Maximum Likelihood Estimation of ADC Parameters

László Balogh, István Kollár
Department of Measurement and Information Systems
Budapest University of Technology and Economics
Email: [lbalogh][kollar]@mit.bme.hu

Attila Sárhegyi
Cypress Semiconductor Corporation
Email: asar@cypress.com

***Abstract*—Dynamic testing of analog-digital converters (ADC) is a complex task. A possible approach is using a sine wave because it can be generated with high precision. However, in the sine wave fitting method for the test of ADC's, all the available information is extracted from the measured data. Therefore, the estimated ADC parameters (ENOB, linearity errors) are not always accurate enough, and not detailed information is gained about the nonlinearity of the ADC.**

Generally, maximum likelihood (ML) estimation is a powerful method for the estimation of unknown parameters. However, currently it is not used for the processing of such data, because of the difficulties of formulating it, furthermore because of the numerically demanding task of the minimization of the ML cost function [9].

We have succeeded in formulating the maximum likelihood function for a sine wave excitation, and in minimizing it. The number of parameters is frightening (all comparison levels of the ADC plus parameters of the sine wave plus variance of an additive input noise), but proper handling allows to determine the best values based on the data.

The proper definition of the ML function and formulation of the numerical method are presented, with results using simulation and measurement data. To our knowledge, this is the first case to solve the full maximum likelihood problem.

I. Introduction

In sine wave testing, the method recommended by the standard IEEE-1241 [1] is least squares fit of the ADC output data by an appropriate sine wave. This works more or less well, but gives no detailed information about the differential and integral nonlinearity of the ADC. Only global measures like the "effective number of bits" (ENOB) are returned.

However, in the measured data there is more information, moreover, least squares fit is not completely proper. Tacitly, it assumes that the quantization noise is independent and is normally (at least symmetrically) distributed [3], [4]. This is not true, therefore a better model of the ADC (quantization) would certainly yield a better results.

If the input signal is indeed sinusoidal, even differential nonlinearity can be evaluated at each level. If we can formulate the likelihood cost function, the "best" estimates of the parameters of the sine wave and of the ADC can be determined.

The idea presented in this paper is that using appropriate parameters (sine wave: sine/cosine amplitudes, frequency; DC level; comparison levels of the ADC; standard deviation of an additive noise in the ADC) the proper maximum likelihood problem can be formulated and solved. One observes the quantized ADC outputs, defines the likelihood function, and maximizes this via all the parameters. The key is use of the simple model of Gaussian noise which allows connection of the samples of the sine to the quantized samples.

At first sight, the problem looks frightening. Even for a 10-bit ADC, there are $2^{10} - 1 = 1023$ comparison levels to be estimated. This sounds large, but numerically, one can make use of the fact that with reasonable noise, the output values are only possible in a relatively small environment of the input values. Therefore, maximization can be handled "locally", at least when the ADC comparison levels are treated; the sine wave can be fitted in "global" steps. Thus, numerically demanding tasks can be reasonably reduced by alternately using optimization steps.

The result is not only a better estimate of the sine wave, but also a set of estimates of the comparison levels. Thus, a more complex description of the ADC behavior is at hand.

In this paper the likelihood function to be maximized and the appropriate cost function to be minimized are derived. The advantage of this is that in statistics, the ML methods have the "best" properties. The disadvantage is computation complexity. In the case of least squares (LS), numerical computation of the solution can be easily handled even if large measurement data sets are used. To compute the maximum of the ML cost function a numerical minimization method is required. A general multidimensional optimization program may be very slow. Therefore, special care needs to be taken in the implementation.

The paper is organized as follows: section II describes the model of measurement and introduces notations. In section III maximum likelihood function of the estimation is derived. The maximum of the ML function can be obtained by the numerical optimization algorithm which is presented in section IV. Section V demonstrates the working of the presented algorithm by using simulated and measurement data.

II. Mathematical Model

In this paper we consider a measurement model which has a finite number of parameters. Our ultimate goal is to estimate the parameters from measurement data. In this section

This work has been supported by OTKA grant #73496.

978-1-4244-2833-5/10/$25.00 ©2010 IEEE

detailed description of the model is given. For all estimations of parameters the symbol $\hat{}$ is applied.

Here, the AD converter is modeled by a function $q : \mathbb{R} \mapsto \{0, \ldots, N-1\}$ which maps the analog input plus noise to the set of digital output codes. The number of elements or possible output codes is N. Moreover, it is also assumed that other imperfections of an AD converter does not present.

Denoting the sampling frequency and time by f_s and T_s, respectively, it means that AD converter produces the lth output code by using values of $x(lT_s)$ and $n(lT_s)$ where $n(t)$ denotes the noise process. The parameters of the model are the code transition levels denoted by $T_k \in \mathbb{R}$ ($k = 1, \ldots, N-1$) and the noise variance σ.

A. Transition Levels

The quantization is modeled by the sign function

$$\mathrm{sgn}(x) = \begin{cases} 1, & \text{if } x \geq 0, \\ -1, & \text{otherwise.} \end{cases}$$

At a particular point of time $t = t_k$ the input-output relationship $y_k = q(x(t_k))$ is defined as

$$y_k = q(x(t_k)) = \frac{N-1}{2} + \frac{1}{2} \sum_{l=1}^{N-1} \mathrm{sgn}(x(t_k) - T_l)$$

$$= \begin{cases} 0, \text{ if } x(t_k) < T_1, \\ m, \text{ if } T_{m-1} \leq x(t_k) < T_m, \\ N-1, \text{ if } T_{N-1} \leq x(t_k). \end{cases} \tag{1}$$

Output sequence denoted by y_k, $k = 1, 2, \ldots, M$ with $\forall k : y_k \in \{0, \ldots, N-1\}$. The negative and positive analog full scale values are noted by $V_{\text{min}} \in \mathbb{R}$ and $V_{\text{max}} \in \mathbb{R}$, respectively. In the case of an ideal quantizer we can write that

$$T_{\text{ideal},k} = Q(k-1) + T_{\text{ideal},1} \tag{2}$$

where $Q \in \mathbb{R}$ is the ideal width of a code which can be calculated as $Q = (V_{\text{max}} - V_{\text{min}})/N$ and $T_{\text{ideal},k}$ denotes the kth transition level of an ideal quantizer. Fig. 1 explains graphically the notations.

B. Additional Noise

The non-ideal measurement environment and the internal noises are modeled by additional noise. The noise appears in the analog domain and is assumed to have Gaussian distribution with zero mean. The noise is denoted by $\mathcal{N}(t)$ and the model is described by

$$y_k = q(x(t_k) + \mathcal{N}(t_k))$$

where y_k is a concrete realization of the stochastic function $q(x(t_k) + \mathcal{N}(t_k))$. The random variables are denoted by Y_k, $k = 0, \ldots, N-1$. In our approach a white noise model is used, so if $t \neq s$ then $\mathcal{N}(t)$ and $\mathcal{N}(s)$ are independent. For all t: $\mathcal{N}(t)$ has normal distribution with zero mean and variance σ. In this paper σ is assumed to be unknown.

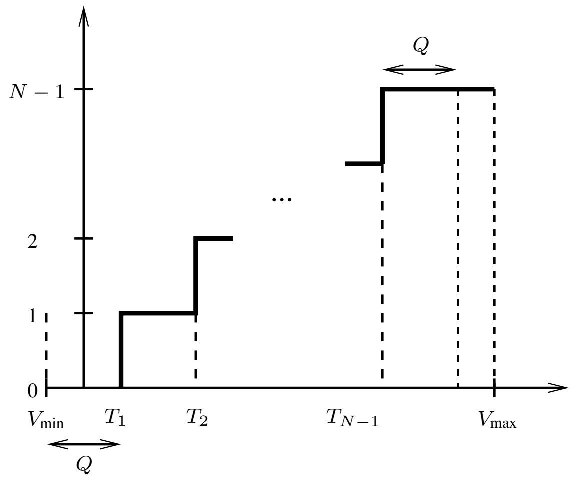

Fig. 1. N-bit ADC transfer function. The notation follows the standard IEEE-1241. T_k, $k = 1, \ldots, N-1$ are the transition levels, Q is the ideal width of a code and V_{min}, V_{max} denote the analog minimum and analog maximum values, respectively.

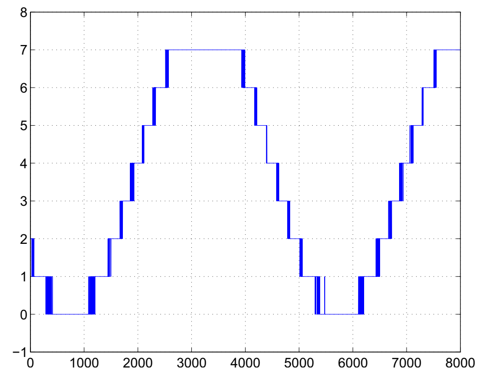

Fig. 2. Example output of a 3-bit AD converter. Simulation parameters: $V_{\text{min}} = 0$, $V_{\text{max}} = 1$, $A = 0.5$, $B = 0$, $C = 0.5$, $f = 0.002$, $\sigma = 0.01$, $T_s = 0.1$.

C. Excitation Signal

Applying sine wave as excitation signal has advantages: can be generated with high precision, determined by four parameters (amplitude, frequency, phase, dc level). In this paper the frequency is supposed to be known, so a linear fitting can be performed. But the method can be easily extended to unknown frequency.

The parametric model is

$$x(t) = A\sin(2\pi f t) + B\cos(2\pi f t) + C \tag{3}$$

where the parameters are $A \in \mathbb{R}$, $B \in \mathbb{R}$, and $C \in \mathbb{R}$.

An example can be seen in Fig. 2. The excitation signal is a sine wave with linear parameters $A = 0.5$ V, $B = 0$ V, $C = 0.5$ V and frequency $f = 0.002$ Hz. Sampling frequency of the 3-bit ideal converter is 10 Hz and $V_{\text{min}} = 0$ V, $V_{\text{max}} = 1$ V. The additive noise has Gaussian distribution with zero mean and variance $\sigma = 0.01$ V. Since the excitation signal frequency is small compared to the sampling frequency the effect of the additive noise close to the transition levels can be observed.

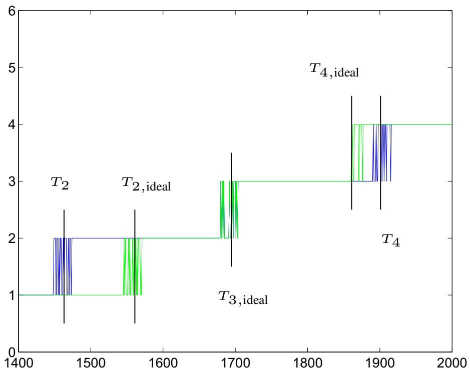

Fig. 3. Demonstration of the effect of non-ideal transition levels. Green line denote the output of an ideal quantizer. Blue line shows the effect where $T_{2,\text{ideal}}$ and $T_{4,\text{ideal}}$ are shifted by $-0.4Q$ and $0.2Q$, respectively.

In Fig. 3 the effect of non-ideal transition levels is demonstrated. The simulation is generated by using the same parameters like in the previous case. Two output sample sequences were generated. In the first one the ideal quantizer was used. Secondly, a non-ideal quantizer was applied. Two transition levels were modified: $T_2 = Q + T_1 - 0.4Q$ and $T_4 = 3Q + T_1 + 0.2Q$ are used instead of (2). In the figure it can be seen that modification of a transition level causes not only shifting of the output samples along the horizontal axes but also a different shape appears because the different value of excitation signal at those time instants.

III. Likelihood Function

The likelihood function is the joint density function for all observations. This probability function cannot be determined in closed form. The result can be calculated from the so called standard normal cumulative distribution function, given by

$$F(x, \mu, \sigma) : \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+ \mapsto \mathbb{R}$$

$$F(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(u-\mu)^2}{2\sigma^2}} du,$$

where $\mu \in \mathbb{R}$ the expected value, and $\sigma \in \mathbb{R}_+$ is the standard deviation of the additive noise. The error function can be expressed with the help of the standard normal cumulative distribution function. Indeed,

$$F(x, \mu, \sigma) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

where

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_{0}^{x} e^{-z^2} dz. \tag{4}$$

It is a very important property that the derivative of the function (4) can be obtained as

$$\frac{d\text{erf}(x)}{dx} = \frac{2}{\sqrt{\pi}} e^{-x^2}.$$

A. The Probability Function

To obtain the likelihood function, we need to evaluate the probability of each sample. From (1) we have

$$\begin{aligned} P(Y_k = 0) &= F(T_1, x(k), \sigma) \\ &= \frac{1}{2}\left[1 + \text{erf}\left(\frac{T_1 - x(t_k)}{\sigma\sqrt{2}}\right)\right] \end{aligned} \tag{5}$$

$$\begin{aligned} P(Y_k = N-1) &= 1 - F(T_{N-1}, x(t_k), \sigma) \\ &= \frac{1}{2}\left[1 - \text{erf}\left(\frac{T_{N-1} - x(t_k)}{\sigma\sqrt{2}}\right)\right] \end{aligned} \tag{6}$$

$$\begin{aligned} P(Y_k = l) &= F(T_l, x(t_k), \sigma) - F(T_{l-1}, x(t_k), \sigma) \\ &= \frac{1}{2}\left[\text{erf}\left(\frac{T_l - x(t_k)}{\sigma\sqrt{2}}\right) - \text{erf}\left(\frac{T_{l-1} - x(t_k)}{\sigma\sqrt{2}}\right)\right] \end{aligned} \tag{7}$$

where $1 < l < N-1$, and $x(t_k)$ denotes the exact value of the input at the kth sample. The function $x(t_k)$ depends also on the input signals. Its value can be numerically evaluated for each parameter set.

B. Cost Function

The cost function can be derived from the probability function. The overall probability is

$$P_{\text{final}} = \prod_{k=1}^{M} P(Y_k = y_k) \tag{8}$$

where y_k, $k = 1, \ldots, M$ are the measured output symbols and M is the number of output symbols. The cost function results by taking the negative logarithm of (8):

$$C_{\text{ML}}(p) = -\log P_{\text{final}} = -\sum_{k=1}^{M} \log P(Y_k = y_k). \tag{9}$$

Maximum likelihood estimation finds the minimum of the cost function $C_{\text{ML}}(p)$ as a function of p. The parameters are the transition levels T_l ($l = 1, \ldots, N-1$), variance of the noise σ^2 and parameters of the sine wave (A, B, C). It can be seen that the cost function is highly non-linear in parameters. In order to simplify the notation the parameter vector p is introduced. It contains all the parameters listed above.

$$p[1] = A, \quad p[2] = B, \quad p[3] = C, p[4] = \sigma,$$

$$p[5:N+3] = [T_1, \ldots, T_{N-1}].$$

The length of the vector p is $N+3$. Is is worth noting that $p \in \mathbb{R}^{N+3}$ but not all points in the parameter space determine a valid parameter vector. In the paper we apply the following restrictions:

- the set of transition levels is ordered, i.e.
$$T_1 \leq T_2 \leq \ldots \leq T_{N-1}$$
with $T_l \in \mathbb{R}$, $l = 1, \ldots, N-1$,
- variance of the noise is not zero, $\sigma > 0$.

These restrictions give the domain of the cost function (9):

$$C_{\text{ML}}(p) : \mathbb{R}^3 \times \mathbb{R}_+ \times \mathbb{R}^{N-1}_{\text{ord}} \mapsto \mathbb{R} \tag{10}$$

where $\mathbb{R}^{N-1}_{\text{ord}}$ denotes the ordered subset of $\mathbb{R}^{N-1}$ and the set $\mathbb{R}_+$ contains the positive real numbers.

IV. Numerical Optimization

Extreme values of the cost function (9), which is highly non-linear in parameters, is determined by a gradient descent method. The algorithm, which is presented below in details, is a special mixture of the so-called conjugate gradient method and backtracking line search [2], [7]. So it requires to have the derivative function during computation of iterative steps. First, the general method is presented, then special modifications which are necessary to handle the restrictions are introduced.

Instead of calculating the full derivative of $\nabla C_{\mathrm{ML}}(p)$ only partial derivatives are used. This is similar to the conjugate gradient method. The backtracking line search is also evaluated with using partial derivatives. Since the number of parameters is very large the whole set is divided into smaller number of subgroups. The subgroups are defined by indexes of the parameter vector p. Because of the construction of p the whole set of the indexes is

$$L = \{1, \ldots, N+3\}. \tag{11}$$

The indexes correspond to the domain of the cost function (10). In every step of the numerical minimization the set L is divided into disjoint subsets such that

$$L = L_1 \cup L_2 \cup \ldots \cup L_m$$

where $m \geq 3$. Using this notation the applied numerical method is the following:

1) Calculate a starting point p.
2) Generate the disjoint subsets of (11), $n = 1$.
3) Determine a descent direction $\Delta p_n = -\nabla_{L_n} C_{\mathrm{ML}}(p)$ based on L_n.
4) Backtracking line search.
5) Update: $p = p + \tau \Delta p$.
6) $n = n + 1$, if $n \leq m$ then goto 3.
7) If stopping criterion is not satisfied goto step 2.
8) Stop.

Assuming that $x \in \mathbb{R}^N$, $X \subset \{1, \ldots, N\}$ and $f(p) : \mathbb{R}^N \mapsto \mathbb{R}$ the notation $\nabla_X f(p)$ is defined as a special gradient vector. $\nabla_X f(p) \in \mathbb{R}^N$ and

$$\nabla_X f(p)[k] = \begin{cases} \dfrac{\partial f(p)}{\partial p[k]}, \text{ if } k \in X, \\ 0, \text{ otherwise.} \end{cases}$$

The backtracking line search is the following. Its input arguments are a descent direction Δp, $\nabla_{L_n} C_{\mathrm{ML}}(p)$ and $\alpha \in (0, 0.5)$, $\beta \in (0, 1)$. The algorithm is very simple:

1) $\tau := 1$.
2) Evaluate
$$C_{\mathrm{ML}}(p + \tau \Delta p) > C_{\mathrm{ML}}(p) + \alpha \tau \nabla_{L_n} C_{\mathrm{ML}}(p)^T \Delta p.$$
3) If it is true, then $\tau = \beta \tau$ and goto 2.
4) Stop.

The derivate function of the logarithm of the cost function (9) can be obtained as

$$\nabla_{L_n} C_{\mathrm{ML}}(p) = -\sum_{l=1}^{M} \frac{1}{P(Y_l = y_l)} \nabla_{L_n} P(Y_l = y_l).$$

The probabilities $P(Y_l = y_l)$ can be calculated using expression (5), (6) and (7). In calculation of $\nabla_{L_n} P(Y_l = y_l)$ the following observation helps: if L_n can be decomposed into two disjoint, non-empty subsets $L_n = L_{n_1} \cup L_{n_2}$ then

$$\nabla_{L_n} P(Y_l = y_l) = \nabla_{L_{n_1}} P(Y_l = y_l) + \nabla_{L_{n_2}} P(Y_l = y_l).$$

It is true because of the construction $\nabla_X f(p)$. We can conclude that $\nabla_{L_n} C_{\mathrm{ML}}(p)[k] = 0$ if $k \notin L_n$. Moreover, it can be seen that the partial derivatives with respect to the corresponding subset of parameters can be determined independently from each other.

Subsets of the parameters are listed here:

- Parameters of the excitation signal: A, B, C. The corresponding indexes are $L_1 = \{1, 2, 3\}$.. Therefore,
$$\nabla_{L_1} P(Y_l = y_l) = \begin{bmatrix} \frac{\partial P(Y_l = y_l)}{\partial p_1} & \frac{\partial P(Y_l = y_l)}{\partial p_2} & \frac{\partial P(Y_l = y_l)}{\partial p_3} & 0 & \ldots & 0 \end{bmatrix}^T.$$
Detailed calculation of the elements of $\nabla_{L_1} P(Y_l = y_l)$ can be found in appendix VII-A.
- Variance of the additive noise: σ. The corresponding index is $L_2 = \{4\}$.. Hence
$$\nabla_{L_2} P(Y_l = y_l) = \begin{bmatrix} 0 & 0 & 0 & \frac{\partial P(Y_l = y_l)}{\partial p_4} & 0 & \ldots & 0 \end{bmatrix}^T.$$
Appendix VII-B contains sub-cases of calculation of the non-zero element of $\nabla_{L_2} P(Y_l = y_l)$.
- In the case of transition levels a different approach is used. The reason is that even in the case of a 8-bit converter the number of transition levels is very high. If every set L_m where $m \geq 3$ contained only one index, the backtracking line-search would have to be evaluated N in the inner cycle of the main algorithm.
A possible way to decrease the computation time, if the number of elements in L_m ($m \geq 3$) increases. We propose a heuristic method which computes disjoint subsets of $L \backslash (L_1 \cup L_2)$ based on the currently estimated noise variance $\hat{\sigma} = p(4)$.
Denoting $\hat{\sigma}_{\mathrm{LSB}} = \hat{\sigma}/Q$ and defining $s = \mathrm{floor}(5\hat{\sigma}_{\mathrm{LSB}})$, the proposed subsets of $\{5, \ldots, N+3\}$ is
$$L_{m+2} = \{m+4, m+4+s, m+4+2s, \ldots\}$$
where $m = 1, \ldots, s$.
$$\nabla_{L_m} P(Y_l = y_l) = \begin{bmatrix} 0 & 0 & 0 & 0 & K \end{bmatrix}^T$$
where the vector $K \in \mathbb{R}^{N-1}$ and
$$K[k] = \begin{cases} \dfrac{\partial P(Y_l = y_l)}{\partial p_{k+4}}, \text{ if } k = m, m+s, \ldots, \\ 0, \text{ otherwise.} \end{cases}$$
Since determining of $K[k]$ depends on the value y_l, it contains nine sub-cases. They are listed and calculated in appendix VII-C.

A. Initial Values

Before starting the whole iterative minimizing procedure a starting value needs to be calculated. In general, one can say that for obtaining a solution of a non-linear minimization problem is difficult and in practice strongly depends on how close the starting value is to the optimum. Our experience is that using the method proposed below calculating starting values has worked fine. Like above, three sub-cases can be distinguished.

First, the initial transition levels are estimated. Since at this step no other information available transition levels of the ideal quantizer (2) is calculated.

Then a three parameter sine fitting is evaluated. Fitting is executed using the least squares (LS) method [6]. The error is in the digital domain and is defined as the difference between the observation y_k (1) and the model $x(t)$ (3) in the digital domain

$$\text{argmin}_{A_d,B_d,C_d} \sum_{m=1}^{M} (y_m - x_d(t_m))^2 .$$

where $x_d(t) = A_d \sin(2\pi f t) + B_d \cos(2\pi f t) + C_d$. The result is in the digital domain and has to transformed back to the analog domain: $A = QA_d$, $B = QB_d$ and $C = Q(C_d + V_{\text{min}}/Q)$.

In the last step of computing the initial values, variance of the noise is estimated. It is based on the paper [10]. The estimation is performed in the digital domain, using the output samples y_k, by trying to eliminate effect of the excitation signal:

$$\hat{\sigma}^2_{\text{init}} = \frac{1}{M-1} \sum_{m=1}^{M} (y_m - q(x(t_m)))^2 .$$

V. Experimental Results

In this section two examples are shown. They demonstrate that the solution of the maximum likelihood problem can be found with using the introduced method. The first one is a simulation example and then an experimental example is presented.

A. Simulated Data

Here, a 8-bit converter is assumed. During generation of the simulated data the following parameters are used: $N = 256$, $V_{\text{min}} = -2$ V, $V_{\text{max}} = +2$ V, $M = 576000$ samples, $f_s = 10$ kHz, amplitude of the sine wave was 2.001 V, its frequency 1 Hz, dc level was 0 V, std of the additive noise: 0.018 V. The ideal quantizer was modified by shifting some transition levels: $T_{10} = T_{\text{ideal},10} + 0.1Q$, $T_{11} = T_{\text{ideal},11} + 0.2Q$ and $T_{12} = T_{\text{ideal},12} + 0.3Q$.

The result can be seen in Fig. 4 where $T[k] - \hat{T}[k]$ is plotted for all transition levels. It can be seen that the absolute value of the estimation error is less than $0.06Q$. Accuracy of the estimation increases, i.e. variance of the estimated value decreases as the number of samples increases. Results w.r.t. the excitation signal and the noise are $|\sigma - \hat{\sigma}| = 0.0017Q$ and $\max_{k=1,\ldots,M} |x(t_k) - \hat{x}| = 0.0442Q$.

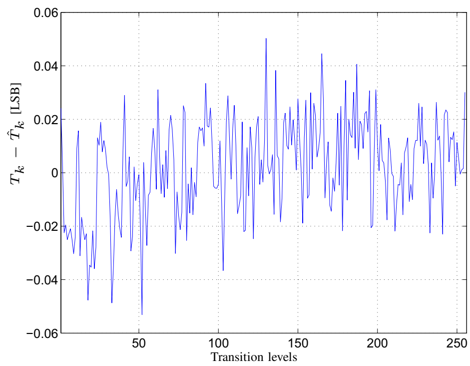

Fig. 4. Difference of the estimated and the real quantization levels of a simulated 8-bit converter.

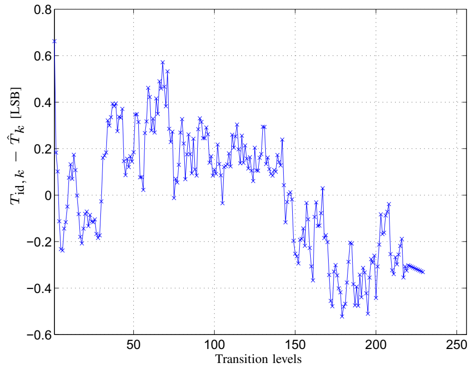

Fig. 5. INL characteristic of the measured AD9245. Difference of the estimated and the ideal quantization levels. $T_{\text{id},k} = \hat{T}_1 + kQ_{\text{id}}$ where $Q_{\text{id}} = (\hat{T}_N - \hat{T}_1)/N$.

B. Measurement Data

A trial measurement was performed to test what is the output of the proposed algorithm. A converter AD9245, a product of Analog Devices, is used in the measurement setup. The converter was on a custom made PCI board which contains also an analog anti-alias filter. The power supply was connected to the PCI bus of the computer, therefore significant noise was injected into the analog components. Because of the limited bandwidth of the PCI bus, most significant 10 bits were used. Therefore, here we characterized the equivalent ADC of the full device. Parameters of the measurement: $f_s = 80$ MHz, $f_{\text{sine}} = 20$ MHz, dc level = 0 V, and amplitude of the sine wave was -40 dBm (generator: Agilent 33250A), $M = 102400$ samples were recorded.

The result can be seen in Fig. 5. Estimated noise std is 2.8549 LSB. This demonstration example shows that the proposed algorithm can indeed determine the INL from measured data.

VI. Conclusions

In this paper it has been shown that maximum likelihood estimation of the parameters of an ADC is possible using sine wave testing. The numerical implementation has been described, and the algorithm has been verified using simulation and measurement. The error analysis of the estimates (Cramer-Rao bound) will be visited next.

VII. Appendix

A. Differentiation w.r.t. A, B, C

If $m \in L_1$, then the corresponding derivatives are

$$\frac{\partial P(Y_l = 0)}{\partial p_m} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(T_1 - x(t_l))^2}{2\sigma^2}} \left(-\frac{\partial x(t_l)}{\partial p_m} \right),$$

$$\frac{\partial P(Y_l = N-1)}{\partial p_m} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(T_{N-1} - x(t_l))^2}{2\sigma^2}} \left(\frac{\partial x(t_l)}{\partial p_m} \right),$$

and if $k \neq 0$ and $k \neq N-1$

$$\frac{\partial P(Y_l = k)}{\partial p_m} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(T_k - x(t_l))^2}{2\sigma^2}} \left(-\frac{\partial x(t_l)}{\partial p_m} \right) - \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(T_{k-1} - x(t_l))^2}{2\sigma^2}} \left(-\frac{\partial x(t_l)}{\partial p_m} \right)$$

where

$$\frac{\partial x(t_l)}{\partial p_m} = \begin{cases} \sin(2\pi t_l), \text{ if } m = 1, \\ \cos(2\pi t_l), \text{ if } m = 2, \\ 1, \text{ if } m = 3. \end{cases}$$

B. Differentiation w.r.t. σ

If $m \in L_2$, i.e. $m = 4$, $p_4 = \sigma$ then the corresponding derivatives are

$$\frac{\partial P(Y_l = 0)}{\partial p_4} = \frac{-1}{\sqrt{2\pi}} e^{-\frac{(T_1 - x(t_l))^2}{2\sigma^2}} \frac{(T_1 - x(t_l))}{\sigma^2},$$

$$\frac{\partial P(Y_l = N-1)}{\partial p_4} = \frac{-1}{\sqrt{2\pi}} e^{-\frac{(T_{N-1} - x(t_l))^2}{2\sigma^2}} \frac{(T_{N-1} - x(t_l))}{\sigma^2},$$

and; if $Y_l \neq 0$ and $Y_l \neq N$

$$\frac{\partial P(Y_l = k)}{\partial \sigma} = \frac{-1}{\sqrt{2\pi}} e^{-\frac{(T_k - x(t_l))^2}{2\sigma^2}} \frac{(T_k - x(t_l))}{\sigma^2} + \frac{1}{\sqrt{2\pi}} e^{-\frac{(T_{k-1} - x(t_l))^2}{2\sigma^2}} \frac{(T_{k-1} - x(t_l))}{\sigma^2}.$$

C. Differentiation w.r.t. transition levels

Derivatives of the probability functions (5), (6) and (7) w.r.t. the first and the last transition levels differ from the case of other transition levels. Therefore, three cases have to be considered.

In the first case differentiation w.r.t. T_1 is performed. From (5) we have

$$\frac{\partial P(Y_l = 0)}{\partial T_1} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(T_1 - x(t_l))^2}{2\sigma^2}}.$$

Since $N > 3$, (6) does not contain T_1, hence

$$\frac{\partial P(Y_l = N-1)}{\partial T_1} = 0.$$

In the general case

$$\frac{\partial P(Y_l = k)}{\partial T_1} = \begin{cases} -\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(T_1 - x(t_l))^2}{2\sigma^2}}, \text{ if } k = 2, \\ 0, \text{ otherwise.} \end{cases}$$

Secondly, we known that (5) does not depend on T_{N-1}, so

$$\frac{\partial P(Y_l = 0)}{\partial T_{N-1}} = 0.$$

From (6), the derivative function is

$$\frac{\partial P(Y_l = N-1)}{\partial T_{N-1}} = -\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(T_{N-1} - x(t_l))^2}{2\sigma^2}}.$$

And if $y_l \neq 0$ and $y_l \neq N-1$ then

$$\frac{\partial P(Y_l = k)}{\partial T_{N-1}} = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(T_{N-1} - x(t_l))^2}{2\sigma^2}}, \text{ if } k = N-2, \\ 0, \text{ otherwise.} \end{cases}$$

And in the third case we assume that $k \neq 1$ and $k \neq N-1$. (5) and (6) give that

$$\frac{\partial P(Y_l = 0)}{\partial T_k} = 0 \text{ and } \frac{\partial P(Y_l = N-1)}{\partial T_k} = 0.$$

Generally, from (7) we have

$$\frac{\partial P(Y_l = m)}{\partial T_k} = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(T_k - x(t_l))^2}{2\sigma^2}}, \text{ if } m = k \\ \frac{-1}{\sqrt{2\pi}\sigma} e^{-\frac{(T_{k-1} - x(t_l))^2}{2\sigma^2}}, \text{ if } m = k+1. \end{cases}$$

References

[1] IEEE Standard 1241-2000, *Standard for Terminology and Test Methods for Analog-to-Digital Converters.*, The Institute of Electrical and Electronics Engineers.

[2] Press, W. H., B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing.*, Cambridge University Press, Cambridge, 1986.

[3] J. J. Blair, *Sine-fitting software for IEEE standards 1057 and 1241*, in Proc. of the 16th IEEE Instrumentation and Measurement Technology Conference, IMTC/99, Venice, Italy, May 1999, vol. 3, pp. 1504-1506.

[4] Kollár and J. J. Blair, *Improved Determination of the Best Fitting Sine Wave in ADC Testing*, IEEE Trans. on Instrumentation and Measurement, Vol. 54 No. 5, pp. 1978-83, 2005.

[5] B. Widrow, and I. Kollár, *Quantization Noise.: Roundoff Error in Digital Computation, Signal Processing, Control, and Communications.*, Cambridge University Press, Cambridge, 2008.

[6] B. Fodor and I. Kollár, *ADC Testing with Verification*, IEEE Trans. on Instrumentation and Measurement, Vol. 57 No. 12, pp. 2762-2768, 2008.

[7] Stephen Boyd, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.

[8] A. Di Nisio, G. Cavone, N. Giaquinto, L. Fabbiano, M. Savino, *Improved Static Testing of A/D Converters for DC Measurements*, Instrumentation and Measurement Technology Conference - IMTC 2007, Warsaw, Poland, May 1-3, 2007.

[9] Attilio Di Nisio, Laura Fabbiano, Nicola Giaquinto, Mario Savino, *Statistical properties of an ML estimator for static ADC testing*, 12th Workshop on ADC Modelling and Testing, Iasi, Romania, September 18-22, 2007. Paper F154.

[10] Antonio Moschitta, Paolo Carbone, *Noise Parameter Estimation From Quantized Data*, IEEE Trans. on Instrumentation and Measurement, Vol. 56 No. 3, pp. 736-742, 2007.