

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

Pasquale Arpaia<sup>1,2</sup>, Carlo Baccigalupi<sup>1,3</sup>, Michele Martino<sup>1</sup>,  
 arpaia@unisannio.it<sup>1,2</sup>, carlo.baccigalupi@cern.ch<sup>1,3</sup>, michele.martino@cern.ch<sup>1</sup>

<sup>1</sup>TE Department, CERN Geneva, 23 Switzerland

<sup>2</sup>EN Department, University of Sannio, 82100 Benevento, Italy

<sup>3</sup>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  $\pm 100$ ppm. A specif-

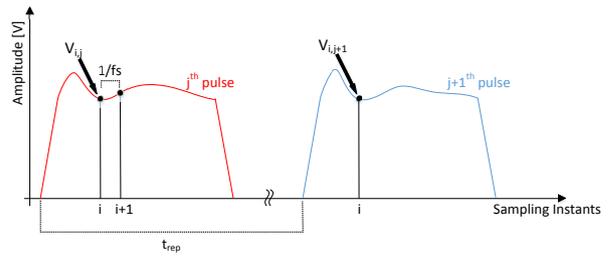


Fig. 1. Pulse-to-Pulse Repeatability Definition.

ically designed reference acquisition system [10] is currently under development at CERN in order to verify this performance. PPR is defined in (1) as:

$$PPR = \max_i |V_{i,j} - V_{i,j+1}|. \quad (1)$$

$V_{i,j}$  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 long term effects (e.g. temperature drift) can be neglected.

In this paper, an analytical method based on 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

Pulse Specification		
Parameter	Symbol	Value
Nominal Pulse Amplitude	$V_n$	150 kV
Nominal Pulse Current	$I_n$	160 A
Pulse Peak Power	$P_{peak}$	24 MW
Positive Going Transition	$t_{rise}$	3 $\mu$ s
Negative Going Transition	$t_{fall}$	3 $\mu$ s
Transition Settling Duration	$t_{set}$	5 $\mu$ s
Flat-Top Duration	$FTD$	150 $\mu$ s
Repetition Period	$RP$	20 ms
Voltage Overshoot	$V_{ovs}$	1 %
Flat-Top Tolerance	$FTT$	0.85 %
Pulse-to-Pulse Repeatability	$PPR$	$\pm 100$ ppm
Frequency Range of Interest	$FRI$	1 kHz – 5 MHz

## 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  $V_{q,k} = \hat{V}_{q,k} + n_{q,k}$  is composed by an ideal flat-top sample (perfectly repeatable),  $\hat{V}_{q,k}$ , 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:

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

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

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

$N_s$  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  $\sigma$ ,  $N_s$ , and  $N_o$ . Once the underlying distribution is known (both in terms of *RMS* noise,  $\sigma$ , and numerosity,  $N_s$ ), 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

Table 2. Model Parameters used in Simulation

Model Parameters		
Parameter	Symbol	Value
Number of Samples	$N_s$	2250
Noise standard deviation	$\sigma$	1 ppm of FS
Test iterations	$N_{test}$	10000
$\chi^2$ test Significance Level	$\alpha$	0.1 %

number of samples  $N_s$  has been derived directly from the *CLIC* pulses specifications: each pulse has a Flat-Top Duration (*FTD*) of 150  $\mu$ s 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:

$$N_s = FTD \cdot SR = 2250 \quad (4)$$

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 *randn* function and equation (3) is applied to calculate both  $PPR_j$  and  $PPR(N_o)$  (*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 *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-

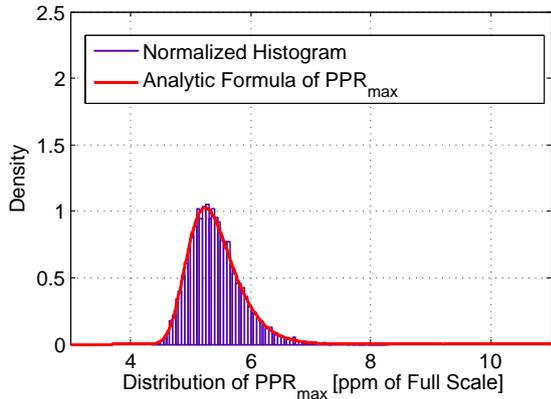


Fig. 2. Distribution of  $PPR_{max}$  for  $N_o = 1$ .

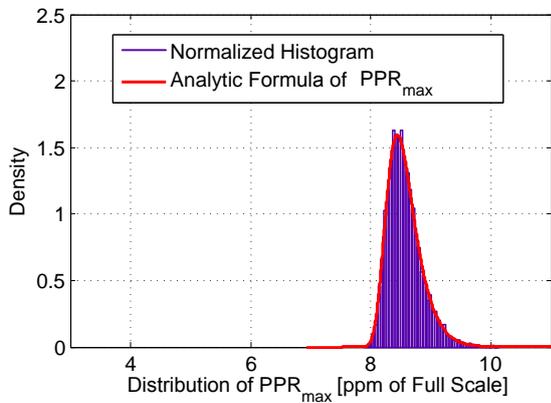


Fig. 3. Distribution of  $PPR_{max}$  for  $N_o = 180000$ .

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  $PPR_{max}$  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  $PPR_{max}$  distribution is depicted in Fig.4 for  $N_o = 1.5768 \cdot 10^9$ , corresponding to 1 year of acquisition considering the CLIC case study ( $1.5768 \cdot 10^9 \times RP = 365$  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 ( $PPR_{max}$ ) has been proposed. Matlab simulations were performed in order to demonstrate its effective capabil-

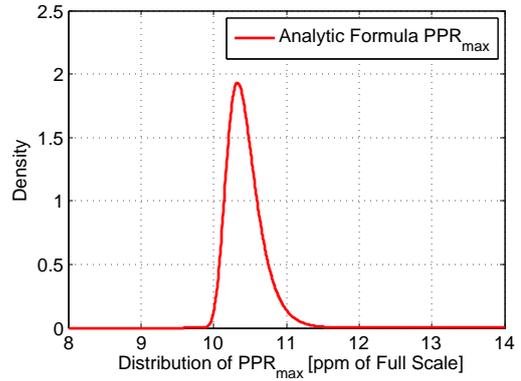


Fig. 4. Estimated  $PPR_{max}$  for  $N_o = 1.5768 \times 10^9$ .

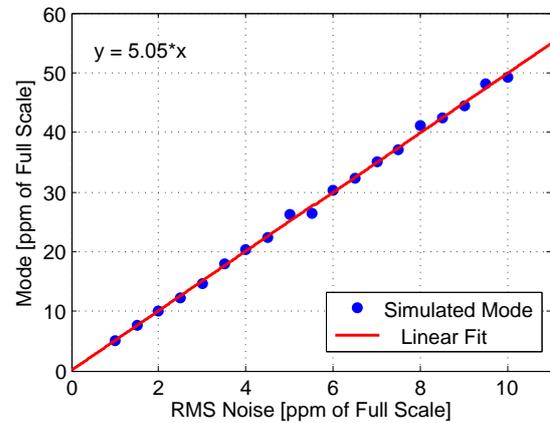


Fig. 5. Mode of  $PPR_{max}$  for different values of RMS Noise.

ity 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

- Stock Market. In *International Conference on Service Systems and Service Management*, volume 2, pages 1073–1077, 2006.
- [2] Nai peng Hu. Research on price risk of soybean market based on extreme value theory. In *Management Science and Engineering, 2009. ICMSE 2009. International Conference on*, pages 824–830, 2009.
- [3] J Einmahl and Smeets S. Ultimate 100 m World Records through Extreme-Value Theory. In *CentER Discussion Paper Series No. 2009-57*, 2009.
- [4] J.A. Tawn. Estimating Probabilities of Extreme Sea-levels. *Appl. Statist.*, 41:77–93, 1992.
- [5] R. Duperrier and D. Uriot. The extreme value theory to estimate beam losses in high power linacs. In *Particle Accelerator Conference, 2007. PAC. IEEE*, pages 815–819, 2007.
- [6] S. Wei, D.L. Goeckel, and P.E. Kelly. A Modern Extreme Value Theory Approach to Calculating the Distribution of the Peak-to-Average Power Ratio in OFDM Systems. In *Communications, 2002. ICC 2002. IEEE International Conference on*, volume 3, pages 1686–1690, 2002.
- [7] W.C Nunnally and T.G. Engel. Review of Klystron Modulator Technology for Particle Accelerators. In *Pulsed Power Conference, 1999. Digest of Technical Papers. 12th IEEE International*, volume 2, pages 967–970, 1999.
- [8] D. Aguglia, C.A. Martins, M. Cerqueira Bastos, D. Nisbet, D. Siemaszko, E. Sklavounou, and P. Viarouge. Klystron Modulator Technology Challenges for the Compact Linear Collider (CLIC). In *Pulsed Power Conference (PPC), 2011 IEEE*, pages 1413–1421, 2011.
- [9] M. Battaglia, A. De Roeck, J. Ellis, and D. Schulte. Physics at the CLIC Multi-TeV Linear Collider. Technical report, Report of the CLIC Physics Working Group, CERN report, ref. hep-ph/0412251; CERN-2004-005, 2004.
- [10] P Arpaia, C Baccigalupi, M Cerqueira Bastos, and M Martino.  $\pm 25$  ppm repeatable measurement of trapezoidal pulses with 5 mhz bandwidth. *Journal of Instrumentation*, 9(06):P06002, 2014.
- [11] P. Arpaia, C. Baccigalupi, and M. Martino. Extreme Value Statistical Definition of Worst-Case Uncertainty. Submitted to Metrologia.
- [12] Valen E Johnson. Revised standards for statistical evidence. *Proceedings of the National Academy of Sciences*, 110(48):19313–19317, 2013.