1^{(i)}+\phi Q_B\bar{C}^{(i)} & \phi Q_A & G_{31}\bar{B}_1^{(i)}\bar{C}_W+\phi Q_B\bar{D}^{(i)}\bar{C}_W+G_{33}\bar{A}_W & G_{31}\bar{B}_1^{(i)}\bar{D}_W+\phi Q_B\bar{D}^{(i)}\bar{D}_W+G_{33}\bar{B}_W \\ * & * & * & I & \bar{C}_1^{(i)}-Q_D\bar{C}^{(i)} & -Q_C & \bar{D}_1^{(i)}\bar{C}_W-Q_D\bar{D}^{(i)}\bar{C}_W & \bar{D}_1^{(i)}\bar{D}_W-Q_D\bar{D}^{(i)}\bar{D}_W \\ * & * & * & * & P_{11}^{(i)} & P_{g12}^{(i)} & P_{13}^{(i)} & 0 \\ * & * & * & * & * & P_{g22}^{(i)} & P_{g23}^{(i)} & 0 \\ * & * & * & * & * & * & P_{33}^{(i)} & 0 \\ * & * & * & * & * & * & * & \gamma^2 I \end{array}\right] > 0$$

and

$$\chi = \alpha \otimes I_{nf}, \quad \phi = \beta \otimes I_{nf}$$

with

$$\alpha = [\alpha_1, \cdots, \alpha_{N_d}]^T \in R^{N_d \times 1}, \quad \beta = [\beta_1, \cdots, \beta_{M_d}]^T \in R^{M_d \times 1}.$$

In the case, the filter is given by

$$F(z) := Q_C \left(zI - M^{-T}Q_A\right)^{-1} M^{-T}Q_B + Q_D \tag{27}$$

Proof: We will invoke Lemma 3.1 to derive the solvability condition for the filter with order nf. To proceed, partition the matrices $P^{(i)}$ and G as follows:

$$P^{(i)} = \begin{bmatrix} P_{11}^{(i)} & P_{12}^{(i)} & P_{13}^{(i)} \\ P_{12}^{(i)T} & P_{22}^{(i)} & P_{23}^{(i)} \\ P_{13}^{(i)T} & P_{23}^{(i)T} & P_{33}^{(i)} \end{bmatrix}, \quad G = \begin{bmatrix} G_{11} & G_{12} & G_{13} \\ G_{21} & G_{22} & G_{23} \\ G_{31} & G_{32} & G_{33} \end{bmatrix} \tag{28}$$

where

$$P_{11}^{(i)} \in R^{(np+mp)\times(np+mp)}, \quad P_{22}^{(i)} \in R^{nf\times nf}, \quad P_{33}^{(i)} \in R^{(nw+mw)\times(nw+mw)},$$
$$G_{11} \in R^{(np+mp)\times(np+mp)}, \quad G_{22} \in R^{nf\times nf}, \quad G_{33} \in R^{(nw+mw)\times(nw+mw)}.$$

Without loss of generality, G_{22} is assumed to be nonsingular. Under the constraint $G_{12} = \chi G_2$, $G_{32} = \phi G_2$, $\chi = \alpha \otimes I_{nf}$, $\phi = \beta \otimes I_{nf}$ with $\alpha = [\alpha_1, \cdots, \alpha_{N_d}]^T \in R^{N_d\times 1}$, $\beta = [\beta_1, \cdots, \beta_{M_d}]^T \in R^{M_d\times 1}$, apply congruence transformation $J = diag\,(T, I, T, I)$ to (25) (i.e., multiplying (25) on the left by J and on the right by J^T) where $T = diag\left(I, G_2G_{22}^{-1}, I\right)$ and define

$$P_{g12}^{(i)} = P_{12}^{(i)}G_{22}^{-T}G_2^T, \quad P_{g22}^{(i)} = G_2G_{22}^{-1}P_{22}^{(i)}G_{22}^{-T}G_2^T, \quad P_{g23}^{(i)} = G_2G_{22}^{-1}P_{23}^{(i)} \tag{29}$$

we obtain (30), i.e.,

$$\begin{bmatrix} G_{11}+G_{11}^T-P_{11}^{(i)} & \Xi_{12} & \Xi_{13} & 0 & \Xi_{15} & \Xi_{16} & \Xi_{17} & \Xi_{18} \\ * & \Xi_{22} & \Xi_{23} & 0 & \Xi_{25} & \Xi_{26} & \Xi_{27} & \Xi_{28} \\ * & * & \Xi_{33} & 0 & \Xi_{35} & \Xi_{36} & \Xi_{37} & \Xi_{38} \\ * & * & * & I & \Xi_{45} & \Xi_{46} & \Xi_{47} & \Xi_{48} \\ * & * & * & * & P_{11}^{(i)} & P_{g12}^{(i)} & P_{13}^{(i)} & 0 \\ * & * & * & * & * & P_{g22}^{(i)} & P_{g23}^{(i)} & 0 \\ * & * & * & * & * & * & P_{33}^{(i)} & 0 \\ * & * & * & * & * & * & * & \gamma^2 I \end{bmatrix} > 0 \tag{30}$$

where

$$\begin{aligned}
\Xi_{12} &= \chi G_2 G_{22}^{-T} G_2^T + G_{21}^T G_{22}^{-T} G_2^T - P_{g12}^{(i)}, \\
\Xi_{22} &= G_2 G_{22}^{-T} G_2^T + G_2 G_{22}^{-1} G_2^T - P_{g22}^{(i)}, \\
\Xi_{13} &= G_{13} + G_{31}^T - P_{13}^{(i)}, \\
\Xi_{23} &= G_2 G_{22}^{-1} G_{23} + G_2 G_{22}^{-1} G_2^T \phi^T - P_{g23}^{(i)}, \\
\Xi_{33} &= G_{33} + G_{33}^T - P_{33}^{(i)}, \\
\Xi_{15} &= G_{11}\bar{A}_1^{(i)} + \chi G_2 B_F \bar{C}^{(i)}, \\
\Xi_{25} &= G_2 G_{22}^{-1} G_{21}\bar{A}_1^{(i)} + G_2 B_F \bar{C}^{(i)}, \\
\Xi_{35} &= G_{31}\bar{A}_1^{(i)} + \phi G_2 B_F \bar{C}^{(i)}, \\
\Xi_{45} &= \bar{C}_1^{(i)} - D_F \bar{C}^{(i)}, \\
\Xi_{16} &= \chi G_2 A_F G_{22}^{-T} G_2^T, \\
\Xi_{26} &= G_2 A_F G_{22}^{-T} G_2^T, \\
\Xi_{36} &= \phi G_2 A_F G_{22}^{-T} G_2^T, \\
\Xi_{46} &= -C_F G_{22}^{-T} G_2^T, \\
\Xi_{17} &= G_{11}\bar{B}_1^{(i)}\bar{C}_W + \chi G_2 B_F \bar{D}^{(i)}\bar{C}_W + G_{13}\bar{A}_W, \\
\Xi_{27} &= G_2 G_{22}^{-1} G_{21}\bar{B}_1^{(i)}\bar{C}_W + G_2 B_F \bar{D}^{(i)}\bar{C}_W + G_2 G_{22}^{-1} G_{23}\bar{A}_W, \\
\Xi_{37} &= G_{31}\bar{B}_1^{(i)}\bar{C}_W + \phi G_2 B_F \bar{D}^{(i)}\bar{C}_W + G_{33}\bar{A}_W, \\
\Xi_{47} &= \bar{D}_1^{(i)}\bar{C}_W - D_F \bar{D}^{(i)}\bar{C}_W, \\
\Xi_{18} &= G_{11}\bar{B}_1^{(i)}\bar{D}_W + \chi G_2 B_F \bar{D}^{(i)}\bar{D}_W + G_{13}\bar{B}_W, \\
\Xi_{28} &= G_2 G_{22}^{-1} G_{21}\bar{B}_1^{(i)}\bar{D}_W + G_2 B_F \bar{D}^{(i)}\bar{D}_W + G_2 G_{22}^{-1} G_{23}\bar{B}_W, \\
\Xi_{38} &= G_{31}\bar{B}_1^{(i)}\bar{D}_W + \phi G_2 B_F \bar{D}^{(i)}\bar{D}_W + G_{33}\bar{B}_W, \\
\Xi_{48} &= \bar{D}_1^{(i)}\bar{D}_W - D_F \bar{D}^{(i)}\bar{D}_W.
\end{aligned}$$

Now, define new variables as follows:

$$\begin{aligned}
&M = G_2 G_{22}^{-1} G_2^T, \quad N = G_2 G_{22}^{-1} G_{23}, \quad S = G_2 G_{22}^{-1} G_{21} \\
&Q_A = G_2 A_F G_{22}^{-T} G_2^T, \quad Q_B = G_2 B_F, Q_C = C_F G_{22}^{-T} G_2^T, \quad Q_D = D_F.
\end{aligned} \tag{31}$$

We obtain the matrix constraints in problem (26). Furthermore, if (26) is feasible, it implies the positive definiteness of the $(2,2)$ block of (26), i.e.,

$$M^T + M > 0. \tag{32}$$

It follows that

$$G_2 G_{22}^{-1} G_2^T + G_2 G_{22}^{-T} G_2^T > 0. \tag{33}$$

Hence, both M and G_2 are invertible. It follows from (10) and (31) that the digital filter is given by

$$\begin{aligned} F(z) &= C_F (zI - A_F)^{-1} B_F + D_F \\ &= Q_C G_2^{-T} G_{22}^T \left(zI - G_2^{-1} Q_A G_{12}^{-T} G_2^T\right)^{-1} G_2^{-1} Q_B + Q_D \\ &= Q_C \left(zM^T - Q_A\right)^{-1} Q_B + Q_D \\ &= Q_C \left(zI - M^{-T} Q_A\right)^{-1} M^{-T} Q_B + Q_D \end{aligned} \tag{34}$$

With the defined change of variables, we obtain the synthesis condition given in (26) and the filter recovery procedure shown in (34). In order to confirm the correctness of the results, we shall further verify that the matrices $P^{(i)}$ and G in (28) can be recovered from any solution of problem (26). Specifically, we need to show that the matrices $P_{12}^{(i)}$, $P_{22}^{(i)}$, $P_{23}^{(i)}$, G_{12}, G_{21}, G_{22}, G_{23}, G_{32} can be recovered since the matrices $P_{11}^{(i)}$, $P_{13}^{(i)}$, $P_{33}^{(i)}$, G_{11}, G_{13}, G_{33} were obtained as part of the solution. For this purpose, we recall that

$$M = G_2 G_{22}^{-1} G_2^T, \quad N = G_2 G_{22}^{-1} G_{23}, \quad S = G_2 G_{22}^{-1} G_{21} \tag{35}$$

where M, N, S can be determined when (26) is feasible. With the solution and let $G_2 G_{22}^{-1}$ be a given nonsingular matrix X, we obtain G_2, G_{23}, and G_{21} via the following formulas:

$$G_2 = M^T X^{-T}, \quad G_{23} = X^{-1} N, \quad G_{21} = X^{-1} S.$$

Then, it is easily found that $G_{22} = X^{-1} G_2$. With a prior determined parameters χ and ϕ, we immediately obtain $G_{12} = \chi G_2$, $G_{32} = \phi G_2$. Next, we can obtain $P_{12}^{(i)}$, $P_{22}^{(i)}$, $P_{23}^{(i)}$ by reversing (29), i.e., $P_{12}^{(i)} = P_{g12}^{(i)} X^{-T}$, $P_{22}^{(i)} = X^{-1} P_{g22}^{(i)} X^{-T}$ and $P_{23}^{(i)} = X^{-1} P_{g23}^{(i)}$.

This completes the proof.

Remark 3.1. *Theorem 3.1 provides a new solvability condition for deriving robust matching filters for uncertain cascaded modulators, where the filter order can be determined beforehand. This overcomes the well-known limitation of the state-space H_∞ loop shaping method where the resulting filters are of the same order as the plant plus weighting functions.*

Remark 3.2. *From (26), one can seen that the scalars χ and ϕ are introduced. The role of the scalars in the condition (26) is to provide extra degrees of freedom. In most cases, the values of α and β in χ and ϕ, respectively, can be easily chosen, for instance, -1, 0, or 1. A systematic way for searching the values of α and β is to employ some numerical optimization searing algorithms, such as the program "fminsearch" in the optimization toolbox of MATLAB. When these parameters are to be fixed, (26) is linear in the variables for the solutions for Theorem 3.1 and can be solved by LMI toolbox [8].*

4. **Simulations.** Nonlinear simulations are carried out and validated with MATLAB/SIMULINK [8,9] for a cascaded 2-1 ΣΔ modulator with 1-bit quantizer. Specifically, the modulator of this experiment is aimed at applying to an audio system. The experimental parameters are set up as follows. The signal bandwidth (BW) is 25 KHz. A 8 KHz sinusoidal wave is used to perform a standard test. The oversampling ratio (OSR) is chosen to be 64. The sampling frequency f_s is 3.2MHz and the number of time points used for FFT is 16384.

In Section 4.1, we consider three weighting functions, each of which has order less than or equal to or greater than the plant order ($np = 2$). We will numerically verify that the resulting filters obtained by applying Theorem 3.1 and Remark 3.2 have order the same as that of the plant, independent of that of the introduced weighting functions. In Section 4.2, we compare the best filter obtained in Section 4.1 with some of the existing results.

4.1. **Filter design with weights of different order.** Our work is to minimize the effect of the leaky quantization noise $E_1(z)$ on the output $Y(z)$ in the signal band. To achieve the goal, it is important to design the digital filter such that the magnitudes of $(NTF_1(z) - F(z))$ is relatively small in the frequency range $[0, 25\text{K}]$ Hz, i.e., we want the shape of $(NTF_1(z) - F(z))$ be a high-pass one. To achieve this, weighting functions are introduced into the design and, theoretically, the low magnitude requirement on the low-pass band can be better achieved by increasing the order of the weights. As far as the noise effect beyond the signal band is concerned, it can be reduced by a subsequent decimation filter [1]. In discrete-time domain, the cut-off frequency of the desired $(NTF_1(z) - F(z))$ can be computed by the following formula [10,p.541]:

$$2\pi \frac{BW}{f_s} = 2\pi \frac{25 \times 10^3}{3.2 \times 10^6} = 0.0490625 \text{ (rad/s)} \tag{36}$$

Accordingly, three low-pass weighting functions (37), (38), and (39) with increasing order are considered.

$$W_1(z) = \frac{z^{-1}}{1 - 0.92z^{-1}} \tag{37}$$

$$W_2(z) = \frac{z^{-2}}{(1 - 0.92z^{-1})(1 - 0.96z^{-1})} \tag{38}$$

$$W_3(z) = \frac{z^{-2}}{(1 - 0.92z^{-1})(1 - 0.96z^{-1})(1 - 0.75z^{-1})} \tag{39}$$

Referring to Table 1, (15), (16) and (18), we can have the dilated weighting functions $W_{1a}(z)$ and $W_{3a}(z)$, i.e.,

$$W_{1a}(z) = W_1(z) \times z^{-1} \tag{40}$$

and

$$W_{3a}(z) = W_3(z) \times z^{-1}, \tag{41}$$

where the delay element is introduced in order to fulfill the premise of Theorem 3.1. As shown in Figure 2, the Bode plots for these weights (37), (38), (39), (40) and (41) are low-pass and have cut-off frequency around 0.0491 (rad/s). In particular, W_1 and W_{1a} (resp. W_3 and W_{3a}) have the same Bode plots for all frequencies. This fact indicates that they have the same effect upon shaping the noise transfer function; hence the design with W_1 and W_3 is equivalent to that using W_{1a} and W_{3a} in which Theorem 3.1 is applicable. Afterward, we suppose that the uncertainties in the gain and pole of the integrator $H_1(z)$ are within the ranges $0 \le \delta_a \le 0.01$, and $0 \le \delta_b \le 0.01$ [3,4]. By mapping the uncertain parameters δ_a and δ_b to (A_1, B_1, C_1, D_1) of (12), these uncertain matrices can be described by a four-vertex polytope, i.e., $m = 4$. By using Theorem 3.1 with $m = 4$ and the relevant information provided in Table 2, we computed the performance index γ and second-order filters via the proposed design. The resulting filters corresponding to the weights W_{1a}, W_2 and W_{3a} are

$$F_{w1}(z) = \frac{1.0054 - 1.9817z^{-1} + 0.9762z^{-2}}{1 + 0.0033z^{-1} - 0.0020z^{-2}}, \tag{42}$$

$$F_{w2}(z) = \frac{0.9083 - 1.7968z^{-1} + 0.8885z^{-2}}{1 - 0.1212z^{-1} - 0.0071z^{-2}}, \tag{43}$$

$$F_{w3}(z) = \frac{0.7428 - 1.4689z^{-1} + 0.7261z^{-2}}{1 - 0.2073z^{-1} - 0.0302z^{-2}}, \quad (44)$$

respectively.

In addition, we also compute a full-order H_∞ filter $F_t(z)$ [11] for problem (24) without using weighting functions.

$$F_t(z) = \frac{1 - 0.9952z^{-1} + 0.9943z^{-2}}{1 + 0.0002z^{-1} + 8.3905 \times 10^{-6}z^{-2}} \quad (45)$$

Figure 3 shows the noise transfer function $(NTF_1(z) - F(z) \times STF_2(z))$ versus frequency for an uncertain modulator matched by filters (42), (43), (44) and (45). It is known that the standard H_∞ filter design (without weights) has the disadvantage that it can not respond to lowering noise transfer function in magnitudes upon any particular band. When compared with other three filters, as expected, the filter $F_t(z)$ does give the worst shape (i.e., the highest magnitude shape) of noise transfer function $(NTF_1(z) - F(z) \times STF_2(z))$ in the signal band. Moreover, one can see that the design with $W_{3a}(z)$ yields filter $F_{w3}(z)$ which performs better than that with $W_{1a}(z)$ and $W_2(z)$

Table 2. Given data and outcome

Case	Parameters	Weighting function		Filter	Performance index	SNR (dB)
		Equation	Order			
A	$\alpha_1 = -0.80$, $\beta_1 = 0.01$	(40)	2	(42)	$\gamma_1 = 0.0505$	76.11
B	$\alpha_1 = -0.72$, $\beta_1 = 0.04$	(38)	2	(43)	$\gamma_2 = 1.2210$	80.59
C	$\alpha_1 = -0.72$, $\beta_1 = 0.05$, $\beta_2 = -0.04$	(41)	4	(44)	$\gamma_3 = 4.9178$	87.90

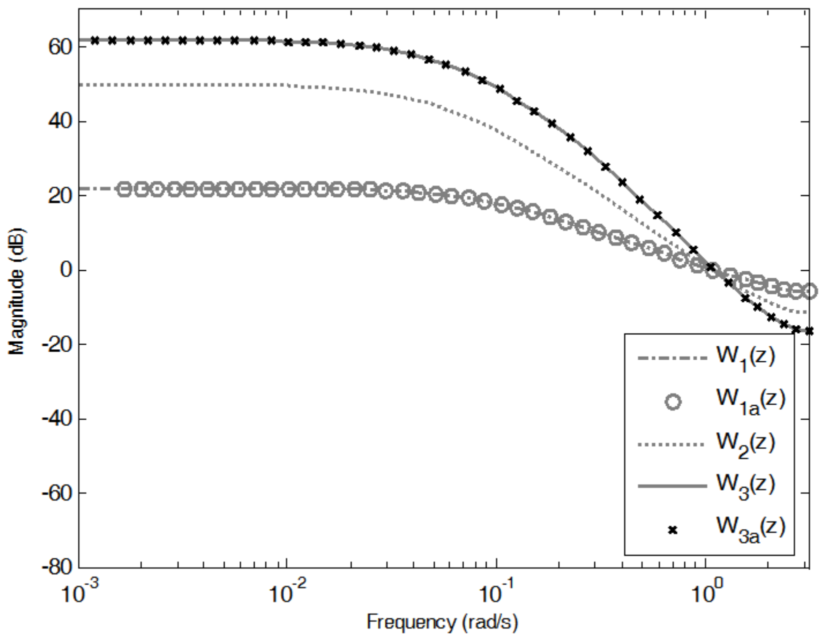

Figure 2. Bode plots of weighting functions

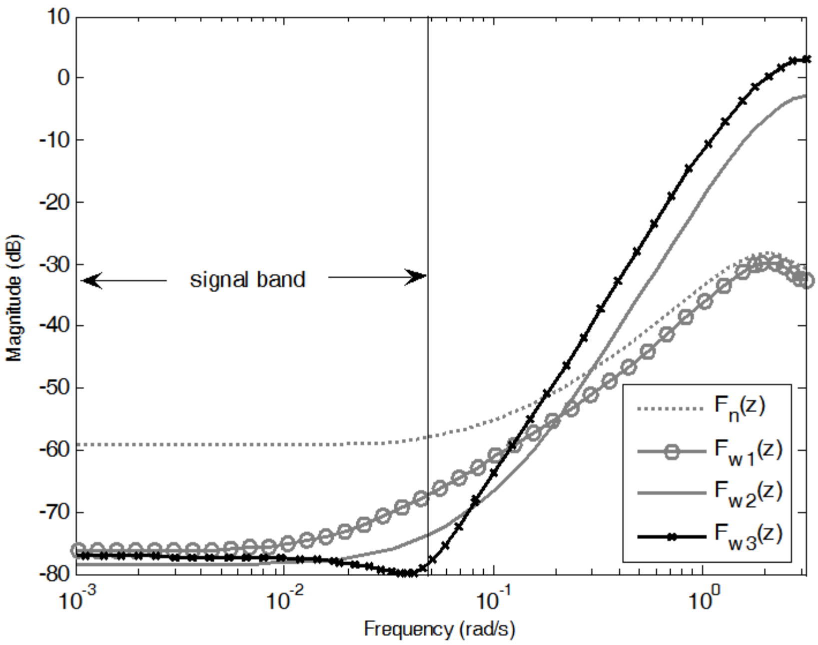

FIGURE 3. Bode plots of $(NTF_1(z) - F(z) \times STF_2(z))$ matched by filters; parameter deviations $\delta_a = \delta_b = 0.01$ in $H_i(z)$ $(i = 1, 2, 3)$

in terms of lower magnitudes for $(NTF_1(z) - F(z) \times STF_2(z))$ at low-frequencies, especially for frequency interval $[0.01, 0.0491]$ (rad/s). This implies that uncertain modulator matched by filter $F_{w3}(z)$, which was obtained by using the proposed design method with highest-order weight, can give the lowest noise power in the signal band. This feature can be checked in the zoom-in figure in Figure 4. More specifically, employing filter $F_{w3}(z)$ produces the best SNR value, 87.90 dB, when a -20 dBFS input signal is given.

From the above simulation results, we conclude that the proposed fixed-order filter design method can be used to design filters whose order is equal to the plant order, even if the order of the employed weighting functions is different. This implies the engineers may introduce high-order weighting functions in the computation for receiving fine performance, because the high-order weights do not increase the hardware implementation and have more chances to search better filters than low-order ones. Moreover, selection of weighting functions is crucial in the $\Sigma\Delta$ modulators filter design using H_∞ loop shaping method, we have applied Formula (36) [10] to choose a class of appropriate weighting functions for obtaining better noise shape. This was not found in the existing works.

4.2. **Comparison with existing methods.** We compare the performance of the proposed filter $F_{w3}(z)$ with the following filters:

Method A [1]: $F_n(z) = 1 - 2z^{-1} + z^{-2}$, i.e., (3);

Method B [3]: $F_c(z) = \dfrac{0.9963 - 1.967z^{-1} + 0.9709z^{-2}}{1 + 0.01815z^{-1} + 0.01194z^{-2}}$;

Method C [4]: $F_o(z) = 0.97465 - 1.9392z^{-1} + 0.9646z^{-2}$.

Similarly, the Bode plots of the $(NTF_1(z) - F(z) \times STF_2(z))$ versus frequency for the uncertain modulator matched by all filters are provided in Figure 5. We can see that $F_{w3}(z)$ can be used to perform better shape than other filters, especially for frequency interval $[0.02, 0.0491]$ (rad/s). With a -20 dBFS input signal, the computed SNR values for $F_n(z)$, $F_c(z)$, $F_o(z)$, and $F_{w3}(z)$ are 66.28, 80.04, 71.38, and 87.90 dB, respectively.

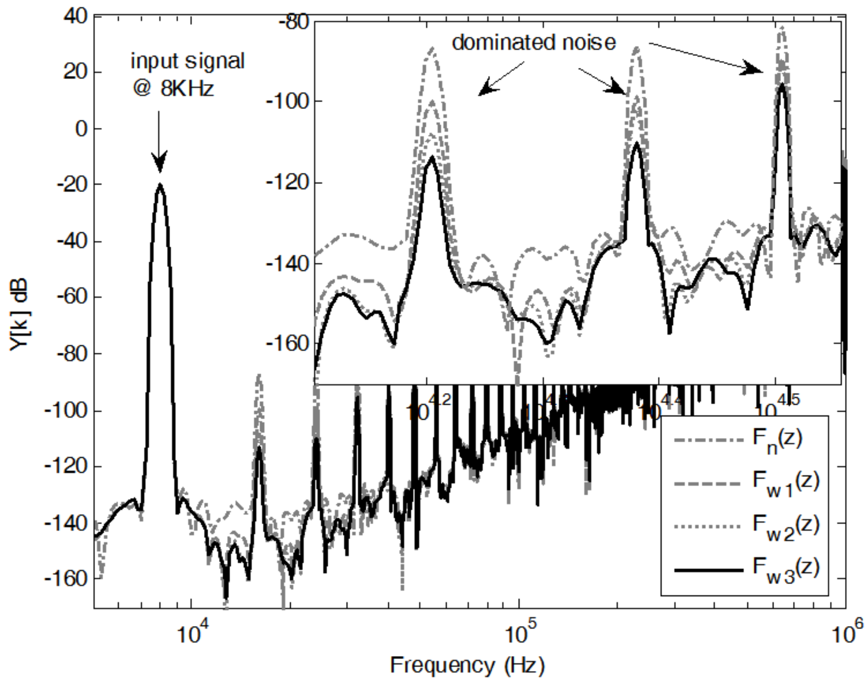

FIGURE 4. Power spectrum density of uncertain modulator matched by filters; parameter deviations $\delta_a = \delta_b = 0.01$ in $H_i(z)$ $(i = 1, 2, 3)$

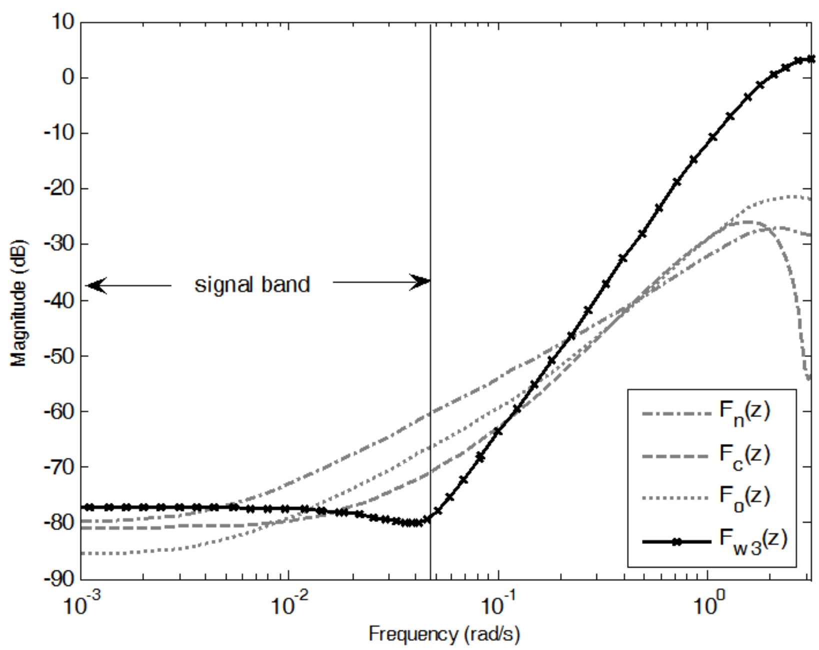

FIGURE 5. Bode plots of $(NTF_1(z) - F(z) \times STF_2(z))$ matched by filters; parameter deviations $\delta_a = \delta_b = 0.01$ in $H_i(z)$ $(i = 1, 2, 3)$

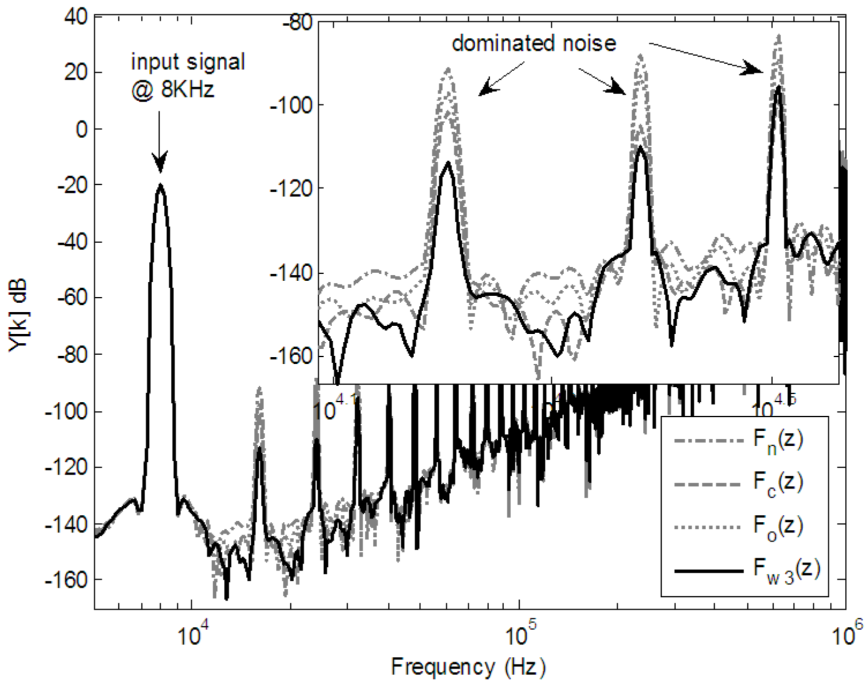

FIGURE 6. Power spectrum density of uncertain modulator matched by filters; parameter deviations $\delta_a = \delta_b = 0.01$ in $H_i(z)$ $(i = 1, 2, 3)$

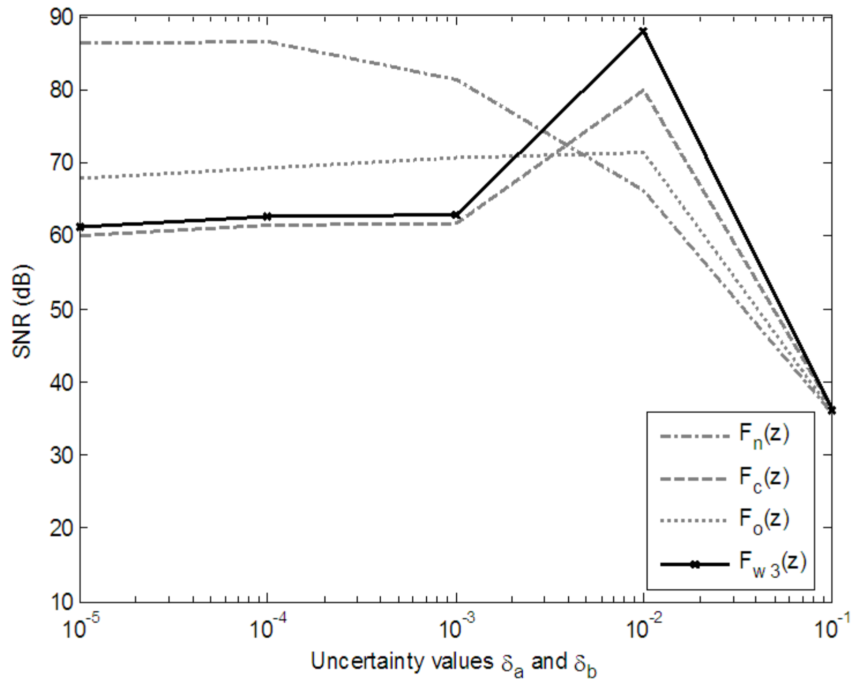

FIGURE 7. Influence of the uncertainty values δ_a, δ_b in $H_i(z)$ $(i = 1, 2, 3)$; uncertainty values are in the interval $[10^{-5}, 10^{-1}]$

Indeed, from a zoom-in figure in Figure 6, filter $F_{w3}(z)$ produced by the proposed method gives the lowest noise power. In Figure 7, the relevant SNR values are provided by selecting several values of δ_a and δ_b in H_i $(i = 1, 2, 3)$. As for $F_n(z)$, one can find that the SNR performance of the uncertain modulator degrades seriously when the values of δ_a and δ_b are large than 0.001. It means that the filter $F_n(z)$ is not suitable to be used for achieving a fine noise cancellation when the uncertainties in the gain and poles of the integrator do not lie in the ranges $0 \leq \delta_a \leq 0.001$, $0 \leq \delta_b \leq 0.001$. It is reasonable because filter $F_n(z)$ is designed by solving an ideal matching condition, i.e., (4). Besides, employing the proposed filter $F_{w3}(z)$ gives the best SNR values when the uncertainty values lie in the ranges $0.005 \leq \delta_a \leq 0.02$, $0.005 \leq \delta_b \leq 0.02$, especially for $\delta_a = \delta_b = 0.01$. Here, we conclude that the proposed method is very suitable for the cost effective production of fine performance $\Sigma\Delta$ modulators.

5. **Conclusion.** In this paper, we have studied the synthesis problem of robust matching filters for uncertain 2-1 cascaded $\Sigma\Delta$ modulators. A new design method which involves minimizing the worst case H_∞ norm of a certain weighted matching error over linearized polytopic model has been presented. In particular, the method overcomes a limitation of the well known H_∞ loop shaping techniques that they yield filters of high order (equal to the sum of the plant order and the order of the weighting function). This feature makes the proposed method very suitable for the design of low cost digital filters. Next, application of the synthesis methods to a cascaded 2-1 sigma-delta modulator with analog imperfections was conducted. Numerical experiments show that the proposed methods can improve the SNR performance for a range of larger parameter derivations. Therefore, it indicates the proposed method is applicable for a coarse cascaded $\Sigma\Delta$ modulator. Extension of the proposed approach to the other cascaded architecture is straightforward.

Acknowledgment. The authors would like to thank the reviewers for the valuable comments. The research was supported by National Science Council, Taiwan, under grant NSC NSC100-2221-E-032-009.

REFERENCES

[1] S. R. Norsworthy, R. Schreirer and G. C. Temes, *Delta-Sigma Data Converters: Theory, Design and Simulation*, IEEE Press, 1997.
[2] F. W. Yang and M. Gani, An H_∞ approach for robust calibration of cascaded sigma-delta modulators, *IEEE Trans. Circuits Systems I*, vol.55, no.2, pp.625-634, 2008.
[3] J. McKernan, M. Gani, D. Henrion and F. W. Yang, Robust filter design for uncertain 2-1 sigma-delta modulators via the central polynomial method, *IEEE Signal Processing Lett.*, vol.15, pp.737-740, 2008.
[4] J. McKernan, M. Gani, D. Henrion and F. W. Yang, Optimal low-frequency filter design for uncertain 2-1 sigma-delta modulators, *IEEE Signal Processing Lett.*, vol.16, pp.362-365, 2009.
[5] D. B. Ribner, A comparison of modulator networks for high-order oversampled $\Sigma\Delta$ analog-to-digital converters, *IEEE Trans. Circuits and Systems*, vol.38, no.2, pp.145-159, 1991.
[6] G. Fischer and A. J. Davis, Alternative topologies for sigma-delta modulators – A comparative study, *IEEE Trans. Circuits and Systems II*, vol.44, no.10, pp.789-797, 1997.
[7] K. Zhou and J. C. Doyle, *Essentials of Robust Control*, Prentice Hall, 1998.
[8] P. Gahinet, A. Nemirovski, A. J. Laub and M. Chilali, *Manual of LMI Control Toolbox*, Math Works, Inc., 1995.
[9] R. Schreier, *The Delta-Sigma Toolbox*, Version 7, http://www.mathworks.com/matlabcentral/fileexchange, 2004.
[10] D. Johns and K. Martin, *Analog Integrated Circuit Design*, Wiley, 1997.
[11] H. J. Gao, J. Lam, L. H. Xie and C. H. Wang, New approach to mixed H_2/H_∞ filtering for polytopic discrete-time systems, *IEEE Trans. Signal Processing*, vol.53, no.8, pp.3183-3192, 2005.