Analysis of powers-of-two calculations of the Allan variance and their relation to the standard variance

Noah Schlossberger* and David Howe†
*† *Time and Frequency Division, NIST* Boulder, CO 80305 USA
*† *Department of Physics, University of Colorado* Boulder, CO 80309 USA
* noah.schlossberger@colorado.edu
† david.howe@nist.gov

***Abstract*—In this work we show that powers of two multiples of the sampling period are the optimal set of averaging period Allan variance (AVAR) calculations for determining power law noise types present in recorded data. The primary merit of AVAR is that it indicates the slopes associated with each of the power law noise types as well as the level of each type included, even for a mixture of noise types. We show that unlike other arbitrary series, the powers-of-two values are spectrally the closest-to-independent set of AVAR values possible, and thus optimally decompose frequencies in such a way as to have the least uncertainty in estimating slopes. We further demonstrate the unique property of this choice of averaging period series by proving the equivalence between the sums of the powers-of-two values of the non-overlapping Allan variance and twice the value of the standard variance.**

***Keywords*—Power Law Noise Determination, Allan Variance, Standard Variance, Statistical Theory, Stability Characterization**

I. Introduction

The Allan variance, $\sigma_y^2(\tau) \equiv 1/2\langle(y(t)-y(t-\tau))^2\rangle$, where $y(t)$ is a τ-average fractional-frequency oscillator error, was the first time-domain characterization that delineated power-law spectral noises, i.e. f^α [1]. The notation AVAR and ADEV for the Allan variance and deviation have persisted from the names for statistical estimators of $\sigma_y^2(\tau)$ from original 1969 FORTRAN programs in NIST's time scale. When α = 0, -1, -2, for white, flicker, and random walk noise, respectively, AVAR responds with power-law of -1, 0, and +1 in terms of τ-averages. For any data run, the square root of AVAR, or ADEV, is convenient for accurately determining the level of these three noise types. Statistics such as the discrete Fourier transform (DFT) are not used to delineate power-law noises because DFT calculates the power spectral density (PSD) using a 1 Hz (constant) band pass filter, whereas ADEV has a proportional-to-f response that readily delineates these noises. On the other hand, the PSD efficiently detects spectral lines. Certainly, a regression analysis can estimate power law noises $\sim f^\alpha$. However, for long-term periodicity near the data-run length, the DFT will contain Fourier-frequency components that are subject to windowing errors, and sampling-Nyquist biases [2], where spectral cycles per-day, per-month, or per-year are important in clocks. Such PSDs can be hard to interpret and relate to in clock applications. Here is where the tunable one-octave mean-square measurement inherent in AVAR's tunable-by-varying τ is a powerful property for characterizing clock stability. Spectral noise identified in a measurement has specific cause and so aids in development of clocks and oscillators at both short and long averaging periods [3].

This variable-averaging-period feature introduces a choice for which values to calculate. It is standard practice to use values of ADEV with averaging periods that are powers of two of the sampling period. Ostensibly, this allows equi-spaced levels vs. τ-averages to span substantial range using logarithmic scaling. However, there are many additional rich features of this choice of averaging periods that we reveal for the first time in this paper.

II. Definitions of AVAR Estimators

For any data set $\{y_1, y_2, , y_{N_y}\}$ where the number of points N_y is a power of 2 (i.e. $N_y = 2^J$ for an integer J), the Allan variance, non-overlapping Allan varianc (AVARnono), and maximally overlapping Allan variance (AVARmaxo) are defined [4] as follows:

$$\begin{cases} \text{AVAR}(2^j) \equiv \frac{1}{2^{J-j+1}} \sum\limits_{k=2}^{2^{J-j}} \left(\bar{y}_{2^j k}(2^j) - \bar{y}_{2^j(k-1)}(2^j)\right)^2 \\ \text{AVAR}_{nono}(2^j) \equiv \frac{1}{2^{J-j}} \sum\limits_{k=1}^{2^{J-j-1}} \left(\bar{y}_{2^{j+1}k}(2^j) - \bar{y}_{2^j(2k-1)}(2^j)\right)^2 \\ \text{AVAR}_{maxo}(2^j) \equiv \frac{1}{2^J - 2^{j+1}+1} \sum\limits_{2^{j+1}}^{2^J} \left(\bar{y}_k(2^j) - \bar{y}_{k-2^j}(2^j)\right)^2 \end{cases} \tag{1}$$

where the bar indicates averaging: $\bar{y}_n(2^j)$ is the average of the last 2^j points at location n

$$\bar{y}_n(2^j) \equiv \frac{1}{2^j} \sum_{l=0}^{2^j-1} y_{n-l}. \tag{2}$$

III. Relationship between Non-Overlapping Allan Variance (AVAR) and Standard Variance (SVAR)

One way that powers-of-two series of Allan variances is unique is that it is known [4], [5] that there exists a relationship between the sum of powers-of-two terms AVAR and the standard variance (when we refer to the standard variance in this paper we do not mean the unbiased standard variance). Both a simulation-based qualitative comparison and a proof-based exact result about this relationship follow.

978-1-5386-8305-7/19/$31.00 ©2019 IEEE

We can characterize the disagreement between the sum over powers of two values of an Allan variance estimator and twice the standard variance with a fractional difference

$$E^{avar} \equiv \frac{\sum_{j=0}^{J-1} \text{AVAR}(2^j)}{2 \cdot \text{SVAR}} - 1. \tag{3}$$

In frequency data for oscillators, we are subject to three types of noise: white noise, flicker, and random walk. Thus, a signal is generally characterized by a linear combination of the three types of noise. We can write for any noise data set Y

$$Y = C_{-2}\{Rw\} + C_{-1}\{Fl\} + C_0\{Wh\}, \tag{4}$$

where $\{Rw\}$ is a data set of random walk noise, $\{Fl\}$ one of flicker noise, and $\{Wh\}$ one of white noise. Here, we use C_α to be the amplitudes of these noise types present in the data (with α being the power of the noise type). Generally, the fractional error E^{avar} will then be a function of the noise amplitudes

$$E^{avar} = E^{avar}(C_{-2}, C_{-1}, C_0).$$

However, E^{avar} is a normalized function of the data set, so it is invariant under scaling the magnitude of the data set, which is equivalent to scaling all noise type coefficients

$$E^{avar}(\zeta C_{-2}, \zeta C_{-1}, \zeta C_0) = E^{avar}(C_{-2}, C_{-1}, C_0) \;\forall \zeta \neq 0.$$

Because of this, we have the freedom to set one of our coefficients to unity and look at E^{avar} as a function of the other two coefficients, i.e.,

$$C_{-1} = 1, \quad E^{avar} = E^{avar}(C_{-2}, C_0).$$

This is an equivalent description of E^{avar} for all linear combinations of noise types as long as $C_{-1} \neq 0$, which would prevent the system from being arbitrarily normalizable. However, most physical systems (for example, all active electronics and passive electronics with DC current flow [6]) exhibit flicker noise, meaning C_{-1} is never zero in most realizable systems. Because of this, the choice of normalizing to C_{-1} is justified and $E^{avar}(C_{-2}, C_0)$ is a valid description.

We can define similar characterization error functions for the non-overlapping and maximally overlapping variants of the statistic (E^{nono}, E^{maxo}). We plot all of these error characterizations for a simulated noise data set of length 512 as a contour plot over C_{-2} and C_0 in Fig. 1.

The features of the contour plots tell us qualitative properties of the statistics. Near the origin, E^{avar} and E^{maxo} are large and positive. Because the origin represents only flicker noise, this means that flicker noise causes sums of the regular and maximally overlapping Allan variance to overestimate the standard variance. Along the axes, E^{maxo} converges to small negative values, meaning random walk and white noise lead to small underestimation. There is also an interesting property of the dip in the corner, indicating that the combination of white noise and random walk leads to more severe underestimation than either on its own. Meanwhile, for E^{avar}, random walk

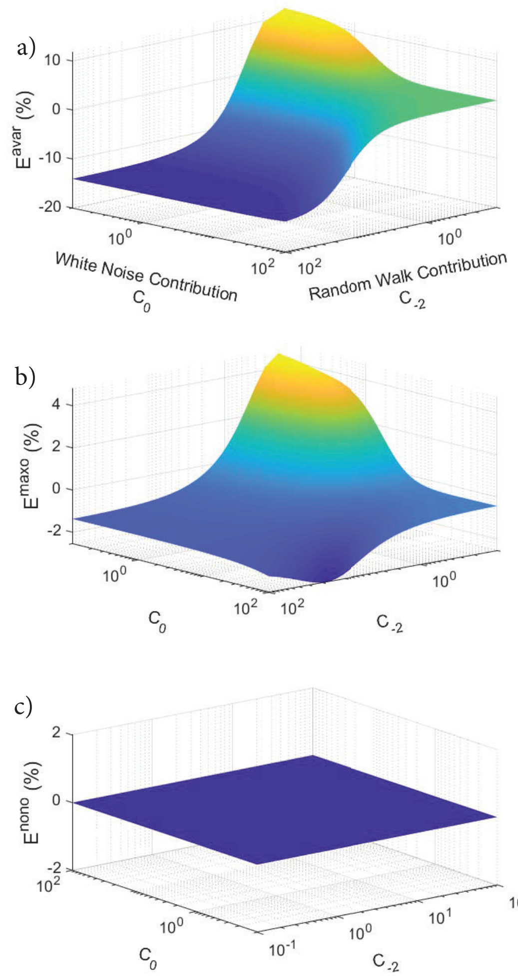

Fig. 1. The error in the estimation of SVAR by the sums over powers of two of (a) the Allan variance, (b) the maximally overlapping Allan variance, and (c) the non-overlapping Allan variance as functions of the ratio of white noise C_0 and random walk noise C_{-2} to flicker. Simulated power law noise samples were obtained from Neil Ashby and were generated using the techniques outlined in [7].

leads to significant underestimation and white noise leads to slight overestimation.

The non-overlapping Allan variance has a flat estimation error curve centered at zero. This is because AVAR_{nono} is an exact analysis of the standard variance. That is,

$$\sum_{j=0}^{J-1} \text{AVAR}_{nono}(2^j) = 2 \cdot \text{SVAR}. \tag{5}$$

As an informative demonstration of this property, let us observe that this is the case for a trivially short data set containing two values, i.e.

$$Y = \{y_1, y_2\}.$$

Then the power of two of the size of the data set is $J = 1$.

Twice the standard variance is given by

$$\begin{aligned} 2\cdot\text{SVAR} &= 2\cdot\tfrac{1}{2}\left(\left(\tfrac{y_1+y_2}{2}-y_1\right)^2+\left(\tfrac{y_1+y_2}{2}-y_2\right)^2\right)\\ &= \left(\tfrac{y_2-y_1}{2}\right)^2+\left(\tfrac{y_1-y_2}{2}\right)^2\\ &= \tfrac{y_1^2+y_2^2-2y_1y_2}{2}. \end{aligned}$$

The sum over the powers of two of AVAR_{nono} is simply one term because the sum is defined from 0 up to $J-1=0$. Thus

$$\begin{aligned} \sum_{j=0}^{J-1}\text{AVAR}_{nono}(2^j) &= \text{AVAR}_{nono}(1)\\ &= \tfrac{1}{2^{1-0}}\sum_{k=1}^{2^{1-0-1}}\left(\bar{y}_{2k}(1)-\bar{y}_{2(2k-1)}(1)\right)^2\\ &= \tfrac{(y_1-y_2)^2}{2}\\ &= \tfrac{y_1^2+y_2^2-2y_1y_2}{2}, \end{aligned}$$

and we can observe that for a trivially small data set the two expressions are exactly the same.

This relation holds for arbitrarily large data sets. An exact, semi-formal derivation of this property is given in Appendix A.

IV. Powers of Two Series of ADEV as a Maximally Independent Statistic

The standard method for characterizing the stability of clocks (and signal sources in general) is to present the Allan deviation as a function of averaging period, particularly those which are powers of two of the sampling period. So far, this decision has not yet been fully motivated.

The reason these powers-of-two values make a powerful basis for describing the noise spectrum of a data set is that they are "near" independent. That is, they are not completely independent, but they are the most independent set of ADEV values that can be chosen. This can be seen by looking at the spectral response of each ADEV to the data being analyzed.

We can re-express the Allan variance as defined in Eq. (1) using a sampling function ${}^js$, which looks like 2^j terms of value 1 followed by 2^j terms of value -1 followed by zeros to fill in the resulting entries (such a sampling function is shown in Fig. 2). Then we can write

$$\text{AVAR}(2^j) \propto \sum_m \left|\sum_n y_n\, {}^js_{m+n}\right|^2 \tag{6}$$

because $\sum_n y_n\, {}^js_{m+n}$ is exactly $\bar{y}_{2^jm}(2^j)-\bar{y}_{2^j(m-1)}(2^j)$. The sum over n can be recognized as a discrete convolution

$$\text{AVAR}(2^j) \propto \sum_m \left|(y * {}^js)_m\right|^2.$$

We can then apply Parseval's theorem:

$$\text{AVAR}(2^j) \propto \sum_{m'} \left|\mathcal{F}(y * {}^js)_{m'}\right|^2,$$

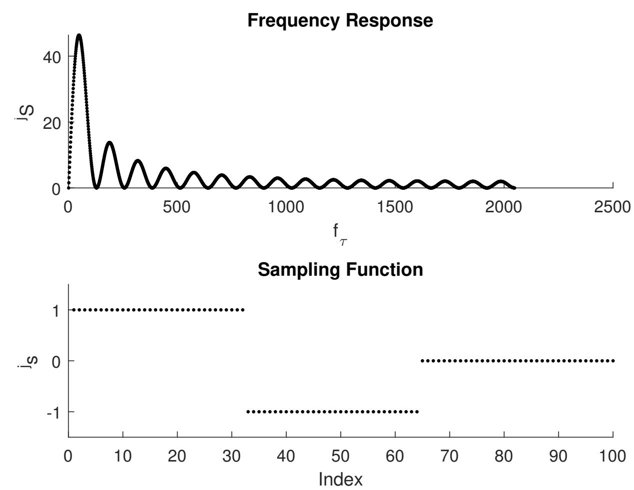

Fig. 2. The frequency response ${}^jS$ and the sampling function ${}^js$ for an ADEV measurement on a system of size 512 with $j=6$.

where the operation $\mathcal{F}$ represents the discrete Fourier transform (DFT) and the prime on m' indicates a summation over the discrete frequencies over the DFT. Applying the transform,

$$\text{AVAR}(2^j) \propto \sum_{m'} \left|(\mathcal{F}y\cdot {}^jS)_{m'}\right|^2, \tag{7}$$

where $\mathcal{F}y$ is the DFT of the data set y and ${}^jS$ is the DFT of the filter ${}^js$. As such, we can interpret $\text{ADEV}(2^j)$ as square root of the power contained in the product of the spectrum of the data and the response of the sampling function, so ${}^jS$ represents the frequency response of $\text{ADEV}(2^j)$.

The sampling function ${}^js$ and the corresponding spectral response function${}^jS$ are shown for $j=6$ on a system of length 512 points is shown in Fig. 2. Note that the transform of the sampling function has a zero at the frequency corresponding to the inverse of averaging period over which the ADEV is being taken because of aliasing effects at the sampling frequency. There are also zeros at each integer multiple of this frequency above this point, as the aliasing occurs periodically. These evenly spaced zeros and the occurrence of a single large peak are the features that make the powers of two avering periods nearly independent.

With each successive power-of-two in averaging period, the sampling frequency decreases by a factor of two, the first zero in the spectral response decreases by a factor of two. Similarly, the maximum of the response function also is scaled down by a factor of two. The result is that the maximum of the response function (up to a bias shift) for a given j value j_0 occurs at a frequency such that the response functions for all $j > j_0$ are zero at that frequency. The response maximum for $j = j_0$ lies on the quickly dying tail of the first peak of the response functions for $j < j_0$. Further, the same is true for all successive *local* maxima.

Because of this, if we consider a given frequency component in the data, the frequency will lie near the maximum of a

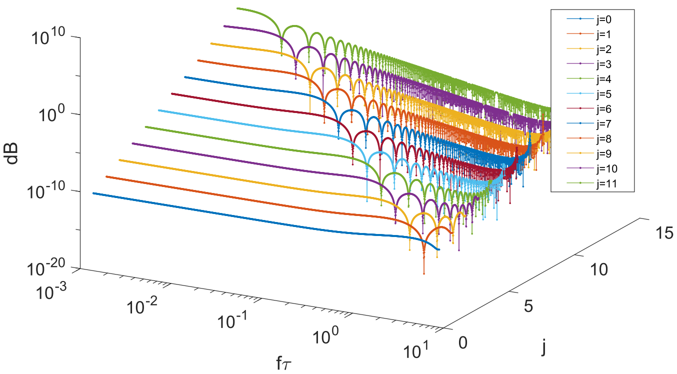

Fig. 3. The spectral response ${}^jS$ of the ADEV of all defined octaves j of averaging period for a system of length 512

single ${}^jS$. Meanwhile, it will be either near a zero or on a quickly dying tail for all other ${}^jS$. As such, the power in this noise frequency will nearly exclusively be picked up by a single ADEV measurement. This effect is demonstrated in Fig. 3, in which all the defined powers-of-two spectral responses for a system of length 512 are displayed in a layered log-log plot. The plot demonstrates the property of the maxima falling on zeros of higher order responses and tails of lower order responses, allowing near-independent spectral decomposition of the data in the different ADEV measurements.

This feature of overlapping zeros occurs when the sampling function is scaled by I^j where I is an integer and j represents successive terms (in practice $I = 2$ is the most useful because it gives the most ADEV values given a finite set of data). To see why this is, first recognize that ${}^jS$ and ${}^js$ are discrete samplings of continuous functions ${}^j_cS$ and ${}^j_cs$, where the leading subscript c denotes continuity. Next, scaling the sampling function by a factor of I^j takes

$$ {}^j_c s \rightarrow {}^j_c s\left(\frac{t}{I^j}\right). $$

Then by the scaling theorem, taking the Fourier transform gives

$$ \mathcal{F}\left\{ {}^j_c s\left(\frac{t}{I^j}\right)\right\} = I^j \cdot {}^j_c S(\omega I^j), \tag{8} $$

so the zeros become a factor of I^j closer together. If the zeros of the sampling function are evenly spaced, then the zeros for j will lie on the zeros for $j+1$ and $j+2$ and so on.

This is generally true for any similar statistic where the frequency response of the sampling function has evenly spaced zeros. For example, the Hadamard variance exhibits this property and is commonly used with powers-of-three averaging periods. [8], [9], which also exhibit this property.

V. Conclusion

We have shown that the set of AVAR/ADEV statistics have useful properties when taken in series where the averaging periods are consecutive powers of two. The near independence of these measurements in terms of frequency domain power sampling makes them a powerful tool in determining power law noise types that exist in data sets. While all powers-of-integer series exhibit this near-independence, powers of two sampling periods allow for the most information to be squeezed out of a finite data set because it produces the most statistical measurements per data point, making it the most useful of these series. This is especially the case when we wish to extract long-term behavior from shorter data sets.

Furthermore, we have proved that the powers-of-two averaging period series exhibit the interesting property that for the non-overlapping variant of the statistic, all the defined powers-of-two measurements sum up to exactly twice the standard variance. In some cases, this can be a useful property in the reduction of error estimates on AVAR and its slope, since the standard variance itself generally has very low fractional uncertainty compared to AVAR measurements. These properties combined help motivate the standard practice in the field and lead to a compelling justification for the use of the power-of-two series when using Allan variance to analyze data.

Appendix A.

Here we will prove the relation given by Eq. (5):

$$ \sum_{j=0}^{J-1} \mathrm{AVAR}_{nono}(2^j) = 2 \cdot \mathrm{SVAR} $$

Let $N = 2^J$ be the number of items in the data set. The definition of the standard variance tells us

$$ 2\cdot\mathrm{SVAR} = \frac{2}{N}\Bigg[\left(y_1 - \frac{1}{N}(y_1+y_2+...+y_N)\right)^2 + \left(y_2 - \frac{1}{N}(y_1+y_2+...+y_N)\right)^2 + ... + \left(y_N - \frac{1}{N}(y_1+y_2+...+y_N)\right)^2\Bigg]. $$

Pulling out a factor of $\frac{1}{N^2}$ and combining like terms,

$$ 2\cdot\mathrm{SVAR} = \frac{2}{N^3}\Big[(y_1(N-1) - y_2 - ... - y_N)^2 + (-y_1 + y_2(N-1) - ... - y_N)^2 + ... + (-y_1 - y_2 - ... + y_N(N-1))^2\Big]. \tag{A.1} $$

Since we only have differences of terms squared in the expression, the result will be a sum over multiples of two terms of the data set. We can write such a sum generally with undetermined coefficients

$$2 \cdot \text{SVAR} = \sum_i A_i y_i^2 + \sum_i \sum_{l<i} B_{il} y_i y_l. \qquad \text{(A.2)}$$

First consider A_i. We can see in Eq. (A.1) that for a given y_i^2, in the sum over squared terms it will appear once in the squared term including $y_i(N-1)$ and $N-1$ more times in the rest of the squared terms. Thus we have

$$A_i = \frac{2}{N^3}\left((N-1)^2 + (N-1)(-1)^2\right) = \frac{2}{N^2}(N-1). \qquad \text{(A.3)}$$

Next consider the cross term coefficients B_{il}. A given $y_i y_l$ will occur once in the squared term with $y_i(N-1)$, once in the squared term with $y_l(N-1)$ and $N-2$ more times in the other squared terms. Thus we have

$$B_{il} = \frac{2}{N^3}(2(N-1)(-1) + (N-2)(-1)^2) = \frac{-2}{N^2}. \qquad \text{(A.4)}$$

Next, let us do the same for AVAR_{nono}. Recall the definition of $\text{AVAR}_{nono}(2^j)$ as given in Eq. (1):

$$\text{AVAR}_{nono}(2^j) \equiv \frac{1}{2^{J-j}} \sum_{k=1}^{2^{J-j-1}} \left(\bar{y}_{2^{j+1}k}(2^j) - \bar{y}_{2^j(2k-1)}(2^j)\right)^2.$$

We aim to analyze the sum over all defined powers of two averaging periods

$$\sum_{j=0}^{J-1} \text{AVAR}_{nono}(2^j).$$

Once again, the difference inside the square is a linear combination of terms, so the sum over these squares will be another sum of multiples of two terms of the data set. We can then write in the same form as Eq. (A.2):

$$\sum_{j=0}^{J-1} \text{AVAR}_{nono}(2^j) = \sum_i A'_i y_i^2 + \sum_i \sum_{l<i} B'_{il} y_i y_l. \qquad \text{(A.5)}$$

First consider the A'_i. For a given $\text{AVAR}_{nono}(2^j)$, the term y_i will appear in a single $\bar{y}(2^j)$ of the sum over k (since the $\bar{y}(2^j)$ are non-overlapping by definition), which is normalized by $\frac{1}{2^j}$, so when squared we get a factor of $\frac{1}{2^{2j}} y_i^2$. Including the outer normalization of $\frac{1}{2^{J-j}}$, we get a total contribution over the entire sum

$$A'_i = \sum_{j=0}^{J-1} \frac{1}{2^{J-j}} \frac{1}{2^{2j}} = 2^{-2J+1}(2^J - 1) = \frac{2}{N^2}(N-1). \qquad \text{(A.6)}$$

Finally, consider B'_{il}. Since we only care about $i \neq l$, there will always be a minimum l such that y_i and y_l both occur in a single $\left(\bar{y}_{2^{j+1}k}(2^j) - \bar{y}_{2^j(2k-1)}(2^j)\right)^2$, which we will call $l = m$. When $l = m$, one of y_i and y_l will be in one each of $\bar{y}_{2^{j+1}k}(2^j)$ and $\bar{y}_{2^j(2k-1)}(2^j)$. Therefore the term $y_i y_l$ will appear as a negative cross term, so the contribution from $\text{AVAR}_{nono}(2^m)$ is $-\frac{1}{2^{2m}}$. Then for $j > m$, both y_i and y_l will occur in the same average term $\bar{y}(2^j)$, so the cross term will be positive, and the contribution is then $+\frac{1}{2^{2j}}$. Including each outer normalization of $\frac{1}{2^{J-j}}$, we get a total contribution over the entire sum

$$\begin{aligned} B'_{il} &= \frac{1}{2^{J-m}}\left(-\frac{1}{2^{2m}}\right) + \sum_{j=m+1}^{J-1} \frac{1}{2^{J-j}}\left(+\frac{1}{2^{2j}}\right) \\ &= -2^{-J-m} + 2^{-2J-m}(2^J - 2^{m+1}) \\ &= -2^{1-2J} \\ &= -\frac{2}{N^2}. \end{aligned} \qquad \text{(A.7)}$$

Thus we can see from Eq. (A.3) and Eq. (A.6) and from Eq. (A.4) and Eq. (A.7) that

$$\begin{cases} A_i = A'_i \\ B_{il} = B'_{il} \end{cases} \qquad \text{(A.8)}$$

so

$$\sum_{j=0}^{J-1} \text{AVAR}_{nono}(2^j) = 2 \cdot \text{SVAR}. \qquad \text{(A.9)}$$

References

[1] J. A. Barnes, A. R. Chi, L. S. Cutler, D. J. Healy, D. B. Leeson, T. E. McGunigal, J. A. Mullen, Jr., W. L. Smith, R. L. Sydnor, R. F. C. Vessot, G. M. R. Winkler, Characterization of frequency stability, IEEE Trans. Instrum. Meas., IM-20, pp. 105-120, 1971.

[2] D.A. Howe, D.W. Allan, J.A. Barnes, Properties of Signal Sources and Measurement Methods, Proc IEEE International Symp on Freq Control, pp. A1-A47, May 1981.

[3] J. McGee and D.A. Howe, TheoH and Allan Deviation as Power-Law Noise Estimators, IEEE Trans. UFFC, vol. 54, no. 2, 448-454, Feb. 2007.

[4] D. B. Percival, D.A. Howe. Total Variance as an Exact Analysis of the Sample Variance, 29th Annual Precise Time and Time Interval (PTTI) Meeting, Dec. 1997

[5] D.A. Howe, The Total Deviation Approach to Long-Term Characterization of Frequency Stability, IEEE Trans. Ultrasonics, Ferroelectrics and Freq. Control., Vol. UFFC-47, No. 5, pp. 1102-1110, Sept. 2000.

[6] P.R. Gray, P.J. Hurst, S.H. Lewis, R.G. Meyer Analysis and design of analog integrated circuits (4th edn.), John Wiley and Sons, New York, USA, 2001

[7] N. Ashby, Probability Distributions and Confidence Intervals for Simulated Power Law Noise, IEEE T. Ultrason. Ferr, Vol. 62, No. 1, pp. 116-128, Jan. 2015

[8] W. J. Riley, Handbook of Frequency and Stability Analysis, NIST Special Publication 1065, 2008.

[9] D. A. Howe, R. Beard, C. Greenhall, F. Vernotte, B. Riley, Total Hadamard Variance: Application to Clock Steering by Kalman Filtering, IEEE T. Ultrason. Ferr, Vol. 52, pp.1253-1261, Aug. 2005