

Initialization of the BIMBO self-test method using binary inputs and outputs

Jérôme Juillard, Eric Colinet

Abstract— This paper deals with the initialization of the BIMBO method, a deterministic identification method based on binary observation, for the (self-) test of integrated electronic and electromechanical systems, such as MEMS. Finding an adequate starting point for the parameter estimation algorithm may be crucial, depending on the chosen model parameterization. We show how this starting point may be obtained using only binary inputs and outputs and a few straightforward calculations. The practical implementation of this method only requires a one-bit digital-to-analog converter (DAC) and a one-bit analog-to-digital converter (ADC). This makes the proposed approach very amenable to integration and leads to no additional cost compared to the BIMBO method. We describe the method from a theoretical point of view, discuss its implementation and illustrate it in some idealized cases.

Index Terms— Parameter estimation, self-test, binary data processing, microsystems, integrated electronics.

I. INTRODUCTION

As the characteristic dimensions of electronic devices and systems become smaller, the performances of these devices become more dispersed. Typical sources of dispersion are variations in the fabrication process or changes in the operating conditions, such as temperature. As a consequence, the capacity to self-test and self-adjust is a very desirable feature for microfabricated devices. However, most existing system identification methods [1-2] do not easily scale down to the micro- or nano-world. This is due to the fact that these methods rely on a high-resolution digital measurement of the system output. Their integration requires the implementation of a high-resolution analog-to-digital converter (ADCs) and, thus, results in longer design times as well as larger silicon areas. These issues have been addressed in [3], where the Basic Identification Method using Binary Observations (BIMBO) is presented and its basic properties are established. The principle of BIMBO is the following (Fig. 1):

1. generate a spectrally rich signal u_k , such as a white noise.
2. measure the system's output y_k using a one bit

ADC, i.e. measure $s_k = \text{sign}(y_k)$.

3. using a discrete parametric model of the unknown system, compute the estimated time response \hat{y}_k and $\hat{s}_k = \text{sign}(\hat{y}_k)$.
4. adjust the parametric model so that a correct estimation \hat{y}_k is produced.

Because of its very low integration cost and its simple practical implementation, it provides designers with a very simple tool for the built-in self-test of micro-electromechanical systems (MEMS) and microelectronic devices. We showed in [3] that the criterion used in the parameter estimation routine (3rd and 4th step) becomes convex in a neighbourhood of the minimum, when the number of samples goes to infinity. As a consequence, it is possible to find the optimal system parameters using the gradient algorithm described in [3], or any other convex optimization method, provided the initial parameters are correctly chosen. Depending on the chosen parameterization and on the regularity of the system, an inadequate starting point may result in the non-convergence of the optimization algorithm. For example, it is very difficult to identify the parameters of an ARX model of a system with a large quality factor without sufficient prior information.

In the present paper, we describe a method for finding these initial parameters in two cases:

- the input signal is a binary white noise.
- the input signal is a gaussian white noise.

The former case is of special practical interest because its practical implementation requires only a one-bit digital-to-analog converter (DAC).

The rest of the paper is organized as follows: first, we describe the initialization method when a binary excitation is used (section II) and when a gaussian excitation is used (section III). In section IV, we discuss the validity of the proposed approach from a mathematical standpoint and establish the statistical properties of our approach. In section V, we describe the practical implementation of the method. Finally, in section VI, we use simulated examples to illustrate our method and we discuss and explain the results.

II. INITIALIZATION IN THE CASE OF A BINARY INPUT

Consider the setup of Fig. 2, where u_k is a symmetric binary white noise with mean m_u and standard deviation 1, so that:

Manuscript received on
J.Juillard is with SUPELEC, Dpt. of Signal Processing and Electronic Systems, France.
E. Colinet is with CEA-LETI, Minattec, France.

$$w_k = u_k - m_u = \text{sign}(u_k - m_u) \quad (1)$$

and

$$\text{P}(w_k = 1) = \text{P}(w_k = -1) = \frac{1}{2}. \quad (2)$$

We suppose that b_k is a noise (not necessarily white), with zero mean and standard deviation σ_b . We also suppose the following two quantities are defined:

$$\Sigma_h = \sum_{l=0}^{\infty} h_l, \quad (3)$$

$$\sigma_h^2 = \sum_{l=0}^{\infty} h_l^2, \quad (4)$$

where the h_l are the coefficients of the system's impulse response, i.e. that the system under consideration is stable.

The cross-covariance function of u_k and s_k may be written as:

$$\begin{aligned} C_{su}(p) &= \text{E}(s_k u_{k-p}) - \text{E}(s_k) \text{E}(u_k) \\ &= \text{E}(s_k (u_{k-p} - m_u)) = \text{E}(s_k w_{k-p}), \end{aligned} \quad (5)$$

where $\text{E}(\cdot)$ denotes the mathematical expectation. Using (1), we have:

$$\begin{aligned} s_k w_{k-p} &= \text{sign}(y_k) w_{k-p} \\ &= \text{sign}(y_k) \text{sign}(w_{k-p}) = \text{sign}(y_k w_{k-p}). \end{aligned} \quad (6)$$

Equation (5) can then be reformulated in terms of probabilities as:

$$\begin{aligned} C_{su}(p) &= \text{P}(y_k w_{k-p} > 0) - \text{P}(y_k w_{k-p} \leq 0) \\ &= 2\text{P}(y_k w_{k-p} > 0) - 1. \end{aligned} \quad (7)$$

Now, we have:

$$\begin{aligned} y_k &= b_k + \sum_{l=0}^{\infty} h_l u_{k-l} \\ &= b_k + m_u \Sigma_h + \sum_{l=0, l \neq p}^{\infty} h_l w_{k-l} + h_p w_{k-p} \end{aligned} \quad (8)$$

and thus:

$$y_k w_{k-p} = b_k w_{k-p} + m_u \Sigma_h w_{k-p} + \sum_{l=0, l \neq p}^{\infty} h_l w_{k-l} w_{k-p} + h_p. \quad (9)$$

Now, the first term on the right-hand side is a random variable with zero mean and standard deviation σ_b , the second term is a Bernoulli random variable with zero mean and standard deviation $m_u \Sigma_h$ and the third term is a weighted sum of Bernoulli random variables with zero means and standard deviations 1. All these variables are independent, thus, provided the conditions of the central limit theorem (CLT) [4] are met, $y_k w_{k-p}$ can be approximated by a gaussian random variable with mean h_p and standard deviation σ_p , where:

$$\sigma_p^2 = \sigma_b^2 + m_u^2 \Sigma_h^2 + \sigma_h^2 - h_p^2 = \sigma^2 - h_p^2. \quad (10)$$

The probability that appears in (7) is then:

$$\text{P}(y_k w_{k-p} > 0) = \frac{1}{2} \left(1 + \text{erf} \left(\frac{h_p}{\sigma_p \sqrt{2}} \right) \right), \quad (11)$$

where erf is the error function, so that

$$C_{su}(p) = \text{erf} \left(\frac{h_p}{\sigma_p \sqrt{2}} \right). \quad (12)$$

Finally, letting erf^{-1} denote the inverse error function, we have:

$$h_p = \sigma \sqrt{2} \frac{\text{erf}^{-1}(C_{su}(p))}{\sqrt{1 + 2(\text{erf}^{-1}(C_{su}(p)))^2}}. \quad (13)$$

An estimation of the system's impulse response can then be found by replacing in (13) C_{su} by an empirical estimation, for example:

$$\hat{C}_{su}(p) = \frac{1}{N} \sum_{k=1}^N s_k u_{k-p}. \quad (14)$$

The statistical properties of this estimation are discussed in section IV, along with the applicability of the CLT.

III. INITIALIZATION IN THE CASE OF A GAUSSIAN INPUT

When u_k is a gaussian white noise, it is difficult to apply the approach described in section II unless m_u equals 0. Let us then suppose that it is the case and let v_k be the sign of u_k .

Then, it is simple to show that:

$$C_{sv}(p) = 2\text{P}(y_k u_{k-p} > 0) - 1. \quad (15)$$

The joint probability density function of y_k and u_{k-p} can be written:

$$f(y, u) = \frac{1}{2\pi\sigma' \sqrt{1-r^2}} \exp \left(\frac{-1}{2(1-r^2)} \left(\frac{y^2}{\sigma'^2} + u^2 - \frac{2ryu}{\sigma'} \right) \right), \quad (16)$$

where it is supposed that the input has variance 1,

$$r = \frac{h_p}{\sigma'}, \quad (17)$$

and

$$\sigma'^2 = \sigma_b^2 + \sigma_h^2. \quad (18)$$

One can then integrate (16) on $\mathfrak{R}^+ \times \mathfrak{R}^+$ to find a closed-form expression of (15). This leads to [5]:

$$h_p = \sigma' \sin \left(\frac{\pi}{2} C_{sv}(p) \right) \quad (19)$$

This result can also be used to initialize the BIMBO algorithm. However, its practical implementation requires a high-resolution DAC and, thus, it is less amenable to integration than the method proposed in section II. On the other hand, this approach can be applied regardless of the properties of the unknown system.

IV. VALIDITY OF THE PROPOSED APPROACH

In this section, we look more deeply into the conditions of applicability of the CLT and we establish the statistical

properties of (14) in the case of binary inputs and outputs.

A. Applicability of the CLT

Intuitively, the applicability conditions of the CLT are met in many cases, provided none of the random variables appearing in (9) dominates too much. In case the input noise has zero mean and the measurement noise is gaussian, this reduces to the condition that the system's impulse response does not vanish quickly.

More accurately, a sufficient condition for applying the CLT is the Lindeberg condition [4]. Letting Q_l be the distribution

of $X_l = h_l w_{k-l} w_{k-p}$ and $s_n^2 = \sum_{l=0}^n h_l^2$, one can show that the

distribution of

$$\mu_n = \frac{1}{s_n} \sum_{l=0}^n X_l \quad (20)$$

converges weakly to a unit normal distribution provided the following condition is met:

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{l=0}^n \int_{\{|x| \geq s_n \varepsilon\}} x^2 Q_l(dx) = 0. \quad (21)$$

It is simple to see that, in the case of Bernoulli random variables, the integrals appearing in (21) are all zero provided

$$\varepsilon > \varepsilon_{\max} = \frac{\max |h_l|}{\sigma_h} = \frac{\|h\|_{\infty}}{\|h\|_2}. \quad (22)$$

Thus, the only way to make sure that the Lindeberg condition is strictly met is to have σ_h go to infinity. However, we shall consider, in practice, that the distribution of μ_n is close to a unit normal distribution if:

$$\frac{\|h\|_{\infty}}{\|h\|_2} \ll 1. \quad (23)$$

This result can easily be extended to the cases when $m_u \neq 0$ or when the measurement noise is not gaussian (for example, it could be a binary dither signal).

B. Statistical properties of (14)

Before moving to the practical implementation of the approach, we must establish the bias and the variance of the cross-covariance estimator (14). For the sake of simplicity, we suppose $m_u = 0$. It is clear that:

$$E(\hat{C}_{su}(p)) = \frac{1}{N} \sum_{k=1}^N E(s_k u_{k-p}) = C_{su}(p). \quad (24)$$

As for the variance of the estimator, we have:

$$\text{var}(\hat{C}_{su}) = E\left(\left(\hat{C}_{su} - C_{su}\right)^2\right) = E(\hat{C}_{su}^2) - C_{su}^2. \quad (25)$$

The first term on the right-hand side can then be written, using (1):

$$\begin{aligned} E(\hat{C}_{su}^2) &= \frac{1}{N^2} \sum_{k=1}^N \sum_{l=1}^N E(s_k u_{k-p} s_l u_{l-p}) \\ &= \frac{2}{N^2} \sum_{k=1}^N \sum_{l=1}^N P(y_k u_{k-p} y_l u_{l-p} > 0) - 1 \end{aligned} \quad (26)$$

Finding a closed-form expression for (26) is feasible but tedious, except in the case when p is large compared to the correlation length of h , i.e.

$$C_{su}(p) \approx 0. \quad (27)$$

It is then simple to verify that:

$$P(y_k u_{k-p} y_l u_{l-p} > 0) \approx \frac{1}{2} + \frac{1}{2} \delta_{kl}, \quad (28)$$

where δ_{kl} is the Kronecker symbol. The variance of (14) can then be expressed as:

$$\text{var}(\hat{C}_{su}) \approx \frac{1}{N}. \quad (29)$$

Equation (29) is valid in the case of any white and centered input u_k .

V. PRACTICAL IMPLEMENTATION

Once an estimation of the cross-covariance between u_k and s_k has been calculated, equation (13) or (19) must be used, depending on whether the inputs are gaussian or binary white noise. In either case, the relation between the coefficients of the impulse response and the cross-covariance function is almost linear (Fig. 3). As a consequence, it may be approximated with sufficient accuracy with a third- or fifth-order Taylor expansion.

We know that, theoretically, when the impulse response vanishes, the variance of the cross-covariance estimator (14) is given by (29). Therefore, it should be possible to truncate the estimated impulse response to an adequate length P_{\max} by dropping the coefficients for which the calculated value of $\hat{C}_{su}(p)$ is below a certain threshold value, say $2/\sqrt{N}$.

Now, consider that σ_b is known and different from zero. We can rewrite (13) or (19) as:

$$\frac{h_p}{\sigma} = F(C_{su}(p)) = F_p. \quad (30)$$

Taking the square of (30) and summing for $p \leq P_{\max}$ leads to:

$$\frac{1}{\sigma^2} \sum_{p=1}^{P_{\max}} h_p^2 = \sum_{p=1}^{P_{\max}} F_p^2, \quad (31)$$

and thus, σ_h^2 may be empirically estimated by:

$$\hat{\sigma}_h^2 = \sigma_b^2 \frac{\sum_{p=1}^{P_{\max}} \hat{F}_p^2}{1 - \sum_{p=1}^{P_{\max}} \hat{F}_p^2}, \quad (32)$$

where \hat{F}_p stands for $F(\hat{C}_{su}(p))$. Note that, as σ_b increases,

the correlation between u_k and s_k tends to decrease and, ultimately, drops below the threshold value $2/\sqrt{N}$, leading to a poor evaluation of P_{\max} . On the other hand, it is clearly impossible to estimate the value of σ_h when σ_b is equal to zero. As a consequence, there should exist a noise level leading to an optimal estimation of σ_h .

If the BIMBO method is to be used to estimate the coefficients of the impulse response, those found with the approach described in the previous sections can directly be used to initialize the optimization algorithm. Otherwise, a reduced-order model (as an ARX model, for instance) must be fitted to the impulse response estimated with (13) or (19) and the resulting model parameters must then be used as the initial parameters of the optimization algorithm.

Some examples are given in the next section.

VI. ILLUSTRATION

In this section, the initialization method described in the previous sections is illustrated in an idealized case. We consider a discrete system, as shown in Fig. 2, where u_k is a symmetric binary white noise with mean 0 and standard deviation 1, so that (13) should apply. $H(z^{-1})$ is chosen as the discrete transfer function of a sampled-and-held second-order system with natural pulsation $\omega_0=1 \text{ rad}\cdot\text{s}^{-1}$ and damping factor ξ . The sampling period T_s is equal to 1s. The static gain of $H(z^{-1})$ is adjusted so that $\sigma_h=1$.

In the absence of measurement noise or of dither, some computer experiments give the results shown in Figs. 4 and 5. By comparing these two figures, we verify that, for a given sampling period, the distribution of y_k becomes more gaussian as the quality factor increases. These simulations are repeated a large number of times, with different trials of u_k and a fixed number of samples, so that we can obtain an empirical estimate of the variance of \hat{h}_p . Using (13), (29) and supposing $\sigma=1$, we have, in theory:

$$\text{var}(\hat{h}_p) \approx \frac{\pi}{2N}. \quad (33)$$

Figs. 4 and 5 show that this approximation is valid for both systems, provided, as mentioned in section IV-B, that p is large compared to the correlation length of h . We also observe that, as expected, this approximation gets better as the quality factor of the system increases. These results validate the practical implementation proposed in section V.

In order to test the validity of (32), the simulations are repeated a large number of times for several values of σ_b . We define the total error on the estimation of σ_h^2 as:

$$\left(\text{E}(\hat{\sigma}_h^2 - \sigma_h^2)^2 + \text{var}(\hat{\sigma}_h^2) \right)^{1/2}. \quad (34)$$

This quantity can be empirically estimated as a function of

σ_b , knowing $\sigma_h=1$. We see in Fig. 6 that, as expected, there exists an optimal noise level for the estimation of σ_h . It is clear that this ‘‘stochastic resonance’’ phenomenon is sharper for the system with the smaller damping factor. The explanation for this is that adding a gaussian white noise at the comparator input improves the gaussianity of y_k . Thus, in the case of a system with a low quality factor or, more generally, a system that does not meet requirement (23), adding a gaussian dither at the comparator input results in a large improvement in the validity of (13) and, therefore, in the accuracy of the estimation of the impulse response. On the other hand, if (23) is met, there is only little improvement to be expected. When σ_b becomes much larger than σ_h , the correlation drops to zero and leads to a poor estimation of σ_h , regardless of the validity of (23). These arguments are also valid in the case of non-gaussian dither: there usually exists an optimal value of σ_b for which the error on σ_h is minimized and this value approximately corresponds to a peak in the gaussianity of y_k , i.e. in a value of kurtosis close to 3 (Fig. 7). Note that this phenomenon stacks with the errors due to the finite number of samples.

VII. CONCLUSION

In this paper, we have shown how the BIMBO self-test method could be initialized in a simple way, without any additional implementation cost. The proposed approach, based on probabilistic considerations, such as whiteness assumptions, is less general in scope than BIMBO but the computational efforts it requires are very small, making it eligible as a standalone self-test method. We have discussed the validity of this approach from a mathematical standpoint and we have described how it could be implemented. Some results based on simulations were given. They were found to be in good agreement with the predicted behaviour of the method. Stochastic resonance phenomena were observed and explained in the presence of a dithering signal at the ADC input.

ACKNOWLEDGMENT

The authors would like to thank Julien Bect, Pascal Bianchi and Walid Hachem for taking time to comment and discuss the contents of this paper.

REFERENCES

- [1] Ljung L., ‘‘System identification - theory for the user’’, Prentice Hall, 1999
- [2] Walter E., Pronzato L., ‘‘Identification of parametric models from experimental data’’, Springer, 1997
- [3] Colinet E., Juillard J., *An identification method based on binary observations*, submitted to IEEE Transactions on Automatic Control, 2006
- [4] Taylor J. C., ‘‘An introduction to measure and probability’’, Springer, New York, 1997
- [5] Papoulis A., Pillai U., ‘‘Probability, random variables and stochastic processes’’, McGraw-Hill, New York, 2002

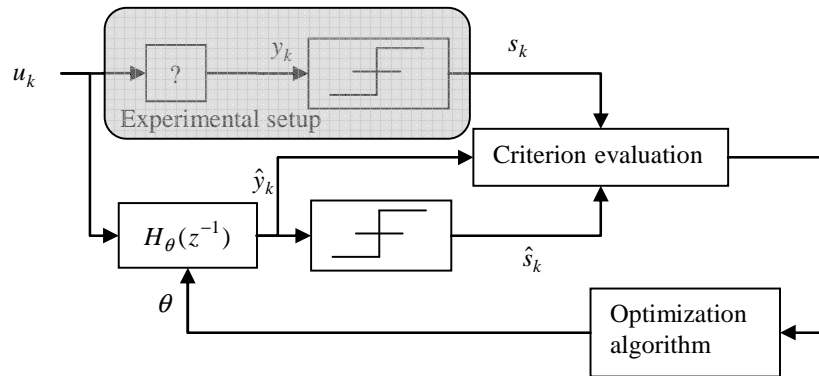


Fig. 1 - Block diagram of the BIMBO method. A parametric model of the real system is built so as to maximize the similarity between s_k and \hat{s}_k .

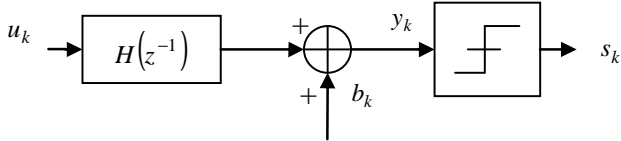


Fig. 2 - Notations.

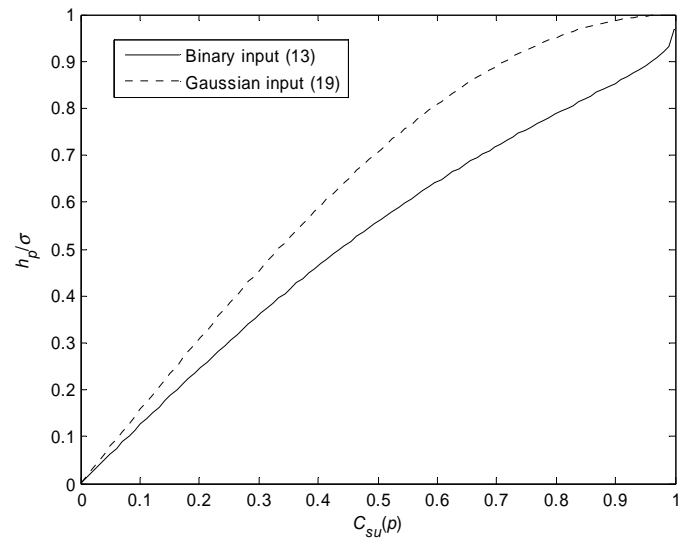


Fig. 3 - $\frac{h_p}{\sigma}$ versus $C_{su}(p)$ for a binary input and for a gaussian input.

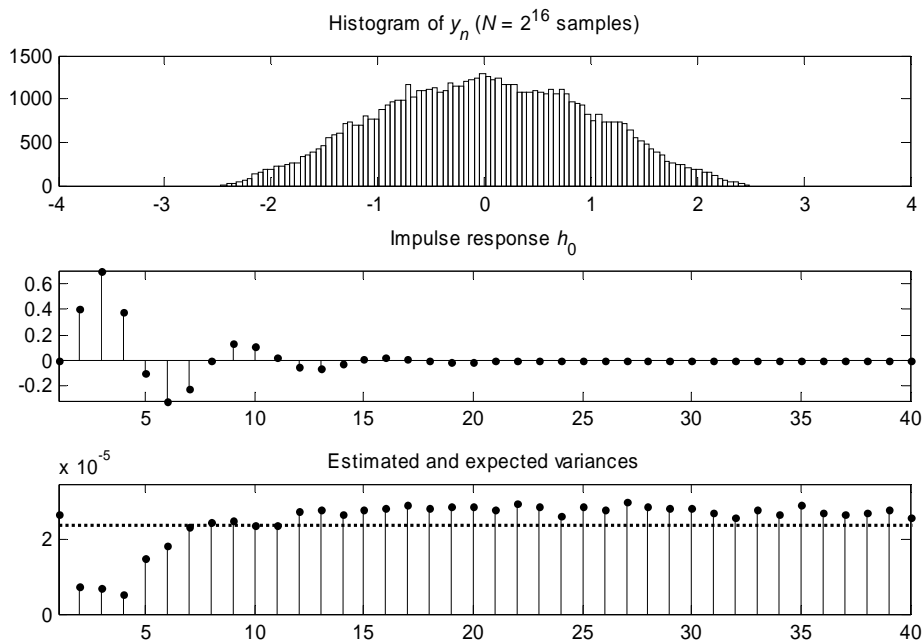


Fig. 4 - Simulation results for $\xi=0.5$ and $N=2^{16}$. The histogram of the output of the simulated system is shown in the upper box. In the lower box, we compare the variance of \hat{h}_p (stem plot) to the theoretical value obtained with (33) (dotted line), based on the results of 500 simulations.

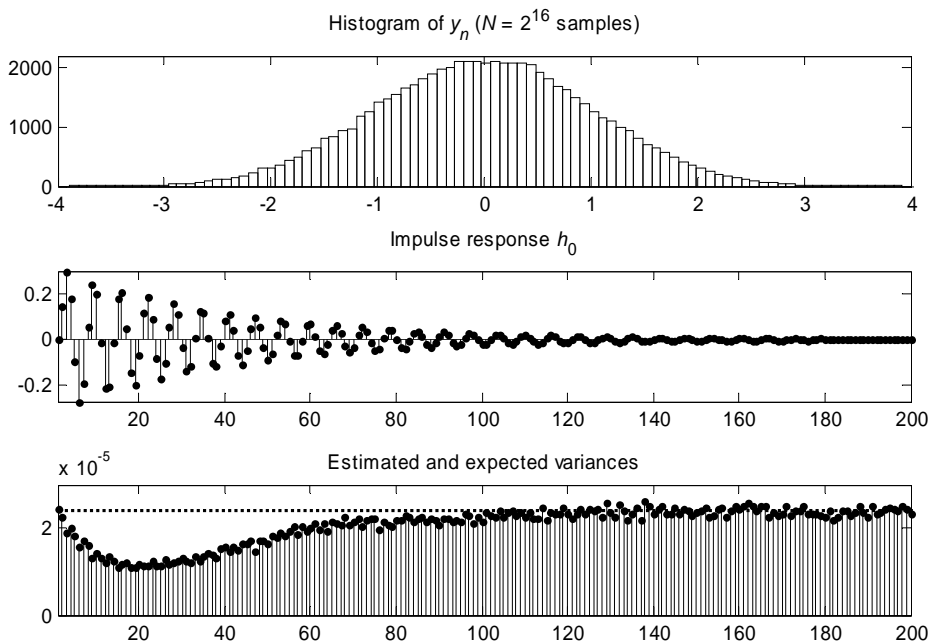


Fig. 5 - Simulation results for $\xi=0.05$ and $N=2^{16}$. The histogram of the output of the simulated system is shown in the upper box: it is clearly more gaussian-shaped than the one in Fig. 4. In the lower box, we compare the variance of \hat{h}_p (stem plot) to the theoretical value obtained with (33) (dotted line), based on the results of 500 simulations.

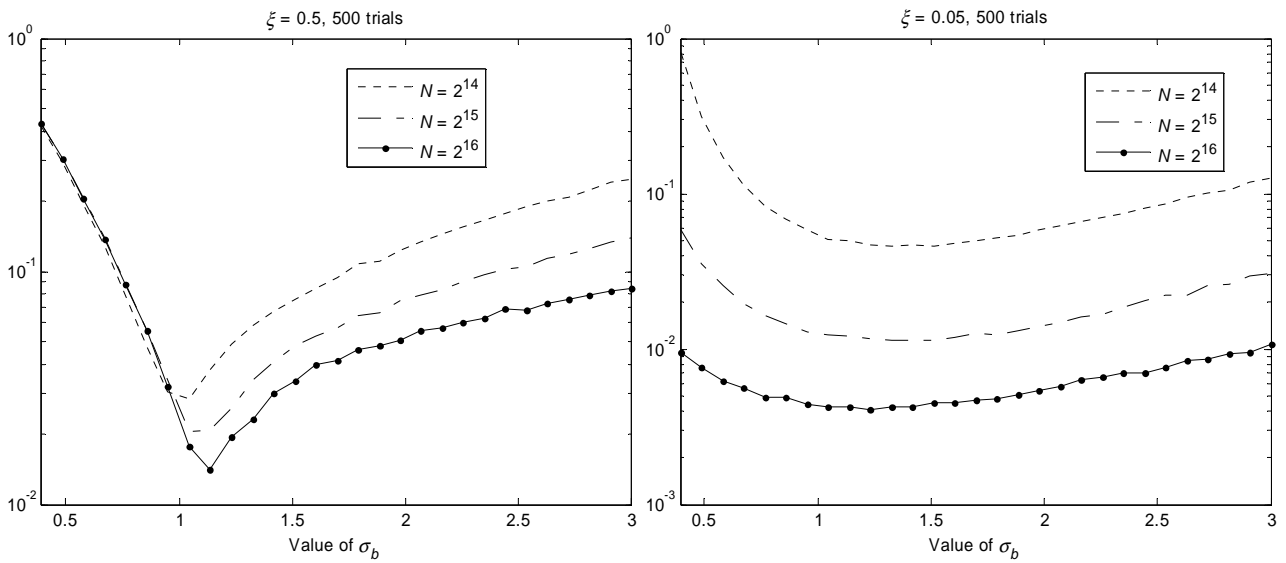


Fig. 6 - Plot of the total error on the estimation of σ_h^2 , as defined in (34), versus the known value of σ_b , for $\xi=0.5$ (left) and $\xi=0.05$ (right).

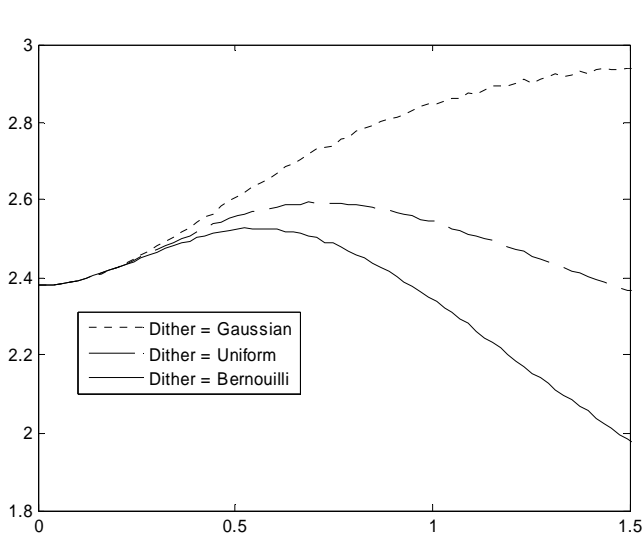


Fig. 7 - Plot of the kurtosis of y_k (for $\xi=0.5$) versus the value of σ_b , for different types of dither. There usually exists a value of σ_b for which the gaussianity (kurtosis = 3) of y_k is maximal.

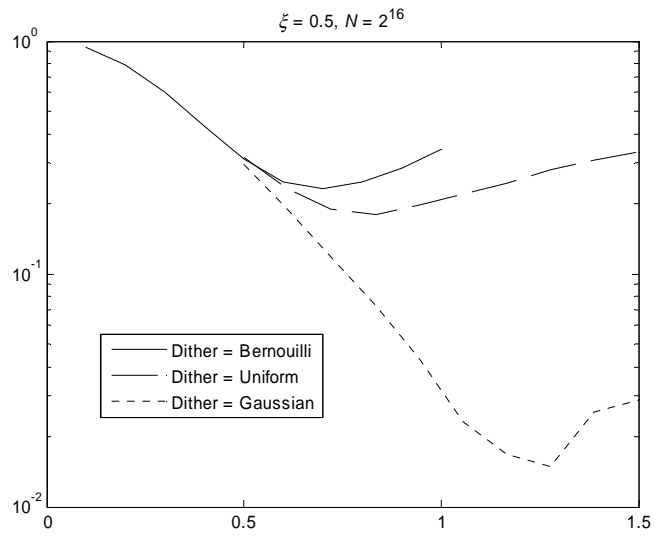


Fig. 8 - Plot of the total error on the estimation of σ_h^2 , as defined in (34), versus the value of σ_b , for $\xi=0.5, N = 2^{16}$ and different types of dither.