Helsinki University of Technology
Espoo, Finland, August 2000

Institute of Intelligent Power Electronics Publications
Publication 4

Round-off Error Free Fixed-Point Design of Polynomial FIR Predictors and Predictive FIR Differentiators

Jarno M. A. Tanskanen and Vassil S. Dimitrov

TEKNILLINEN KORKEAKOULU
TEKNISKA HÖGSKOLAN
HELSINKI UNIVERSITY OF TECHNOLOGY
TECHNISCHE UNIVERSITÄT HELSINKI
UNIVERSITE DE TECHNOLOGIE D'HELSINKI

Helsinki University of Technology
Espoo, Finland, August 2000

Institute of Intelligent Power Electronics Publications
Publication 4

Round-off Error Free Fixed-Point Design of Polynomial FIR Predictors and Predictive FIR Differentiators

Jarno M. A. Tanskanen[1] and Vassil S. Dimitrov[2]

[1] Institute of Intelligent Power Electronics
Department of Electrical and Communications Engineering
Helsinki University of Technology
P.O. Box 3000, FIN-02015 HUT, FINLAND
E-mail: jarno.tanskanen@hut.fi

[2] VLSI Research Group
Department of Electrical and Computer Engineering
University of Windsor
Ontario, N9B 3P4, CANADA

Institute of Intelligent Power Electronics
Espoo, Finland, August 2000

Distributor:
Helsinki University of Technology
Institute of Intelligent Power Electronics and Signal Processing Laboratory
P.O.Box 3000
FIN-02015 HUT
FINLAND
http://www.hut.fi/Units/PowerElectronics/ipe/
http://wooster.hut.fi/

Published only as an electronic publication, available from the Internet
URL: http://wooster.hut.fi/publications/tanskanen/round_off_error_free_predictors.pdf

ISBN 951-22-5140-X
ISSN 1456-0445

Jarno M. A. Tanskanen and Vassil S. Dimitrov

Round-off Error Free Fixed-Point Design of Polynomial FIR Predictors and Predictive FIR Differentiators

Espoo, Finland, August 2000
16 pages
Helsinki University of Technology Institute of Intelligent Power Electronics Publications
Publication 4
ISBN 951-22-5140-X
ISSN 1456-0445

Abstract:

In this report, we present a novel method for designing polynomial FIR predictors (PFP) and polynomial-predictive FIR differentiators (PPFD) for fixed-point environments. Our method yields filters that perform exact prediction and differentiation even with short coefficient word lengths. Under ordinary coefficient truncation or rounding, prediction and differentiation capabilities of these filters degrade considerably, or may even be totally lost. With the proposed method, the filters are designed so that *the prediction and differentiation properties are exactly preserved in fixed-point implementations*. The presented filter design method is based on integer programming (IP) and can be directly applied to fixed-point FIR design specifications which can be formulated in a form of linear constraints on the filter coefficients.

Keywords: error free coefficient quantization
Diophantine equation
FIR design
fixed-point filter design
integer programming
perfect quantization
polynomial differentiation
polynomial prediction

1. INTRODUCTION

By their nature, digital devices handle numbers using a finite number of bits per digit [7]. On the other hand, digital filters are typically designed using general-purpose computers. When the target application has the same computational precision as the filter design environment, there are usually no implementation problems if the filter itself was appropriately designed. Many times this is not the case, however, but the filters are operating within inexpensive, fixed-point processors, or in embedded applications using highly optimized, compact and less power consuming application specific integrated circuit (ASIC) designs. In these cases, there might be a great difference between the calculation precisions of the filter design environment and the final operation platform. This obviously results in filter quality degradation and possibly even in a totally unintended kind of filtering operation. In practice, even if the finite word length effects are paid adequate attention to, there has really been no other way to design these filters but to check if the filter with rounded or truncated coefficients is still within the design specifications [8]. In this paper, we present a novel method for designing polynomial FIR predictors (PFP) [4] and polynomial-predictive FIR differentiators (PPFD) [9] whose quantized coefficients exactly fulfill the set constraints for prediction, and prediction and differentiation, respectively, even with short word lengths, e.g., with six bits in some cases.

In many engineering disciplines, accurate control of physical processes is highly desirable. Many of the real world process parameters exhibit more or less smooth transitions. Noisy measurements of these parameters are typically used for process control after signal propagation, signal processing, and actuating delays. The research presented in this paper has spawned from the needs of application oriented research work; our examples of closed loop control include motion control of an elevator car [9], and mobile phone power control [3]. In the latter, the inherent closed loop control delays make it a lucrative environment to apply polynomial predictive techniques since the received power fluctuations can in many cases be modeled as Rayleigh distributed signals, which in turn can be accurately modeled as piece-wise low degree polynomials. Transmitter power control is regarded as one of the key issues in the third generation of mobile communications systems [3]. Besides, accurate control of an elevator car can effectively utilize, not only predicted position, but also predicted velocity and acceleration information. This information can be made available for the controller by a predictive differentiator. Here again, the position and velocity of the elevator car can be accurately modeled as piece-wise polynomials. Should these controllers be implemented in low-precision fixed-point environments, the properties of the actual quantized-coefficient filters are crucial. As the methods presented in this paper yield quantized-coefficient filters that exactly fulfill the given design constraints, these filters are naturally safe to use even in low-precision fixed-point environments. The filters designed by the proposed method to exactly fulfill the set constraints and to minimize the noise gain within a limited region of the quantized coefficient space, are here called ideally quantized coefficient filters. In this paper, two’s complement presentation is used for fixed-point presentation of filter coefficients. Magnitude truncation is applied as the conventional quantization method, and ‘infinite precision’ means the computational precision of Matlab, i.e., the long number format specified by the IEEE floating-point standard.

In Section 2, PFPs and PPFDs are shortly reviewed along with the constraints that are to be exactly fulfilled by the coefficients to provide for the desired filter properties. Integer programming interpretation of fixed-point PFP and PPFD design, and the proposed design algorithm are presented in Section 3. Characteristics of the conventionally quantized-coefficient and ideally quantized-coefficient PFPs and PPFDs are illustrated in Section 4, and Section 5 concludes the paper.

2. POLYNOMIAL FIR PREDICTORS AND POLYNOMIAL-PREDICTIVE FIR DIFFERENTIATORS

Polynomial predictive and differentiative filtering theory has been well established [4,5,9,10] but the applicability of both PFPs and PPFDs have suffered from the practical constraint of finite coefficient precision, which may cause severe degradation of filter characteristics. The fixed-point effects on the PPFDs characteristics have been found severe (cf. Figs. 4, 5, and 6) [8], and similar degradation is observed in fixed-point PFPs (cf. Figs. 2 and 3). By selecting a filter according to the responses with the infinite precision coefficients, it is most certain that the same filter with truncated or rounded coefficients will function inaccurately, in some cases the desired filtering properties are totally lost. To some extend these fixed-point effects can be avoided by carefully selecting the filters [8], but by employing the filter design method described in this paper the effects are eliminated altogether.

2.1. Polynomial FIR Predictors

PFPs, derived in [4], assume a low-degree polynomial input signal contaminated by white Gaussian noise. Filter output at a discrete time instant *n*, *x*(*n*), is defined to be a *p*-step-ahead predicted input,

$$x(n+p)=\sum_{k=1}^{N}h(k)x(n-k+1) \tag{1}$$

where *h*(*k*) are filter coefficients, *N* is filter length, and *p* is prediction step. After providing for exact prediction, the rest of the degrees of freedom are used to minimize the white noise gain,

$$NG=\sum_{k=1}^{N}\left|h(k)\right|^{2} . \tag{2}$$

In [5], a feedback extension to FIR predictors is given to provide considerable noise attenuation while maintaining the prediction property set forth by the underlying PFP (PPFD). For the feedback extension to function properly, it is necessary that the underlying PFP (PPFD) basis filters are implemented exactly. Until now, this has been rarely possible in short word length fixed-point environments.

A set of linear constraints can be derived from the definition of the filter output (1) [4]:

$$g_0=\sum_{k=1}^{N}h(k)-1=0\,, \tag{3}$$

$$g_1=\sum_{k=1}^{N}kh(k)=0\,, \tag{4}$$

$$g_2=\sum_{k=1}^{N}k^2h(k)=0\,, \tag{5}$$

$$\vdots$$

$$g_I=\sum_{k=1}^{N}k^Ih(k)=0\,. \tag{6}$$

The constraints (3)-(6) give the prediction of the polynomial degrees 0, …, *I*, and from them can closed form solutions for the FIR coefficients for low-degree polynomial input signals be solved by the method of Lagrange multipliers [1]. The closed form solutions for FIR coefficients for the first, second, and third degree polynomial input signals can be found in [4]. In this paper, we consider the case with the highest polynomial input signal component degree of two, *I* = 2, as an example. In this case, we have to fulfill the constraints (3), (4) and (5), and use the remaining degrees of freedom to minimize the noise gain (2). The exact, i.e., the infinite precision coefficients for the one-step-ahead, *p* = 1, second degree, *I* = 2, PFPs are given by [4]

$$h(k)=\frac{9N^2+(9-36k)N+30k^2-18k+6}{N^3-3N^2+2N},\ k\in[1,\,2,\,...,\,N]. \tag{7}$$

2.2. Polynomial-Predictive FIR Differentiators

PPFDs are derived in the similar way as the PFPs. For the PPFDs, the filter input-output relation is written as [9]

$$\dot{x}(n+p)=\sum_{k=1}^{N}h(k)x(n-k+1) \tag{8}$$

where the dot denotes time derivative, and the linear constraints on the filter coefficients are now given by [9]

$$g_0 = \sum_{k=1}^{N} h(k) = 0\,, \tag{9}$$

$$g_1 = \sum_{k=1}^{N} (N-k)h(k) = 1\,, \tag{10}$$

$$g_2 = \sum_{k=1}^{N} (N-k)^2 h(k) = 2(N-1+p)\,, \tag{11}$$

$$\vdots$$

$$g_I = \sum_{k=1}^{N} (N-k)^I h(k) = I(N-1+p)^{I-1}\,. \tag{12}$$

The constraints (9)-(12) give prediction and differentiation for the polynomial degrees 0, …, *I*, and the closed form solutions for the FIR coefficients for low-degree polynomial input signals are again obtained by the method of Lagrange multipliers [1]. The closed form solution for FIR coefficients for the second degree polynomial input signals is given in [9]. Again, we use the case with the highest polynomial input signal component degree of two, *I* = 2, as an example, and now have to fulfill the constraints (9), (10) and (11), and use the remaining degrees of freedom to minimize the noise gain (2). The exact, i.e., infinite precision, coefficients for the one-step-ahead *p* = 1, second degree, *I* = 2, PPFDs are given by [9]

$$h(k) = \frac{6\left[(30N+30)(k-1)^2 + (-32N^2+38)(k-1) + 6N^3 - 11N^2 - 9N + 14\right]}{(N-2)(N-1)N(N+1)(N+2)},\ k \in [1,\ 2,\ ...,\ N]\,. \tag{13}$$

It is worth noting that the coefficients for the filter length *N* = 3 are still exact if quantized to six bits or more for both PFPs (7) and PPFDs (13) with *I* = 2 and *p* = 1, but the noise gains (2) of these filters are unpractically high; 19 and 24.5, respectively. However, the filters of length *N* = 3 are good basis filters for the feedback extension [5,9] which relieves the noise gain problem. Otherwise, longer filters are to be used for achieving acceptable noise gains, and the method described in this paper is to be used to obtain correctly functioning fixed-point coefficient filters.

3. LINEAR DIOPHANTINE EQUATION BASED PFP AND PPFD DESING

3.1. Linear Diophantine Equation Formulation of the Filter Design Problem

The optimization problem that has to be solved can be reformulated as an integer programming (IP) problem. Suppose that all coefficients *h*(*k*) of the filter are multiplied by 2^n where *n* is the number of bits available, and truncated to yield the integer coefficients $h^*(k)$. Then the PFP design task can be defined as an algorithm with the following input and output:

Input: Function $\mathrm{F}(h^*(1), h^*(2), \cdots, h^*(N)) = \sum_{k=1}^{N} h^{*2}(k)$, with integer variables $h^*(k)$, that is to be minimized while having the constraints

$$g_0 = \sum_{k=1}^{N} h^*(k) - 2^n = 0\,, \tag{14}$$

$$g_1 = \sum_{k=1}^{N} kh^*(k) = 0\,, \tag{15}$$

$$g_2 = \sum_{k=1}^{N} k^2 h^*(k) = 0\,, \tag{16}$$

$$\vdots$$

and

$$g_I = \sum_{k=1}^{N} k^I h^*(k) = 0\,, \tag{17}$$

on the variables. The constraints (14)-(17) correspond to the constraints (3)–(6) with both sides multiplied by 2^n and with integers variables.

Output: An integer vector $\mathbf{h}^* = [h^*(1), h^*(2), \ldots, h^*(N)]$ that minimizes F(·) and satisfies *exactly* the constrains (14)-(17) above.

For designing PPFDs, the integer input constraints corresponding to the constraints (9)-(12) are given by

$$g_0 = \sum_{k=1}^{N} h^*(k) = 0\,, \tag{18}$$

$$g_1 = \sum_{k=1}^{N} (N-k)h^*(k) = 2^n\,, \tag{19}$$

$$g_2 = \sum_{k=1}^{N} (N-k)^2 h^*(k) = 2^{n+1}(N-1+p)\,, \tag{20}$$

$$\vdots$$

and

$$g_I = \sum_{k=1}^{N} (N-k)^I h^*(k) = 2^n I (N-1+p)^{I-1}\,. \tag{21}$$

The output of the algorithm is again an integer coefficient vector $\mathbf{h}^*$ that now exactly fulfills the constraints (18)-(21) and thereafter minimizes the cost function F(·).

The solution we offer is based on the following considerations:

1. As the filter coefficients are to be presented with short word-length fixed-point numbers, the task in hand is a quadratic *integer programming* problem, which is well-known to be an NP-complete problem; therefore it is unrealistic by any means to find the best solution in a reasonable amount of time, especially for long filters. Designing these filters with floating-point coefficients would present us with a quadratic *real programming* problem, which is solvable in polynomial time [6].
2. Without restricting the variables to be integers, we have closed form solutions of the problem, which are given for PFPs and PPFDs, for the case I = 2 and p = 1, by (7) and (13), respectively. Although the values computed by these formulas are not integers, these expressions are here used as initial approximations.
3. To make sure that the conditions (14)-(17), or (18)-(21), are met exactly, one has to solve a desired system above in integers. This problem has been a subject of deep investigations in number theory and the theory of Diophantine equations. By variable elimination, the problem can be presented as a single linear equation of the form

$$A_1x_1 + A_2x_2 + \cdots + A_lx_l = B \tag{22}$$

where A_1, A_2, …, A_l, , B, and x_1, x_2, …, x_l are integers. Equations of the form (22) with given integers A_i and B, and with unknown integers x_i, i = 1, …, l, are called Diophantine equations. For example, for the PPFD with I = 1, one could do the following elimination of variables. Solving for $h^*(1)$ in (18) yields

$$g_0 = \sum_{k=1}^{N} h^*(k) = 0 \Rightarrow h^*(1) = -\sum_{k=2}^{N} h^*(k)\,. \tag{23}$$

Taking $h^*(1)$ out of the summation in (19) gives

$$g_1 = \sum_{k=1}^{N} (N-k)h^*(k) = 2^n \Rightarrow \tag{24}$$

$$(N-1)h^*(1) + \sum_{k=2}^{N} (N-k)h^*(k) = 2^n \tag{25}$$

and substituting $h^*(1)$ (23) into (25) yields a Diophantine equation of the form (22):

$$\sum_{k=2}^{N}(1-k)h^*(k) = 2^n\,. \tag{26}$$

A solution $[h^*(2), \ldots, h^*(N)]$ of (26) is then substituted back into (23) to yield the filter coefficient vector $\mathbf{h}^*$ that exactly satisfies the conditions (18) and (19). Similar variable elimination can be applied for the cases when $I > 1$.

3.2. Solving the Diophantine Equations by a Search Algorithm

Solutions of (22) are usually obtained by multidimensional continued fraction algorithms. The approach we use here is based on Clausen-Fortenbacher algorithm [2] which in our case is an exhaustive search within a limited region around an initial guess. The reasons why we chose this particular technique are: First, with a 166 MHz Pentium PC programmed with C, the algorithm succeeds in matter of seconds for $N = 8$ and in less than one hour with $N = 16$, to find the solutions of (22), amongst whom the optimal one, i.e., the one that minimizes the noise gain (2), or, the function F(·), can be found in a fraction of a second, c.f. Table 3 for the numbers of solutions whose noise gains are to be calculated and compared. Secondly, the program provided in [2] can be easily generalized to more than 16 variables (the largest case analyzed by Clausen and Fortenbacher). Thirdly, we have an initial approximation for the algorithm. Without any initial approximations, the whole quantized coefficient space would need to be searched, which would be prohibited by the required computation time (of the order of 10^{14} hours for 8-bit precision, 10^{33} hours for 16-bit coefficients). Since the infinite precision solution is known, it is intuitive to search for the quantized solutions within a vicinity of it, though it there is no reason why the very best quantized coefficient solution should lie close to it. Exhaustive search for ideal quantization is performed within a band of ±2 from the coefficients $h(k)$, given by (7) or (13), presented in integer form with a given number of bits, i.e, the search space for each coefficient consists of the four integers closest to $2^n h(k)$. The search band width ±2 is selected ad hoc to give the search more degrees of freedom than the minimum search band of ±1 while still being computationally feasible. The found quantized coefficients that exactly fulfill the constraints (14)-(17), or (18)-(21), and minimize the noise gain (2) within the search band, are called ideally quantized coefficients. The search band with the conventionally and ideally quantized coefficients of the PPFD with $I = 2$, $p = 1$, and $N = 16$ is illustrated in Fig. 1 for the coefficient precision of 8 bits.

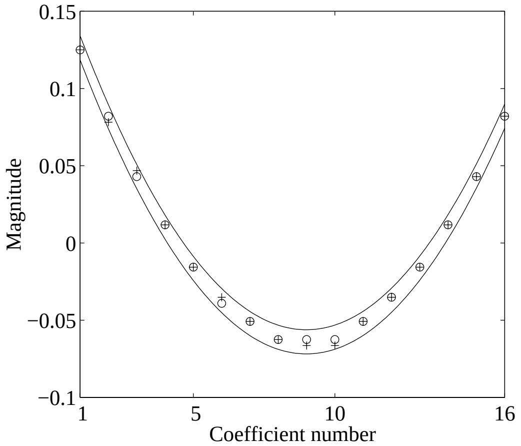

Fig. 1. Ideal quantization search band (between solid lines) for the $I = 2$, $p = 1$, $N = 16$, PPFD with the coefficient precisions of 8 bits. Circles 'o' denote the magnitude truncated, and plusses '+' the ideally quantized coefficients.

In Tables 1 and 2, the results of ideal quantization are shown for the $I = 2$, $p = 1$, PPFD of length $N = 16$ with coefficient precisions of 8, and 16 bits, respectively, along with the real number form (infinite precision rounded to six decimals) solutions of (13). The ideally quantized coefficients shown in Tables 1 and 2 satisfy the constraints (18)-(20) exactly and minimize the noise gain (2) within the search band. It is worth noting that simple truncation or rounding of the infinite precision coefficients never but once in our experiments ($N = 3$ is an exception) produced a solution of the system of the Diophantine equations (18)-(20). This demonstrates the necessity of special techniques aimed at solving the integer optimization problem. Table 1 shows that with the PPFD length $N = 16$ and coefficient precision of 8 bits, for five out of sixteen coefficients one has to approximate the real coefficient with an integer that is not the rounded infinite precision coefficient given by (13). In this case, also six out of sixteen ideally quantized coefficients are not equal to the corresponding truncated infinite precision coefficients, and two ideally quantized coefficients lie on the search band boundary, i.e., are not the closest integers on either side of the real number form coefficients. The 16-bit filter coefficients for the same filter are shown in Table 2, and they show similar results are seen in Table 1. In Tables 1 and 2, the ideally quantized coefficients that differ from the rounded or truncated real number form coefficients are marked. As the search yields several coefficient vectors $\mathbf{h}^*$ that exactly fulfill the constraints (18)-(20), the one that minimizes the noise gain (2) is shown in Tables 1 and 2.

Table 1. The infinite precision presentations of the $I = 2$, $p = 1$, $N = 16$, PPFD coefficients computed by (13) and rounded to six decimals (real number form), and the corresponding ideally quantized coefficients with the coefficient precision of 8 bits.

Coefficient	Real number form	Ideally quantized form	Coefficient	Real number form	Ideally quantized form
256 $h(0)$	32.313725	32	256 $h(8)$	-16.376471	-17[*†]
256 $h(1)$	20.894118	20[*]	256 $h(9)$	-15.605602	-17[**†]
256 $h(2)$	10.998319	12[**†]	256 $h(10)$	-13.310924	-13
256 $h(3)$	2.626331	3[†]	256 $h(11)$	-9.492437	-9
256 $h(4)$	-4.221849	-4	256 $h(12)$	-4.150140	-4
256 $h(5)$	-9.546218	-9[*]	256 $h(13)$	2.715966	3[†]
256 $h(6)$	-13.346779	-13	256 $h(14)$	11.105882	11
256 $h(7)$	-15.623529	-16[†]	256 $h(15)$	21.019608	21

[*] The ideally quantized coefficient is not the rounded infinite precision coefficient.

[**] The ideally quantized coefficient is not an integer on either side of the infinite precision coefficient.

[†] The ideally quantized coefficient is not the truncated infinite precision coefficient.

Table 2. The infinite precision presentations of the $I = 2$, $p = 1$, $N = 16$, PPFD coefficients computed by (13) and rounded to six decimals (real number form), and the corresponding ideally quantized coefficients with the coefficient precision of 16 bits.

Coefficient	Real number form	Ideally quantized form	Coefficient	Real number form	Ideally quantized form
65536 $h(0)$	8272.313725	8272	65536 $h(8)$	-4192.376471	-4193[*†]
65536 $h(1)$	5348.894118	5348[*]	65536 $h(9)$	-3995.034174	-3997[**†]
65536 $h(2)$	2815.569418	2816[†]	65536 $h(10)$	-3407.596639	-3408[†]
65536 $h(3)$	672.340616	673[*†]	65536 $h(11)$	-2430.063866	-2430
65536 $h(4)$	-1080.793277	-1080[*]	65536 $h(12)$	-1062.435854	-1062
65536 $h(5)$	-2443.831933	-2444[†]	65536 $h(13)$	695.287395	696[*†]
65536 $h(6)$	-3416.775350	-3416[*]	65536 $h(14)$	2843.105882	2843
65536 $h(7)$	-3999.623529	-3999[*]	65536 $h(15)$	5381.019608	5381

[*] The ideally quantized coefficient is not the rounded infinite precision coefficient.

[**] The ideally quantized coefficient is not an integer on either side of the infinite precision coefficient.

[†] The ideally quantized coefficient is not the truncated infinite precision coefficient.

Table 3 lists the numbers of quantized coefficient solutions that exactly satisfy the $I = 2$, $p = 1$, PFP or PPFD coefficient constraints (14)-(16) or (18)-(20), respectively, for coefficient precisions 6, 8, 10, 12, 14, and 16 bits for the filter lengths $N = 8$ and $N = 16$. From Table 3 it is seen that there are several quantized coefficient combinations within the search band that exactly satisfy the constraints (14)-(16) for PFP, or (18)-(20) for PPFD, respectively, and that there thus are some degrees of freedom left for noise gain (2) minimization. To find the optimum solution, it is necessary to search all of the solutions in the search band and to select the one which minimizes the noise gain (2). For the filter length $N = 8$, there are altogether $N_{tot} = 4^8 = 65\,536$ candidate quantized coefficient vectors $\mathbf{h}^*$ within the search band. For this case the search and selecting the solution with the smallest noise gain takes less than one second on a 166 MHz Pentium processor programmed with C language, while for the filter length $N = 16$, $N_{tot} = 4^{16} = 4\,294\,967\,296$, with the total algorithm run time of 47 minutes. For many applications, also the first-found solution could be adequate taken that the search is organized to start from the candidates which have the smallest noise gains, i.e., whose positive coefficients are on the lower search band boundary and negative coefficients on the upper search band boundary, cf. Fig. 1. The noise gain (2) of the first found solution should be checked against the design specifications.

Table 3. The number of ideally quantized solutions that exactly satisfy the $I = 2$, $p = 1$, PFP constraints (14)-(16), or PPFD constraints (18)-(20), for the filter lengths $N = 8$ and $N = 16$ with coefficient precisions of 6, 8, 10, 12, 14 and 16 bits. N_{tot} is the total number of candidate quantized coefficient vectors within the search band.

Coefficient precision (bits)	6	8	10	12	14	16
PFP, $N = 8$, $N_{tot} = 4^8 = 65\,536$	15	14	15	15	14	15
PFP, $N = 16$, $N_{tot} = 4^{16} = 4\,294\,967\,296$	55086	54760	49164	54394	54760	49164
PPFD, $N = 8$, $N_{tot} = 4^8 = 65\,536$	21	14	14	21	14	14
PPFD, $N = 16$, $N_{tot} = 4^{16} = 4\,294\,967\,296$	56326	53633	58791	55027	58287	57341

4. CHARACTERISTICS OF THE QUANTIZED AND IDEALLY QUANTIZED COEFFICIENT PSPS AND PPFDS

In this section, frequency response and group delay properties of the infinite precision, quantized-coefficient, and ideally quantized-coefficient PFPs and PPFDs are illustrated. The quantization effects on the filter responses can be seen in Figs. 2, 4 and 5. From these Figs. it is clearly seen that as the coefficients are quantized, the prediction an/or differentiation properties are lost or at least degraded from their exact desired values. As seen comparing Figs. 4 and 5, differentiation property is generally more robust to coefficient quantization than prediction property which can be lost already with the coefficient word length of 16 bits. The one-step-ahead prediction property is identified as the negative unity group delay in Figs. 2b, 4b, and 5b. Differentiation of an input signal consisting of polynomial signal components of 0th, 1st and 2nd degree is set forth by the zero magnitude response at zero frequency along with the ramp-shaped frequency response within a desired differentiation band as explicitly stated by the constraints (9)-(11), Figs. 4a and 5a. As the ideal quantization yields several filters that exactly satisfy the PFP constraints (3)-(5), or (9)-(11) for the PPFD, the ones that minimize the noise gain (2) are shown in Figs. 2, 4 and 5. In all these figures, the curves for the infinite precision and ideally quantized coefficient filters are hardly recognizable since they are exactly on the top of each other at zero frequency, as they should since ideal quantization yields exactly the properties of the infinite precision filters at zero frequency. Also the responses are seen to be close to each other at higher frequencies but it is to be remembered that it is the behavior at and near zero frequency that actually defines the predictive and/or differentiative properties, the rest is only additional spectral shaping.

In Fig. 2, the frequency response and group delay of the $I = 2$, $p = 1$, $N = 8$, PFP are shown with the coefficients quantized both conventionally and ideally to 8 bits, along with the infinite precision coefficient filter. As seen in Fig. 2b, the $I = 2$, $p = 1$, $N = 8$, PFP with conventionally quantized 8-bit coefficients does not provide for exact prediction whereas the predictor with the ideally quantized coefficients does, as it should, since it satisfies the constraints (3)-(5) exactly. For a polynomial predictor, it is crucial that the dc-gain is exactly unity, Fig. 2a, since polynomial signal prediction by its nature operates on the signal amplitude while the noise suppression operates in the frequency domain; the same applies to the zero dc-gain of the polynomial-predictive differentiators. In Fig. 2a, the conventionally quantized coefficient PFP is seen to have a bias problem whereas the ideally quantized coefficient filter shows exact unity dc-gain. Filter degradation effects are seen in Fig. 3 in which a polynomial signal is presented in time domain. In Fig. 3, it is seen that as the one-step-delayed polynomial signal is fed into a truncated coefficient PFP, the filter output is useless, while the ideally quantized PFP is able to recover (predict) the desired signal. With 16 bit coefficients, the filters of length $N = 8$ and $N = 16$ behave very much like their infinite precision counterparts and for many applications it is sufficient to use the truncated coefficients calculated with (7). It is to be noted that generally the deviation from the exact prediction due to coefficient quantization gets larger as the filter length increases and for practical applications longer filters are necessary to provide for lower noise gains.

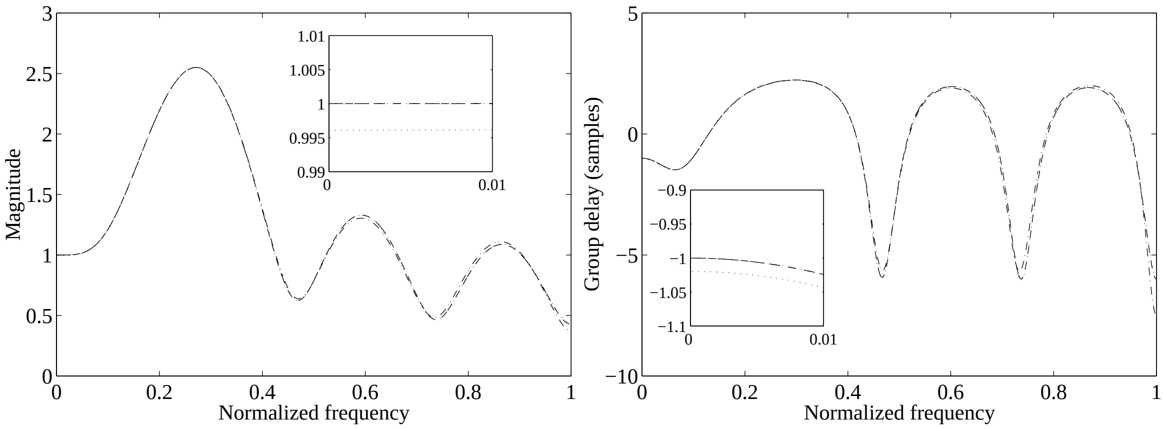

Fig. 2. Magnitude responses (a) and group delays (b) of the infinite precision (dashed), conventionally quantized (dotted), and optimally quantized (dash-dot) coefficient second-degree one-step-ahead PFP of length $N = 8$ with the coefficient precision of 8 bits.

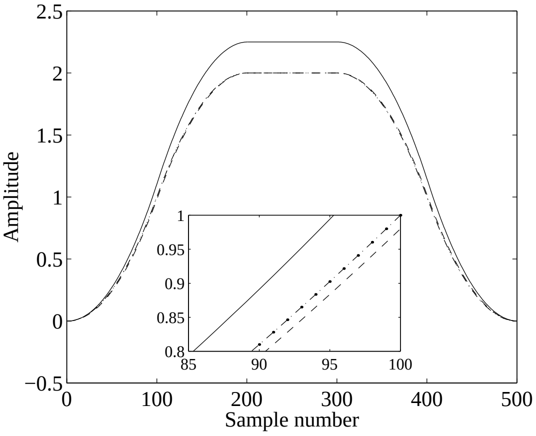

Fig. 3. Time domain example of prediction of a one-step-delayed polynomial signal with second and zeroth degree polynomial sections; signal, i.e., the desired filter output (dash-dot), delayed signal, i.e., the filter input (dashed), prediction with the PFP of length $N = 8$ with truncated 8-bit coefficients (solid), and prediction with the corresponding ideally quantized 8-bit coefficient filter (dotted).

In Figs. 4 and 5, the frequency response and group delay of the $I = 2$, $p = 1$, $N = 16$, PPFD are shown with both conventional and ideal coefficient quantization to 8 and 16 bits, respectively. Also the infinite precision coefficient filter is shown in the Figs. The filter shown in Fig. 4 corresponds to the coefficients in Table 1, and that in Fig. 5 to the coefficients in Table 2. It is seen that the prediction property is totally lost in conventional coefficient quantization with both precisions, Figs. 4b and 5b, while the ideally quantized coefficient PPFDs provide for exact prediction. With 8-bit coefficients, the differentiation property is degraded, Fig. 4a, while with 16-bit conventional coefficient quantization, the differentiation property is practically undisturbed, Fig. 5a. The ideal quantization is seen to yield perfect differentiation with both precisions, as expected. Also generally, the PPFD differentiation property is more robust to the coefficient quantization than the prediction property. In Fig. 6, the time domain polynomial signal in Fig. 3 is fed into both truncated coefficient and ideally quantized coefficient PPFDs with sampling frequency of 100 Hz. From Fig. 6 it is seen that the output of the truncated coefficient PPFD is useless while the ideal quantization yields predicted differentiation that very closely follows the desired filter output. It is also seen in Fig. 6 that after an abrupt change in the derivative, the filter needs $N = 16$ samples to find out the new derivative, as natural.

The noise gains of the ideally quantized filters that minimize the noise gain (2) are shown in Table 4, for the PFPs and PPFDs of length $N = 16$, with coefficients quantized to 8, 10, 12, 14 and 16 bits, along with the noise gains of their infinite precision counterparts. The noise gains of the $I = 2$ PFP and PPFD of length $N = 8$ are greater than unity and thus they are not very practical unless a feedback extension [10] is applied. From Table 4 it can be seen that as the coefficient precision is increased, the noise gain approaches that of the corresponding infinite precision filter, and that the loss in noise gain is not substantial with any ideally quantized coefficient precision.

Table 4. Noise gains of the infinite precision and ideally quantized coefficient PFPs and PPFDs with I = 2, p = 1, and N = 16 with coefficient precisions of 8, 10, 12, 14 and 16 bits.

Coefficient precision (bits)	8	10	12	14	16	Inf.
Noise gain, PFP	0.732421875	0.730468750	0.730363846	0.730359077	0.730357163	0.730357143
Noise gain, PPFD	0.053619385	0.053543091	0.053536773	0.053536430	0.053536416	0.053536415

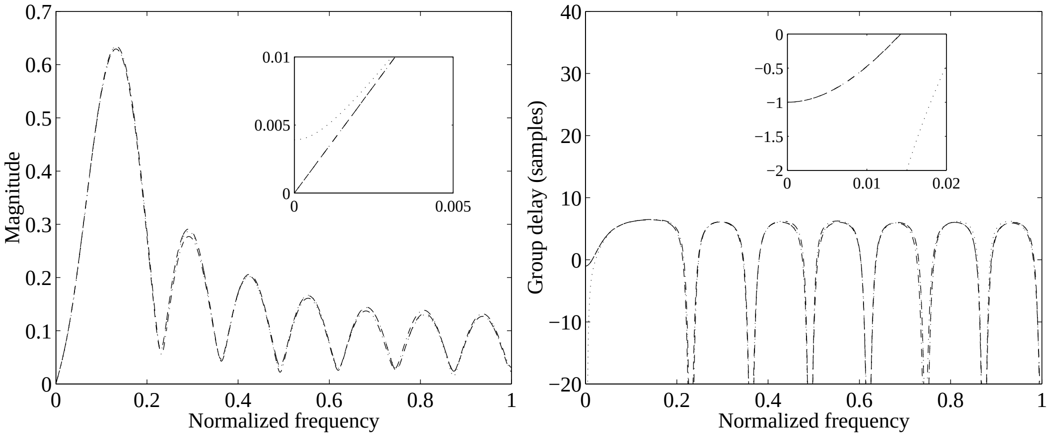

Fig. 4. Magnitude responses (a) and group delays (b) of the infinite precision (dashed), conventionally quantized (dotted), and optimally quantized (dash-dot) coefficient second-degree one-step-ahead PPFD of length N = 16 with the coefficient precision of 8 bits.

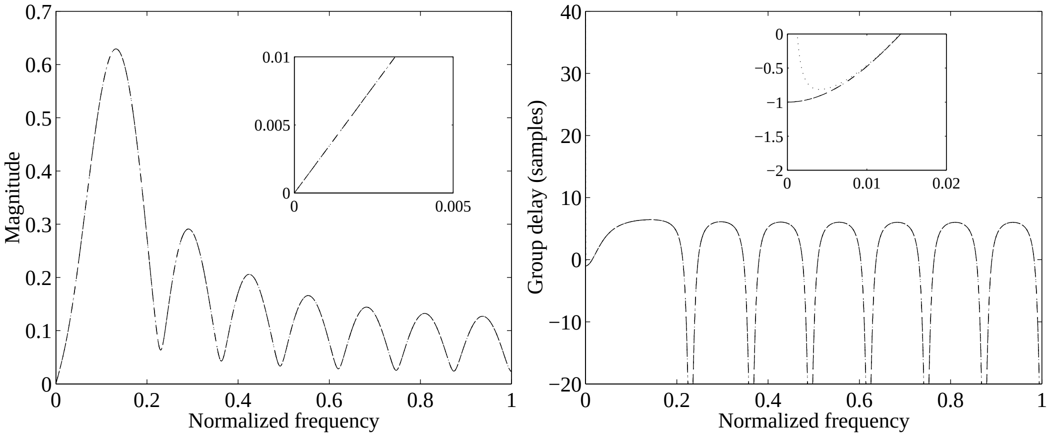

Fig. 5. Magnitude responses (a) and group delays (b) of the infinite precision (dashed), conventionally quantized (dotted), and optimally quantized (dash-dot) coefficient second-degree one-step-ahead PPFD of length N = 16 with the coefficient precision of 16 bits.

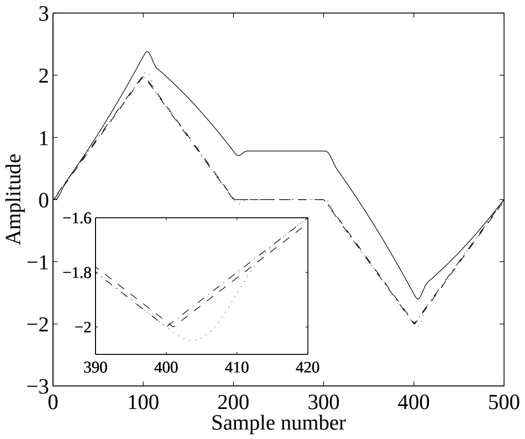

Fig. 6. Differentiation of the one-step-delayed polynomial signal in Fig. 3; result of exact differentiation, i.e., the desired filter output (dash-dot), delayed differentiation, i.e., the filter input (dashed), differentiation by the PPFD of length $N = 16$ with truncated 8-bit coefficients (solid), and differentiation by the ideally quantized 8-bit coefficient PPFD (dotted).

5. CONCLUSIONS

A new technique for perfect digital polynomial FIR predictor and polynomial-predictive FIR differentiator coefficient quantization has been proposed. Our method uses an exhaustive search. For designing longer filters efficient number-theoretic tools would be needed for solving the Diophantine equation associated with the filter design problem. As it is demonstrated in the paper, the given filter design constraints giving the filters their polynomial signal prediction and/or differentiation properties, can be exactly satisfied with low-precision fixed-point coefficients, and thus, the influence of round-off errors is eliminated. For the second degree one-step-ahead polynomial FIR predictors and polynomial-predictive FIR differentiators used as examples in this paper, the conditions can be exactly satisfied with even as low as 6-bit coefficient precision, with still some degrees of freedom available to minimize the noise gain of the designed fixed-point coefficient filter. The proposed integer programming based search method for fixed-point filter design may be applied also to other filter design tasks in which the design criteria can be formulated in a form of linear constraints on the filter coefficients.

REFERENCES

[1] Bertsekas, D. *Constrained Optimization and Lagrange Multipliers Methods*. Academic Press, New York, NY, 1982.

[2] Clausen, M., and Fortenbacher, A. Efficient solution of linear Diophantine equations. *Journal of Symbolic Computation* **8** (1989), 201–216.

[3] Harju, P. T., Laakso, T. I., and Ovaska, S. J. Applying IIR predictors on Rayleigh fading signal. *Signal Processing* **48** (1996) 91–96.

[4] Heinonen, P., and Neuvo, Y. FIR-median hybrid filters with predictive FIR substructures. *IEEE Trans. Acoustics, Speech, and Signal Processing* **36** (1988) 892–899.

[5] Ovaska, S. J., Vainio, O., and Laakso, T. I. Design of predictive IIR filters via feedback extension of FIR forward predictors. *IEEE Trans. Instrumentation and Measurement* **46** (1997) 1196–1201.

[6] Papadimitriou, C. R., and Steiglitz, K. Combinatorial Optimization: Algorithms and Complexity. *Prentice Hall*, Englewood Cliffs, NJ, 1982.

[7] Proakis, J. G., and Manolakis, D. G. *Digital Signal Processing: Principles, Algorithms, and Applications*, Macmillan Publishing Company, New York, NY, 1992.

[8] Tanskanen, J. M. A., and Ovaska, S. J. Coefficient sensitivity of polynomial-predictive FIR differentiators: Analysis. in *Proc. 42nd IEEE Midwest Symposium on Circuits and Systems, Las Cruces, NM*, Aug. 1999, to appear.

[9] Väliviita, S., and Ovaska, S. J. Delayless recursive differentiator with efficient noise attenuation for control instrumentation. *Signal Processing* **69** (1998) 267–280.

[10] Väliviita, S., Ovaska, S. J., and Vainio, O. Polynomial predictive filtering in control instrumentation: a review. *IEEE Trans. Industrial Electronics* **46** (1999) 876–888.

ACKNOWLEDGMENT

The work of Jarno M. A. Tanskanen has been partially funded by the Technology Development Centre of Finland, Nokia Corporation, Sonera Ltd., Finland, and Omnitele Ltd., Finland, and supported by Jenny ja Antti Wihuri Foundation, Finland, Walter Ahlström Foundation, Finland, The Finnish Society of Electronics Engineers, and by Foundation of Technology, Finland.

The work of Dr. Vassil S. Dimitrov has been funded by the Technology Development Centre of Finland, Nokia Corporation, Sonera Ltd., Finland, and Omnitele Ltd., Finland.

THE AUTHORS

JARNO M. A. TANSKANEN was born in Lahti, Finland in 1968. He received his M.Sc. in technical physics, and Licentiate of Technology (E.E.) degrees from Helsinki University of Technology (HUT), Finland in 1995, and 1998, respectively. 1994–1999 he was a researcher in Signal Processing and Computer Technology Laboratory of HUT. Currently he is with Institute of Intelligent Power Electronics, HUT, where he is engaged in digital predictor research, and with Signal Processing Laboratory, HUT, where he takes part in mobile CDMA power control research.

VASSIL S. DIMITROV was born in Plovdiv, Bulgaria in 1964. He received his Ph.D. degree in mathematics in 1995 from the Mathematical Institute of the Bulgarian Academy of Sciences, Sofia, Bulgaria. Since then he has spent two years (Jan. 1996–Dec. 1997) as a Postdoctoral fellow at the VLSI Research Group, Department of Electrical and Computer Engineering, University of Windsor, Canada, one year (Jan. 1998–Feb. 1999) as a research scientist at Reliable Software Technologies Corporation, Virginia, USA, and one year (March 1999–June 2000) as a chief research scientist at the Signal Processing and Computer Technology Laboratory, Helsinki University of Technology, Finland. Since July 2000 Dr. Dimitrov is holding the position of Associate Professor at the Department of Electrical and Computer Engineering in the University of Windsor. His main research interests are in the area of number theoretic algorithms, computational complexity theory, cryptography and information security, global optimization, fast algorithms for digital signal processing and related topics. Dr. Dimitrov is a member of the New York Academy of Sciences.

Helsinki University of Technology Institute of Intelligent Power Electronics Publications

1. Sami Väliviita,
   Predictive Filtering Methods for Motor Drive Instrumentation, 1998

2. Xia-Zhi Gao,
   Soft Computing Methods for Control and Instrumentation, 1999

3. Jorma Kyyrä (editor)
   Cost Effective Reducing of Line-Current Distrotion in Rectifiers, 2000

4. Jarno M. A. Tanskanen and Vassil S. Dimitrov,
   Round-off Error Free Fixed-Point Design of Polynomial FIR Predictors and Predictive FIR Differentiators, 2000

ISBN 951-22-5140-X
ISSN 1456-0445