Available online at www.sciencedirect.com
SCIENCE DIRECT
Control Engineering Practice 14 (2006) 337–350
CONTROL ENGINEERING PRACTICE
www.elsevier.com/locate/conengprac

Improving the accuracy of analog encoders via Kalman filtering ☆

Yaron Zimmerman[a,1], Yaakov Oshman[b,*], Amit Brandes[c,2]

[a]Technion—Israel Institute of Technology, Department of Mechanical Engineering, Technion City, Haifa 32000, Israel
[b]Technion—Israel Institute of Technology, Department of Aerospace Engineering, Technion City, Haifa 32000, Israel
[c]University of Rome 'La Sapienza', Computer and Systems Science Department, Via Eudossiana 18, 00184, Rome, Italy

Received 16 May 2004; accepted 24 January 2005
Available online 26 February 2005

Abstract

A new decoding method is presented for analog encoders enabling major improvements in both accuracy and resolution. A simulation study and experiments with real, industrial-grade, equipment demonstrate the performance improvement of the proposed method, revealing that the new method can generate position estimates with accuracy about three times better than that of standard methods. Moreover, in some special cases, the resulting position accuracy can reach sub-nanometer levels, thus enabling further size reduction in the semiconductor industry. The proposed algorithm also yields velocity estimates better by about two orders of magnitude than those obtained with standard methods.

Keywords: Encoders; Kalman filters; Estimation algorithms; Nonlinear filters; Accuracy; Measurement noise

1. Introduction

Linear analog encoder is a very common sensor in the semiconductor industry. Its main use is as a feedback sensor in X–Y or X–Y–Z tables that hold the wafer against a vision instrument like a high-resolution camera or an electronic microscope. In such applications the resolution requirements are extremely stringent, reaching few nanometers or even less than one nanometer. Existing decoding algorithms are limited because they are based on solving the geometrical problem only and do not attempt to improve the accuracy by applying some filtering strategy.

The analog encoder generates sine and cosine signals that are related to its linear translation. Like in any other real-world device, the encoder's measured signals are corrupted by common error sources, such as electronic noise or quantization error, which, in turn, reduce the accuracy of the encoder. The noise issue becomes even more important when bearing in mind that the encoder, a position sensor by its nature, is also commonly used to estimate velocities by a straightforward numerical differentiation of its output. To the best of the authors' knowledge, the idea of implementing state filtering techniques to confront the effects of the measurement error sources has not appeared in the open literature up to this date. Furthermore, only a few published works relate to the general problem of encoder measurement noise alleviation. Yang, Rees, and Chuter (2002) address only deterministic error sources, arising from mechanical installation errors, in an analog encoder. They develop a mathematical model of the resulting error and use a Kalman filter, implemented off-line, to estimate the model parameters. This model, in turn, is used to recalculate the measured position. The work of Venema (1994) is of a similar

☆Based on a paper presented at the 5th IFAC International Symposium on Intelligent Components and Instruments for Control Applications (SICICA-2003), Aveiro, Portugal, July 9–11, 2003.

*Corresponding author. Tel.:/fax: +97248293803.
E-mail addresses: yaron.zimmerman@kla-tencor.com (Y. Zimmerman), yaakov.oshman@technion.ac.il (Y. Oshman), amit_brandes@amat.com (A. Brandes).
[1]Currently with KLA-Tencor Corporation (Israel), Hatikshoret St., P.O.Box 143, Migdal Haemek 23100, Israel.
[2]Currently with Applied Materials, PDC business group, 9 Oppenheimer St., Rehovot 76705, Israel.

0967-0661/$ - see front matter
doi:10.1016/j.conengprac.2005.01.008

nature, using the filtering algorithm off-line for the sole purpose of calibrating the analog encoder. Other researchers (Hagiwara, Suzuki, & Murase, 1992; Mayer, 1994) deal only with the issue of calculating the encoder angle from the sine and cosine signal by applying inverse trigonometric functions.

This paper deals with the accuracy issue of analog encoders by applying on-line filtering to the measured signals. Using the extended Kalman filtering (EKF) framework, a novel decoding algorithm is developed and verified. The new algorithm enables improved position estimation and a dramatic improvement in velocity estimation, relative to commercially available existing algorithms.

The remainder of this paper is organized as follows. In the next section the operational principles of an analog encoder are described and the measurement model is constructed. The system model is then developed and both models are put together within the framework of an EKF. The following section presents a simulation study through which the filter's performance is verified, and a discussion of the results. Some of the more interesting and promising aspects of the proposed decoding method are pointed out. Next, the proposed algorithm is experimented with real industrial hardware and confronted with a conventional algorithm. Concluding remarks are offered in the final section.

2. Encoder model

2.1. Principle of operation

The encoder's principle of operation is schematically depicted in Fig. 1. A scale, having gold-made triangular facets, reflects the light of an infra-red LED light source, through an index grating, unto a photo-detector. Due to the periodic pattern of both the scale and the index grating, sinusoidal interference fringes are produced on the detection plane of the photo-detector. Whenever the moving parts are in motion, i.e., the light source, the index grating and the photo detector move relative to the static scale, the fringes move along the detection plane. Using a special assembly, the photo detectors provide signals related to the motion of the fringes. An electronic circuit amplifies and combines the photo detectors' signals to generate two sinusoidal waveforms of equal amplitude and period (equaling the scale period), separated by a 90 degrees phase shift. The interested reader is referred to (Webster, 1999; Renishaw, 1998) for further details.

2.2. Measurement model

The most accurate scales now commercially available have resolutions of a few micrometers. To obtain sub-micron resolution, as required by production systems, the analog signals must be sampled and decoded. The measurement vector $\boldsymbol{z} \in \mathbb{R}^2$ can be described as follows:

$$\boldsymbol{z} \triangleq \begin{bmatrix} z(1) \\ z(2) \end{bmatrix} = \begin{bmatrix} V \sin(2\pi f_s x(t)) \\ V \cos(2\pi f_s x(t)) \end{bmatrix}, \tag{1}$$

where f_s is the scale spatial frequency, V is the signal amplitude and $x(t)$ is the position.

Most manufacturers use the following equation to estimate the position from the measured signals (Venema, 1994):

$$x = \frac{a\tan 2(z(1), z(2))}{2\pi f_s}, \tag{2}$$

where $a\tan 2(\cdot,\cdot)$ is the 4 quadrant inverse tangent function. Some manufacturers increase the estimation accuracy by normalizing the signals and applying an arcsine or an arccosine function, depending on the signal absolute value, as follows:

$$x = \begin{cases} \dfrac{\arcsin(z(1)/\|\boldsymbol{z}\|)}{2\pi f_s}, & |z(1)| \leqslant |z(2)|, \\ \dfrac{\arccos(z(2)/\|\boldsymbol{z}\|)}{2\pi f_s}, & |z(2)| < |z(1)|. \end{cases} \tag{3}$$

This procedure increases the signal-to-noise ratio (SNR) by taking the information from the more sensitive regions in the signal harmonic.

All of the above algorithms do not apply any filtering procedure to the signal, i.e., they do not use any prior information regarding the signal to weigh it against the current measured information. In the presence of measurement noise, the algorithm's output, i.e., the estimated position, becomes noisy as well. Furthermore, the estimated velocity signal, calculated by numerically differentiating the (noisy) position signal, is subject to much higher noise levels.

To account for the sampling operation that transforms the continuous signal $x(t)$ to its sampled version $x(k+1)$, and for the contamination of the measurement by the measurement noise, Eq. (1) is rewritten as

$$\begin{aligned} \boldsymbol{z}(k+1) &= \boldsymbol{h}(x(k+1)) + \boldsymbol{v}(k+1) \\ &= \begin{bmatrix} V \sin(2\pi f_s x(k+1)) \\ V \cos(2\pi f_s x(k+1)) \end{bmatrix} + \boldsymbol{v}(k+1), \end{aligned} \tag{4}$$

where $\boldsymbol{v}(k+1)$ is the measurement noise. There are several sources for measurement noise (Venema, 1994), but the most dominant one is introduced by an electronic noise within the sampling operation (ground noise of the analog-to-digital (A/D) converter). The measurement noise is assumed to be a white, zero-mean Gaussian stationary sequence:

$$\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}). \tag{5}$$