2014 IEEE 28-th Convention of Electrical and Electronics Engineers in Israel

MSE Reduction in Digital Compensation for Non-Linear Analog Channels

Amir Weiss, Arie Yeredor and Mark Shtaif
School of Electrical Engineering, Tel-Aviv University, Israel
amirwei2@mail.tau.ac.il, arie@eng.tau.ac.il, shtaif@eng.tau.ac.il

***Abstract*—Unlike linear analog channels, for which a significant number of well-studied, mostly simple equalization methods exist, non-linear analog channels pose a much more challenging task regarding the channel equalization, be it by means of analog and / or of digital operations. In this work we present a method for digital pre-distortion compensation for memoryless non-linear channels, which yields a reduction, compared to naïve digital methods, in the Time-Averaged Mean Squared Error (TA-MSE) between the desired and actual pre-distorted signals.**

***Index Terms*—Channel Equalization, Channel Compensation, Pre-distortion, Non-Linear Channels, Mean Squared Error (MSE) Reduction.**

I. Introduction

Quite often in communication systems the transmitted analog information signal experiences a physical channel, which, if sufficient knowledge exists, can be modeled by a mathematical operator, either linear or non-linear. While linear channels have been studied intensively and many effective and efficient methods are known for their equalization, the equalization of non-linear channels is considerably more difficult, even for the apparently simplest types of nonlinearities, such as memoryless nonlinearities, which are addressed in this paper.

Theoretically, invertible memoryless nonlinearities can be easily compensated for by applying the inverse nonlinear function to the modulated signal (i.e., in the analog domain). The feasibility of this approach is however very limited in practice because of the limited repertoire of non-linearities that can be applied using high-speed analog devices at acceptable costs. It is therefore desirable to design a solution that can be implemented in the digital domain prior to the digital to analog converter (DAC). Nonetheless, this task is not entirely trivial because (memoryless) non-linear operations on an analog signal generally tend to increase its bandwidth. Therefore, an analog signal generated from samples which have been distorted by a given nonlinear operation does not coincide with the analog, continuous-time version of the original analog signal distorted by the same nonlinear operator, except at the sample instances. The differences in the analog signals in-between the sample instances might be significant and undermine the compensation scheme. Of course, a straightforward remedy can be to increase the sampling rate, but such a solution cannot be applied in systems which already push the sample rate to its upper feasibility limit, such as digital optical communication systems, for which a compensation for non-linearities is indeed needed and widely discussed, e.g., in [1], [2], [3], [4], [5].

Quite a few approaches for pre-compensation for memoryless nonlinear channels have been proposed over the past few decades. For example, Tsimbinos and Lever [6] suggest a compensation technique based on orthogonal polynomial inverses, which overcomes some of the problems associated with power series inverses, though a specific implementation scheme is not discussed in there. Another example can be found in [7], where Chang and Powers present a simplified pre-distortion scheme to compensate nonlinear distortion introduced by the high power amplifier in the context of Orthogonal Frequency-Division Modulation (OFDM) systems. In this paper we consider a Volterra-like pre-distortion system, which generates each sample (of the pre-distorted digital version of the analog signal) as a linear combination of products (up to fixed orders) of samples of the original signal within a fixed sliding window frame. The fixed parameters (coefficients) of this linear combination are optimized under a criterion of minimum Time-Averaged Mean Squared Error (TA-MSE), which will be defined and explained in the sequel. We note that this criterion is chosen due to the cyclostationarity property of the error process. The derivation assumes the following framework:

- The memoryless non-linear channel $h(\cdot)$ has a known inverse $g(\cdot) \triangleq h^{-1}(\cdot)$;
- The information signal $s(t)$ is a (strict-sense) stationary random process with a known distribution (e.g., a zero-mean stationary Gaussian process with a known Power Spectrum Density (PSD)).

These assumptions are sufficient in order to define a minimization problem (in terms of the proposed compensation system's parameters), in which the objective function is the TA-MSE between the desired and actual pre-distorted signal produced by the system. Under several simplifying assumptions, a closed-form analytical expression for the solution is presented, along with a simple example.

II. Problem Formulation

A. Preliminaries

Let us assume an information signal $s(t)$ and a memoryless, invertible non-linear analog channel $h(\cdot)$, such that the channel's response to $s(t)$ is $y(t) = h(s(t))$. Ideally, in order to invert the channel, we would like to apply $g(\cdot)$, either to the information signal $s(t)$ before transmission (pre-distortion) or to the received signal $y(t)$. This work is focused on pre-distortion.

Quite often, however, $g(\cdot)$ is not a trivial function, and hence might be too difficult to apply by means of analog circuits. We would therefore like to design a system in the digital domain, which would receive a sampled version of $s(t)$, denoted here by $s[n]$, as an input, and would produce an output $\hat{d}[n]$, a sampled version of an analog signal $\hat{d}(t)$ (to be reproduced from $\hat{d}[n]$ using an ideal DAC), such that $\hat{d}(t)$ is the best approximation, in some sense, to the desired signal $d(t) \triangleq g(s(t))$. We assume that $s(t)$ is a band-limited signal sampled at its Nyquist rate (which equals its two-sided bandwidth). Alternatively, $s(t)$ may already be given in terms of its Nyquist-rate samples $s[n]$, e.g., in the context of OFDM. Assuming ideal Nyquist rate reconstruction we have

$$s(t) = \sum_{n=-\infty}^{\infty} s[n]p(t - nT_s) \tag{1}$$

$$\hat{d}(t) = \sum_{n=-\infty}^{\infty} \hat{d}[n]p(t - nT_s), \tag{2}$$

where T_s denotes the sampling interval and $p(t) \triangleq \mathrm{sinc}(t/T_s)$ is the ideal reconstruction kernel. Here the $\mathrm{sinc}(\cdot)$ function is defined as $\mathrm{sinc}(x) = \sin(\pi x)/\pi x$.

For simplicity of the exposition we shall assume, from now on, that all signals involved are real-valued.

B. Cyclostationarity of the Error Process

Ideally, we would like to get $\hat{d}(t) = d(t) = g(s(t))$ for perfect pre-distortion. However, since non-linear operations generally tend to increase the signal's bandwidth, one cannot hope to perfectly reconstruct $d(t)$, which generally has a wider bandwidth, from its samples $d[n]$ generated at the Nyquist rate of $s(t)$, which is generally smaller than the Nyquist rate of $d(t)$. Let us define the error process $e(t) \triangleq \hat{d}(t) - d(t)$. Given the (strict-sense) stationarity of $s(t)$ and the time-invariance of the memoryless operation $g(\cdot)$, the desired $d(t)$ is also strict-sense stationary. We further assume that the generation of samples $\hat{d}[n]$ from the samples of $s[n]$ takes a discrete-time-invariant form, namely, the signal $\hat{d}[n]$ is generated by applying a general (possibly non-instantaneous) operator $G(\{s(n)\})$, such that if the operator's response to a signal $s_1[n]$ is $\hat{d}_1[n] = G(\{s_1[n]\})$, its response to a time-shifted signal $s_2[n] = s_1[n - \ell]$ is the similarly shifted $\hat{d}_2[n] = G(\{s_2[n]\}) = \hat{d}_1[n - \ell]$ (for all ℓ). Under this assumption (combined with the stationarity of $s(t)$), it is easy to show that the reconstructed $\hat{d}(t)$ in (2) is cyclostationary with cyclic period T_s, and is therefore jointly cyclostationary (with the same cyclic period) with $s(t)$ and thus also with $d(t)$ (recall that any stationary signal is also cyclostationary with any cyclic period). Therefore, the error process $e(t)$ is also cyclostationary with the same cyclic period T_s. In the following subsection we will see that this property plays a key role in defining the optimality criterion.

C. Optimality Criterion

Let us denote the autocorrelation function of the error process $e(t)$ as

$$R_{ee}(t;\tau) = E[e(t + \tau)e(t)]. \tag{3}$$

The instantaneous power of $e(t)$ (which is also the instantaneous MSE) is given by $E[e^2(t)] = R_{ee}(t; 0)$. Due to the cyclostationary with cyclic period T_s, we have $R_{ee}(t;\tau) = R_{ee}(t + T_s;\tau)$ for all t and τ. Obviously, it is possible to get $R_{ee}(nT_s; 0) = 0$ for all integer n (namely, at the sampling instances) by choosing $G(\{s(n)\}) = g(s[n])$, however this may come at the expense of a larger MSE for other values of t, in-between samples. We therefore propose to choose the operator $G(\{s(n)\})$ (out of a pre-defined family of operators under consideration) so as to minimize the Time-Averaged MSE (TA-MSE)

$$P_e \triangleq \frac{1}{T_s}\int_0^{T_s} R_{ee}(t; 0)dt. \tag{4}$$

We say that the operator minimizing P_e yields the best pre-distorted signal in terms of TA-MSE. In the next section we consider a particular parameterized family of operators and derive a solution to get the optimal pre-distortion attainable by this family.

III. Optimal Digital Pre-Distortion

Generally, there is an infinite repertoire of (time-invariant) operations that the operator $G\{s[n]\}$ can represent. In order to decrease the complexity of the problem, i.e., reduce the solution space, but at the same time preserve efficient, effective and satisfying solutions (significantly reducing the TA-MSE), we propose to use the following Volterra-like structure: The output at time-instant n of the operator $G\{s[n]\}$ consists of a linear combination of all possible products (up to a fixed order) among the group of samples consisting of $s[n]$, as well as of symmetrically neighboring samples up to a fixed distance. Note that for the sake of simplicity we allow $G\{s[n]\}$ to possibly be non-causal - in practice this can be realized by incorporating a fixed delay. More specifically, let $K+1$ denote the desired memory-span of the operator, and let $K_- \triangleq \lceil \frac{K}{2} \rceil$ and $K_+ \triangleq \lfloor \frac{K}{2} \rfloor$ denote the "past" and "future" lengths, respectively, such that the operator's output at time-instant n depends on the $K+1$ samples $s[n-K_-], \ldots, s[n], \ldots, s[n+K_+]$. To this end, let us define the $(K + 2) \times 1$ vector

$$\bar{\mathbf{s}}[n] \triangleq \left[1 \;\; s[n - K_-] \;\; s[n - K_- + 1] \;\; \cdots \;\; s[n + K_+]\right]^T. \tag{5}$$

Now, let D denote the operator's order and let $\mathbf{s}_n$ denote a vector consisting of products of all possible (unordered) combinations of D elements (with repetitions) of $\bar{\mathbf{s}}[n]$.

For example, for $K = 1$ and $D = 2$ we have

$$\bar{\mathbf{s}}[n] = \left[1 \;\; s[n-1] \;\; s[n]\right]^T, \tag{6}$$

$$\mathbf{s}_n = \begin{bmatrix}1 & s[n-1] & s[n] & s^2[n-1] & s^2[n] & s[n-1]s[n]\end{bmatrix}^T. \quad (7)$$

Likewise, for $K=2$ and $D=2$ we have

$$\bar{\mathbf{s}}[n] = [1 \;\; s[n-1] \;\; s[n] \;\; s[n+1]]^T, \quad (8)$$

$$\mathbf{s}_n = \begin{bmatrix}1 & s[n-1] & s[n] & s[n+1] & s^2[n-1] & s^2[n] & s^2[n+1] \ldots \\ s[n-1]s[n] & s[n-1]s[n+1] & s[n]s[n+1]\end{bmatrix}^T. \quad (9)$$

The operator $G\{s[n]\}$ is then defined by a linear combination of the elements of $\mathbf{s}_n$,

$$\hat{d}[n] = G\{s[n]\} = \boldsymbol{\theta}^T \mathbf{s}_n \quad (10)$$

where $\boldsymbol{\theta}$ is a vector of fixed parameters (coefficients), to be optimized so as to minimize the TA-MSE subject to the particular structure,

$$\boldsymbol{\theta}_{opt} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \frac{1}{T_s} \int_0^{T_s} E[(\hat{d}(t) - d(t))^2]dt \triangleq \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \, \mathcal{E}(\boldsymbol{\theta}). \quad (11)$$

Substituting (2) and (10), by simple expansion the objective function can also be expressed as

$$\mathcal{E}(\boldsymbol{\theta}) = \boldsymbol{\theta}^T \mathbf{A} \boldsymbol{\theta} - 2\mathbf{b}^T \boldsymbol{\theta} + \gamma \quad (12)$$

with

$$\mathbf{A} \triangleq \frac{1}{T_s} \int_0^{T_s} \mathbf{A}(t)dt, \;\; \mathbf{b} \triangleq \frac{1}{T_s} \int_0^{T_s} \mathbf{b}(t)dt, \quad (13)$$

where

$$\mathbf{A}(t) \triangleq \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} E[\mathbf{s}_n \mathbf{s}_m^T] p(t-nT_s) p(t-mT_s), \quad (14)$$

$$\mathbf{b}(t) \triangleq \sum_{n=-\infty}^{+\infty} E[\mathbf{s}_n d(t)] p(t-nT_s) \quad (15)$$

and γ is a constant independent of $\boldsymbol{\theta}$.

Thus, the objective function in (11) is quadratic in $\boldsymbol{\theta}$. Therefore, if $\mathbf{A}$ is positive-definite, a unique minimizer is given by

$$\boldsymbol{\theta}_{opt} = \mathbf{A}^{-1}\mathbf{b}, \quad (16)$$

yielding a closed form solution of the optimal coefficients vector (under a fixed polynomial order D and memory span $K+1$) for the pre-distortion DSP operation.

The principal difficulties in finding the optimal coefficients are in obtaining the required means $E[\mathbf{s}_n \mathbf{s}_m^T]$ and $E[\mathbf{s}_n d(t)]$ and in applying the integrations. As we shall see, under reasonable assumptions on the signal's probability distribution, the computations of the required means, as well as the required integrations, take a convenient form. We note, however, that when the required integrations in (13) are too complicated, a fall-back option, which involves more simple calculations, is to define the optimality criterion as

$$\boldsymbol{\theta}_{opt} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \, E[e^2(T_s/2)], \quad (17)$$

motivated by the reasonable assumption that the deviation of $\hat{d}(t)$ from the desired $d(t)$ would usually take its maximal mean squared value in the mid-point between samples. This hypothesis was verified by simulations and can be observed in Fig. (2).

In the next section we demonstrate our method in a simple case which shows, theoretically and by simulations, a significant reduction in the TA-MSE compared to naïve application of the non-linear, memoryless inverse operator $\hat{d}[n] = g(s[n])$ as pre-distortion.

IV. A Solution for Gaussian Signals and Polynomial Channel Inverses

We examine a simple case assuming a polynomial as the non-linear inverse (memoryless) operator

$$g(x) = \sum_{m=0}^{M} a_m x^m \quad (18)$$

and a zero mean, unit variance Gaussian process with a rectangular PSD as an input signal to our system,

$$R_{ss}(\tau) = \operatorname{sinc}(\tau) \overset{\mathcal{F}}{\longleftrightarrow} S_{ss}(f) = \operatorname{rect}(f) \quad (19)$$

where $\mathcal{F}$ denotes the Fourier transform and $\operatorname{rect}(f)$ equals 1 for $|f| \le \frac{1}{2}$ and 0 otherwise.

Fortunately, the Gaussianity assumption simplifies (and enables) the analytical computation of the matrix $\mathbf{A}(t)$ and thus brings us closer to a closed form analytical solution for $\boldsymbol{\theta}_{opt}$. The PSD function also has a significant influence on the calculations, and the rectangular function simplifies them even further, facilitating the integration and yielding a closed form expression for the matrix $\mathbf{A}(t)$. Since $s(t)$ has a rectangular PSD, sampling at the Nyquist rate we get $s[n] = s(n \cdot T_s) = s(n \cdot 1)$, with

$$R_{ss}[\ell] = \delta[\ell] \overset{\mathcal{F}}{\longleftrightarrow} S_{ss}(e^{j\omega}) = 1 \quad \forall \omega \quad (20)$$

where $\delta[\ell]$ stands for Kronecker's delta function. The Gaussianity implies that the samples of the discrete-time process $s[n]$ are independent, identically distributed (iid), which simplifies the computation of $\mathbf{A}$. In Appendix A we show that in this particular case we get

$$\mathbf{A} = E[\mathbf{s}_n \mathbf{s}_n^T], \quad (21)$$

and the elements of this matrix can be easily obtained by exploiting the iid structure of the process $s[n]$, as well as the well-known high-order moments of a zero-mean, unit-variance Gaussian distribution: $E[s^m[n]]$ equals $1 \cdot 3 \cdot 5 \cdots (m-1)$ when m is even and vanishes when m is odd.

For a particular example we consider $g(\cdot)$ to be $g(x) = \alpha x + \beta x^2$, hence the desired ideal pre-distorted signal is

$$d(t) = g(s(t)) = \alpha s(t) + \beta s^2(t) \quad (22)$$

For simplicity, we set $\alpha = 1$ since we are interested in the effect of the non-linear element anyhow. We apply our operator with $D=2$ and $K=2$, namely second-order terms with memory span of three samples ($s[n-1]$, $s[n]$ and $s[n+1]$). As indicated above, in this case $\mathbf{s}_n$ is a ten-elements vector specified in (9).

From (21), the computation of $\mathbf{A}$ for this case is quite simple. We will now address the computation of the vector $\mathbf{b}$. Each of the elements of the vector $\mathbf{b}(t)$ requires computation of the expectation of the product of jointly Gaussian Random Variables (RV), since $s[n]$ is actually $s(n \cdot 1)$, which is jointly Gaussian with $s(t_0)$ for any t_0. Using Isserlis' theorem, regarding the computation of higher-order moments of the multivariate Gaussian distribution [8], each of the elements can be expressed in terms of the autocorrelation function of $s(t)$. Due to space limitations we omit the general derivation; However, as an example, for obtaining the 5^{th} element of $\mathbf{b}$ we first obtain

$$\begin{aligned} E[s^2[n-1]d(t)] &= E[s^2[n-1](s(t)+\beta s^2(t))] \\ = \beta \left(E[s^2((n-1)\cdot 1)]E[s^2(t)] + 2E^2[s((n-1)\cdot 1)s(t)]\right) \\ = \beta \left(R_{ss}^2(0) + 2R_{ss}^2(t-(n-1)\cdot 1)\right) \\ = \beta \left(1 + 2R_{ss}^2(t-n+1)\right). \end{aligned} \tag{23}$$

Substituting into the second term of (13) we get

$$(\mathbf{b})_5 = \int_0^1 \sum_{n=-\infty}^{+\infty} \beta \left(1 + 2R_{ss}^2(t-n+1)\right) \operatorname{sinc}(t-n)\, dt. \tag{24}$$

Further substituting the integration variable to $\tau = t - n$, using $R_{ss}(\tau) = \operatorname{sinc}(\tau)$, and writing the infinite sum of integrals of adjacent intervals as an infinite integral we get

$$(\mathbf{b})_5 = \beta \int_{-\infty}^{+\infty} \left(1 + 2\operatorname{sinc}^2(\tau+1)\right) \operatorname{sinc}(\tau) d\tau. \tag{25}$$

This integral has a closed-form analytic solution, yielding

$$(\mathbf{b})_5 = \beta \left(1 + \frac{1}{\pi^2}\right). \tag{26}$$

The other elements are computed in a similar way.

Fig. (1) shows the analytic and simulation results for the TA-MSE comparison of a naïve DSP operation (memoryless application of the inverse non-linearity to the sample, i.e., $\hat{d}[n] = g(s[n])$) vs. our optimized solution, for three different values of K. When the product of the non-linear element's coefficient β with the signal's standard deviation ($\sigma_{s(t)} = \sqrt{R_{ss}(0)} = 1$) is comparable to α, a significant relative improvement is achieved as shown in Fig. (1), e.g., for $\beta = 1$, where a reduction of up to $\sim$30% in the TA-MSE is attained by the proposed method. As expected, the simulation results validate the analytical solution. It can be seen that as the system's memory span $K+1$ grows, the performance of the optimized system improves, and a more significant reduction is the TA-MSE is attained. Fig. (2) shows a single time period of the MSE, i.e., a single sampling period of the MSE, produced by Monte Carlo simulation. As can be seen, the maximum value of the MSE of all four systems is exactly in the middle of the sampling period, as expected. Moreover, the profile of a single period of the MSE can provide some intuition about the proposed method: Unlike the naïve digital operation, which yields an analog signal at the DAC's output that coincides exactly with the ideal desired pre-distorted signal (only) at the sample instances but suffers from a significant MSE in the middle of the sampling period (the farthest time-distance from each sample), the optimal system "prefers" the cost of a non-zero, though minor, MSE at the samples instances for the profit of a smaller TA-MSE (the area under the curve).

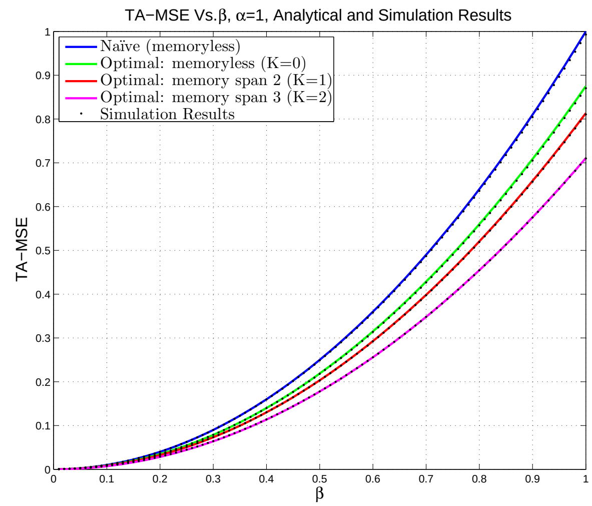

Fig. 1. Analytical and simulation results for the TA-MSE of the naïve method and the optimal method with 3 different system memory spans: memoryless ($K = 0$), two ($K = 1$), and three ($K = 2$). Simulation results are based on averaging 30000 independent trials.

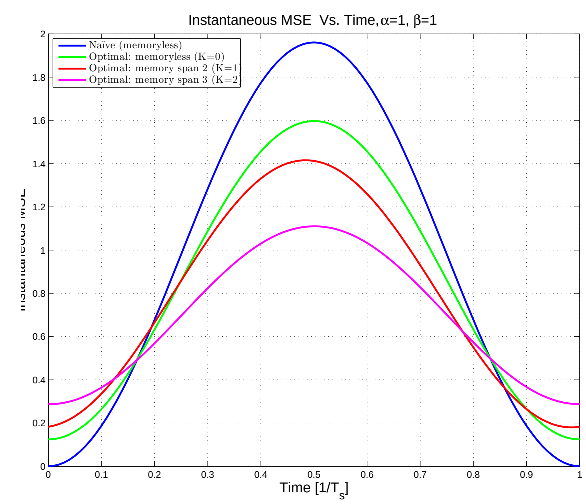

Fig. 2. A single time period of the MSE $R_{ee}(t;0)$, based on averaging 30000 independent trials.

V. Conclusion

We proposed a method for MSE reduction in a digital operation for the compensation of memoryless non-linear analog channels. A closed form analytical optimal solution has been presented, under fixed system parameters which were discussed in section II, and a proof of concept has been given in section IV for the case of a 2^{nd} order polynomial as a model for the inverse channel, and a Gaussian signal with a rectangular PSD function as a model for the signal of interest. We note that the polynomial model of the inverse non-linear operation is an important and practical model since many non-linear functions can be approximated to sufficient precision by a polynomial function, at least over a certain interval (which can be characterized in many real systems). A Gaussian model for a signal of interest is also a widely used, often reasonable assumption. Notice that as the signal's autocorrelation function's effective length increases, it is desirable and useful to take into account farther time-distance samples for $\mathbf{s}_n$ (namely, increase K), using information that lies in the relations between the samples and the values of the interpolated analog signal at time instances nearby. Likewise, it is desirable and useful to include higher-order moments in $\mathbf{s}_n$ (namely, increase D) when approximating an analytic function, which can be represented as a power series. Nevertheless, the system parameters K and D should be chosen conservatively, since the computational complexity of the vector $\boldsymbol{\theta}_{opt}$ grows combinatorially.

Appendix A
The Matrix $\mathbf{A}$ for a Gaussian Signal with Rectangular PSD Function

Since $s[n]$ is Gaussian and stationary $E[\mathbf{s}_n\mathbf{s}_m^T]$ is a function of $n-m$. Define the matrix

$$\boldsymbol{\Psi}[n-m] \triangleq E[\mathbf{s}_n\mathbf{s}_m^T] \tag{27}$$

Now, after the summation index substitution $\ell = n-m$, (14) can be rewritten as

$$\mathbf{A}(t) \triangleq \sum_{n=-\infty}^{+\infty}\sum_{\ell=-\infty}^{+\infty} \boldsymbol{\Psi}[\ell]\mathrm{sinc}\left(\frac{t-nT_s}{T_s}\right)\mathrm{sinc}\left(\frac{t-(n-\ell)T_s}{T_s}\right), \tag{28}$$

where T_s is the Nyquist-rate sampling period, which equals the reciprocal of the signal's double-sided bandwidth.

Substituting (28) into the first expression in (13), and substituting the integration variable by

$$\tau = \frac{t-nT_s}{T_s} \tag{29}$$

we obtain

$$\mathbf{A} = \sum_{\ell=-\infty}^{+\infty} \boldsymbol{\Psi}[\ell] \int_{-\infty}^{+\infty} \mathrm{sinc}(\tau)\mathrm{sinc}(\tau+\ell)d\tau = \sum_{\ell=-\infty}^{+\infty} \boldsymbol{\Psi}[\ell]\mathrm{sinc}(\ell) \tag{30}$$

(where we exploited the orthogonality of integer-shifted $\mathrm{sinc}(\cdot)$ functions in the transition). Since

$$\ell \in \mathbb{Z} : \mathrm{sinc}(\ell) = \begin{cases} 1 & \text{if } \ell = 0 \\ 0 & \text{if } \ell \neq 0 \end{cases}$$

we finally get

$$\mathbf{A} = \boldsymbol{\Psi}[0] = E[\mathbf{s}_n\mathbf{s}_n^T]. \tag{31}$$

References

[1] Arthur J Lowery. Fiber nonlinearity pre-and post-compensation for long-haul optical links using ofdm. *Optics Express*, 15(20):12965–12970, 2007.

[2] Yanir London and Dan Sadot. Nonlinear effects mitigation in coherent optical ofdm system in presence of high peak power. *Lightwave Technology, Journal of*, 29(21):3275–3281, 2011.

[3] Arthur J Lowery. Fiber nonlinearity mitigation in optical links that use ofdm for dispersion compensation. *Photonics Technology Letters, IEEE*, 19(19):1556–1558, 2007.

[4] Robert Killey. Dispersion and nonlinearity compensation using electronic predistortion techniques. 2005.

[5] René-Jean Essiambre, Peter J Winzer, Xun Qing Wang, Wonsuck Lee, Christopher A White, and Ells C Burrows. Electronic predistortion and fiber nonlinearity. *IEEE photonics technology letters*, 18(17-20):1804–1806, 2006.

[6] John Tsimbinos and Kenneth V Lever. Nonlinear system compensation based on orthogonal polynomial inverses. *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, 48(4):406–417, 2001.

[7] Sekchin Chang and Edward J Powers. A simplified predistorter for compensation of nonlinear distortion in ofdm systems. In *Global Telecommunications Conference, 2001. GLOBECOM'01. IEEE*, volume 5, pages 3080–3084. IEEE, 2001.

[8] Leon Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, pages 134–139, 1918.