B. S.T. J. BRIEFS

Estimation of the Variance of a Stationary Gaussian Random Process by Periodic Sampling

By J. C. DALE

(Manuscript received February 14, 1967)

I. INTRODUCTION

This paper* applies previous work[1] on estimation of the mean of a stationary random process by periodic sampling to estimation of the variance with the added restriction that the process under consideration be Gaussian with known mean.

The samples are taken from a sample function of the random process, in a closed interval $(0, T)$ and are in general correlated. The estimator used is the average of equally-weighted squared samples. The variance of this estimator is derived and its behavior is predicted as a function of the number of samples and length of record.

II. THEORY

2.1 *General*

Let $x(t)$ be a sample function from a stationary, Gaussian random process $\{x(t)\}$ with known mean.†

An unbiased estimator of the variance is given by

$$\hat{\sigma}_x^2 = \frac{1}{N+1} \sum_{k=0}^{N} x^2\left(\frac{kT}{N}\right), \tag{1}$$

where T/N is the sampling period.

By invoking the Gaussian assumption, the variance of this estimator follows directly and is given by

$$\text{var}\,(\hat{\sigma}_x^2) = \frac{2}{N+1} \sum_{k=-N}^{N} \left(1 - \frac{|k|}{N+1}\right) R_x^2\left(\frac{kT}{N}\right),$$

* This work was supported by the U.S. Navy, Bureau of Ships under Contract N00 600-67-CO549.

† So long as the mean of $\{x(t)\}$ is assumed known, no generality of the derivation is lost by letting it be zero.

or in an equivalent form (2)

$$\text{var}\,(\hat{\sigma}_x^2) = \frac{2}{N+1}\int_{-\infty}^{\infty} q_{(N+1/N)T}(\tau)R_x^2(\tau)\sum_{k=-\infty}^{\infty}\delta\left(\tau - \frac{kT}{N}\right)d\tau.$$

$q_{(N+1/N)T}(\tau)$ is the triangular weighting function (See Ref. 1) and $R_x(\tau)$ is the autocorrelation function of $\{x(t)\}$.

By comparing the var $(\hat{\sigma}_x^2)$ to (3) in the previous paper[1] it is seen that (2) gives the variance of the sample mean of a stationary random process $\{x^2(t)\}$, whose autocovariance function is given by $2R_x^2(\tau)$. The spectrum of $\{x^2(t)\}$ is $2S_x(\omega) * S_x(\omega)$.

At this point the previous theory[1] applies directly. Using the same notation, the spectrum of the squared samples can be written as

$$G(\omega) = \frac{2}{N+1}\int_{-\infty}^{\infty} q_{(N+1/N)T}(\tau)R_x^2(\tau)e^{-i\omega\tau}\sum_{k=-\infty}^{\infty}\delta\left(\tau - \frac{kT}{N}\right)d\tau,$$

which is equivalent to (3)

$$G(\omega) = \frac{N}{T}\sum_{k=-\infty}^{\infty} F\left(\omega - k\,\frac{2\pi N}{T}\right).$$

In this case

$$F(\omega) = Q(\omega) * 2[S_x(\omega) * S_x(\omega)], \tag{4}$$

and $Q(\omega)$ is the transform of the weighting function.

$G(\omega)$ can be interpreted as $F(\omega)$, shifted by integral multiples of the sampling frequency, $2\pi N/T$. As before to obtain the variance of the estimate we need only be concerned with the value of $G(\omega)$ at $\omega = 0$. To minimize the variance of the estimate, the sampling frequency should be high enough to prevent overlapping of the sideband at $\omega = 0$. Satisfying this condition results in

$$\text{var}\,(\hat{\sigma}_x^2) = \frac{N}{T}F(0).\dagger \tag{5}$$

To answer the question of how many samples to take in time T to obtain minimum variance, consider (4). $Q(\omega)$ is approximately zero for

† Equation (5) is not quite true when both end points of the time record are included as samples. This is because $(N/T)\,F(0)$ is a function of N namely,

$$(N/T)\,F(0) = \frac{1}{2\pi}\int_{\infty}^{-\infty} 2[S_x(y) * S_x(y)]\left[\frac{\sin y(N + 1/2N)T}{y(N + 1/2N)T}\right]^2 dy.$$

Increasing N beyond the value given in (8) actually results in a higher variance on the estimate. This is apparent in the two examples, particularly for T small. This same effect was discussed in Ref. 1.

$|\omega| \geqq (2\pi/T)[N/(N + 1)]$. If $S_x(\omega)$ is zero for $|\omega| \geqq 2\pi B$, then $S_x(\omega) * S_x(\omega)$ will be zero for $|\omega| \geqq 4\pi B$ and $F(\omega)$ will be approximately 0 for

$$| \omega | \geqq 2\pi\left[2B + \frac{1}{T}\left(\frac{N}{N + 1}\right)\right]. \tag{6}$$

Therefore, choosing the sampling frequency so that

$$\frac{2\pi N}{T} \geqq 2\pi\left[2B + \frac{1}{T}\left(\frac{N}{N + 1}\right)\right] \tag{7}$$

results in (5) being satisfied.

Solving (7) for N yields the required number of samples taken in time T to approximately minimize the variance of the estimate, namely

$$N = 2BT\left[\frac{1 + \sqrt{1 + 2/BT}}{2}\right]. \tag{8}$$

For BT $\gg 1$, N is approximately equal to $2BT$. Thus, twice the number of samples are required to obtain a minimum variance estimate of the variance than was previously shown to obtain a minimum variance estimate of the mean.

2.2 *Variance For Large T*

If T is allowed to become large $Q(\omega)$ will approach a delta function,

$$\lim_{T\to\infty} Q(\omega) = \frac{2\pi}{N + 1}\,\delta(\omega). \tag{9}$$

This results in

$$F(\omega) = \frac{1}{\pi(N + 1)} \int_{-\infty}^{\infty} S_x(y)S_x(\omega - y)\,dy. \tag{10}$$

If $S_x(\omega)$ is zero for $| \omega | \geqq 2\pi B$, and the sampling frequency satisfies (7), then the minimum value of variance of the estimate is given by

$$\operatorname{var}(\hat{\sigma}_x^2)\Big|_{\substack{\min \\ T\text{ large}}} = \frac{N}{T}\,F(0) = \frac{N}{T(N + 1)\pi} \int_{-\infty}^{\infty} S_x^2(y)\,dy, \tag{11}$$

$$\approx \frac{1}{\pi T} \int_{-\infty}^{\infty} S_x^2(y)\,dy.$$

This is the same value obtained by continuous sampling.

III. EXAMPLES

The variance of the estimate of variance as a function of number of samples $(N+1)$ and length of record (T) has been computed for two examples.

The computation was done using an expression equivalent to (2).

3.1 *Rectangular Spectrum*

$$S_x(\omega) = \begin{cases} \frac{1}{2}; & -2\pi < \omega < 2\pi, \\ 0; & \text{elsewhere.} \end{cases} \tag{12}$$

Fig. 1 shows var $(\hat{\sigma}_x^2)$ plotted against number of samples. Each curve represents a different length of record as indicated by the values shown on the figure. It should be noted that the minimum value of var $(\hat{\sigma}_x^2)$ occurs at the number of samples predicted by (8). Also for small values of T the var $(\hat{\sigma}_x^2)$ reaches a minimum and then increases as more samples are taken. This is due to including both end points of the time record as samples.

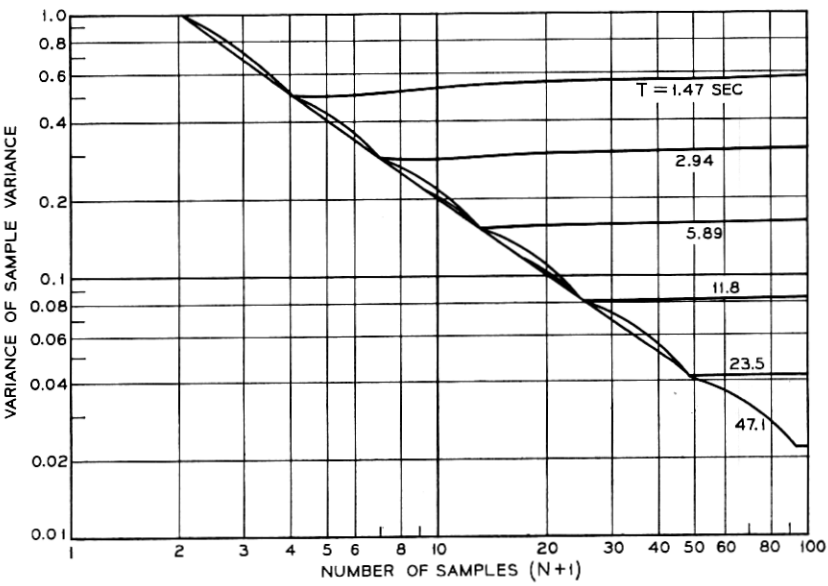


Fig. 1 — Variance of the sample variance as a function of the number of samples and length of record for a process with rectangular spectral density.

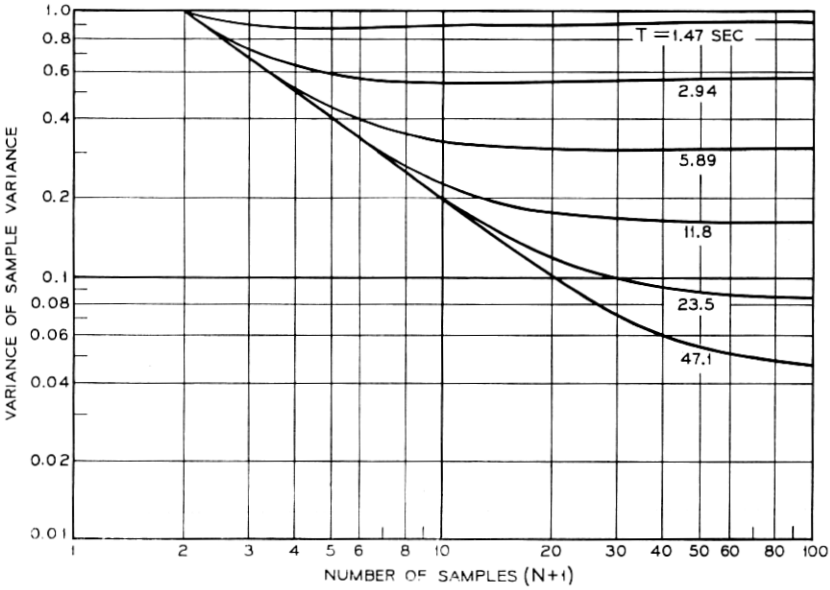


Fig. 2 — Variance of the sample variance as a function of the number of samples and length of record for a process with Markoff spectral density.

3.2 *Markoff Spectrum*

$$S_x(\omega) = \frac{2}{\omega^2 + 1}. \tag{13}$$

This sample shows a nonbandlimited spectrum. The results are shown in Fig. 2.

IV. CONCLUSION

By making the assumption that the random process $\{x(t)\}$ was Gaussian, it was possible to express var $(\hat{\sigma}_x^2)$ into an array of terms containing $R_x^2(kT/N)$, (2). In this form it is possible to apply the theory developed in the work on estimation of the mean.[1] The interesting result from this derivation was that when $BT \gg 1$, the variance of the sample variance is essentially minimized when $2BT$ samples are taken. This is in contrast to the BT samples required to minimize the variance of sample mean.

REFERENCES

1. Balch, H. T., Dale, J. C., Eddy, T. W., and Lauver, R. M., Estimation of the Mean of a Stationary Random Process by Periodic Sampling, B.S.T.J., *45*, May-June, 1966, pp. 733–741.

A Floating Gate and Its Application to Memory Devices

By D. KAHNG and S. M. SZE

(Manuscript received May 16, 1967)

A structure has been proposed and fabricated in which semi-permanent charge storage is possible. A floating gate is placed a small distance from an electron source. When an appropriately high field is applied through an outer gate, the floating gate charges up. The charges are stored even after the removal of the charging field due to much lower back transport probability. Stored-charge density of the order of $10^{12}/cm^2$ has been achieved and detected by a structure similar to an metal-insulator-semiconductor (MIS) field effect transistor. Such a device functions as a bistable memory with nondestructive read-out features. The memory holding time observed was longer than one hour. These preliminary results are in fair agreement with a simple analysis.

It has been recognized for some time that a field-effect device, such as that described by Shockley and Pearson,[1] can be made bistable utilizing switchable permanent displacement charges on ferroelectric material.[2] Subsequent studies of ferroelectric material have revealed,[3] however, that the inherent speed capability of a device incorporating a ferroelectric material is limited by domain motion, whose highest speed is limited by the acoustic velocity. In the absence of highly ordered, near-ideal thin film ferroelectric material, the speed capability of a bistable device, therefore, is in the microsecond range at best.[4] In addition, many ferroelectric materials suffer from irreversible mechanical disorder after many cycles of polarization switching,[2] rendering some uncertainty on the long term device reliability aspect.

An alternative to a ferroelectric gate is a floating gate chargeable by field emission, which hopefully circumvents the above mentioned difficulties. Consider a sandwich structure, metal $M(1)$, insulator $I(1)$, metal $M(2)$, insulator $I(2)$, and finally metal $M(3)$. (See Fig. 1). If the thickness of $I(1)$ is small enough so that a field-controlled electron transport mechanism such as tunneling or internal tunnel-hopping are possible, a positive bias on $M(3)$ with respect to $M(1)$ with $M(2)$ floating [$M(2)$ is called the floating gate henceforth], would cause electron accumulation in the floating gate, provided electron transport

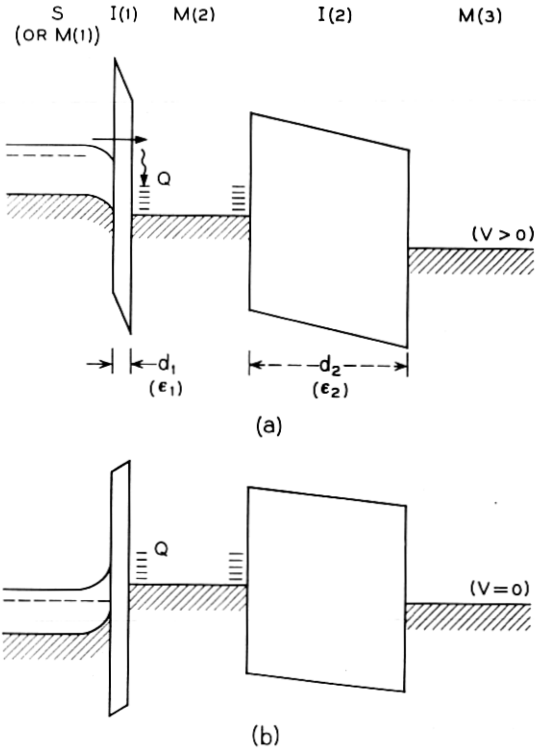


Fig. 1 — Energy band diagram of a floating gate structure with a semiconductor-insulator-metal-insulator-metal sandwich. For calculation of the stored charge, the semiconductor is replaced by a metal $M(1)$. (a) When a positive voltage step is applied to the outer gate. (b) When the voltage is removed. The stored charge Q causes an inversion of the semiconductor surface.

across $I(2)$ is small. These conditions can be met by choosing $I(1)$ and $I(2)$ such that the ratio of dielectric permittivity ϵ_1/ϵ_2 is small and/or the barrier height into $I(1)$ is smaller than that into $I(2)$. The sandwich structure is somewhat similar to the tunnel emitter metal-base transistor proposed by Mead[5] in its structure but with the following essential differences.

(*i*) $M(2)$ is much thicker than the hot electron range, so that emitted electrons are close to the Fermi-level of $M(2)$ before reaching $I(2)$.

(*ii*) No carrier transport is allowed across $I(2)$.

(*iii*) $M(2)$ is floating.

The stored charge Q, as a function of time when a step voltage function with amplitude V is applied across the sandwich, is given by

$$Q(t) = \int_0^t j \, dt' \qquad \text{coul/cm}^2. \tag{1}$$

When the emission is of Fowler-Nordheim tunneling type, then the current density, j, has the form

$$j = C_1 E^2 \exp(-E_0/E), \tag{2a}$$

where C_1 and E_0 are constants in terms of effective mass and the barrier height. (We have neglected the effects due to the image force lowering[6] of the barrier, etc., but the essential feature is expected to be retained even after detailed corrections are made). This type of current transport occurs in SiO_2 and Al_2O_3.

When the field emission is of the internal Schottky or Frankel-Poole type, as occurs in Si_3N_4,[7] then j follows the form

$$j = C_2 E \exp[-q(\Phi_1 - \sqrt{qE/\pi\epsilon_1})/kT], \tag{2b}$$

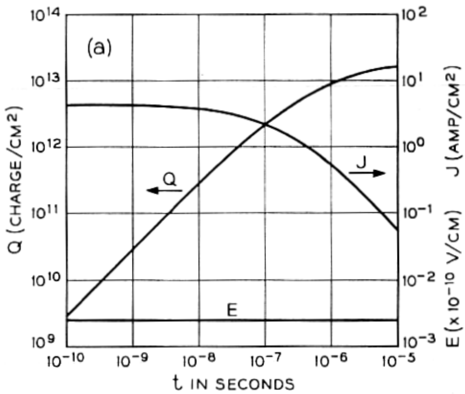
where c_2 is a constant in terms of trapping density in the insulator, Φ_1 the barrier height in volts, ϵ_1 the dynamic permittivity.

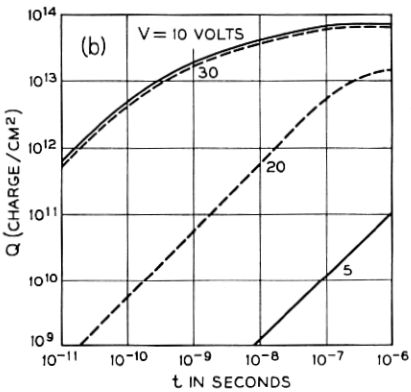
The electric field in $I(1)$ at all times is a function of the applied voltage V and $Q(t)$, and is obtainable from the displacement continuity requirement as

$$E = \frac{V}{d_1 + d_2(\epsilon_1/\epsilon_2)} - \frac{Q}{\epsilon_1 + \epsilon_2(d_1/d_2)}, \tag{3}$$

where d_1 and d_2 are the thickness of $I(1)$ and $I(2)$, respectively.

Fig. 2(a) shows the results of a theoretical computation using (1), (2a), and (3) with the following parameters: $d_1 = 50$ Å, $\epsilon_1 = 3.8\,\epsilon_0$ (for SiO_2), $d_2 = 1000$ Å, $\epsilon_2 = 30\,\epsilon_0$ (for ZrO_2), and $V = 50$ volts. One notes that the stored charge initially increases linearly with time and then saturates. The current is almost constant for a short time and then decreases rapidly. The field in $I(1)$ decreases slightly as the time increases. The above results can be explained as follows: When a voltage pulse is applied at $t = 0$, the initial charge Q is zero, and the initial electric field across $I(1)$ has its maximum value, $E_{max} = V/[d_1 + (\epsilon_1/\epsilon_2)d_2]$. As t increases, Q will first increase linearly with time. This is because of the fact that for small Q such that E remains essentially the same, the current will in turn remain the same, so $Q = j(E_{max}) \cdot t$. Eventually, when Q is large enough to reduce the





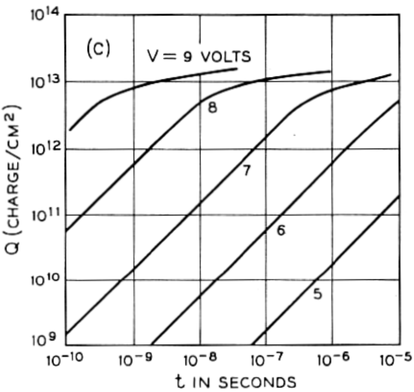


Fig. 2(a) — Theoretical results of the stored-charge density (Q), the current density (J), and the electric field across $I(1)$ as a function of time. V = 50 volts, d_1 = 50 Å, ϵ_1/ϵ_0 = 3.8 (for SiO_2), d_2 = 1000 Å, ϵ_2/ϵ_0 = 30 (for ZrO_2). (b) Theoretical results of the stored charge density as a function of time with the same ϵ_1 and ϵ_2 as in (a), and d_1 = 10 Å, d_2 = 100 Å (solid lines), d_1 = 30 Å, d_2 = 300 Å (dotted lines). (c) Theoretical results of the stored density as a function of time with d_1 = 20 Å, ϵ_1/ϵ_0 = 60 (for Si_3N_4), d_2 = 200 Å, ϵ_2/ϵ_0 = 30 (for ZrO_2) and various applied voltages.

value of E substantially, then the current will decrease rapidly with time and Q increases slowly.

Fig. 2(b) shows the stored charge as a function of time for the time ϵ_1 and ϵ_2 but different d_1, d_2, and V. It is clear that for a given structure, in order to store a given amount of charge, one can either increase the applied voltage or increase the charging time (pulse width) or both. Fig. 2(c) shows the calculated stored charge for the current transport described by (2b). Here $I(1)$ is a 20 Å thick Si_3N_4 film. There are