ANZIAM J. **67, e14** (2025), 1–10
doi:10.1017/S1446181125000070

ESTIMATION OF MEAN SQUARED ERRORS OF NON-BINARY A/D ENCODERS THROUGH FREDHOLM DETERMINANTS OF PIECEWISE-LINEAR TRANSFORMATIONS II: GENERAL CASE

KATSUTOSHI SHINOHARA[1]

(Received 4 September, 2024; accepted 16 December, 2024)

Abstract

We conduct a theoretical analysis of the performance of β-encoders. The β-encoders are A/D (analogue-to-digital) encoders, the design of which is based on the expansion of real numbers with noninteger radix. For the practical use of such encoders, it is important to have theoretical upper bounds of their errors. We investigate the generating function of the Perron–Frobenius operator of the corresponding one-dimensional map and deduce the invariant measure of it. Using this, we derive an approximate value of the upper bound of the mean squared error of the quantization process of such encoders. We also discuss the results from a numerical viewpoint.

2020 *Mathematics subject classification*: primary 94C05; secondary 37A44, 37N99.

Keywords and phrases: β-expansion, Perron–Frobenius operator, analogue-to-digital converter.

1. Introduction

Beta encoders (written as β-encoders in the following) are a new kind of A/D (analogue-to-digital) encoder based on the expansion of real numbers with noninteger radix [1–3]. They have special properties that are more advantageous than conventional ones from the engineering viewpoint, such as the auto-correction property and robustness against fluctuation of environmental conditions (see [6, 8] for more information about the implementation).

As such, it is an important problem to analyse the performance of such encoders theoretically. Makino et al. [4] discussed the upper bound of the mean squared error (MSE) of the quantization process by β-encoders. Although the problem is fairly straightforward for conventional binary encoders, for β-encoders it turns out to be a

[1]Hitotsubashi University, Graduate School of Commerce and Management, Naka 2-1, Kunitachi, Tokyo, Japan; ka.shinohara@r.hit-u.ac.jp
© The Author(s), 2025. Published by Cambridge University Press on behalf of Australian Mathematical Publishing Association Inc.

https://doi.org/10.1017/S1446181125000070 Published online by Cambridge University Press

complicated problem due to the nonintegrality of the base β, together with the chaotic nature of the β-map, which describes the encoding process.

Makino et al. obtained an upper bound of the MSE based on the concept called *segments* (see [4]). The results are rigorous and satisfactory for realistic use of β-encoders. However, some parts of their arguments rely on brute-force calculations by computers and, as such, their results have some limitations on the validity of the upper bound.

The same problem was discussed with a different idea by Kobayashi and the first author [7]. More precisely, to analyse the combinatorics of the segments, instead of doing the direct calculations, we used the spectral analysis of the Perron–Frobenius operators of β-maps. This technique has the advantage that the results are free from the limitations in [4].

This paper is a sequel to [7]. In the previous paper, to demonstrate the power of the spectral method more clearly, we put some assumptions to the β-encoders that we analysed. In this paper, we explain how the technique can be extended to the general cases. Let us briefly discuss the novelty of this paper. To analyse the MSE of the β-encoder, we need to understand the growth of the population of segments. For this purpose, we investigate the generating function of the characteristic function of segments by means of the renewal equation. This is done by carefully investigating the behaviour of the point of discontinuity under the iteration of the β-map, assuming that the point of discontinuity is located at a convenient position. In general cases, we do not have the assumption and the direct application of the same strategy does not seem to work. We circumvent this problem by introducing another convenient characteristic function ψ_α (see Section 3). These are simple functions whose linear combination gives the segments. Thus, their understanding tells us about the MSE of the encoders. Also, they behave nicely under the iteration of the Perron–Frobenius operator, and their analysis is simpler than dealing with segments directly. The use of ψ_α is the novel and important ingredient for carrying out the spectral analysis in the general cases.

Please note that this is a continuation of the paper [7]. Also, as this paper is intended for a special edition, the number of pages is limited. As such, we do not repeat detailed explanations of basic concepts that appear in the previous paper. We recommend the readers to read [7] first.

2. Description of the problem

In this section, we formulate the problem (see [4, 7] for more information).

Let $\beta \in (1, 2)$. Given a real number $x \in (0, 1)$, we are interested in converting it to a 0-1 sequence $(e_i(x))_{i\geq 1}$ (where $e_i(x) = 0$ or 1) such that

$$x = (\beta - 1) \sum_{i=1}^{\infty} \frac{e_i(x)}{\beta^i}$$

https://doi.org/10.1017/S1446181125000070 Published online by Cambridge University Press

holds. This expression is called the *β-expansion* of x. The coefficient $(\beta - 1)$ appears so that the expansion $e_i(x) \equiv 1$ corresponds to the real number 1. Note that, given x, the choice of $(e_i(x))$ is not unique in general.

Let us explain how the β-encoders transform signals into the digits with the base β. Let $\nu \in [1 - 1/\beta, 1/\beta]$. Suppose we have a circuit that converts a given signal of magnitude $x \in [0, 1]$ into $T(x)$, where

$$T(x) = \begin{cases} \beta x & (x \le \nu) \\ \beta(x-1) + 1 & (x > \nu). \end{cases}$$

Note that, as $T : [0, 1] \to [0, 1]$, we have $T^n(x) \in [0, 1]$ for any $n \ge 1$. Let ϕ be the indicator function given by

$$\phi(x) = \begin{cases} 0 & (x \le \nu) \\ 1 & (x > \nu). \end{cases}$$

Then, $T(x) = \beta x - (\beta - 1)\phi(x)$. Set $d_i(x) = \phi(T^{i-1}(x))$. We have

$$T^K(x) = \beta^K x - (\beta - 1) \sum_{i=1}^{K} \beta^{K-i} d_i(x) \iff x = (\beta - 1) \sum_{i=1}^{K} \frac{d_i(x)}{\beta^i} + \frac{T^K(x)}{\beta^K}. \tag{2.1}$$

Since $|T^K(x)| \le 1$, we see that the sequence $(d_i(x))$ gives a β-expansion of x by letting $K \to \infty$. This shows that if we have a circuit described as above, then by circulating the signal around this circuit and recording $d_i(x)$, we obtain a β-expansion of x. This is the mechanism of the β-encoders that we are interested in.

Now let us explain the problem we discuss. If all the digits $(d_i(x))_{i \ge 1}$ are available, then there would be no loss of information in the encoding–decoding process. However, in real-world use, only a finite number of digits $(d_i(x))_{1 \le i \le K}$ are available (K is the number of digits that we record) and there is some loss of information. That is, given $x \in (0, 1)$, we have its incomplete β-expansion $(d_i(x))_{1 \le i \le K}$. This sequence is decoded into

$$\rho_{\beta,\nu,K}(x) = \rho(x) = (\beta - 1) \sum_{i=1}^{K} \frac{d_i(x)}{\beta^i} + \frac{\theta}{\beta^K}, \tag{2.2}$$

where θ is some real number. The last term is an auxiliary constant added so as to decrease the loss of information on average. In this article, we assume that $\theta = 0.5$. (For the choice of θ, see [4, Section 2] for more information.) We are interested in how much $\rho(x)$ differs from x on average. More precisely, we want to estimate the *MSE*

$$\mathrm{MSE}(\beta, \nu, K) := \int_{[0,1]} (\rho(x) - x)^2 \, dx.$$

It is reasonable to expect that, for fixed β and ν, the MSE would be proportional to β^{-2K} (see also equation (2.3)). Thus, we are interested in estimating the calibrated MSE (abbreviated to CMSE), that is, $\mathrm{CMSE} := \beta^{2K}\mathrm{MSE}$.

https://doi.org/10.1017/S1446181125000070 Published online by Cambridge University Press

The direct analysis of the MSE is a difficult task for large K. The number of values of $\rho(x)$ increases exponentially rapidly and, for general β, the points of discontinuity distribute unevenly. As such, the direct calculation of the MSEs requires much effort.

In [4], the concept of *segments* is used to analyse the MSEs. By changing the variable, the calculation of the MSEs is reduced to counting the population of intervals grouped with regard to their images under T^K. They have some Markovian property and this reduces the cost of the calculations.

The technique introduced by Shinohara and Kobayashi [7] has similar flavour. First, recall that given a piecewise C^1-map $S : I \to I$ (where I denotes the unit interval $[0, 1]$), we have the *Perron–Frobenius operator* P associated to S (see [7] for more information). This is a linear operator acting on the space of finite measures on I. The action of P is characterized by

$$\int f(x)P(\mu) = \int f(S(x))\mu,$$

for every L^1-function $f(x)$ and a finite measure μ.

For our map T, using equations (2.1)–(2.2) one can check that

$$x - \rho(x) = \beta^{-K}(T^K(x) - \theta).$$

Thus,

$$\mathrm{MSE} = \beta^{-2K} \int (x - \theta)^2 P^K(dx), \tag{2.3}$$

where dx denotes the Lebesgue measure. Hence, to understand the MSE, we need to understand $P^n(dx)$. Analysis of Perron–Frobenius operators was carried out by Mori (see [5]). In [7], assuming that $\nu = 1/\beta$ (which makes the calculation easier), we obtained an explicit formula of $\eta\, dx := \lim_{n\to\infty} P^n(dx)$, which turns out to be the invariant measure of T (see [7] for more information). Also, by the spectral analysis we can have a rigorous estimate of the difference between $\eta\, dx$ and $P^n(dx)$ for large n.

In this article, we consider the case where ν is not necessarily equal to $1/\beta$. In the case $\nu = 1/\beta$, the generating function of $P^n(dx)$ is tightly related to the characteristic function of the segments (see [7]). In our case, it seems difficult to establish the same relationship due to the fact that $\nu \neq 1/\beta$. Instead of dealing with segments, we use different simple functions. By introducing these functions, we can obtain the explicit formula of the invariant measure and the generating function of the Perron–Frobenius operator. We show this in the next section.

3. Generating function

Given a sequence $(A_n)_{n\geq 0}$, its *generating function* is a formal power series $\sum_{k=0}^{\infty} A_k z^k$. We assume that A_n are elements of some additive group. In our case, $A_n = P^n(dx)$ (we define $A_0 = dx$). The purpose of considering the generating function is that sometimes we can derive a closed expression of it. Then, by investigating it, we can conclude

https://doi.org/10.1017/S1446181125000070 Published online by Cambridge University Press

some unexpected properties of (A_n). This strategy was demonstrated in [7]. By a similar analysis, we can deduce the closed formula for our case, but it will be more involved. Put

$$s(z) := \sum_{k=0}^{\infty} P^k(dx) z^k.$$

We want to obtain a closed expression of $s(z)$.

We prepare some notation. Let J denote some sub-interval of I. By $\mathbf{1}_J = \mathbf{1}_J(x)$, we denote the characteristic function of J. Given $\alpha \in [0,1]$, we set $\psi_\alpha(x) = \mathbf{1}_{[0,\alpha]} - \mathbf{1}_{(\alpha,1]}$. Note that $\mathbf{1}_{[0,\alpha]} = \frac{1}{2}(\psi_\alpha(x) + \mathbf{1}_{[0,1]})$.

We set $P_+ = \beta\nu$, $P_- = \beta\nu - (\beta - 1)$. These are the "upper" and the "lower" values of $T(\nu)$. By $\operatorname{sgn}(x)$, we denote the sign of x: that is, $\operatorname{sgn}(x) = 1$ if $x > 0$, $\operatorname{sgn}(x) = -1$ if $x < 0$ and $\operatorname{sgn}(0) = 0$.

In the following, for simplicity, we assume that $T^i(P_\pm) \neq \nu$ for every $i \geq 0$. Now we are ready to state our conclusion.

THEOREM 3.1. *We have*

$$s(z) = \frac{1}{1 - z/\beta}\mathbf{1}_{[0,1]} + \frac{z}{2}\frac{1}{\beta - z}\frac{\chi_+ - \chi_-}{1 - E_+ + E_-}, \tag{3.1}$$

where

$$\chi_\pm = \sum_{i=0}^{\infty} \psi_{T^i(P_\pm)}\left(\frac{z}{\beta}\right)^i, \quad E_\pm = \frac{1}{2}\sum_{i=0}^{\infty} \operatorname{sgn}(T^i(P_\pm) - \nu)\left(\frac{z}{\beta}\right)^{i+1}.$$

Owing to the limitation in the number of pages, we do not give the proof of this here. The proof is similar to that in [7]. The only difference is that instead of considering the generating function of the intervals, we consider the generating function of ψ_α (see also [5]), that is,

$$s^{\psi_\alpha}(z) = \psi_\alpha + P(\psi_\alpha)z + \cdots + P^n(\psi_\alpha)z^n + \cdots.$$

Equation (3.1) can be used to estimate $P^k(dx)$. Recall that $P^k(dx)$ is the coefficient of z^k. Then, by considering the Taylor expansion of $s(z)$, we can deduce the formula of $P^k(dx)$. Another interesting piece of information that we can read from equation (3.1) is the asymptotic behaviour of $P^k(dx)$. As a rational function of z, one can see that $s(z)$ has no pole on $|z| < 1$, $z = 1$ is a simple pole and there are no other poles on $|z| = 1$. The residue at $z = 1$ corresponds to the limit of $P^k(dx)$. Furthermore, by the technique given in [7], it is possible to obtain the rigorous error between the invariant measure and $P^k(dx)$. In fact, the error decays exponentially rapidly. (For more information, see [7].)

Applying this for sufficiently large n, we can obtain an approximated value of $\mathrm{CMSE}(\beta, \nu, n)$ by the integral (2.3) calculated with respect to the invariant measure. We give more details in the next section.

https://doi.org/10.1017/S1446181125000070 Published online by Cambridge University Press

4. Residue and invariant measure

We continue the calculation. In equation (3.1), $s(z)$ is holomorphic in the disk $|z| < 1$, it has a simple pole at $z = 1$ and this is the only pole in $|z| \le 1$. The pole $z = 1$ corresponds to the invariant measure. More precisely, $s(z)$ can be written as

$$\frac{A}{1-z}(\chi_+ - \chi_-) + M(z),$$

where A is some constant and $M(z)$ is some meromorphic function which is holomorphic on some disc $|z| < 1 + \varepsilon$. Then one can deduce the following theorem.

THEOREM 4.1. *In the above formula,*

$$A = \frac{1}{(\beta - 1) + G_+ - G_-} \quad \text{where } G_\pm = \sum_{i=1}^{\infty} \frac{T^i(P_\pm)}{\beta^i}. \tag{4.1}$$

The calculation is straightforward (see [7]). Equations (3.1) and (4.1) give us the formula of the invariant measure $\eta\, dx = \lim_n P^n(dx)$. In fact, if one considers the Taylor expansion of $s(z)$, the terms from $M(z)$ decay exponentially rapidly, while the terms from the other one do not. As such, the terms from the first one gives the invariant measure.

THEOREM 4.2. *We have*

$$\eta = \sum_{i=0}^{\infty} \frac{A}{\beta^i}(\psi_{T^i(P_+)} - \psi_{T^i(P_-)}). \tag{4.2}$$

Using equation (4.2), we can obtain an estimate of the MSE (in equation (2.3)). In fact, for n large, $P^n(dx)$ is close to $\eta\, dx$. As the formula of η in equation (4.2) is given by an infinite series which decays exponentially fast, by taking some finite terms we have an approximation of η, which also approximates $P^n(dx)$.

5. Numerical results

We now show the results of the previous section numerically. First, Figures 1 and 2 show the distribution function η of the invariant measure of the map T calculated by the formula in equation (4.2). Here, x denotes the coordinate of $[0, 1]$ and y denotes the magnitude of η.

This is a form of infinite sum of simple functions. We have drawn the graph by adding terms up to $i = 20$. We set $\beta = 1.8$, $\nu = 0.5, 0.51$, respectively. Note that, for $\nu = 0.5$, we can observe the symmetry with respect to $x = 0.5$.

Using the invariant measure, we can calculate the estimated value of the CMSE. Set

$$\mathrm{CMSE}^\infty = \int (x - \theta)^2\, \eta\, dx.$$

In Figures 3 and 4, we show the calculated CMSE^∞ (by the invariant measure, calculated in the same way) and compare it with the $\mathrm{CMSE}(\beta, \nu, n)$ (by definition).

https://doi.org/10.1017/S1446181125000070 Published online by Cambridge University Press

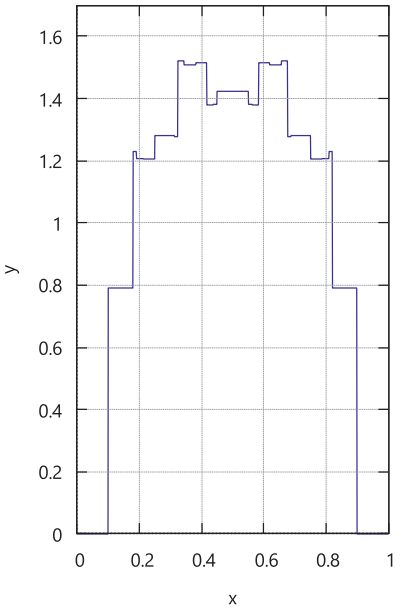


FIGURE 1. The distribution function η of the invariant measure of T, $\beta = 1.8$, $\nu = 0.5$.

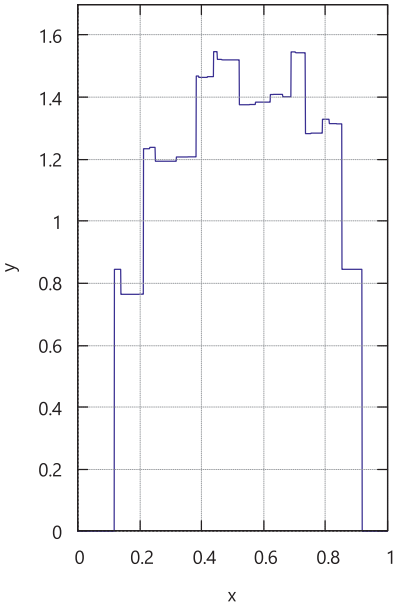


FIGURE 2. The distribution function η of the invariant measure of T for $\beta = 1.8$, $\nu = 0.51$.

https://doi.org/10.1017/S1446181125000070 Published online by Cambridge University Press

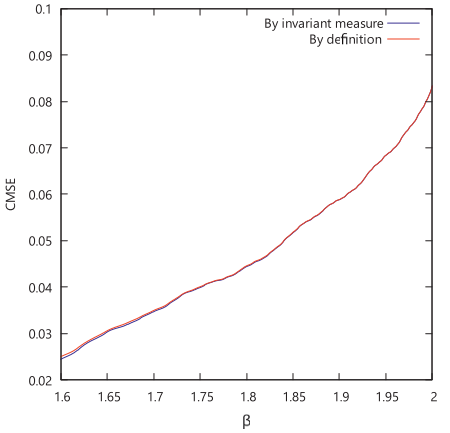


FIGURE 3. CMSE(β, 0.5, 10) for $1.6 \leq \beta \leq 2.0$ and CMSE$^{\infty}$.

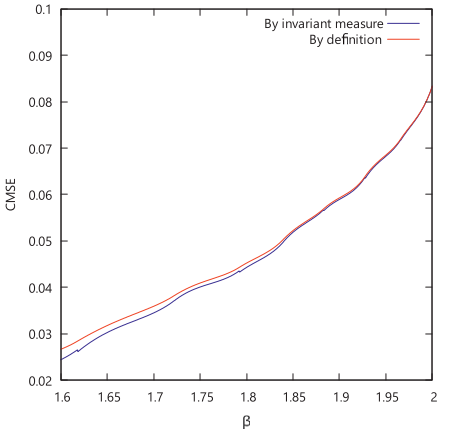


FIGURE 4. CMSE(β, 0.5, 7) for $1.6 \leq \beta \leq 2.0$ and CMSE$^{\infty}$.

Recall that the CMSE is the β^{2K}MSE and that the CMSE$^{\infty}$ is a novel quantity which is expected to approximate the CMSE.

For $K = 10$, these two quantities are almost the same. In the range calculated, their difference is less than 3%. Note that there is some difference observable for the case $K = 7$. In the actual use of the β-encoders, K is set around 12. Thus, for practical use, our method gives satisfactory results.

We also calculated the CMSE by varying the value of ν for $\beta = 1.8, 1.7$ (see Figures 5 and 6). By varying ν, the CMSE seems to vary continuously, but it looks like it has complicated local behaviour.

https://doi.org/10.1017/S1446181125000070 Published online by Cambridge University Press

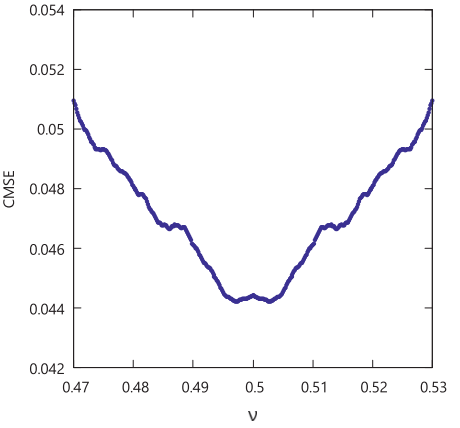


FIGURE 5. The graph of CMSE^{∞}. We set $\beta = 1.8$ and $0.47 \leq \nu \leq 0.53$.

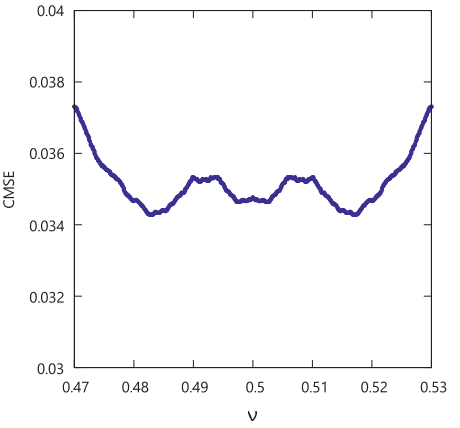


FIGURE 6. The graph of CMSE^{∞}. We set $\beta = 1.7$ and $0.47 \leq \nu \leq 0.53$.

6. Conclusion

We explained the problem of bounding MSEs of β-encoders. We proposed a method of obtaining an approximated value of the CMSE by means of the spectral analysis of Perron–Frobenius operators of one-dimensional maps. We gave some numerical results concerning the values of CMSEs for different parameters.

We propose a few future problems. First, it would be interesting to estimate the difference between the CMSE and CMSE^{∞}. In a previous paper [7], by means of interval arithmetics and the further spectral analysis of the generating function, we derived a rigorous upper bound for the error of these two quantities. In principle, it is possible to establish similar results, although it would be difficult.

Another interesting problem is to investigate the behaviour of the convergence. For fixed β and ν, as $K \to \infty$ we know that $P^K(dx)$ converges to $\eta\, dx$. As such, we know that

https://doi.org/10.1017/S1446181125000070 Published online by Cambridge University Press

the CMSE converges to CMSE^{∞}. Numerical calculations suggest that the convergence is monotone for sufficiently large K, but we do not have any proof of it. Finally, it would be also interesting to investigate the local property of CMSE^{∞} as a function of β and ν. It would not be hard to confirm that it is a continuous function. The question is how regular it is. Is it, for instance, Hölder continuous for some exponent?

Acknowledgements

The author is grateful to Dr. Yutaka Jitsumatsu for his support with the calculation of the generating functions. This work was supported by JSPS KAKENHI Grant Numbers JP16K00333 and JP19K12156. Also, this work was supported by the Institute of Mathematics for Industry, Joint Usage/Research Center in Kyushu University (FY2017 Short-term Visiting Researcher "Analog-to-Digital converter and dynamical systems" (Reference No. 20170003)).

References

[1] I. Daubechies, R. DeVore, C. S. Gunturk and V. A. Vaishampayan, "Beta expansions: a new approach to digitally corrected A/D conversion", *IEEE 2002 Int. Sympo. Circ. and Syst. ISCAS (2002) II*; doi:10.1109/ISCAS.2002.1011470.

[2] I. Daubechies and O. Yilmaz, "Robust and practical analog-to-digital conversion with exponential precision", *IEEE Trans. Inform. Theory* **52** (2006) 3533–3545; doi:10.1109/TIT.2006.878220.

[3] T. Kohda, Y. Horio, Y. Takahashi and K. Aihara, "Beta encoders: symbolic dynamics and electronic implementation", *Internat. J. Bifur. Chaos Appl. Sci. Engrg.* **22** (2012) Article ID: 1230031; doi:10.1142/S0218127412300315.

[4] T. Makino, Y. Iwata, K. Shinohara, Y. Jitsumatsu, M. Hotta, H. San and K. Aihara, "Rigorous estimates of quantization error for A/D converters based on Beta-map", *NOLTA* **6** (2015) 99–111; doi:10.1587/nolta.6.99.

[5] M. Mori, "Fredholm determinant for piecewise linear transformations", *Osaka J. Math.* **27** (1990) 81–116; doi:10.18910/10418.

[6] H. San, T. Kato, T. Maruyama, K. Aihara and M. Hotta, "Non-binary pipeline analog-to-digital converter based on β-expansion", *IEICE Trans. Fundam. Electron Commun. Comput. Sci.* **96** (2013) 415–421; doi:10.1587/transfun.E96.A.415.

[7] K. Shinohara and K. Kobayashi, "Estimation of mean squared errors of non-binary A/D encoders through Fredholm determinants of piecewise-linear transformations", *NOLTA* **9** (2018) 243–258; doi:10.1587/nolta.9.243.

[8] R. Suzuki, T. Maruyama, H. San, K. Aihara and M. Hotta, "Robust cyclic ADC architecture based on β-expansion", *IEICE Trans. Electron.* **96** (2013) 553–559; doi:10.1587/transele.E96.C.553.

https://doi.org/10.1017/S1446181125000070 Published online by Cambridge University Press