

# A Unified Framework for Over-Clocking Linear Projections on FPGAs under PVT Variation

Rui Policarpo Duarte and Christos-Savvas Bouganis

Imperial College London, United Kingdom

{r.duarte09, christos-savvas.bouganis}@imperial.ac.uk

**Abstract.** Linear Projection is a widely used algorithm often implemented with high throughput requisites. This work presents a novel methodology to optimise Linear Projection designs that outperform typical design methodologies through a prior characterisation of the arithmetic units in the data path of the circuit under various operating conditions. Limited by the ever increasing process variation, the delay models available in synthesis tools are no longer suitable for performance optimisation of designs, as they are generic and only take into account the worst case variation for a given fabrication process. Hence, they heavily penalise the optimisation strategy of a design by leaving a gap in performance. This work presents a novel unified optimisation framework which contemplates a prior characterisation of the embedded multipliers on the target device under PVT variation. The proposed framework creates designs that achieve high throughput while producing less errors than typical methodologies. The results of a case study reveal that the proposed methodology outperforms the typical implementation in 3 real-life design strategies: high performance, low power and temperature variation. The proposed methodology produced Linear Projection designs that were able to perform up to 18 dB better than the reference methodology.

**Keywords:** FPGA, embedded multiplier, over-clocking, PVT optimisation, Linear Projection.

## 1 Introduction

The Linear Projection algorithm, also known known as Karhunen-Loeve Transformation (KLT), is used in many scientific areas to compress data, i.e. face recognition. Additionally, the advent of *big data* and near real-time performance requirements has propelled an increased demand in performance for implementations of this algorithm [5].

The continuous increase of resources and performance in Field-Programmable Gate Arrays (FPGAs), along with low power consumption and highly specialised embedded blocks, has made them attractive to implement the Linear Projection algorithm, which relies on multiplications and additions, with high throughput requirements. In real-life applications, where usually there are hard area requirements, or when the Linear Projection targets a large number of dimensions, unrolling or deeply pipelining the design is unfeasible. Thus, the multiplication is assigned to the embedded multipliers and the convolution is folded. To overcome the performance limitation of the embedded multipliers, which can't be deeply pipelined, the Linear Projection circuit is over-clocked to



characterisation of the device with respect to the degradation of the performance of the computational units. This characterisation suits the use of FPGAs because of their reconfigurability property.

The proposed framework automatically selects the best designs to improve the performance of Linear Projection implementations, while minimising errors without the expense of extra circuit resources. The obtained performance information (i.e. errors that are expected at the output of the multipliers) is injected to a Bayesian formulation of the problem in order to obtain the coefficients for the Linear Projection.

This framework uses an error model (for the operating conditions), and later automatically combines this information with high-level parameter selection of the algorithm, generating designs less prone to error, when compared to typical implementations of the KLT algorithm.

The main contributions in this paper are:

- Extension of the characterisation and optimisation frameworks to support embedded multipliers;
- Introduction of the support for PVT variation in the characterisation framework and in the optimisation algorithm;
- Creation of a new error model for the embedded multipliers under a range of operating conditions;
- Optimisation of Linear Projection designs for performance targeting different scenarios (i.e. low power, high performance, temperature variation resilience).

## 2 Background

Usually the implementation and performance of Linear Projections designs in a digital system is bound to the number of bits used in quantisation and the depth of pipelining. However, these methodologies are unable to cope with the aforementioned operating conditions, consequence of the adopted design strategies, and the intra-die and inter-die variation.

The ever increasing process variation and the fact that circuits need to support different voltages and temperatures makes the designs to operate at lower clock frequencies than the maximum offered by the fabrication process. Synthesis tools use conservative models which set the maximum performance of a circuit below the performance of the worst transistor for the family of the device.

One of the most well known techniques that can be applied to address the problem of performance variability in a Linear Projection design within a device is Razor [13]. It is a generic time-redundant method proposed for Dynamic Voltage Scaling (DVS) of CPUs. More recently, [6] proposed a methodology to recover from errors due to PVT. Both strategies can be applied to any path prone to errors due to time violations, and the recovery is performed at the expense of extra latency, which heavily penalise Linear Projections applications processing streams of data.

In [10] the authors present two strategies to compensate for *intra-die* performance variability by providing a generic characterisation step for the performance of the device followed by a reconfiguration step, where parts of the design are mapped to specific locations of the device given their performance requirements.

A novel approach to improve the performance of Linear Projection designs, using Constant Coefficient Multipliers (CCMs), relied on over-clocking of the design [7]. Notwithstanding this work is restricted to CCMs and doesn't consider different operating conditions demanded by different designs strategies. In this paper, this constraint is lifted to make the proposed framework practical in a wider range of real-life applications. Furthermore, a new optimisation framework that utilises information from a prior characterisation for a Linear Projection design optimisation is presented. It includes a new algorithm for design space exploration that utilises an objective function tuned for the utilisation of embedded multiplier modules within that Linear Projection framework under different operating conditions. Moreover, since the routing inside the embedded multipliers doesn't change, it means that the designs are optimised on a per device basis.

The following sections of this paper detail the proposed methodology to accelerate Linear Projection designs, while being resilient to PVT variation, based on the prior characterisation of the device.

The proposed framework breaks new ground proposing:

- graceful degradation of results at the output of the Linear Projection with the increase in variation of the working conditions;
- a methodology to push forward the performance of embedded multipliers without using extra circuitry;
- a methodology to optimise a design over a set of varying conditions performing better than accounting for the worst case scenario.

### 3 Linear Projection Revisited

The Linear Projection, also known as KLT, or Principal Component Analysis (PCA), transform is formulated as follows. Given a set of  $N$  data  $x^i \in R^P$ , where  $i \in [1, N]$  an orthogonal basis described by a matrix  $A$  with dimensions  $P \times K$  can be estimated that projects these data to a lower dimensional space of  $K$  dimensions. The projected data points are related to the original data through the formula in (1), written in matrix notation, where  $X = [x^1, x^2, \dots, x^N]$  and  $F = [f^1, f^2, \dots, f^N]$ , where  $f^i \in R^K$  denote the factor coefficients.

$$F = A^T X. \quad (1)$$

The original data can be recovered from the lower dimensional space via (2):

$$X = AF + D \quad (2)$$

where  $D$  is the error of the approximation. The objective of the transform is to find a matrix  $A$  such as the Mean-Square Error (MSE) of the approximation of the data is minimised. A standard technique is to evaluate the matrix  $A$  iteratively as described in steps (3) and (4), where  $\lambda_j$  denotes the  $j^{th}$  column of the  $A$  matrix.

$$\lambda_j = \arg \max E\{(\lambda_j^T X_{j-1})^2\} \quad (3)$$

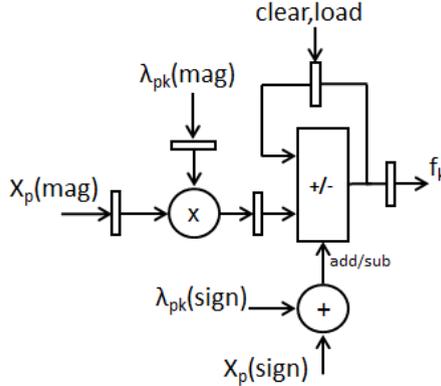
$$X_j = X - \sum_{k=1}^{j-1} \lambda_k \lambda_k^T X \quad (4)$$

where  $X = [x^1 x^2 \dots x^N]$ ,  $X_0 = X$ ,  $\|\lambda_j\| = 1$  and  $E\{\cdot\}$  refers to expectation.

## 4 Optimisation of Linear Projection Designs for Over-Clocking

In the circuit to implement the Linear Projection design, the data path holds the most critical paths. The main purpose of this work focus on over-clocking embedded multipliers, as they're the components with the largest delay in the data path of the design.

Figure 2 shows a rolled architecture of a circuit to implement the data path of one projection vector from a  $Z^p$  to  $Z^k$  Linear Projection. This was preferred instead of the unrolled one due to the amount of area taken by it.



**Fig. 2.** Schematic of the data path for one projection vector of a Linear Projection circuit

The input stream of the circuit is identified with  $X_p$ . The input sample from each dimension is multiplied by the corresponding projection vector  $\lambda_{pk}$ . This multiplication only concerns the magnitude of the input and  $\lambda_{pk}$  values. The inputs and outputs of the multiplier are registered. The output of the multiplier is connected to an adder to do the accumulation. The accumulation sign depends on the signs of the input and  $\lambda_{pk}$ . The output stream is identified with  $f_k$ .

The calculation of the projection matrix  $A$  and its hardware mapping onto FPGAs are often considered as two independent steps in the design process. However, considerable area savings can be achieved by coupling these two steps as shown in [4,3]. The Bayesian formulation presented in [3] considers the subspace estimation and the hardware implementation simultaneously, allowing the framework to efficiently explore the possibilities of custom design offered by FPGAs. This framework generates Linear Projection designs which minimise errors and circuit resources, when compared to the standard approach of the KLT transform application followed by the mapping to the FPGA.

The key idea in [3] is to inject information about the hardware (i.e. in this case about the required hardware resources of a CCM) as a prior knowledge in the Bayesian formulation of the above optimisation problem. In more detail, the proposed framework in [3] estimates the basis matrix  $A$ , the noise covariance  $\Psi$ , and the factors using Gibbs

sampling algorithm [8] from the posterior distribution of the variables, having injecting knowledge about the required hardware resources for the implementation of the CCMs through a prior distribution. Thus, a probability density function is generated for the unknown  $\Lambda$  matrix, which is used to for generation of samples, where the prior distribution tunes this posterior distribution, and thus accommodating the impact the required hardware resources.

[7] provides an extension of the above work for the optimisation of Linear Projection designs using CCMs combating the effects for circuit area as well as performance variation due to over-clocking. This work is focused on the extension of the previous work to support embedded multipliers and PVT variation in the characterisation, error modeling, and generation of designs to implement (1). The framework selects the multipliers used for the implementation of each dot product in (1) along with the coefficients of the  $\Lambda$  matrix that define the lower dimension space.

#### 4.1 Objective Function

The objective function is formed by the MSE of the reconstructed data in the original space, and errors that are produced due to the over-clocking, under PVT variation, of the utilised embedded multipliers.

Let's denote with  $\hat{X}$  the result of the reconstruction of the projected data in a matrix form. Then, the objective function  $U$  is defined as in (5), where both reconstruction errors and variation errors are captured.  $\mathbb{E}$  denotes the expectation and  $\text{Tr}$  the trace operator. The matrix formulation is defined as:

$$U = \text{Tr} \left( \mathbb{E} \left[ \left( X - \hat{X} \right)^T \left( X - \hat{X} \right) \right] \right) \quad (5)$$

By expressing the reconstructed data as a function of the  $\Lambda$  matrix, and the variation error with  $\varepsilon$  such as  $\hat{X} = \Lambda(F + \varepsilon)$ .

By imposing  $\varepsilon$  to have zero mean, which is achieved by subtracting a constant in the circuit, and using the fact that the  $\Lambda$  matrix is orthogonal and orthonormal, the objective function is expressed as:

$$\begin{aligned} U &= \text{Tr} \left( \mathbb{E} \left[ \left( X - \Lambda F \right)^T \left( X - \Lambda F \right) \right] \right) + \text{Tr} \left( \mathbb{E} \left[ \varepsilon^T \varepsilon \right] \right) \\ &= \text{Tr} \left( \mathbb{E} \left[ \left( X - \Lambda F \right)^T \left( X - \Lambda F \right) \right] \right) + \sum_j \text{var}(\varepsilon_j) \end{aligned}$$

Here  $j$  denotes the columns of the  $\Lambda$  matrix. By assuming that the errors at the output of the multipliers are uncorrelated, then the first term in the final expression relates to the approximation of the original data from the Linear Projection without any variation errors, where the second term relates to the variance of the errors at the output of the embedded multipliers due to over-clocking, under PVT variation. Thus, the errors due to dimensionality reduction and variation are captured by one objective function without any need to formulate a problem using a multi-objective function.

## 4.2 Prior Distribution Formation

The proposed framework utilises information regarding the performance characterisation of the embedded multipliers for a given device and their respective resource utilisation, by suitably constructing a prior distribution function for the coefficients of the  $\Lambda$  matrix. The utilised models for the over-clocking errors, under PVT variation, are described below.

**Error Models.** The proposed framework utilises the performance characterisation framework for CCMs, introduced in [7] and now extended to support embedded multipliers and capture PVT variation. By executing that framework, a profile of the errors expected at the output of the embedded multipliers when one of the operands is fixed (i.e. representing a coefficient of the  $\Lambda$  matrix) for various frequencies can be obtained. As indicated by the objective function formulation, the objective is to capture the variance of the error at the output of the multiplier which models the uncertainty of the result. As such, a data structure is formed,  $Err(m, f, p, v, T)$ , that holds information regarding the variance at the output of a multiplier when a stream of data is multiplied by a constant  $m$ , the circuit is clocked at frequency  $f$ , placed on  $p$  coordinates on the FPGA, using core voltage  $v$ , and temperature  $T$ .

**Prior Distribution.** The formation of the prior distribution  $p(\cdot)$  of the  $\Lambda$  matrix is a key part of the framework as it injects hardware information to the framework for the estimation of the  $\Lambda$  matrix. The aim of the prior distribution is to penalise  $\Lambda$  matrix instances with high errors, due to the use of coefficients that generate high errors due to over-clocking or due to poor description of the original space, by assigning low probabilities to them. As no information regarding the distribution of the coefficients is available on their suitability in representing the original space, this part of the prior distribution is uninformative and results to a flat prior. Thus, the prior distribution reflects solely information about the errors at the output of the over-clocked multipliers as  $p(\lambda_{pk}, f, p, v, T) = g(Err(\lambda_{pk}, f, p, v, T))$ , where the performance of every coefficient in the  $\Lambda$  matrix is dictated by the targeted clock frequency, the placement on the FPGA, the core voltage, and the temperature of the device; and  $g(\cdot)$  denotes a user defined function. In this work, the following  $g(\cdot)$  function is selected as it provides good results, without any claim on its optimality.

$$g(Err(\lambda_{pk}, f, p, v, T)) = c_E(1 + Err(\lambda_{pk}, f, p, v, T))^{-\beta} \quad (6)$$

$c_E$  is a constant used to ensure that  $\sum_{\lambda_{pk}} g(Err(\lambda_{pk}, f, p, v, T)) = 1$ .  $\beta$  is a *Hyper-Parameter* that allows the tuning of the contribution of errors in the prior distribution.  $Err(\lambda_{pk}, f, p, v, T)$  is the variance of the error observed from the performance characterisation of the multiplier.

## 4.3 Design Generation

The proposed framework uses Gibbs sampling [8] to extract, from the design space, a set of designs that minimise the selected objective function  $U$ . The resulting designs

are the ones that minimise the value of the objective function. The proposed framework estimates each dimension (i.e. column) of the  $\Lambda$  matrix in a sequential manner. The user supplies the targeted dimensions  $K$ , the targeted clock frequency  $f$ , the coordinates on the FPGA  $p$ , the core voltage  $v$ , the temperature  $T$ , and the  $\beta$  parameter. The pseudocode of the new version for the optimisation algorithm is given in Alg. 1.

---

**Algorithm 1.** Linear Projection Design Unified Framework for Over-Clocking
 

---

**Require:**  $K \geq 1 \wedge \beta > 0 \wedge f, p, v, T > 0$

**Ensure:** 1 Linear Projection design

$X \leftarrow \text{input}$  {original data N cases}

**for**  $d = 1$  **to**  $K$  **do**

  Create new empty *Candidate\_Projs* list

$\text{prior} \leftarrow \text{generate\_prior}(\beta, f, p, v, T)$

$\lambda_d \leftarrow \text{sample\_projection}(X, \text{prior})$

$F \leftarrow (\lambda_d^T \lambda_d)^{-1} \lambda_d^T X$

$\text{error} \leftarrow X - \sum_{j=1}^d \lambda_j F$

$\text{MSE}_d \leftarrow \sum \sum \text{error}^2 / PN$

$\text{Proj} \leftarrow (\lambda_d, \text{MSE}_d)$

  Add *Proj* to *Candidate\_Projs* list

  Extract candidate projections {min MSE}

**end for**

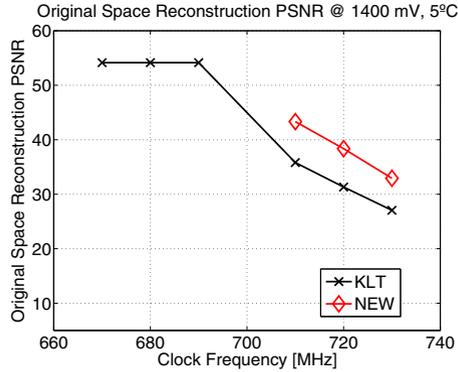
**return** The Linear Projection design with minimum *MSE*

---

## 5 Evaluation of Over-Clocked Linear Projections Circuits under PVT Variation

The performance of the proposed methodology was compared contra the performance of the reference design, which is based on a typical implementation of the KLT algorithm. All designs were implemented on a Cyclone III EP3C16 FPGA from Altera [2], attached to a DE0 board from Terasic [11]. The core voltage of the FPGA was provided by a digital power supply PL303QMD-P [1] from TTI. The temperature of the FPGA was set by a thermoelectric cooler placed on top of the FPGA. The temperature controller was calibrated using a digital thermometer from Lascar Electronics [9] and its deviation is below 1 °C.

The aim of this case study is to demonstrate that an optimisation of a Linear Projection design targeting different design strategies, under different operating conditions, using the same framework is achievable. The effectiveness of the framework is demonstrated with a case study implementing a Linear Projection from  $Z^6$  to  $Z^3$ . The characterisation of the FPGA, the training of the framework and the test used different sets of data from a uniform pseudo-random distribution, quantised with 9 bits. After synthesis, the tool reported a resource usage of 126 logic cells and  $3 \ 9 \times 9$  embedded multipliers, and a maximum clock frequency of 342 MHz. Examining the timing report revealed that the critical paths belong to the embedded multiplier and the delay for the remaining



**Fig. 3.** Comparison of the performance of the two methodologies for the particular case of 1400 mV and 5 °C

components in the data path, i.e. accumulator, and the Finite State Machine (FSM), are out of reach for the selected over-clocking frequencies. The results from the characterisation of the embedded multipliers were verified using Transition Probability from [12]. To better demonstrate the impact of variation for each design strategy only one setting has been changed. The framework supports variation of many operating conditions simultaneously. The results for the reference implementation without information about the characterisation of the device are identified with **KLT**, whereas the results for the proposed framework are identified with **NEW**. They are compared in terms of Peak Signal-to-Noise Ratio (PSNR) of the reconstructed data in the original space.

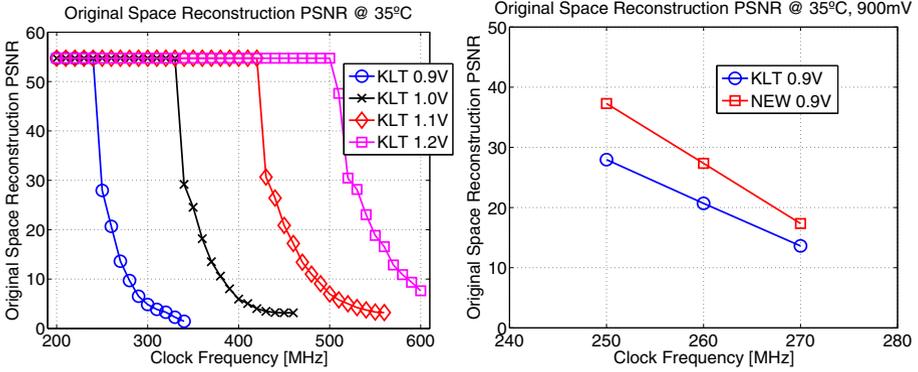
### 5.1 Optimisation Targeting Maximum Performance

Optimising a Linear Projection design aiming for the maximum performance implies an increased FPGA core voltage and active cooling of the device. During the test, the device was kept at 5 °C and supplied with 1400 mV, instead of the 1200 mV specified by the manufacturer.

With a clock frequency twice as much as the maximum specified by the synthesis tool for the normal working conditions, the designs generated by the proposed framework exhibited a reconstruction PSNR up to 15 dB better than the KLT designs for the same working conditions, as can be observed in figure 3. On the other hand, if a target PSNR of 30 dB is to be met, then the designs generated by the framework can operate up to 20 MHz higher than the KLT designs.

### 5.2 Optimisation Targeting Low Voltage

Linear Projection circuits operating under limited power budgets, or battery operated, tend to operate using the least core voltage possible and be without any active cooling components. Figure 3 (left) shows the results for the KLT designs when operating at 35 °C with different FPGA core voltages. This design strategy considered 900 mV as



**Fig. 4.** Performance of the KLT Linear Projection application under different core voltages (left), and a comparison between the two methods for the particular case of 900mV (right)

the minimum core voltage for the FPGA. Figure 3 (right) shows that the designs created by the framework achieve a better PSNR up to 10 dB for the same clock frequency, or for similar PSNR, a clock frequency up to 10 MHz higher than the reference designs.

### 5.3 Optimisation Targeting Device Temperature Tolerance

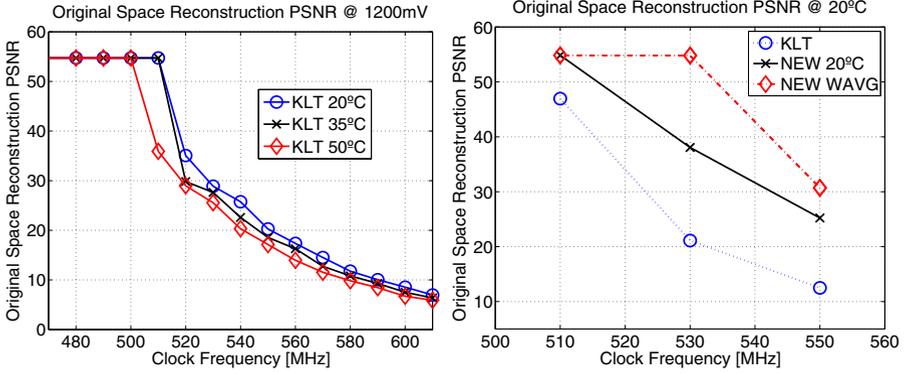
It is well established that temperature affects the performance of silicon devices. Implementing Linear Projection designs without any active cooling components, and operating them in environments prone to large temperature variation can compromise their correct functioning. Usually, if an implementation of a Linear Projection has to consider a wide range of temperatures, then it will have to cope with the worst performance of them.

To go beyond with the optimisation methodology, it was considered a scenario where a single design could offer better performance than the reference designs for a range of temperatures, instead of a design per temperature. Seeing that the errors increase with the temperature, optimising a design for the worst-case temperature can restrict the coefficients available to implement the Linear Projection design, hence placing a ceiling on the best reconstruction MSE that a particular design could achieve, even without errors in its data-path.

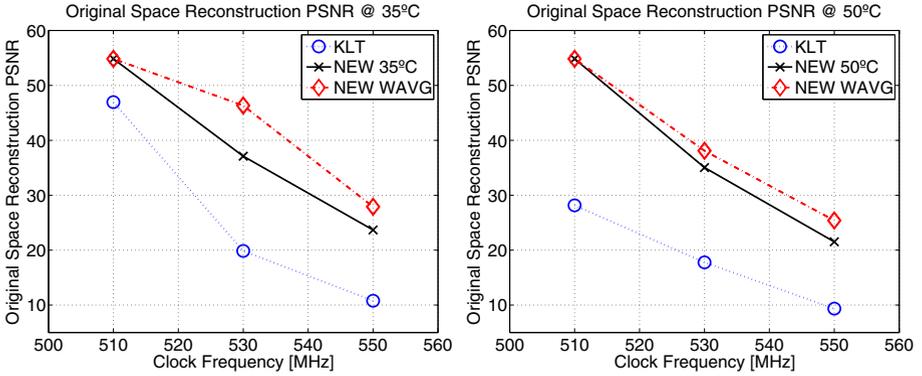
The new idea is focused on sampling Linear Projection designs using the information from the characterisation of the device at specific temperatures along with its probability to operate under those temperatures. To accomplish this, it was investigated the weighted average of the characterisation errors for a range of operating temperatures in the generation of Linear Projection designs. As follows, the prior distribution from equation (7) is now:

$$g(Err(\lambda_{pk}, f, p, v, T)) = \sum_i \alpha_i c_E (1 + Err(\lambda_{pk}, f, p, v, T_i))^{-\beta} \quad (7)$$

Here  $i$  iterates over all contributing temperatures, and  $\sum_i \alpha_i = 1$ . The different weights represent the significance of the errors at a particular temperature. This



**Fig. 5.** Performance of the KLT Linear Projection application depending on the temperature of the device (left), and a comparison between the three methods at 20 °C (right)



**Fig. 6.** Performance of the three methods at 35 °C (left) and 50 °C (right)

particular test case used temperatures 20, 35 and 50 °C and  $\alpha_{20} = 0.3$ ,  $\alpha_{35} = 0.5$  and  $\alpha_{50} = 0.2$ . In practice, the proposed framework generates circuit designs per clock frequency, covering all the temperatures within the expected range. They are identified with **NEW WAVG** in the results.

Figure 5 (left) shows the dependency of the performance of the reference Linear Projection circuit with the temperature of the device, with a supply voltage of 1200 mV. Figure 5 (right) shows in detail the comparison between the reference and the optimised designs for a specific temperature and a range of temperatures. Figure 6 holds the results for 35 and 50 °C.

The figures show that the designs generated by the framework always outperformed the KLT designs for all temperatures. Furthermore, at 510 MHz the PSNR is more than 10 dB better than the KLT design, and at 530 MHz the performance of the NEW design at 35 °C is better than the KLT design at 20 °C. The NEW WAVG designs perform significantly better than the NEW ones since they incorporate more information about uncertainty of the results under variation, up to the maximum temperature considered.

## 6 Conclusion

This paper proposes a novel unified methodology for implementation of extreme over-clocked Linear Projection designs on FPGAs. It couples the problem of data approximation and error minimisation under variation of the operating conditions. It was demonstrated that the proposed methodology optimises Linear Projection designs for performance and resilience simultaneously, by inserting information regarding the performance of the arithmetic units when operating under variation.

**Acknowledgments.** RPD would like to acknowledge “Fundação para a Ciência e Tecnologia” (Lisbon, Portugal) for the support through grant SFRH/BD/69587.

## References

1. Aim & Thurlby Thandar Instruments. The new pl-p series - advanced bus programmable DC power supplies, <http://www.tti-test.com/products-tti/pdf-brochure/psu-npl-series-8p.pdf>
2. Altera. Cyclone III device handbook, [http://www.altera.co.uk/literature/hb/cyc3/cyclone3\\_handbook.pdf](http://www.altera.co.uk/literature/hb/cyc3/cyclone3_handbook.pdf)
3. Bouganis, C.-S., Pournara, I., Cheung, P.: Exploration of heterogeneous FPGAs for mapping linear projection designs. *IEEE Trans. VLSI Syst.* 18(3), 436–449 (2010)
4. Bouganis, C.S., Pournara, I., Cheung, P.Y.K.: Efficient mapping of dimensionality reduction designs onto heterogeneous FPGAs. In: *Proc. 15th Annual IEEE Symp. Field-Programmable Custom Computing Machines FCCM 2007*, pp. 141–150 (2007)
5. Chu, J.-U., Moon, I., Seong Mun, M.: A real-time EMG pattern recognition system based on linear-nonlinear feature projection for a multifunction myoelectric hand. *IEEE Transactions on Biomedical Engineering* 53(11), 2232–2239 (2006)
6. Das, S., Tokunaga, C., Pant, S., Ma, W.-H., Kalaiselvan, S., Lai, K., Bull, D.M., Blaauw, D.T.: RazorII: In situ error detection and correction for PVT and SER tolerance. *IEEE J. Solid-State Circuits* 44(1), 32–48 (2009)
7. Duarte, R.P., Bouganis, C.-S.: High-level linear projection circuit design optimization framework for FPGAs under over-clocking. In: *FPL*, pp. 723–726 (2012)
8. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6(6), 721–741 (1984)
9. Lascar Electronics. El-usb-tc k, j, and t-type thermocouple temperature usb data logger, <http://www.lascarelectronics.com/temperaturedatalogger.php?datalogger=364>
10. Sedcole, P., Cheung, P.Y.K.: Parametric yield modeling and simulations of fpga circuits considering within-die delay variations. *ACM Trans. Reconfigurable Technol. Syst.* 1(2), 10:1–10:28 (2008)
11. Terasic Technologies. Terasic DE0 board user manual v. 1.3 (2009)
12. Wong, J.S.J., Sedcole, P., Cheung, P.Y.K.: A transition probability based delay measurement method for arbitrary circuits on FPGAs. In: *Proc. Int. Conf. ICECE Technology FPT 2008*, pp. 105–112 (2008)
13. Ziesler, C., Blaauw, D., Austin, T., Flautner, K., Mudge, T.: Razor: A low-power pipeline based on circuit-level timing speculation (2003)