On Correct Understanding and Classification of Saleh’s and Related Models of AM/AM and AM/PM Conversions

Andrzej Borys

Abstract—In this paper, some of the existing classifications of the Saleh’s and related models of the AM/AM and AM/PM conversions occurring in communication power amplifiers (PA) are reviewed. It is shown that these classifications are inconsistent and must be refined. Obviously, carrying out such a refinement properly needs a good knowledge and correct understanding of the mechanisms leading to the AM/AM and AM/PM conversions. This was achieved in this paper by performing a thorough analysis of the PA behavior using an analytical tool, the Volterra series. The main points of this analysis are presented here in great detail. Among others, it is shown that the influence of the PA memory on the AM/PM conversion is two-fold: direct and indirect. The former can be however fully neglected. On the other hand, the indirect influence caused by “the interaction of the carrier with the PA memory” cannot be neglected when the PA has not enough wideband frequency characteristics. The latter effect mentioned causes changes in the carrier phase that are received as the phase changes of the baseband modulating signal.

Keywords—Saleh’s and related models of AM/AM and AM/PM conversions and their classifications, power amplifiers, memory effects

I. INTRODUCTION

POWER amplifiers (PAs) used in wireless communications, such as satellite communication systems [1], produce nonlinear distortions in their operating regimes, which are selected so as to be slightly nonlinear. These distortions manifest themselves in the occurrence of the so-called amplitude-to-amplitude (AM/AM) and amplitude-to-phase (AM/PM) conversions, among other distortion products. Our focus here is on the AM/AM and AM/PM distortions.

One of the most popular models used to model the AM/AM and AM/PM distortions of PAs was published by Saleh in [2]. It is one of the representatives of a whole class of descriptions modeling the AM/AM and AM/PM conversions, which are listed and classified as “system-level models without memory effect” in a tutorial paper [3].

The work presented in this paper was supported partly by the grant AMG DS/450/2018.

The author is with the Department of Marine Telecommunications, Faculty of Electrical Engineering, Gdynia Maritime University, Gdynia, Poland, (email: a.borys@we.am.gdynia.pl).

In this paper, we consider the problem of whether all the members of the above class mentioned in [3] are really models without memory effect. Our answer to this question is negative.

Once again, we show here that such models as the following ones named in [3]: soft limiter (SL) model and Rapp (R) model do really belong to the class of models without memory effect. However, opposite to the above, the next two mentioned in [3], that is Saleh’s (S) model and Ghorbani’s (G) model, are representations incorporating memory effect. (For more details regarding the SL, R, S, and G models, see [3]). We remember that all the four models mentioned above, as describing nonlinear distortions, are nonlinear ones. However, the memory effect is not present in two of them (in SL and R models). That is as if it was “erased from” these models. But, in two others (in S and G models), it is present (it is “not erased from” them).

In the first two models mentioned above, the value of the AM/PM conversion is identically equal to zero [3] that demonstrates the lack of any memory effect. But, opposite to this, the latter two models exhibit nonzero values of the AM/PM conversion [3], witnessing the presence of memory effect. And the reverse conclusions are of course also valid. That is, shortly, the existence of the memory in a PA leads to appearance of the nonzero values of AM/PM conversion in it.

These facts do not seem to be known to many engineers working in the area of telecommunications. The prominent example are the authors of the publication [3]. Because of this reason the above points were discussed recently in a series of papers [4-6].

On the other hand, another classification regarding the SL, R, S, and G models was published [7]. The S and G models are assumed in [7] to be quasi-memoryless, but the R model is classified as memoryless (obviously the latter property will also regard the SL model). In [7], the models exhibiting identically zero values of the AM/PM conversion are called memoryless. On the other hand, the models that demonstrate nonzero values of the AM/PM conversion as well as have the nonlinear function describing the AM/AM conversion are named quasi-memoryless. So, in fact, they possess a “small memory”. But, what does it mean to have a “small memory”? This was not explained in [7].

The next section presents a thorough analysis of the PA behavior that was performed with the use of the Volterra series [8]. Conclusions, which we draw from this analysis, allow to set the aforementioned example classifications in order and/or to refine them as well as many others in which the S model is considered to be memoryless, as for example in the papers [9-11] and the book [12].

The paper concludes with some final remarks.

II. DESCRIPTION OF NONZERO AM/PM CONVERSION MODELS VIA VOLTERRA SERIES

As already mentioned in Introduction, the Saleh’s (S) and Ghorbani’s (G) representations belong to a class of models in which the AM/PM conversion is nonzero. These are the models obtained experimentally by adjusting the measurement data to the preconceived functions. So, because of this restriction, they are not suitable for explanation of the mechanisms which lead to appearance of the AM/AM and AM/PM distortions in a nonlinear circuit (system) with memory driven by a bandpass signal of the form

\[ x(t) = r(t) \cos(\omega_t \tau + \psi(t)), \]  

where \( \omega_t = 2\pi f_c \) with \( f_c \) meaning the carrier frequency, \( t \) denotes a time variable, and \( j = \sqrt{-1} \). Further, it is assumed that \( x(t) \) in (1) contains a slowly varying real-valued bandbase signal \( r(t) \) that modulates the carrier amplitude, and the carrier phase changes with time according to a function \( \psi(t) \). The latter function, similarly as \( r(t) \), represents also a slowly varying bandbase signal.

Example mathematical tool, which can be used for the above objective, is the Volterra series [8]. It was used in [4] to get the output signal of the circuit mentioned above (with the excitation given by (1)) in the following form:

\[ y(t) = A(r(t))\cos(\omega_t \tau + \psi(t) + \Phi(r(t))), \]  

where the functions \( A(r(t)) \) and \( \Phi(r(t)) \) are generally the nonlinear functions of \( r(t) \) and can be expressed by the nonlinear transfer functions [8] of the nonlinear circuit considered. For more details, see [4] and [5]. Furthermore, comparison of (2) with (1) shows that \( A(r(t)) \) and \( \Phi(r(t)) \) describe the amplitude modulation and phase modulation, respectively, caused by the characteristics of the circuit considered and amplitude bandbase signal \( r(t) \). Therefore, they are referred to as the AM/AM characteristic (conversion) and the AM/PM characteristic (conversion), respectively.

Moreover, it was found in [4] and [5] that when at least some of the complex-valued nonlinear transfer functions of the circuit being nonlinear and possessing memory differ from the identically zero functions its AM/PM characteristic is a nonzero function. Further, it was also shown that the reverse conclusion is true, too.

Really, the results presented in [4] and [5] convince that the Volterra series is a proper mathematical tool for investigation of a class of the nonzero AM/PM models. In what follows, we use it to study the mechanism of producing the AM/PM distortion in a wideband PA whose nonlinear impulse responses [4], [5], [8] are given by

\[ h_{12}(t) = b_1 \exp(-t/a), \]  

\[ h_{22}(t_1, t_2) = b_2 \exp(-t_1/a) \exp(-t_2/a), \]  

\[ h_{33}(t_1, t_2, t_3) = b_3 \exp(-t_1/a) \exp(-t_2/a) \exp(-t_3/a), \]  

and so on, for the values of the time variables \( t \) and \( t_i, i = 1,2,3,..., \) greater and equal to zero, and identically zero otherwise. Further, the upper index in the successive responses in (3): \( h_{12}^{(1)}(t) \cdot h_{12}^{(2)}(t_1, t_2) \cdot h_{12}^{(3)}(t_1, t_2, t_3) \), and so on, means their order (degree). That is they are, respectively, the first order (linear), second order, third order, and so on, nonlinear impulse responses (Volterra kernels) of the PA considered. Moreover, the coefficients \( a \) and \( b_i, i = 1,2,3,..., \) in (3) denote some constants. Observe further that the constant \( a \) corresponds to the time constant \( RC \) of a simple low-pass RC filter. And we assume here that this time constant is very small making the PA a wideband amplifier. Moreover, note that the constant \( a \) can be treated as a measure of the amplifier memory length.

Substituting (1) and (3) into the Volterra defined as [8]

\[ y(t) = \int_{-\infty}^{\infty} h_{12}^{(1)}(\tau) x(t - \tau) d\tau + \int_{-\infty}^{\infty} h_{22}^{(2)}(\tau_1, \tau_2) x(t - \tau_1) x(t - \tau_2) d\tau_1 d\tau_2 + \int_{-\infty}^{\infty} h_{33}^{(3)}(\tau_1, \tau_2, \tau_3) x(t - \tau_1) x(t - \tau_2) x(t - \tau_3) d\tau_1 d\tau_2 d\tau_3 + \cdots, \]

we get

\[ y(t) = b_1 \int_{-\infty}^{\infty} \exp(-\tau/a) \cdot \Re \left[ r(t - \tau) \exp\left( j(\omega_t (t - \tau) + \psi(t - \tau)) \right) \right] d\tau + \cdots, \]

\[ + b_2 \int_{-\infty}^{\infty} \exp(-\tau_1/a) \exp(-\tau_2/a) \cdot \Re \left[ r(t - \tau_1) \exp\left( j(\omega_t (t - \tau_1) + \psi(t - \tau_1)) \right) \right] \]

\[ + b_3 \int_{-\infty}^{\infty} \exp(-\tau_1/a) \exp(-\tau_2/a) \exp(-\tau_3/a) \cdot \Re \left[ r(t - \tau_1) \exp\left( j(\omega_t (t - \tau_1) + \psi(t - \tau_1)) \right) \right] \cdot \]
Let us now restrict consideration of the characteristics of the PA discussed to only the range involved in its memory. This can be done by neglecting these parts of the impulse responses (3) that lie outside the time interval \(<0, a>\) (that is by equating them to zero). In other words, it means that we perform then the following approximations:

\[
h^{(3)}(t) \approx \left\{ \begin{array}{ll} b_i \exp(-t/a) & \text{for } t < 0, a > \\ 0 & \text{otherwise} \end{array} \right. \quad (6a)
\]

\[
h^{(2)}(t, t_1, t_2) \approx \left\{ \begin{array}{ll} b_i \prod_{i=1}^{2} \exp(-t_i/a) & \text{for } t, t_1, t_2 < 0, a > \\ 0 & \text{otherwise} \end{array} \right. \quad (6b)
\]

and so on.

Taking into account the above approximations in (5), we can rewrite it as

\[
y(t) \approx b_i \left[ \int_{0}^{a} \exp(-\tau/a) \right] d\tau + 
\]

\[
+b_i \left[ \int_{0}^{a} \prod_{i=1}^{2} \exp(-\tau_{i}/a) \right] d\tau_{1,2} + 
\]

\[
\left( \begin{array}{l} \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \quad \right. \]

\[
(b) h^{(3)}(t, t_1, t_2) \approx \left\{ \begin{array}{ll} b_i \prod_{i=1}^{2} \exp(-t_i/a) & \text{for } t, t_1, t_2 < 0, a > \\ 0 & \text{otherwise} \end{array} \right. \quad (6d)
\]

\[
(b) h^{(5)}(t, t_1, t_2, t_3) \approx \left\{ \begin{array}{ll} b_i \prod_{i=1}^{3} \exp(-t_i/a) & \text{for } t, t_1, t_2, t_3 < 0, a > \\ 0 & \text{otherwise} \end{array} \right. \quad (6e)
\]
\[
\begin{align*}
\text{Re}\left\{r(t-\tau)\exp\left[j\left(\omega_i(t-\tau)+\psi(t-\tau)\right)\right]\right\} \\
\text{Re}\left\{r(t-\tau)\exp\left[j\left(\omega_i(t-\tau)+\psi(t-\tau)\right)\right]\right\} \\
\text{Re}\left\{r(t-\tau)\exp\left[j\left(\omega_i(t-\tau)+\psi(t-\tau)\right)\right]\right\} \\
\text{Re}\left\{r(t-\tau)\exp\left[j\left(\omega_i(t-\tau)+\psi(t-\tau)\right)\right]\right\} \\
\text{Re}\left\{r(t-\tau)\exp\left[j\left(\omega_i(t-\tau)+\psi(t-\tau)\right)\right]\right\} \\
\cdot d\tau_d\tau_d\tau_d\tau_d\tau_d + \ldots .
\end{align*}
\]

In the next step of our explanations, we assume that the slowly varying baseband signals \(r(t)\) and \(\psi(t)\) occurring in (1) do not approximately change in the interval \(<0, a>\) of the integrations indicated in the definite integrals in (7) for the time variables \(\tau, \tau_i, i=1,2,3,...\). In other words, we assume now that
\[
r(t-\tau) \equiv r(t) \quad \text{and} \quad \psi(t-\tau) \equiv \psi(t) \quad (8)
\]
for \(\tau = \tau_i, \tau_i, i=1,2,3,...\), taking on the values from the range \(<0, a>\). Furthermore, note that this assumption is valid if the maximal frequency of the amplitude characteristics of the baseband signals \(r(t)\) and \(\psi(t)\), say \(f_m\), fulfills the following inequality: \(f_m << \sqrt{a}\).

So, taking into account (8) in (7), we can rewrite the latter as
\[
y(t) \equiv b_1 \cdot r(t) \int_0^B \exp(-\tau/a) \cdot \text{Re}\left\{\exp\left[j\left(\omega_i(t+\psi(t))\right)\right]\right\} d\tau + \\
+ b_2 \cdot (r(t))^2 \prod_{a_j} \exp(-\tau_j/a) \exp(-\tau_j/a) \cdot \\
\text{Re}\left\{\exp\left[j\left(\omega_i(t+\psi(t))\right)\right]\right\} + \\
\text{Re}\left\{\exp\left[j\left(\omega_i(t+\psi(t))\right)\right]\right\} d\tau_d\tau_d\tau_d + \\
+ b_3 \cdot (r(t))^3 \prod_{a_j} \exp(-\tau_j/a) \exp(-\tau_j/a) \cdot \\
\text{Re}\left\{\exp\left[j\left(\omega_i(t+\psi(t))\right)\right]\right\} + \\
\text{Re}\left\{\exp\left[j\left(\omega_i(t+\psi(t))\right)\right]\right\} d\tau_d\tau_d\tau_d + \ldots .
\]

In what follows, we restrict ourselves to retaining only the first \(N\) components in the Volterra series description (9) and include a passband filter with the centre frequency \(f_c = \omega_c / (2\pi)\) at the PA output. In further analysis, we will treat this filter as a part of the PA model.

Note now that by virtue of the passband filter all the products in (9) related with the frequencies different from \(\pm f_c\) will be filtered out. So, in effect, we obtain (for more details, see [4] and [5]) the following expression:
\[
y_{b}(t) \equiv \sum_{n=\text{odd}}^{N} \int_{0}^{B/2} \int_{0}^{B/2} b_1^n \cdot C(n.(n-1)/2) \cdot \\
\text{exp}\left\{\left(\omega_i t + \psi(t)\right) - \omega_0 \sum_{r=1}^{n} C(r,n-r)/2 \right\} + C(n.(n+1)/2) \cdot \\
\text{exp}\left\{-\left(\omega_i t + \psi(t)\right) + \omega_0 \sum_{r=1}^{n} C(r,n-r)/2 \right\} \prod_{j=1}^{n} \exp(-\tau_j/a)d\tau_j .
\]

where we denoted by \(y_{b}(t)\) the PA output signal after passing through its bandpass output filter. Moreover, the symbol \(C(n,m)=n!/\left[m!(n-m)\right]!\) in (10) is the so-called binomial coefficient. Furthermore, the symbol \(\sum_{r=1}^{n} \tau_j\) stands for a sum of the time variables \(\tau_j\) related with one of the distinct product frequencies. Note that there are in each case \(C(n,m)\) of such combinations what is indicated here by using this symbol beneath the summation symbol \(\sum\). For more details, see [4] and [5].

To remove the integrals from (10), we use now the multidimensional Fourier transforms [8] to the nonlinear
impulse responses occurring in (10). These transforms are
called the nonlinear transfer functions $H^{(n)}(f_1,...,f_n)$ of the
corresponding orders $n=1$ (linear case), 2, 3,..., and are given by

$$H^{(n)}(f_1,...,f_n) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h^{(n)}(\tau_1,...,\tau_n) \cdot \exp(-j2\pi f_1 \tau_1) \cdot \prod_{i=1}^{n} \exp(-j2\pi f_i \tau_i) d\tau_1 \cdots d\tau_n$$

where $f_1,...,f_n$ mean the frequencies forming the $n$-th
dimensional frequency space [8]. Let us also use below, for
simplicity, the notation $H^{(n)}(\omega_1,...,\omega_n)$ with $\omega_i = 2\pi f_i$, $i=1,2,...,n$, instead of $H^{(n)}(f_1,...,f_n)$.

We apply (11) in (10) in the following way:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} b_n \cdot \exp \left( -j2\pi f_1 \tau_1 \sum_{i=1}^{n} \tau_i \right) \prod_{i=1}^{n} \exp(-j\tau_i / a) d\tau_1 \cdots d\tau_n \equiv$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} b_n \cdot \exp \left( -j2\pi f_1 \tau_1 \sum_{i=1}^{n} \tau_i \right) \prod_{i=1}^{n} \exp(-j\tau_i / a) d\tau_1 \cdots d\tau_n \equiv$$

$$\equiv H^{(n)}(\chi^{+}_n(\pm \omega_1)) = b_n \cdot \frac{a}{1+j\chi^{+}_n(\pm \omega_1) a} \frac{a}{1+j\chi^{+}_n(\pm \omega_1) a} \cdots \frac{a}{1+j\chi^{+}_n(\pm \omega_1) a}.$$  

where $\chi^{+}_n(\pm \omega_1)$ and $\chi^{+}_n(\pm \omega_1)$ below denote such the
angular frequency sets $\{\omega_1,...,\omega_n\}$ whose elements
$\omega_i$, $i=1,2,...,n$, can assume only the values $+\omega_1$ or $-\omega_1$,
and whose sums give the value $+\omega_k$ or $-\omega_k$, respectively.
Furthermore, $\chi^{+}_n(\pm \omega_1)$ in (12) means the first element of the set
$\chi^{+}_n(\pm \omega_1)$, and so on.

Obviously, the components in (10) related with the
expressions involving the sums $\sum_{i \leq \omega_i (n(n+1)/2)} \tau_i$ can be
transformed in the same way as those in (12) with $H^{(n)}(\chi^{+}_n(\pm \omega_1))$ meaning

$$H^{(n)}(\chi^{+}_n(\pm \omega_1)) = b_n \cdot \frac{a}{1+j\chi^{+}_n(\pm \omega_1) a} \frac{a}{1+j\chi^{+}_n(\pm \omega_1) a} \cdots \frac{a}{1+j\chi^{+}_n(\pm \omega_1) a}.$$  

Further, taking into account this and (12) in (10), we get

$$y_s(t) \equiv \sum_{n=1, \text{odd}}^{N} \left( \frac{r(t)}{2} \right)^{n} \left[ C(n,(n-1)/2) \right]^{H^{(n)}(\chi^{+}_n(\pm \omega_1))}.$$

We identify the signal $y_s(t)$ given by (14) with $y(t)$ in (2). And by comparison of the components in (14) with the corresponding ones in (2), we can deduce the expressions describing the AM/AM and AM/PM conversions in the PA discussed. This was done in [4] and [5]. However, we do not continue discussing this aspect here; our objective in this paper is different.

Let us now summarize the main results of the analysis presented above:

1. We can identify two important points in the above analysis in which the influence of the PA memory on the AM/PM conversion was considered. These are the following: a) relations (8) associated with fulfilment of the condition $f_m << 1/a$ (point of a direct influence) and b) relations (12) and (13) when the expressions on their right hand sides cannot be assumed to be real numbers (this influence is called here an indirect one).

2. Note that the direct influence can be fully neglected.

3. Further, note that we can interpret the indirect influence as the one following from “the interaction of the carrier with the PA memory”. Moreover, it follows from consideration of (12) and (13) that it cannot be neglected when $f_i \geq 1/(2\pi a)$ holds. That is when the PA frequency characteristic preceding the output bandpass filter is not enough wideband. Evidently, the nonzero values of the AM/PM conversion will then occur.

III. Final Remarks

Note that the results of the analysis presented in the previous section provide a simple and clear proposition for classification of the PAs (at least for those which can be approached in a way that was implemented above). That is when $f_i < 1/(2\pi a)$ holds for a PA considered, it means that this PA is a circuit without memory. Further, for the carrier frequencies $f_i$ around $1/(2\pi a)$, it can be assumed to be approximately quasi-memoryless. However, when the carrier frequencies $f_i$ are chosen to be clearly greater than $1/(2\pi a)$, then this PA must be considered as a circuit with memory. In the latter case, the PA produces the AM/PM distortion which cannot be neglected.

Moreover, a plenty of other interesting conclusions can be drawn from the material presented in section II. For example, note that the coefficients of the models $S$ and $G$, which foresee the occurrence of the AM/PM distortion, on the other hand, do not depend upon the carrier frequency $f_i$. This bears witness to their limited applicability.
REFERENCES


