CONFIDENTIAL. Limited circulation. For review only.

High Resolution Sigma-Delta Modulator implementations via Suboptimal Second Order Sliding Mode Control

Alessandro Pilloni, Mauro Franceschelli, Alessandro Pisano and Elio Usai

***Abstract*— Two novel Sigma-Delta modulator algorithms based on second order sliding mode concepts for high resolution analog-to-digital conversion purposes are proposed. Let f_s be the finite over-sampling frequency of the modulator, novel schemes that convert to a 1-bit signal, resp., the first, or the second time-derivative of the signal to be modulated are suggested to reduce the error of the conversion process, at the demodulation side, to an $\mathcal{O}(1/f_s^2)$ boundary layer, instead of $\mathcal{O}(\sqrt{1/f_s})$ of the traditional scheme. The algorithms are discussed and analyzed in the discrete-time setting, and a comparison analysis illustrate the effectiveness of these solutions.**

I. Introduction

The modern data-driven world demands efficient ways of interacting with diverse physical environments in the so-called Internet-of-Things (IoT) by means of smart sensors, actuators, signal conditioning and analog-to-digital interfaces [1]. Since the performance of digital signal processing is limited by the precision of the digital input signal, in recent years, data compression and conversion tools have become of great interest to researchers. Among different analog-to-digital conversion (ADC) techniques, such as Pipeline, Successive Approximation Register (SAR) and Sigma-Delta ("ΣΔ", or Delta-Sigma, "ΔΣ"), the latter covers most of the market's segments up to 40MHz applications, as shown in Figure 1. Table I further shows pros/cons of these ADC techniques, as well as their operating ranges [2].

ΣΔ techniques were introduced in the 40s as a simplified Pulse Code Modulation (PCM). However, their effectiveness were realized with the advancement of integrated circuit technologies [3]. In particular, their low cost, simplicity, and very high precision due to noise-shaping properties make them the best choice in mixed-signal applications to convert either analog or digital signals to single-bit signals, thus not only facilitating the integration on cheap hardware with limited resources, but also reducing the cost of cables and interfaces [4]. ΣΔ are constantly and extensively studied in several fields, e.g., in signal processing [5], communications [6], networked control systems [4], and in power applications to drive switching actuators and dc-dc converters [7].

The research leading to these results has received the finantial support form the Italian Ministry of Research and Education (MIUR), project "CoNetDomeSys", from code RBSI14OF6H under call SIR 2014, and from "CONACYT" under Grant n. 165958.

A. Pilloni, M. Franceschelli, A. Pisano, and E. Usai are with the Department of Electrical and Electronic Engineering (DIEE), University of Cagliari, Cagliari 09123, Italy.

E-mail addresses: {alessandro.pilloni, mauro.franceschelli, pisano, eusai}@diee.unica.it

TABLE I

Operating Ranges and Features of ADC Techniques [1].

ADC Topology	Sample Rate	Resolution (bit)	Main Features Pros (+) / Cons (−)
SAR	$\leq$ 1.25Msps $\leq$ 4Msps	$\leq$ 18-bit $\leq$ 16-bit	(+) low latency, low power, good accuracy. (−) cost, low speed.
ΔΣ	$\leq$ 4ksps $\leq$ 4Msps $\leq$ 10Msps	$\leq$ 31-bit $\leq$ 24-bit $\leq$ 16-bit	(+) resolution, compression, moderate cost, low power. (−) delay, latency.
Pipeline	$\leq$ 200Msps $\leq$ 250Msps $\leq$ 550Msps	$\leq$ 16-bit $\leq$ 14-bit $\leq$ 12-bit	(+) speed and bandwidth. (−) cost, resolution, delay, latency, power consumption.

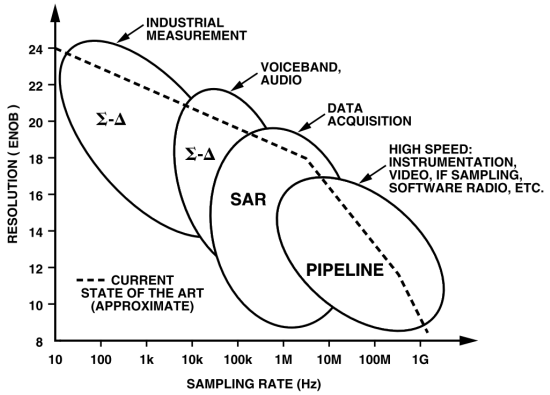


Fig. 1. ADC architectures, applications, resolution, and sampling rates [2].

From a control perspective a ΣΔ modulator (ΣΔ-M), whose traditional implementation in the discrete-time setting is given in Figure 2, can be understood as a Sliding Mode-based Control (SMC) systems, in which the sliding manifold "s" is the argument of the 1-bit quantizer element, i.e., 2 quantized levels. Connections between SMCs and ΣΔ-Ms were firstly discussed by Sira-Ramírez in [8]. There, performance of Δ-, ΣΔ-, and Exponential ΣΔ-Ms, was analyzed under the infinite sampling frequency assumption. Conditions for the existence of Quasi-SMC (Q-SMC) regime in discrete-time Δ controlled, linear, stable system were outlined in [9]. A step beyond is the work in [10], there it was shown that ZOH discretization of the equivalent control of a SMC system gives rise to a Δ-modulation in the Utkin sense. More recently in [11], taking the advantages of the results in [12] for the switched implementation of average designed feedback laws, conditions for the existence of the equivalent mode, i.e. where the modulator's outputs are ideally equivalent to the inputs [13], are derived by using concepts of Q-SMCs [14]. It

Preprint submitted to 15th International Workshop on Variable Structure Systems and Sliding Mode Control. Received February 8, 2018.

is worth to note that, all the reviewed works, share the same first-order SMC implementation given in Figure 2. Thus, they provide the same conversion's error, that lies within an $\mathcal{O}(1/f_s)$ boundary layer of the sliding manifold, whose value coincides with chattering amplitude due to the finite switching frequency "f_s" or real SMCs [13].

Although this, we point out that the 1-bit conversion process is not only affected by errors due to the chattering phenomenon, but it is also affected by errors introduced at the demodulation site. In particular, since a low-pass filter is required to extract the "equivalent information" of the switching signal (see e.g. Figure 2), then errors proportional to $\mathcal{O}(\sqrt{f_s})$ need to be also taken into account [15].

Main result: In this paper we propose two novel discrete-time ΣΔ-Ms based on the Suboptimal Second-order SMC (2-SMC) [15] that allow to reduce the conversion error, due to the finite switching frequency, to a boundary layer of the sliding surface that is an order of $\mathcal{O}(1/f_s^2)$ (instead of $\mathcal{O}(1/f_s)$ of the current state of the art). Furthermore, the proposed methods completely remove the error due to low pass filtering at the demodulation side, that is proportional of an $\mathcal{O}(1/f_s)$. In particular, by adopting schemes to convert to a 1-bit signal, resp., the first, or the second time-derivative of the signal to be converted, at the demodulation side the original signal can be obtained by a cascade of integrations processes, that do not introduce errors. Follows that the error at the demodulation side remain an $\mathcal{O}(1/f_s^2) \ll \mathcal{O}(\sqrt{1/f_s})$.

For the sake of completeness, we remark that, although the proposed scheme considerably increase the resolution of the conversion process, higher gains in the ΣΔ converters should be used, to track the derivatives of the signal. However, since we are considering software-based discrete-time implementations, this is not a practical problem. On the contrary, since the magnitude of the boundary layer depends on the magnitude of the discontinuous gains, a compromise between the bandwidth of the signal to be converted and precision to be achieved is necessary.

Paper organization: In Section II preliminaries on the Suboptimal 2-SMC algorithm are given in the discrete-time setting. In Section III a novel performance analysis of the traditional ΣΔ-M implementation in the discrete-time setting is given. In Section IV the proposed high resolution ΣΔ implementations are illustrated and their performance discussed. In Section V some numerical simulations and a comparison analysis with the traditional ΣΔ scheme are discussed. Finally, Section VI provides concluding remarks and outline next directions for future research.

II. Preliminaries on Suboptimal 2-SMC

Consider the nonlinear second-order dynamical system

$$\begin{array}{rcl} \dot{\boldsymbol{s}}(\mathrm{t}) & = & \boldsymbol{\omega}(\mathrm{t}) \\ \dot{\boldsymbol{\omega}}(\mathrm{t}) & = & \boldsymbol{\phi}(\boldsymbol{s},\boldsymbol{\omega},\mathrm{t}) + \boldsymbol{\gamma}(\boldsymbol{s},\boldsymbol{\omega},\mathrm{t}) \cdot \boldsymbol{u}(\mathrm{t}) \end{array} \tag{1}$$

where $\boldsymbol{\phi} : \mathbb{R}^2 \times \mathbb{R} \mapsto \mathbb{R}$ and $\boldsymbol{\gamma} : \mathbb{R}^2 \times \mathbb{R} \mapsto \mathbb{R}$ are uncertain functions satisfying standard boundedness conditions, i.e., $\|\boldsymbol{\phi}(\boldsymbol{s},\boldsymbol{\omega},\mathrm{t})\|_\infty \leq \boldsymbol{\Phi} \ \forall \ \boldsymbol{s},\boldsymbol{\omega},\mathrm{t} \in \mathbb{R}$ and $0 < \boldsymbol{\Gamma}_{\min} \leq \boldsymbol{\gamma}(\cdot,\mathrm{t}) \leq \boldsymbol{\Gamma}_{\max}$, and $\boldsymbol{u} : \mathbb{R}^+ \to \mathbb{R}$ is the control input. The "Suboptimal" 2-SMC is a well known algorithm within the robust control community because it solves the stabilization problem for (1) in finite-time, by exploiting information only on the sliding variable $\boldsymbol{s}$, see [16], [17] for details. Its continuous-time representation is

$$\boldsymbol{u}(\mathrm{t}) = -\boldsymbol{\mathcal{U}} \cdot \mathrm{sign}(\boldsymbol{s}(\mathrm{t}) - \boldsymbol{\beta} \cdot \boldsymbol{s}_{\mathrm{sp}}), \ \boldsymbol{\mathcal{U}} > \frac{\boldsymbol{\Phi}}{\boldsymbol{\Gamma}_{\min}}, \ \boldsymbol{\beta} \in [0,1) \tag{2}$$

where $\boldsymbol{\beta}$ is the "anticipation factor", usually chosen such that $\boldsymbol{\beta} = 1/2$, and $\boldsymbol{s}_{\mathrm{sp}}$ is the most recent "Singular Point" of the sliding variable $\boldsymbol{s}(\mathrm{t})$ (i.e., the value of $\boldsymbol{s}$, at the most recent time $\mathrm{t}_{\mathrm{sp}} : \dot{\boldsymbol{s}}(\mathrm{t}_{\mathrm{sp}}) = 0$). An estimation of $\boldsymbol{s}_{\mathrm{sp}}$ can be achieved by simply inspecting the sign of the first backward difference approximation of $\dot{\boldsymbol{s}}$, as follows

$$\mathrm{sign}(\dot{\boldsymbol{s}}(\mathrm{t})) \approx \mathrm{sign}(\boldsymbol{s}[k] - \boldsymbol{s}[k-1]), \tag{3}$$

where $T_s = 1/f_s > 0$ is the measurement interval, and $\boldsymbol{s}[k] = \boldsymbol{s}(kT_s) = \boldsymbol{s}(\mathrm{t})$, $\mathrm{t} \in [kT_s, (k+1)T_s)$. From [18], it follows that (2) with the approximation (3) steers the variable $\boldsymbol{s}$ to zero in finite-time with an $\mathcal{O}(T_s^2)$ accuracy. Discrete-time versions of algorithm (2) are proposed in [15], [19], and [20], resp., for control and derivative estimation purposes. Let $\boldsymbol{\gamma}(\boldsymbol{s},\boldsymbol{w},\mathrm{t}) = 1$, problem (1)-(2) can be directly discretized by using the first forward differences as follows

$$\boldsymbol{s}[k+1] = \boldsymbol{s}[k] + T_s \cdot \boldsymbol{\omega}[k] \tag{4}$$

$$\boldsymbol{\omega}[k+1] = \boldsymbol{\omega}[k] + T_s \cdot (\boldsymbol{\phi}[k] - \boldsymbol{\mathcal{U}} \cdot \mathrm{sign}(\boldsymbol{s}[k] - \boldsymbol{\beta}\boldsymbol{s}_{\mathrm{sp}})). \tag{5}$$

From a similar analysis to that of [20, Section 3], it yields that the discrete-time trajectories of system (4)-(5) will establish a Q-SMC regime such that they provides the fulfilment of $|\boldsymbol{s}[k]| \leq \varepsilon_1 \cdot \mathcal{O}(T_s^2)$, $|\boldsymbol{\omega}[k]| \leq \varepsilon_2 \cdot \mathcal{O}(T_s)$ in finite time, if $\boldsymbol{\mathcal{U}}$ is chosen to dominate $\boldsymbol{\phi}(\cdot,\mathrm{t})$, i.e., $\boldsymbol{\mathcal{U}} = \boldsymbol{\Phi}$, and the estimation of $\boldsymbol{s}_{\mathrm{sp}}$ are evaluated according to Algorithm 1.

Algorithm 1 Approximate digital peak-detector [15].

Initialization:
$\boldsymbol{s}[-1] = \boldsymbol{s}_{\mathrm{sp}}[-1] = \boldsymbol{s}(0), \ \boldsymbol{s}[-2] = 0, \ i = -1, \ k = 0$
$\boldsymbol{\Lambda}[k] = (\boldsymbol{s}[k] - \boldsymbol{s}[k-1]) \cdot (\boldsymbol{s}[k-1] - \boldsymbol{s}[k-2])$
Repeat:
if $\boldsymbol{\Lambda}[k] \leq 0$ **then** $\left\{ \begin{array}{rcl} \mathrm{i} & = & \mathrm{i}+1 \\ k_{\mathrm{sp_i}} & = & k-1 \\ \boldsymbol{s}_{\mathrm{sp}} & = & \boldsymbol{s}[k_{\mathrm{sp_i}}] \end{array} \right.$

III. Performance analysis of ΣΔ Modulator

A block diagram of a first-order ΣΔ-M, followed by a low-pass filter and a digital decimator for demodulation purposes is shown in Figure 2. From a control perspective the ΣΔ-M can be viewed as a feedback-based tracking control system where $\boldsymbol{x}[k] = \boldsymbol{x}[kT_s]$, $\mathrm{t} \in [kT_s, (k+1)T_s)$ is the output of a ZOH filter which takes as input the reference to be tracked $\boldsymbol{x}(\mathrm{t})$, T_s is the sampling period, $k \in \mathbb{Z}$, whereas $\boldsymbol{y}_{\mathrm{d}}[k]$ is the 1-bit modulator's digital output. Notice that the signal that is quantized is not the reference $\boldsymbol{x}[k]$ but the integral of the error between the $\boldsymbol{x}[k]$ and an analog representation, $\boldsymbol{y}[k]$, of the quantized output $\boldsymbol{y}_{\mathrm{d}}[k]$. If the DAC is ideal, it can replaced by a unity gain. Then, let the 1-bit quantizer be

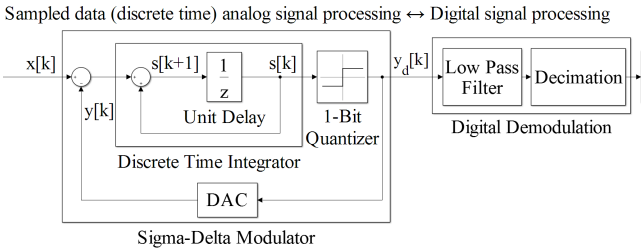


Fig. 2. Block of a first order ΣΔ modulator A/D system [3].

bipolar, and let $\mathcal{K}$ be a finite gain, the 2-level quantized output satisfies $\boldsymbol{y}_{\rm d}[k] \in \{-\mathcal{K}, +\mathcal{K}\}$. The resulting closed loop dynamic for the ΣΔ-M in Figure 2 is

$$\boldsymbol{s}[k+1] = \boldsymbol{s}[k] + T_s \cdot (\boldsymbol{x}[k] - \mathcal{K} \cdot \mathrm{sign}(\boldsymbol{s}[k])) \tag{6}$$

where $\boldsymbol{s}[k]$ is the internal state of the discrete integrator. In the remainder, the following assumption on $\boldsymbol{x}$ is in force.

Assumption 1: *Let $\boldsymbol{x}(\mathrm{t})$ be the signal to be quantized, $\boldsymbol{X}(f)$ be its spectrum, and let $\boldsymbol{x}[k] = \boldsymbol{x}(kT_s)$ be the sampled representation of $\boldsymbol{x}(t)$. We assume that there exist two known positive constants $\mathcal{X}$ and $\mathcal{B}_{\boldsymbol{x}}$, such that $\boldsymbol{x}$ is bounded in magnitude $\boldsymbol{x}[k] \leq \mathcal{X}$, and spectrum $\boldsymbol{X}(f) = 0 \ \forall \ f > \mathcal{B}_{\boldsymbol{x}}$.* ■

Theorem 1: *Consider the ΣΔ modulator of Figure 2. Given a bounded signal $\boldsymbol{x}[k]$, satisfying Assumption 1 and let $\mathcal{K} > \mathcal{X}$. Then, the closed loop switched system* (6)*, from any arbitrary initial value $\boldsymbol{s}[0]$, exhibits a quasi-sliding motion and the switching manifold $\boldsymbol{s}[k]$ converges to an $\mathcal{O}(T_s)$-boundary layer after a finite number of steps.*

Proof of Theorem 1: Let us consider the following positive definite, discrete-time Lyapunov function

$$V(\boldsymbol{s}) = (\boldsymbol{s}[k])^2. \tag{7}$$

According to the Lyapunov stability paradigm for discrete-time systems, we are going to study under which conditions $V(\boldsymbol{s})$ has contraction along the trajectories of system (6). Let us first compute

$$\begin{aligned} \Delta V(\boldsymbol{s}) &= (\boldsymbol{s}[k+1])^2 - (\boldsymbol{s}[k])^2 \\ &= T_s^2 \cdot (\boldsymbol{x}[k] - \mathcal{K} \cdot \mathrm{sign}(\boldsymbol{s}[k]))^2 \\ &\quad + 2T_s \cdot (\boldsymbol{x}[k] - \mathcal{K} \cdot \mathrm{sign}(\boldsymbol{s}[k])) \cdot \boldsymbol{s}[k] \end{aligned} \tag{8}$$

by solving $\Delta V(\boldsymbol{s}) < 0$ w.r.t. $\boldsymbol{s}[k]$, and reminding that $\mathcal{K}$ dominates the time-varying reference signal $\boldsymbol{x}[k]$, it follows that the $V(\boldsymbol{s})$ is contractive outside the invariant set

$$\Theta = \left\{ \boldsymbol{s}[k] \ : \ |\boldsymbol{s}[k]| \leq \frac{(\mathcal{K} + \boldsymbol{X})^2}{2(\mathcal{K} - \boldsymbol{X})} \cdot T_s \right\}. \tag{9}$$

We point out that inside the invariant set Θ, it yields $\boldsymbol{s}[k] \leq \varepsilon_3 \cdot \mathcal{O}(T_s)$, $\varepsilon_3 \in \mathbb{R}^+$. Thus, the trajectories of $\boldsymbol{s}[k]$ may deviate from the manifold $\boldsymbol{s}[k] = 0$ with an $\mathcal{O}(T_s)$ error. Notice that, as expected, (6) produces the "zig-zag motion" typical of sampled-data Q-SMCs, due to the chattering. ■

Let us known focus our attention at the demodulation side of ΣΔ-M in Figure 2. It can be noted that, in order to reconstruct $\boldsymbol{x}[k]$, before the digital decimation process, a low pass filtering process is done, with the aim to extract the slows averaged components, $\boldsymbol{y}_{\rm av}[k]$ associated with the discontinuous signal $\boldsymbol{y}_{\rm d}[k]$. Assumed the dynamics of the low-pass filter be as follows

$$\boldsymbol{y}_{\rm av}[k+1] = e^{-\frac{T_s}{\tau}} \cdot \boldsymbol{y}_{\rm av}[k] + \left(1 - e^{-\frac{T_s}{\tau}}\right) \cdot \boldsymbol{y}_{\rm d}[k] \tag{10}$$

$$\boldsymbol{y}_{\rm av}[0] = \boldsymbol{y}_{\rm d}[0] = \mathcal{K} \cdot \mathrm{sign}(\boldsymbol{s}[0]), \quad k = 0, 1, 2, \ldots \tag{11}$$

the following theorem quantify the accuracy to be expected at the demodulation side of the ΣΔ conversion process.

Theorem 2: *Let us assume the hypotheses of Theorem 1. Let* (10)-(11) *be equations describing the digital low-pass filter at the demodulation side of the ΣΔ modulator in Figure 2. Once system* (6) *is in quasi-sliding mode regime, then signal $\boldsymbol{x}[k]$ can reconstructed with a $\mathcal{O}(\sqrt{T_s})$ precision.*

Proof of Theorem 2: Let us first note that, from a SMC's perspective, the low-pass process at the demodulation side of the ΣΔ system in Figure 2 corresponds to the online evaluation of the so-called "equivalent control" [13], i.e., that "ideal" continuous signal $\boldsymbol{y}_{\rm eq}$ that, if replaced to the discontinues term and under the infinite switching frequency assumption, can establish the following condition

$$[\boldsymbol{s}(\mathrm{t}) = \dot{\boldsymbol{s}}(\mathrm{t}) = 0] \equiv [\boldsymbol{s}[k+1] = \boldsymbol{s}[k] = 0].$$

According to [13], $\boldsymbol{y}_{\rm eq}$ is the solution of the algebraic equation $\boldsymbol{s}[k+1] - \boldsymbol{s}[k] = 0$ w.r.t. the discontinues term. Here, from (6), it results that

$$\boldsymbol{y}_{\rm eq} : T_s \cdot \left(\boldsymbol{x} - \mathcal{K} \cdot \boldsymbol{y}_{\rm eq}\right) = 0 \Rightarrow \mathcal{K} \cdot \boldsymbol{y}_{\rm eq} \equiv \boldsymbol{x}. \tag{12}$$

In the literature, the problem of quantify the accuracy of the on-line computation of the equivalent control when the system motion is confined within a boundary layer of the sliding manifold $\boldsymbol{s} = 0$ (as in (9)), was firstly studied in the continuous-time setting by Utkin in [13, Chapter 2.4]. There, provided that $\boldsymbol{s}$ and $\dot{\boldsymbol{s}}$ remains within a $\mathcal{O}(T_s)$ vicinity of zero of magnitude $\boldsymbol{D} = \varepsilon_3 \cdot \mathcal{O}(T_s)$, and that a certain small known constant $\boldsymbol{M}$ is available such that, in a neighborhood of the sliding manifold $\left|\boldsymbol{y}_{\rm eq} - \dot{\boldsymbol{s}}/\mathcal{K}\right| \leq \boldsymbol{M}$. Let $\boldsymbol{y}_{\rm av}$ be state of a continuous-time first-order filter

$$\tau \cdot \dot{\boldsymbol{y}}_{\rm av}(\mathrm{t}) + \boldsymbol{y}_{\rm av}(\mathrm{t}) = \mathcal{K} \cdot \mathrm{sign}(\boldsymbol{s}(\mathrm{t})) \tag{13}$$

and by neglecting exponentially vanishing terms, it results that (see [13, Chapter 2.4] for details):

$$\left|\boldsymbol{y}_{\rm av}(\mathrm{t}) - \boldsymbol{y}_{\rm eq}(\mathrm{t})\right| \leq \boldsymbol{M} \cdot (\tau + \boldsymbol{D}) + \frac{3\boldsymbol{M}\boldsymbol{D}}{\tau}. \tag{14}$$

Noticed that (14) can be minimized if the time constant τ of (13) is coordinated with the width of the area $\boldsymbol{D}$ where Quasi-Sliding Mode takes place. By solving $(\mathrm{d}/\mathrm{d}\tau)(\boldsymbol{M} \cdot (\tau + \boldsymbol{D}) + 3\boldsymbol{M}\boldsymbol{D}/\tau) = 0$, it yields $\tau = \sqrt{3\boldsymbol{M}\boldsymbol{D}}$, which leads to

$$\left|\boldsymbol{y}_{\rm av}(\mathrm{t}) - \boldsymbol{y}_{\rm eq}(\mathrm{t})\right| \leq \varepsilon_4 \cdot \sqrt{T_s} + \varepsilon_5 \cdot T_s \leq \varepsilon_6 \cdot \sqrt{T_s} \tag{15}$$

where ε_4, ε_5 and ε_6 are proper positive constants, and T_s is sufficiently small such that the high frequency components of $\boldsymbol{y}_{\rm d}[k]$ do not overlap with the spectrum of $\boldsymbol{x}(\mathrm{t})$ in (12).

Let us now note that, since the input to the filter (13) is piecewise constant between subsequent samplings, thus it can be discretized, without any error as in (10)-(11).

Preprint submitted to 15th International Workshop on Variable Structure Systems and Sliding Mode Control. Received February 8, 2018.

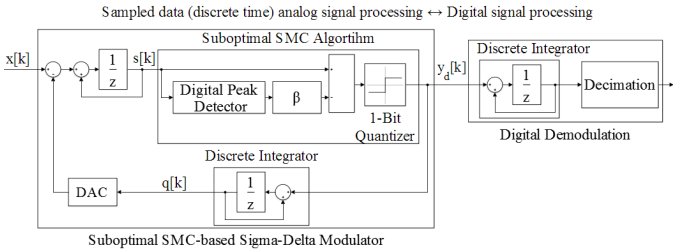


Fig. 3. Proposed First-order Differentiator-based ΣΔ ADC scheme.

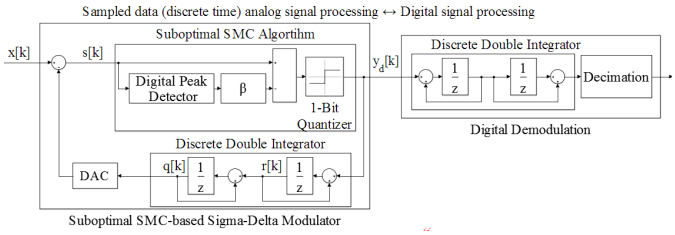


Fig. 4. Proposed Second-order Differentiator-based ΣΔ ADC scheme.

Thus, it yields that (15) is still in force. Finally, since the down-sampling process ahead of the low-pass filter in Figure 2 simply re-adapt $\boldsymbol{y}_{\rm av} \equiv \boldsymbol{x} + \varepsilon_6 \cdot \mathcal{O}(\sqrt{T_s})$ to the original sampling rate of the reference $\boldsymbol{x}$, then no further errors are introduced. This conclude the proof. ■

IV. Problem Formulation and Main Results

From the performance analysis provided in Section III, the following facts are emerged:

- From Theorem 1: The maximum accuracy of the ΣΔ A/D system during the 1-bit conversion is $\propto \mathcal{O}(T_s)$;
- From Theorem 2: The maximum expected error during the demodulation process of $\boldsymbol{y}_{\rm d}[k]$ to reconstruction the reference signal $\boldsymbol{x}[k]$ is $\propto \mathcal{O}(\sqrt{T_s}) > \mathcal{O}(T_s)$.

Moreover, since the analysis provided in Theorem 2 takes into account also the error due to the 1-bit conversion, we conclude that the maximum overall error expected by the traditional ΣΔ ADC process is, at least, within an $\mathcal{O}(\sqrt{T_s})$.

In the following we will present two novel ΣΔ modulator schemes able to reduce the error of the whole 1-bit conversion process up to an $\mathcal{O}(T_s^2)$ boundary layer. Both these schemes, shown resp. in Figure 3 and Figure 4, are inspired to concepts of feedback-based derivative estimation [20] and take the advantages from the digital Suboptimal 2-SMC introduced in Section II.

A. First-order Differentiator-based ΣΔ Modulator

Consider the block diagram shown in Figure 3. The resulting closed loop dynamic for the First-order Differentiator-based ΣΔ Modulator (FD ΣΔ-M) is given below

$$\boldsymbol{s}[k+1] = \boldsymbol{s}[k] + T_s \cdot (\boldsymbol{x}[k] - \boldsymbol{q}[k]) \tag{16}$$
$$\boldsymbol{q}[k+1] = \boldsymbol{q}[k] + T_s \cdot \boldsymbol{\mathcal{U}} \cdot \operatorname{sign}(\boldsymbol{s}[k] - \beta \boldsymbol{s}_{\rm sp}) \tag{17}$$

Let $\boldsymbol{w}[k] = \boldsymbol{x}[k] - \boldsymbol{q}[k]$ and $\Delta\boldsymbol{x}[k] = \boldsymbol{x}[k+1] - \boldsymbol{x}[k]$, and by adding and subtracting $\boldsymbol{x}[k]$ in (17), the discrete-time model (16)-(17) can be reformulated as follows

$$\boldsymbol{s}[k+1] = \boldsymbol{s}[k] + T_s \boldsymbol{w}[k] \tag{18}$$
$$\boldsymbol{w}[k+1] = \boldsymbol{w}[k] + T_s \cdot (\dot{\boldsymbol{x}}[k] - \boldsymbol{\mathcal{U}} \cdot \operatorname{sign}(\boldsymbol{s}[k] - \beta \boldsymbol{s}_{\rm sp})) \tag{19}$$

where $\dot{\boldsymbol{x}}[k] = \Delta\boldsymbol{x}[k]/T_s$ can be thought as a first forward difference approximation of $\mathrm{d}\boldsymbol{x}(\mathrm{t})/\mathrm{dt}$. Let $2\boldsymbol{\mathcal{B}}_{\boldsymbol{x}}$ be the Nyquist Frequency for signal $\boldsymbol{x}$, we firstly remind that ΣΔ modulators use to operate with very high over-sampling ratio [3], e.g.,

$$f_s = T_s^{-1} > 2^{\rm n} \cdot (2\boldsymbol{\mathcal{B}}_{\boldsymbol{x}}) \quad \text{with} \quad {\rm n} > 5. \tag{20}$$

Thus, from (20) and by noticing that from Assumption 1 $|\boldsymbol{x}| \leq \boldsymbol{\mathcal{X}}$ and that it has a limited bandwidth $\boldsymbol{\mathcal{B}}_{\boldsymbol{x}}$, we have that also the spectrum of $\dot{\boldsymbol{x}}(t) \equiv \dot{\boldsymbol{x}}[k]|_{T_s \to 0}$ satisfies $\dot{\boldsymbol{X}}(f) = 0 \ \forall \ f > \boldsymbol{\mathcal{B}}_{\boldsymbol{x}}$. Since according to (20) the high-frequency components of the 1-bit term in (19) do not overlap with the spectrum $\dot{\boldsymbol{X}}(f)$, we can state the following upper-estimation

$$|\dot{\boldsymbol{x}}[k]| \leq \|\dot{\boldsymbol{x}}[k]\|_\infty \leq \|\boldsymbol{x}\|_\infty \cdot (2\pi\boldsymbol{\mathcal{B}}_{\boldsymbol{x}}) = \boldsymbol{\mathcal{X}} \cdot (2\pi\boldsymbol{\mathcal{B}}_{\boldsymbol{x}}). \tag{21}$$

Let us now note that system (18)-(19) mets the structure of system (4)-(5). Thus, let $\boldsymbol{\mathcal{U}}$ be chosen as in Section II, such that $\mathcal{K} > \|\dot{\boldsymbol{x}}[k]\|_\infty = \boldsymbol{\mathcal{X}} \cdot \boldsymbol{\mathcal{B}}_{\boldsymbol{x}}$., then $\boldsymbol{s}[k]$ will converges within an $\mathcal{O}(T_s^2)$ boundary layer in finite-time.

Finally, by exploiting Equivalent Control's concepts [13], it further results that the 1-bit output of the FD ΣΔ-M scheme in Figure 3 encodes the information of the time-derivative of the reference signal $\boldsymbol{x}(t)$, i.e.,

$$\boldsymbol{y}_{\rm d} \equiv \dot{\boldsymbol{x}}[k] = \left.\frac{\boldsymbol{x}[k+1] - \boldsymbol{x}[k]}{T_s}\right|_{T_s \to 0} \equiv \frac{\mathrm{d}\boldsymbol{x}(\mathrm{t})}{\mathrm{dt}}. \tag{22}$$

Moreover, since $\boldsymbol{y}_{\rm d}[k] \equiv \dot{\boldsymbol{x}}[k]$, then the low-pass filtering process needed at the demodulation side of the traditional ΣΔ-M scheme in Figure 1 is no more required. Here, to demodulate the original signal $\boldsymbol{x}[k]$, we can simply integrate $\boldsymbol{y}_{\rm d}[k]$ as in Figure 3. Since both the discrete-time integration process and the down-sampling process at the demodulation side of the scheme in Figure 3 are errorless computation processes, the proposed scheme allow to reduce to an $\mathcal{O}(T_s^2)$ the precision of the whole ADC process, with respect to the traditional schemes, i.e., $\mathcal{O}(T_s^2) \ll \mathcal{O}(\sqrt{T_s})$.

B. Second-order Differentiator-based ΣΔ Modulator

Consider the block diagram shown in Figure 4. The resulting closed loop dynamic for the Second-order Differentiator-based ΣΔ Modulator (SD ΣΔ-M) is given below

$$\boldsymbol{q}[k+1] = \boldsymbol{q}[k] + T_s \cdot \boldsymbol{r}[k] \tag{23}$$
$$\boldsymbol{r}[k+1] = \boldsymbol{r}[k] + T_s \cdot \boldsymbol{\mathcal{U}} \cdot \operatorname{sign}(\boldsymbol{s}[k] - \beta \boldsymbol{s}_{\rm sp}) \tag{24}$$

where $\boldsymbol{s}[k] = \boldsymbol{x}[k] - \boldsymbol{q}[k]$. Thanks to (23) we compute

$$\boldsymbol{s}[k+1] = \boldsymbol{x}[k+1] - \boldsymbol{q}[k] - T_s \cdot \boldsymbol{r}[k]. \tag{25}$$

Let $\Delta\boldsymbol{x}[k] = \boldsymbol{x}[k+1] - \boldsymbol{x}[k]$, and let, in accordance with (30), $\dot{\boldsymbol{x}}[k] = (\Delta\boldsymbol{x}[k]/T_s)$. By the change of variable $\boldsymbol{w}[k] =$

$(\Delta \boldsymbol{x}[k]/T_s) - \boldsymbol{r}[k] = \dot{\boldsymbol{x}}[k] - \boldsymbol{r}[k]$ and by adding and subtracting, resp., $\boldsymbol{x}[k]$ to the right-hand side of (25), and $\dot{\boldsymbol{x}}[k]$ to the right-hand side of (24), it yields

$$\begin{aligned}\boldsymbol{s}[k+1] =& \boldsymbol{s}[k] + T_s \cdot \boldsymbol{w}[k] \quad (26)\\ \boldsymbol{w}[k+1] =& (\Delta \boldsymbol{x}[k+1])/T_s - \boldsymbol{r}[k+1] \\ =& \dot{\boldsymbol{x}}[k+1] - \boldsymbol{r}[k] - T_s \cdot \boldsymbol{\mathcal{U}} \cdot \operatorname{sign}(\boldsymbol{s}[k] - \boldsymbol{\beta}\boldsymbol{s}_{\mathrm{sp}}) \\ =& \boldsymbol{w}[k] + T_s \cdot (\ddot{\boldsymbol{x}}[k] - \boldsymbol{\mathcal{U}} \cdot \operatorname{sign}(\boldsymbol{s}[k] - \boldsymbol{\beta}\boldsymbol{s}_{\mathrm{sp}})) \quad (27)\end{aligned}$$

where

$$\begin{aligned}\ddot{\boldsymbol{x}}[k] =& (\dot{\boldsymbol{x}}[k+1] - \dot{\boldsymbol{x}}[k])/T_s = (\Delta \boldsymbol{x}[k+1] - \Delta \boldsymbol{x}[k])/T_s^2 \\ =& (\boldsymbol{x}[k+2] - 2\boldsymbol{x}[k+1] + \boldsymbol{x}[k])/T_s^2 \quad (28)\end{aligned}$$

Then, analogously with Section IV-A, thank to Assumption 1, the over-sampling condition (20), and relation (21), we can upper-estimate $\ddot{\boldsymbol{x}}[k]$ as next

$$|\ddot{\boldsymbol{x}}[k]| \leq \|\ddot{\boldsymbol{x}}[k]\|_\infty \leq \|\dot{\boldsymbol{x}}[k]\|_\infty \cdot (2\pi \boldsymbol{\mathcal{B}}_{\boldsymbol{x}}) = \boldsymbol{\mathcal{X}} \cdot (2\pi \boldsymbol{\mathcal{B}}_{\boldsymbol{x}})^2. \quad (29)$$

Let us now note that, after the previous manipulations, the SD ΣΔ-M in (23)-(24) mets the (4)-(5), cfr. with (26)-(27). Thus, let $\boldsymbol{\mathcal{K}} > \boldsymbol{\mathcal{X}} \cdot (2\pi \boldsymbol{\mathcal{B}}_{\boldsymbol{x}})^2$, according to Section II, $\boldsymbol{s}[k]$ will converges in finite-time within an $\mathcal{O}(T_s^2)$ boundary layer.

Finally, by exploiting Equivalent Control's concepts [13], it further results that the 1-bit output of the SD ΣΔ-M scheme in Figure 4 encodes the information of the second time-derivative of the reference signal $\boldsymbol{x}(\mathrm{t})$, i.e.,

$$\boldsymbol{y}_{\mathrm{d}} \equiv \ddot{\boldsymbol{x}}[k] = \left. \frac{\boldsymbol{x}[k+2] - 2\boldsymbol{x}[k+1] + \boldsymbol{x}[k]}{T_s^2} \right|_{T_s \to 0} \equiv \frac{\mathrm{d}^2 \boldsymbol{x}(\mathrm{t})}{\mathrm{dt}^2}. \quad (30)$$

Thus, also the SD ΣΔ-M scheme does not need to the low-pass filtering process at the demodulation side, cfr. Figure 2 and Figure 4. Thus, to demodulate $\boldsymbol{x}[k]$, we can simply integrate twice $\boldsymbol{y}_{\mathrm{d}}[k]$ as in Figure 4. Since both the integrations process and the down-sampling processes at the demodulation side of the scheme in Figure 3 are errorless computation processes, follows that also the SD ΣΔ-M may reduce to an $\mathcal{O}(T_s^2)$ the precision of the whole ADC process, w.r.t. the traditional scheme, i.e., $\mathcal{O}(T_s^2) \ll \mathcal{O}(\sqrt{T_s})$.

Remark 1: The proposed schemes reduce not only the boundary-layer of the error associated with the ΣΔ conversion to an $\mathcal{O}(T_s^2)$, instead of $\mathcal{O}(T_s)$ for the traditional scheme, but they provide also an improvement in terms of accuracy at the demodulation side. There, the error in demodulation is reduced to an $\mathcal{O}(T_s^2)$, instead of $\mathcal{O}(\sqrt{T_s})$. However, since the size ε_1 and ε_6 of these boundary layers depend on the magnitude of the discontinuous gains $\boldsymbol{\mathcal{K}}$ in (17) and (24), and both the proposed scheme requires high gain, is reasonable that a compromise between the bandwidth of the signal to be converted and precision to be achieved must be found. ■

V. Simulation Results

Hereinafter the accuracy features of the proposed FD ΣΔ-M scheme in Figure 3 and the SD ΣΔ-M given in Figure 4, are compared with that of the traditional implementation in Figure 2. Simulations were performed in the MATLAB®/Simulink® environment by using the "ode1" (Euler) integration method. The use-case considers the 1-bit high-resolution ADC process in the Voice-Frequency band, i.e., in the range $[0 \div 4\mathrm{kHz}]$. Let $\boldsymbol{\mathcal{B}}_{\boldsymbol{x}} = 4\mathrm{kHz}$ be the highest component frequency allocated for a single transmission channel including guard bands, the corresponding Nyquist sampling frequency is twice, i.e., 8kHz. For the three implementations: a) According to the noise-shaping paradigm we select a high over-sampling ratio, equal to 64, i.e. $\mathrm{n} = 6$ in (20), thus $f_s = 512\mathrm{kHz}$, $T_s = (f_s)^{-1} \approx 2\mu\mathrm{s}$; b) The DACs in Figures 2, 3 and 4 have been assumed to be ideal for simplicity, and thus replaced by a unity gain; c) The decimation processes at the demodulation side have been implemented by a cascade of three FIR Decimation stages available in DSP System Toolbox™. Each stage reduces the sampling rate by a factor of four, thus the at the demodulation side the sampling rate back to 8kHz. The latency of each stage is of 16 samples.

TABLE II
Parameters for the three ΣΔ implementations under test

	ΣΔ-M (6),(10),(11)	FO ΣΔ-M (16),(17)	SO ΣΔ-M (23),(24)
f_s	512kHz	512kHz	512kHz
$\boldsymbol{\mathcal{K}}$	$\boldsymbol{\mathcal{X}} + \boldsymbol{\epsilon}$	$\boldsymbol{\mathcal{X}} \cdot (2\pi \boldsymbol{\mathcal{B}}_{\boldsymbol{x}}) + \boldsymbol{\epsilon}$	$\boldsymbol{\mathcal{X}} \cdot (2\pi \boldsymbol{\mathcal{B}}_{\boldsymbol{x}})^2 + \boldsymbol{\epsilon}$
$\boldsymbol{\beta}$	–	0.5	0.5
τ	$5.134 \times 10^{-5}\mathrm{s}$	–	–

The reference input $\boldsymbol{x}(\mathrm{t})$ considered for the tests is a reverse sawtooth wave of frequency 200Hz. To avoid aliasing, the Fourier series of $\boldsymbol{x}(\mathrm{t})$ is truncated at the 15-th term as

$$\boldsymbol{x}(\mathrm{t}) = \frac{2}{\pi} \cdot \sum_{\ell=1}^{15} (-1)^{\ell} \cdot \frac{\sin((2\pi f \cdot \mathrm{t}) \cdot \ell)}{\ell} \quad (31)$$

From (31), and in accordance with Assumption 1, $\boldsymbol{x}(\mathrm{t})$ is a bounded signal, with a bounded bandwidth, such that

$$|\boldsymbol{x}(\mathrm{t})| \leq \|\boldsymbol{x}\|_\infty = \boldsymbol{\mathcal{X}} = 1.1158 \quad (32)$$

$$\boldsymbol{X}(f) \leq = 0 \quad \forall \quad f > 15 \cdot 200\mathrm{Hz} = 3\mathrm{kHz} > \boldsymbol{\mathcal{B}}_{\boldsymbol{x}}. \quad (33)$$

The behavior of the signal $\boldsymbol{x}[k]$, down-sampled at 8kHz, is depicted in Figure 5. Let $\boldsymbol{\epsilon} = 1$, the parameters for the three ΣΔ implementations under test are provided in Table II.

In Figure 6 the output variable at the demodulation side for the three implementations under test in, resp., Figure 2, 3 and 4, are displayed. Notice that a latency is evident for all the structures, cfr. with Figure 5, however this phenomenon is due to the delays introduced by the decimation stages. The latency in a known problem in ΣΔ techniques as highlighted in Table I.

In Figure 7 is shown the conversion error at the demodulation side after having compensate the latency on signal $\boldsymbol{x}[k]$. It is evident that the two novel schemes provide a better accuracy w.r.t. to the traditional scheme.

Finally, in Figure 8 is shown the temporal behavior of the 1-bit (2 symbols) output $\boldsymbol{y}_{\mathrm{d}}[k]$ for the three ΣΔ implementations. Since the three ADCs converts different information, thus their output are different. We further note that the new schemes use to switch less frequently w.r.t. the traditional

Preprint submitted to 15th International Workshop on Variable Structure Systems and Sliding Mode Control. Received February 8, 2018.

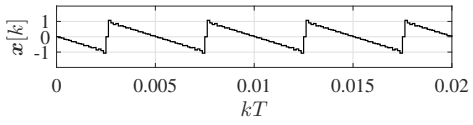


Fig. 5. Actual behavior of the signal $\boldsymbol{x}[k]$ down-sampled at 8kHz.

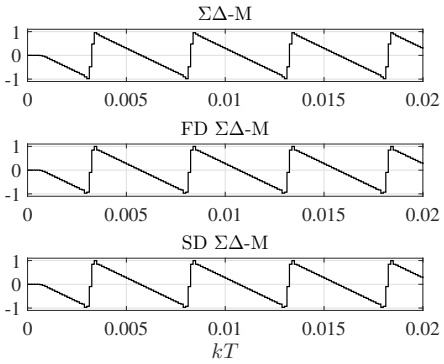


Fig. 6. Actual behavior of the ΣΔ modulators's output, at the demodulation side, down-sampled at 8kHz.

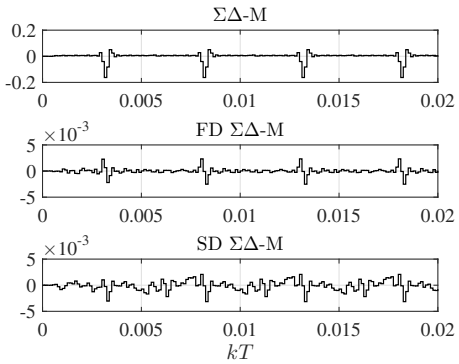


Fig. 7. Actual behavior of the accuracy error $\boldsymbol{e}[k]$, at the demodulation side, down-sampled at 8kHz.

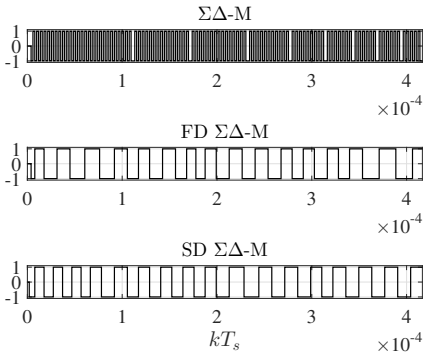


Fig. 8. Temporal behavior of the 1-bit output $\boldsymbol{y}_{\mathrm{d}}[k]$ of the three ΣΔ implementations at 512kHz of sampling rate.

scheme. Further studies will be devoted to better understand this phenomenon.

VI. Conclusions

Here two novel ΣΔ-Ms that may reduce the error associated with the 1-bit conversion process to an $\mathcal{O}(T_s^2)$ boundary-layer, instead of $\mathcal{O}(T_s)$ for the traditional scheme, and that also provide better accuracy at the demodulation side, to an $\mathcal{O}(T_s^2)$ instead of $\mathcal{O}(\sqrt{T_s})$, were presented.

Further researches will be devoted to better characterize the features of these schemes in terms accuracy, in both the time and the frequency domain, and to study their noise-shaping capabilities. Further direction for next activities consist to evaluate their application in noisy environments where a compromise between the over-sampling frequency, the bandwidth of the reference signal, and thus the magnitude of the discontinuous terms is needed to achieve optimality in the 1-bit conversion and transmission process.

References

[1] M. José, R. Schreier, K.-P. Pun, and S. Pavan, "Next-generation Delta-Sigma converters: Trends and perspectives," *IEEE J. on Emerging and Selected Topics Circ. Sys., 5(4), pp. 484–499*, 2015.

[2] W. Kester, "Which ADC architecture is right for your application?" 2005.

[3] P. M. Aziz, H. V. Sorensen, and J. Vn der Spiegel, "An overview of Sigma-Delta Converters," *IEEE Signal Processing Magazine, 13(1), pp. 61–84*, 1996.

[4] T. Li and Y. Fujimoto, "Control system with high-speed and real-time communication links," *IEEE Trans. Ind. Electron., 55(4), pp. 1548–1557*, 2008.

[5] S. Ohno, T. Shiraki, M. R. Tariq, and M. Nagahara, "Rate-distortion analysis of Delta-Sigma modulators," *IEEE Int. Conf. Acoustics, Speech and Signal Processing, pp. 4581–4585*, 2017.

[6] S. Liu, W. Jiang, and M. Zhang, "Dual-channel multiplexing technology and its realization in interpolation filter in stereo audio Sigma-Delta DAC," *Springer, Analog Integrated Circuits and Signal Processing*, vol. 81(2), pp. 487–494, 2014.

[7] A. Dashtestani and B. Bakkaloglu, "A fast settling oversampled digital sliding-mode DC–DC converter," *IEEE Trans. Power Electron., 30(2), pp. 1019–1027*, 2015.

[8] H. Sira-Ramírez, "Sliding regimes in analog signal encoding and Delta modulation circuits," pp. 78–87, 1987.

[9] X. Xia and A. S. I. Zinober, "Periodic orbits from Δ-modulation of stable linear systems," *IEEE Trans. Automatic Control, 49, (8), pp. 1376–1380*, 2004.

[10] X. Xia and A. S. Zinober, "Δ-modulated feedback in discretization of sliding mode control," *Elsevier Automatica, 42(5), pp. 771–776*, 2006.

[11] D. Almakhles, A. K. Swain, and A. Nasiri, "The dynamic behaviour of data-driven Δ-M and ΔΣ-M in sliding mode control," *Int. J. of Control, 90(11), pp. 2406–2414*, 2017.

[12] H. S. Ramırez and V. F. Battle, "A generalized PI sliding mode and PWM control of switched fractional systems," *Springer, Modern Sliding Mode Control Theory: New Perspectives and Applications,375, pp.201-221*, 2008.

[13] V. I. Utkin, "Sliding modes in control and optimization," 2013.

[14] A. Bartoszewicz, "Discrete-time quasi-sliding-mode control strategies," *IEEE Trans. Ind. Electron., 45(4), pp. 633–637*, 1998.

[15] G. Bartolini, A. Pisano, and E. Usai, "Digital second-order sliding mode control for uncertain nonlinear systems," *Elsevier, Automatica, 37(9), pp. 1371–1377*, 2001.

[16] G. Bartolini, A. Ferrara, and E. Usai, "Sub-optimal sliding mode control of uncertain second order dynamical systems," *Elsevier IFAC Proc. Volumes, 29(1), pp. 2254–2259*, 1996.

[17] ——, "Chattering avoidance by second-order sliding mode control," *IEEE Trans. Automa. Control, 43(2), pp. 241–246*, 1998.

[18] G. Bartolini, A. Ferrara, A. Levant, and E. Usai, "On second order sliding mode controllers," 1999.

[19] G. Bartolini, A. Pisano, and E. Usai, "Digital second order sliding mode control of siso uncertain nonlinear systems," 1998.

[20] ——, "First and second derivative estimation by sliding mode technique," *Journal of Signal Processing, 4(2), pp. 167–176*, 2000.