Extreme Value Statistical Characterization of Time Domain Pulse-to-Pulse Measurements

Pasquale Arpaia$^{1,2}$, Carlo Baccigalupi$^{1,3}$, Michele Martino$^1$
$^1$TE Department, CERN Geneve, 23 Switzerland
$^2$EN Department, University of Sannio, 82100 Benevento, Italy
$^3$PH Department, University of Calabria, 87036 Rende (CS), Italy

Keywords:

Abstract

An analytical method, based on Extreme Value Theory (EV T), for predicting the worst case repeatability of time domain pulse-to-pulse measurements, modeled as independent and identically distributed random variables, is proposed. The method allows the use of the noise level of a measurement system for predicting the upcoming peak values over a given number of independent observations. The proposed analytical model is compared against simulated distributions generated in Matlab, highlighting satisfying match for any sample size. The simulations are based on a case study on the characterization of a pulsed power supply for the klystron modulators of the Compact Linear Collider (CLIC) under study at CERN.

Presented at: IMEKO 2014

Geneva, Switzerland
October 2015
Extreme Value Statistical Characterization of Time Domain Pulse-to-Pulse Measurements

Pasquale Arpaia\textsuperscript{1,2}, Carlo Baccigalupi\textsuperscript{1,3}, Michele Martino\textsuperscript{4}, arpaia@unisannio.it\textsuperscript{1,2}, carlo.baccigalupi@cern.ch\textsuperscript{1,3}, michele.martino@cern.ch\textsuperscript{1}

\textsuperscript{1}TE Department, CERN Geneve, 23 Switzerland
\textsuperscript{2}EN Department, University of Sannio, 82100 Benevento, Italy
\textsuperscript{3}PH Department, University of Calabria, 87036 Rende (CS), Italy

Abstract – An analytical method, based on Extreme Value Theory (EVT), for predicting the worst case repeatability of time domain pulse-to-pulse measurements, modeled as independent and identically distributed random variables, is proposed. The method allows the use of the noise level of a measurement system for predicting the upcoming peak values over a given number of independent observations. The proposed analytical model is compared against simulated distributions generated in Matlab, highlighting satisfying match for any sample size. The simulations are based on a case study on the characterization of a pulsed power supply for the klystron modulators of the Compact LInear Collider (CLIC) under study at CERN.

I. INTRODUCTION

The Univariate Extreme Value Theory or simply Extreme Value Theory (EVT) is widely used in different fields, such as risk management [1],[2], finance, economics, or even for estimating the fastest human time on the 100 m sprint [3]. In addition, recently, the EVT has also found important applications in some engineering fields, e.g., studies ranging from tides[4] to accelerator physics [5]. For example, in telecommunications, an accurate expression for the peak distribution of the Orthogonal Frequency Division Multiplexing (OFDM) envelope was determined by the University of Massachusetts [6]. In fact, the main problem about the applicability of OFDM systems in low-power wireless systems is the highly variable amplitude of transmitted signals. The above study defined a rigorous method to predict the upcoming peak values of the envelope by the previous observation and data collection of the precedent peaks.

A new particle accelerator, the Compact LInear Collider (CLIC), is currently under study at CERN. Its klystron modulators [7] will be operated in pulsed mode with a pulse length of 150 \(\mu s\) [8],[9]. In Tab.1, the CLIC klystron modulators pulses specifications are shown. To meet the very challenging requirements for the RF power quality, derived directly from the accelerator performance specifications, the modulator’s flat-top Pulse-to-Pulse Repeatability \(PPR\) needs to be better than \(\pm 100ppm\). A specifically designed reference acquisition system [10] is currently under development at CERN in order to verify this performance. \(PPR\) is defined in (1) as:

\[ PPR = \max_i |V_{i,j} - V_{i,j+1}|. \]  

\(V_{i,j}\) and \(V_{i,j+1}\), defined in Fig.1, are the instantaneous voltage values at the same (in equivalent time) sampling instant \(i\) within two consecutive pulses flat-tops (\(j\) and \(j+1\)). By assuming perfectly repeatable pulses, these differences should be exactly equal to zero; however, even under this assumption, the reference acquisition system will introduce some additive noise on the flat-tops affecting the \(PPR\) measurement. If the acquisition system has suitable stability within the Repetition Period of the train of pulses (20 \(ms\), its noise is the only factor that could affect \(PPR\) because all the \textit{long term} effects (e.g. temperature drift) can be neglected.

In this paper, an analytical method based on \textit{EVT} for predicting the worst case repeatability of time domain pulse-to-pulse measurements, modeled as independent and identically distributed random variables, is proposed. In particular, in section ii. the equation describing the model is presented whereas in section iii. the problem is formalized in terms of the CERN case study. Finally, in section iv., the results of two Matlab simulations of \(PPR\), with different sample sizes, are fitted with the proposed model in order to validate it.
TABLE 1. Typical Pulse Characteristics

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal Pulse Amplitude</td>
<td>$V_{pa}$</td>
<td>150 kV</td>
</tr>
<tr>
<td>Nominal Pulse Current</td>
<td>$I_{pa}$</td>
<td>100 A</td>
</tr>
<tr>
<td>Pulse Peak Power</td>
<td>$P_{peak}$</td>
<td>21 MW</td>
</tr>
<tr>
<td>Positive Going Transition</td>
<td>$t_{rise}$</td>
<td>3 μs</td>
</tr>
<tr>
<td>Negative Going Transition</td>
<td>$t_{fall}$</td>
<td>3 μs</td>
</tr>
<tr>
<td>Transition Settling Duration</td>
<td>$t_{set}$</td>
<td>5 μs</td>
</tr>
<tr>
<td>Flat-Top Duration</td>
<td>$FTD$</td>
<td>150 μs</td>
</tr>
<tr>
<td>Repetition Period</td>
<td>$RP$</td>
<td>20 ms</td>
</tr>
<tr>
<td>Voltage Overshoot</td>
<td>$V_{ov}$</td>
<td>1 %</td>
</tr>
<tr>
<td>Flat-top Tolerance</td>
<td>$FTT$</td>
<td>0.85 %</td>
</tr>
<tr>
<td>Pulse-to-Pulse Repeatability</td>
<td>$PPR$</td>
<td>±100 ppm</td>
</tr>
<tr>
<td>Frequency Range of Interest</td>
<td>$FR$</td>
<td>1 kH z – 5 M H z</td>
</tr>
</tbody>
</table>

II. THEORY

According to the definition given in (1), three operations should be performed on each pair of flat-top samples, assumed to be independent and identically distributed (i.i.d.) random variables, in order to obtain the $PPR$: (i) the difference, (ii) the absolute value, and (iii) the maximum. Each pair of counterparts flat-top samples $V_{i,j}$ and $V_{i,j+1}$ are subtracted each other obtaining another random variable $Y_i = V_{i,j} - V_{i,j+1}$. The generic $V_{q,k} = V_{q,k} + n_{q,k}$ is composed by an ideal flat-top sample (perfectly repeatable), $V_{q,k}$, and a sample of the time domain noise $n(t)$ which is assumed to be Additive White Gaussian Noise (AWGN) with zero mean and standard deviation $\sigma$, therefore:

$$\begin{cases} 
\mu Y_i = \hat{V}_{i,j} - \hat{V}_{i,j+1} = 0 \\
\sigma^2 Y_i = \sigma^2_{t,j} + \sigma^2_{t,j+1} = 2\sigma^2 
\end{cases} \quad (2)$$

The pdf of the $PPR$ distribution (a detailed explanation of the underlying mathematics is presented in [11]) is expressed by equation (3).

$$f_{PPR}(z) = \begin{cases} 
2N_o N_o f_{Y_i}(z) [2F_{Y_i}(z) - 1]^{N_o - 1}, z \geq 0 \\
0, z < 0 
\end{cases} \quad (3)$$

$N_o$ is the number of considered samples, whereas $N_o$ represents the number of observations made on each variable. Equation (3) is a function of only $\sigma$, $N_o$, and $N_o$. Once the underlying distribution is known (both in terms of RMS noise, $\sigma$, and numerosity, $N_o$), it only depends on the desired sample size specification $N_o$.

III. PROBLEM DEFINITION FOR CLIC

REFERENCE ACQUISITION SYSTEM

The reported theory can be applied to the CERN experimental case study, where a reference acquisition system for measuring the fast voltage pulses used for the CLIC klystron modulators power supply is required. In Tab.2, the model parameters used in simulation are reported. The number of samples $N_s$ has been derived directly from the CLIC pulses specifications: each pulse has a Flat-Top Duration ($FTD$) of 150 μs and it is acquired by the reference acquisition system at Sampling Rate ($SR$) of 15 MS/s. By considering the flat-top samples as iid random variables the numerosity $N_s$ should be chosen as follows:

$$N_s = FTD \cdot SR = 2250$$

(4)

Under the ideal hypotheses of perfectly repeatable flat-tops, two counterparts samples will have exactly the same value; however the reference acquisition system will introduce some additive noise on the flat-tops, therefore the considered standard deviation is only due to the acquisition system’s noise.

IV. SIMULATION RESULTS

The validity of equation (3) was verified by means of two simulation trials in MATLAB. Two sets of samples, each of them composed by $N_s$ Normal random numbers, are generated by the MATLAB’s $\text{randn}$ function and equation (3) is applied to calculate both $PPR_i$ and $PPR(N_o)$ ($\text{randn}$ is not guaranteed to generate a zero mean sample, or a negligible mean value, so this effect is always compensated before being processed further). The statistical sample is generated by reiterating the simulation $N_{test}$ times as a whole. A $\chi^2$ test, implemented in MATLAB with the $\text{chi2gof}$ function, is then carried out in order to verify that the data in the sample under study are distributed according to the proposed distribution. The $\chi^2$ test, performed with $\alpha = 0.1\%$ (as recently suggested in [12]), was repeated for different numbers of bins and never rejected the null hypothesis resulting in $p-values$ always higher than $\alpha$. An AWGN with $\sigma = 1 ppm$ of Full Scale is assumed as the underlying distribution (as reported in table 2).

The maxima distribution of $PPR$, estimated with a sample size $N_o = 1$ (corresponding to two pulses), is depicted in Fig.2 where both the normalized histogram obtained in simulation and the analytic formula (3) are shown (the quality of the fit, confirmed by the $\chi^2$ test can also be assessed visually).

For the second simulation trial the distribution of $PPR$ over a sample size $N_o = 180000$ was estimated (corresponding to 1 hour of observation). Also in this case, the proposed model fits the simulated distribution accord-
ing to the result of another $\chi^2$ test. A satisfying match can be also assessed visually by confirming the effectiveness of the proposed model (Fig.3). The excellent agreement shown during these tests allows the $PPR_{\text{max}}$ of the system to be predicted for larger $N_o$, when the actual measurement, or even the simulation, would be unfeasible due to the extremely high test duration. Assuming the total noise of the measurement system to be 1 ppm of Full Scale, the estimated $PPR_{\text{max}}$ distribution is depicted in Fig.4 for $N_o = 1.5768 \times 10^9$, corresponding to 1 year of acquisition considering the CLIC case study ($1.5768 \times 10^9 \times RP = 365$ days). However the model can be generalized for higher noise by simply characterizing its standard deviation to be given as model input. In Fig.5 the trend of the PPR’s mode is depicted for different amounts of noise introduced by the measurement system and for $N_o = 1$ showing an approximately linear behavior.

V. CONCLUSION AND FUTURE DEVELOPMENTS

An accurate statistical model for characterizing the maximum values of the Pulse-to-Pulse Repeatability ($PPR_{\text{max}}$) has been proposed. Matlab simulations were performed in order to demonstrate its effective capability of fitting the actual distribution and, thus, validate the model. The proposed model allows the $PPR$ of the Acquisition System to be predicted for any sample size and this analytic tool can be used to formalize the specification for $PPR$ of the CLIC Reference Acquisition System. When the development of the Reference Acquisition System will be finalized, the proposed model, already validated in simulation, will be verified with the actual measured distributions. In addition, a further step of this research work will be to determine a noise budget (in terms of its standard deviation) in order to guarantee the required $PPR$ over a given sample size within the required confidence level.

ACKNOWLEDGMENTS

The authors would like to thank Miguel Cerqueira Bastos, whose collaboration and precious advice were essential for the realization of this work.

REFERENCES

[1] Li Xiu-Min and Li Fa-Chao. Extreme Value Theory: An Empirical Analysis of Equity Risk for Shanghai


