# The Misconception of Exponential Tail Upper-Bounding in Probabilistic Real-Time

Federico Reghenzani, Giuseppe Massari, and William Fornaciari

*Abstract*—Measurement-Based Probabilistic Timing Analysis, a probabilistic real-time computing method, is based on the Extreme Value Theory (EVT), a statistical theory applied to Worst-Case Execution Time analysis on real-time embedded systems. The output of the EVT theory is a statistical distribution, in the form of Generalized Extreme Value Distribution or Generalized Pareto Distribution. Their cumulative distribution function can asymptotically assume one of three possible forms: light, exponential or heavy tail. Recently, several works proposed to upper-bound the light-tail distributions with their exponential version. In this paper, we show that this assumption is valid only under certain conditions and that it is often misinterpreted. This leads to unsafe estimations of the worst-case execution time, which cannot be accepted in applications targeting safety critical embedded systems.

## I. INTRODUCTION

**P**ROBABILISTIC real-time computing has been proposed to overcome the complexity of using traditional Worst-Case Execution Time (WCET) analyses on modern processor architectures. In fact, advanced micro-architectural features – such as pipelines, multi-level caches or system management interrupts – complicate the WCET estimation [2], being sources of non determinism. The timing analyses become even more challenging on COTS platforms [3] and in time-sensitive HPC applications [1]. In critical embedded systems, the violation of timing constraints is not acceptable. To guarantee their satisfaction, it is mandatory to get a *safe*, i.e. non-underestimated, value for the tasks WCET.

The widely used probabilistic approach is called *Measurement-Based Probabilistic Timing Analyses (MBPTA)*. This method has still several open issues [4] and it has recently gained considerable research interest. It is based on using direct measurements of the tasks execution times to predict the occurrence of *extreme events* at run-time, i.e. the probability of observing an execution time that is larger than the maximum value previously observed. The output of probabilistic real-time analyses is the so-called *probabilistic-WCET* (pWCET), i.e. a statistical distribution used to derive the WCET, given the violation probability, and vice versa. Some works upper-bound this probability with the exponential version of its distribution. In this paper, we argue that this assumption is valid only under certain conditions. A

comprehensive and basic tutorial on probabilistic real-time can be found in [5].

### A. Extreme Value Theory

The Extreme Value Theory (EVT) [6] is a well-known and assessed statistical theory, traditionally used for natural disaster prediction, e.g. the forecast of catastrophic floods. Provided that the input values are iid[1], the output of the EVT is a distribution that represents the probability of observing extreme minimums or maximums.

In real-time computing, the EVT input values, identified by the random variables $X_1, X_2, ..., X_n$, are the time measurements of a given task's jobs. Following a well-defined estimation and testing process [7], the EVT can estimate the pWCET cumulative distribution function (cdf): $F(x) = P(X < x)$. Its complementary (ccdf), i.e. $1 - F(x) = P(X \geq x)$, represents the probability of experiencing execution times larger than a fixed value $x$.

### B. Extreme distributions

From the results of the *Fisher-Tippett-Gnedenko theorem* [8] [9], the pWCET distribution is asymptotically equivalent to the *Generalized Extreme Value Distribution* (*GEV* or *GEVD*):

$$F(x) = \begin{cases} e^{-(1+\xi \frac{x-\mu}{\sigma})^{-\frac{1}{\xi}}} & \text{if } \xi \neq 0 \\ e^{-e^{-\frac{x-\mu}{\sigma}}} & \text{if } \xi = 0 \end{cases}$$

The GEV has three parameters: the location $\mu$, the scale $\sigma$ and the shape $\xi$. The value of the shape parameter specializes the distribution: for $\xi < 0$, the distribution is a reversed Weibull; for $\xi = 0$, it is a Gumbel; for $\xi > 0$, it is a Fréchet. Traditionally, the estimation of the GEV parameters is performed by using the Block-Maxima (BM) method: given the original sequence of random variables $X_1, X_2, ..., X_n$ and a block size $B$, BM generates the following sequence $Y_1, Y_2, ..., Y_{n/B}$ of random variables:

$$Y_i = \max(X_{B \cdot (i-1)+1}, X_{B \cdot (i-1)+2}, ..., X_{B \cdot (i-1)+B-1})$$

The Peak-Over-Threshold (PoT) approach is an alternative to BM:

$$(Y_1, Y_2, ..., Y_m) = \{X \text{ s.t. } X > u\}$$

where $u$ is a predefined threshold and $m$ is the size of the right term set. The resulting random variables $Y_i$ are distributed according to the Generalized Pareto Distribution

All authors are affiliated with Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milano, 20133 Italy e-mail: [name].[surname]@polimi.it.

[1]Independent and Identically Distributed. This assumption can be relaxed, but this discussion is outside the scope of this paper.
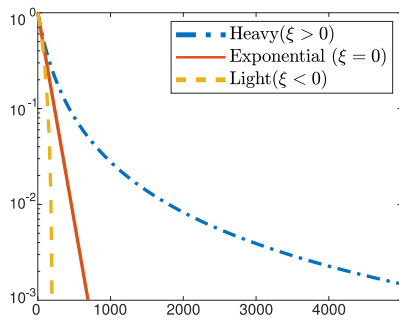
Fig. 1: The asymptotic behaviours of the $GP3(0, 100, \xi)$ cumulative distribution function $F(x)$.

(GPD) in the PoT case. The GPD can be provided in three-parameter form (GP3) or in two-parameter form (GP2). In order to simplify the subsequent notations, we consider the two-parameter form lacking the location parameter $\mu$, i.e. $GP2(\sigma, \xi) = GP3(0, \sigma, \xi)$. This form does not reduce the generality of this work, as discussed in II-C. The cdf of $GP2(\sigma, \xi)$ is therefore defined as:

$$F(x) = \begin{cases} 1 - \left(1 + \xi \frac{x}{\sigma}\right)^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - e^{-\frac{x}{\sigma}} & \text{if } \xi = 0 \end{cases} \quad (1)$$

The GPD is asymptotically equivalent to GEV [10] and it can consequently be used for the extreme probability computation.

## II. EXPONENTIAL TAIL-BOUNDING PROBLEM

In probabilistic real-time research, some authors [11] [12] [13] argued that the exponential-tail distributions, i.e. GEV/GPD with $\xi = 0$, are good candidates for fitting the pWCET distribution. These claims are motivated by empirical demonstrations, as it is almost impossible to provide formal ones. Conversely, other experiments [14] [15] showed that it is worth considering the $\xi < 0$ and $\xi > 0$ cases. The exponential distribution may in fact not be representative of all the scenarios. In literature, the $\xi > 0$ case is controversial: considering it would mean that the WCET can also get a infinite value. However, discussing the validity of this case is out of the scope of this work.

Figure 1 shows the tails of the extreme distributions, depicted by the ccdf for different values of the shape parameter $\xi$. This result has been recently used [16] [17] [18] to state that exponential-tail distribution upper-bounds the light-tail distribution. Formally:

$$1 - F_{\mathcal{G}_1}(x) > 1 - F_{\mathcal{G}_2}(x)$$

where $\mathcal{G}_1 \sim GP3(\mu, \sigma, 0)$ and $\mathcal{G}_2 \sim GP3(\mu, \sigma, \xi < 0)$. This relation is equivalent for GEV. In the probabilistic real-time context, this means that using a distribution with exponential-tail ($\xi = 0$) to upper-bound a light-tail ($\xi < 0$) should not lead to pWCET underestimation. This is true because, for a certain WCET $x$, the probability to incur in a longer execution time ($1 - F_{\mathcal{G}_1}(x)$) is always higher than the one computed with light-tail ($1 - F_{\mathcal{G}_2}(x)$). Vice versa, for a given probability $p$, the WCET estimated by exponential tail ($F_{\mathcal{G}_1}^{-1}(1 - p)$) is always higher than the one computed by light-tail ($F_{\mathcal{G}_2}^{-1}(1 - p)$).

These results are valid only if the other parameter(s) of the distribution remains unchanged. For example, assume to fit the complete distribution $GP3(\mu, \sigma, \xi)$ (with $\xi < 0$) and then enforce $\xi = 0$ obtaining $GP3(\mu, \sigma, 0)$. In this case, the second distribution upper-bounds the first one and the pWCET is not under-estimated. However, enforcing $\xi = 0$ before performing the distribution fitting may lead to different values of $\mu$ and $\sigma$, w.r.t. the real distribution of the data. This would invalidate the previous result, carrying out potentially unsafe pWCET estimations. The estimation procedure in fact usually provides the $\mu$ and $\sigma$ values that best fit the input data. These in general are different from the ones that would have been computed without enforcing $\xi = 0$. Moreover, if the estimator is unbiased w.r.t. the mean of the extreme population, the resulting distribution is always unsafe, as proven below.

### A. Parameters shift effects on the first moment

To prove the statements above, we initially use a two-parameters Generalized Pareto Distribution $GP2(\sigma, \xi)$ (this to simplify the calculus). The extension to GP3 and GEV is then discussed in Section II-C.

Let $Y_1, ..., Y_m$ be the $m$ maximum time measurements distributed under $\mathcal{G} \sim GP2(\sigma, \xi)$, as result of the EVT, e.g. by using the Peak-over-Threshold algorithm, $Y_i > u$. If the estimator is unbiased, the mean value of $Y_i$ matches the expected value of the GP2 distribution:

$$E[Y] = \frac{\sigma}{1 - \xi}$$

It follows that, for the exponential distribution case ($\xi = 0$) the expected value is:

$$E[Y] = \sigma_{\xi=0}$$

Upper-bounding with the exponential-tail distribution means forcing $\xi = 0$ for the same set of data, while maintaining the same expected value $E[Y]$, that is:

$$\sigma_{\xi=0} = \frac{\sigma_{\xi<0}}{1 - \xi} \quad (2)$$

In case the data are distributed with $\xi < 0$, the simplification $\xi = 0$ leads to estimate $\sigma_{\xi=0} > \sigma_{\xi<0}$. However, the following section provides the proof that this scale parameter skew may lead to unsafe pWCET estimations.

### B. Proof of failure of exponential-tail upper-bounding

Given the definition of cdf $F(x) = P(X \leq x)$, upper-bounding a distribution in the context of MBPTA means that the relation $F'(x) < F(x)$ holds for any $x$.

*Proof.* Let $\bar{F} = 1 - F(x) = P(X > x)$ be the complementary cdf. A safe upper-bound for pWCET has to guarantee the conservative relation $P'(X > x) > P(X > x) \quad \forall x$. From this, it is possible to obtain $1 - F'(x) > 1 - F(x)$ and, in turn, $F'(x) < F(x)$. $\square$

In our scenario, $F'(x)$ corresponds to the upper-bound with $\xi = 0$, while $F(x)$ is the real distribution, with $\xi$ assumed to
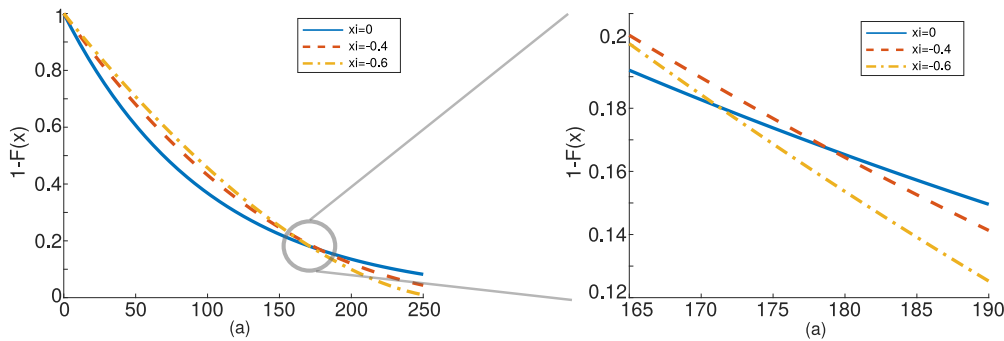
Fig. 2: The complementary cdf varying $\xi$ of two-parameter GPD. The reference distribution GPD$(100,0)$ is represented by the solid blue line. The (b) plot is the zoom of (a) plot at the ccdf intersection.

be set to an unknown negative value $\xi < 0$. By expanding the cdfs we obtain:

$$1 - e^{-\frac{x}{\sigma_{\xi=0}}} < 1 - \left(1 + \xi \frac{x}{\sigma_{\xi<0}}\right)^{-1/\xi} \quad (3)$$

This inequality must hold for any $x$, but since we are dealing with a positive variable (execution time), this holds only for $x > 0$. As a consequence of Equation 2, it is possible to state that this inequality is not true in general.

*Proof.* Removing the constant term and multiplying by $-1$:

$$e^{-\frac{x}{\sigma_{\xi=0}}} > \left(1 + \xi \frac{x}{\sigma_{\xi<0}}\right)^{-1/\xi} \quad (4)$$

Let us now replace $\sigma_{\xi<0}$ according to Equation 2:

$$e^{-\frac{x}{\sigma_{\xi=0}}} > \left(1 + \frac{\xi}{1-\xi} \frac{x}{\sigma_{\xi=0}}\right)^{-1/\xi} \quad (5)$$

The equation corresponding to this inequality has a trivial zero for $\bar{x}_1 = 0$, but it has another solution $\bar{x}_2$ for $x > 0$:

$$\bar{x}_2 = \frac{\sigma_{\xi=0}}{\xi}\left(-W\left[(\xi-1)\left(e^{\frac{1}{\xi}-1}\right)^{-\xi}\right] + \xi - 1\right) \quad (6)$$

where $W[\cdot]$ is the *Lambert W function*. Since $\xi < 0$, the argument of $W[\cdot]$ is negative, as well as $W[\cdot]$. While, being $|W[\cdot]| < 1$, $\bar{x}$ assumes a positive value. This means that there is a second zero $(\bar{x}_2 > 0)$ and consequently at least a value $(\bar{x}_2)$ that violates the inequality of Equation 3. $\square$

As discussed in Section III, there is actually a continuous interval, i.e. infinite points, that violates the inequality. Providing an analytical proof for it is not simple because of the complexity introduced by the *Lambert W function*. Luckily, this is not necessary to demonstrate that the exponential-tail upper-bounding is unsafe. The counterexample obtained by numerical evaluation is in fact sufficient to prove this.

### C. Applicability to GEV and GPD 3-parameters

The problem of exponential-tail bounding exists also in GP3 and GEV since both share the same tail behaviour presented in Figure 1. The previous proofs can be easily ported to GP3 and GEV distributions. Following the same approach of the GP2

version, the mean value of GP3$(\mu, \sigma, \xi)$ is $E[Y] = \mu + \frac{\sigma}{1-\xi}$. If the mean value does not change once the estimation run with $\xi = 0$, the results are exactly the same of the provided proof. If $\mu_{\xi=0} < \mu_{\xi<0}$, then the error is higher and the estimation becomes unsafe. If $\mu_{\xi=0} > \mu_{\xi<0}$, then nothing can be said without an accurate analysis of the specific case.

Similarly to the GPD case, the GEV condition for safe upper-bounding is:

$$e^{1-e^{-\frac{x}{\sigma_{\xi=0}}}} < e^{1-\left(1+\xi\frac{x}{\sigma_{\xi<0}}\right)^{-1/\xi}}$$

and since $e^{f(x)} < e^{g(x)} \leftrightarrow f(x) < g(x)$, it results that:

$$1 - e^{-\frac{x}{\sigma_{\xi=0}}} < 1 - \left(1 + \xi\frac{x}{\sigma_{\xi<0}}\right)^{-1/\xi}$$

that is exactly the same of Equation 3. For this reason, the previous analysis can be applied also to GEV.

### III. NUMERICAL EVALUATION

In order to clarify the previous equations and to provide a counterexample to the exponential upper-bounding claim, we use the GP2 distribution with $\sigma = 100, \xi = 0$ as a reference. To show the violation of the upper-bounding rule, we compare it with other two GP2 distributions with $\xi = -0.4$ and $\xi = -0.8$. The scale parameter $\sigma$ is computed according to Equation 2. The respective ccdfs are depicted in Figure 2a. The exponential GPD ($\xi = 0$) upper-bounds both distributions only starting from the value $\bar{x}_2 \approx 179$. The absolute value of $\bar{x}_2$ is not negligible: in this case we consider $\sigma = 100$, that is the mean value of the extremes, and the exponential tail upper-bound becomes safe only after nearly the double of it. Figure 2b zooms in the intersection point of the previous Figure 2a. As expected, increasing the value of $\xi$ towards $0$ produces a smaller error in the difference of cdf between $\xi < 0$ and $\xi = 0$. However, it shifts also the intersection point, i.e. the point at which the upper-bound is safe, towards $+\infty$.

To investigate better the last result, we compute the maximum error between a GP2$(\sigma_{\xi<0}, \xi < 0)$ and the reference distribution GP2$(100, 0)$, by varying the value of $\xi$ from $-1$ to $0$. The result is depicted in Figure 3a. The blue (solid) line represents the value $\bar{x}_2$ after which the upper-bound is
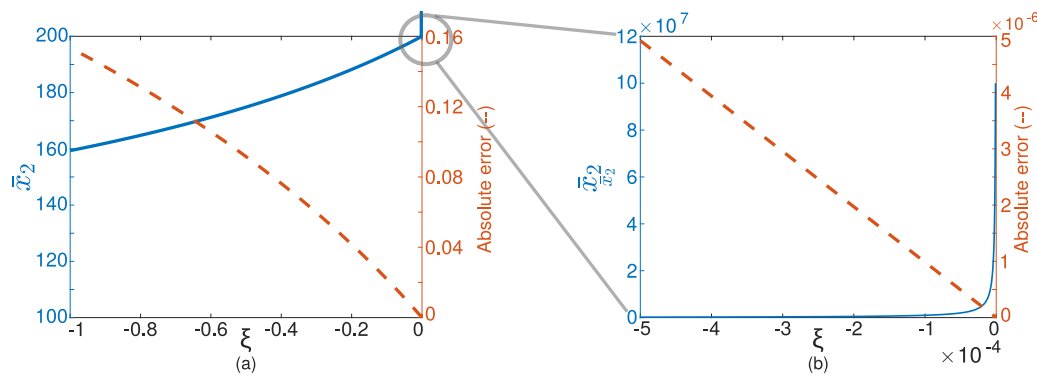
Fig. 3: Analysis of the safety point $\bar{x}_2$ (blue solid line) and the maximum absolute error (red dashed line), varying the real $\xi < 0$ and compared to $\xi = 0$ case. The (b) plot is the zoom of the (a) plot when $\xi$ has a magnitude of $10^{-4}$.

safe, while the red (dashed) line shows the maximum error compared to the reference distribution. The $\bar{x}_2$ value increases with a peculiar slope, for which we provide Figure 3b to show the trend for small values of $\xi$, while the maximum error has a quasi-linear trend. The key point here is to observe that there is no upper-bound for $\bar{x}_2$. Even if there exists a point $\bar{x}_2$ from which the upper-bounding is safe, it is not possible to know it without knowing the real value of $\xi$. This leads to an uncertainty on the pWCET estimation that can not be accepted in hard real-time systems. On the other hand, when $\xi$ is close to 0 and $\bar{x}_2$ increases towards infinite, the error decreases, but we still need to know $\xi$ in order to estimate both.

## IV. Conclusion

In hard real-time embedded systems, the reliability of the WCET analysis is essential in order to guarantee the timing constraints. Probabilistic real-time computing is a promising solution to deal with complex architectures, since it offers probabilistic-WCET estimations. Some recent works proposed to upper-bound the pWCET extreme value distribution when it has a light-tail ($\xi < 0$) with its exponential tail version ($\xi = 0$). While it simplifies the overall process, it may underestimate the pWCET value, if the $\xi$ value is bounded *a priori* with respect to the estimation phase. Previous works neglected this aspect by assuming the upper-bounding safe. We showed instead, that its validity holds only for WCET values greater than a certain unknown value $\bar{x}_2$. For these reasons, we should not blindly consider exponential-tail distributions on critical systems, without taking in account the $\xi$ parameter. Rather, the error magnitude must be estimated or the WCET must be proven to be large enough to guarantee a safe upper-bound.

## References

[1] W. Fornaciari, G. Agosta, D. Atienza, and e. a. Carlo Brandolese, "Reliable power and time-constraints-aware predictive management of heterogeneous exascale systems," in *2018 International Conference on Embedded Computer Systems: Architectures, Modeling, and Simulation (SAMOS)*, July 2018.

[2] C. Berg, J. Engblom, and R. Wilhelm, "Requirements for and design of a processor with predictable timing," in *Perspectives Workshop: Design of Systems with Predictable Behaviour*. Dagstuhl Seminar Proceedings IBFI, 2004.

[3] D. Dasari, B. Akesson, V. Nlis, M. A. Awan, and S. M. Petters, "Identifying the sources of unpredictability in cots-based multicore systems," in *2013 8th IEEE International Symposium on Industrial Embedded Systems (SIES)*, June 2013, pp. 39–48.

[4] S. J. Gil, I. Bate, G. Lima, L. Santinelli, A. Gogonel, and L. Cucu-Grosjean, "Open challenges for probabilistic measurement-based worst-case execution time," *IEEE Embedded Systems Letters*, vol. 9, no. 3, pp. 69–72, Sept 2017.

[5] F. Reghenzani, G. Massari, and W. Fornaciari, "The extreme value theory in probabilistic real-time computing," Politecnico di Milano, Tech. Rep., oct 2018, available online at https://doi.org/10.5281/zenodo.1462957.

[6] L. De Haan and A. Ferreira, *Extreme value theory: an introduction*. Springer Science & Business Media, 2007.

[7] F. Reghenzani, G. Massari, and W. Fornaciari, "chronovise: Measurement-based probabilistic timing analysis framework," *Journal of Open Source Software*, vol. 3, no. 28, 2018, available online at https://doi.org/10.21105/joss.00711.

[8] R. A. Fisher and L. H. C. Tippett, "Limiting forms of the frequency distribution of the largest or smallest member of a sample," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 24, no. 2, pp. 180–190, 1928.

[9] B. Gnedenko, "Sur la distribution limite du terme maximum d'une serie aleatoire," *Annals of Mathematics*, vol. 44, no. 3, pp. 423–453, 1943.

[10] R.-D. Reiss and M. Thomas, *Statistical Analysis of Extreme Values*. Birkhäuser, Basel, 2007.

[11] J. Hansen, S. Hissam, and G. A. Moreno, "Statistical-Based WCET Estimation and Validation," in *9th International Workshop on Worst-Case Execution Time Analysis*, N. Holsti, Ed., vol. 10. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2009, pp. 1–11.

[12] L. Cucu-Grosjean, L. Santinelli, M. Houston, C. Lo, T. Vardanega, L. Kosmidis, J. Abella, E. Mezzetti, E. Quiones, and F. J. Cazorla, "Measurement-based probabilistic timing analysis for multi-path programs," in *24th Euromicro Conference on Real-Time Systems*, July 2012.

[13] J. Abella, E. Quiones, F. Wartel, T. Vardanega, and F. J. Cazorla, "Heart of gold: Making the improbable happen to increase confidence in mbpta," in *26th Euromicro Conference on Real-Time Systems*, July 2014, pp. 255–265.

[14] L. Santinelli, F. Guet, and J. Morio, "Revising measurement-based probabilistic timing analysis," in *IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS)*, April 2017, pp. 199–208.

[15] G. Lima, D. Dias, and E. Barros, "Extreme value theory for estimating task execution time bounds: A careful look," in *2016 28th Euromicro Conference on Real-Time Systems (ECRTS)*, July 2016, pp. 200–211.

[16] J. Abella, M. Padilla, J. D. Castillo, and F. J. Cazorla, "Measurement-based worst-case execution time estimation using the coefficient of variation," *ACM Trans. Des. Autom. Electron. Syst.*, vol. 22, no. 4, pp. 72:1–72:29, Jun. 2017.

[17] L. F. Arcaro, K. P. Silva, and R. S. D. Oliveira, "On the reliability and tightness of gp and exponential models for probabilistic wcet estimation," *ACM Trans. Des. Autom. Electron. Syst.*, vol. 23, no. 3, pp. 39:1–39:27, Mar. 2018.

[18] I. Agirre, F. J. Cazorla, J. Abella, C. Hernandez, E. Mezzetti, M. Azkarate-askatsua, and T. Vardanega, "Fitting software execution-time exceedance into a residual random fault in iso-26262," *IEEE Transactions on Reliability*, pp. 1–14, 2018.