# A Statistical Model for System Components Selection

Varuna Gupta, Abdelkader Mazouz, Ankur Agarwal, Georgiana Hamza-Lup
College of Engineering & Computer Science
Florida Atlantic University
Boca Raton, Florida
{vgupta1, mazouzk, aagarwa2, ghamzal}@fau.edu

***Abstract*— The proposed research is aimed at enhancing system design productivity by exploiting the principle of "design and reuse" to its full potential. Specifically, we present a statistical model for selecting from a component library the optimal components for a network-on-chip architecture such that to satisfy certain system performance requirements. Our model is based on regression analysis and Taguchi's optimization technique. The model estimates the relationship between system performance and component attributes, to help the architect in the component selection process. Having such a model in the system design phase will allow the architect not only to make informed decisions when selecting components but also to exchange components with similar characteristics to fine tune system performance.**

***Keywords - regression analysis, Taguchi optimization, performance evaluation, component selection, network-on-chip***

## I. INTRODUCTION

System level design, long the domain of a few expert senior engineers, is becoming a new field of study [1]. For this discipline to grow and enhance engineering design productivity, many key concepts have to be adopted from other domains [2]. In this paper we present the use of statistical models and regression analysis in system design, specifically in component selection.

Nowadays, companies must speculate on what the consumer preferences will be in the following two years and start developing a product in advance to meet the predicted demand, which may however not become reality. A wrong prediction may adversely hurt the fortune of a company. To lessen this problem, the product design cycle needs to be further shortened from its current length of 24 to 30 months [3]. This requires integration of new technologies in the product development cycle, especially at the earlier stages, since errors introduced there get magnified significantly at later stages. As an example, NASA estimates that to fix an error introduced at the specification stage will require 138 times more effort at the prototyping stage and 536 times more effort in the field, then at the specification stage [4]. To minimize these errors system designers should exploit the “design & reuse” principle to its full potential, by designing systems (sub-systems) from reusable sub-systems (components). For example, a Network-on-Chip (NOC) architecture can serve as a reusable communication sub-system for an embedded device [5]. This NOC architecture may in turn comprise of several components such as, routers, input and output buffers, network interfaces, switches, virtual channel allocators, schedulers and switch allocators. To develop a system from such reusable components, one will have to design and develop variants of each component with different parameters. For example, buffer size is a customizable parameter for an input buffer. Similarly, scheduling criteria provides customizability for a scheduler. A system architect must then estimate the performance of various configurations of components and select the optimal configuration that satisfies the system performance requirements.

To allow the architect to make an informed decision during the selection process, each customized component needs to encapsulate performance parameters [6]. However, components performance cannot be measured in isolation but only in an integrated system. For this purpose the system must be modeled in advance at an abstract level that hides the detailed system functionality but allows analyzing system performance [7]. This model can be used for predicting system performance and estimating component requirements [8], [9]. Specifically, in this paper we propose to estimate the relationship between system performance and component attributes, to help the architect in the component selection process. Having such a model in the system design phase will allow the architect not only to make informed decisions when selecting components but also to exchange components with similar characteristics to fine tune system performance.

Current approaches for addressing component selection are mainly driven by system requirements that depend directly on the attributes of the individual components. For example, in [10] and [11] the authors assume that every component satisfies one or more requirements and the goal is to select the minimal set of components that together satisfy a set of given system requirements. While [10] and [11] aim at minimizing the number of components of the resulting system, in [12] the cost of the final system is minimized. This system requirement is directly dependent on the attributes of the system components. Specifically, the cost of the final system is the sum of the costs of the individual components. In [13] reliability and delivery time requirements are specified in addition to cost requirements. Again, these requirements depend directly on the attributes of the system components. In contrast to existing approaches, our proposed component selection technique is driven by system requirements, whose dependencies on the components attributes cannot be expressed by exact mathematical formulas. Therefore, our goal is to extract and approximate these dependencies, so that we can predict with certain accuracy the performance of a system, from the attributes of the individual components.

978-1-4244-9493-4/11/$26.00 ©2011 IEEE

The remainder of this paper is organized as follows. Section 2 describes the design of the experiment method used to determine the regression curve showing the relationship between component attributes and system performance. Section 3 describes the results we obtained for our selected example, the NOC architecture. Section 4 presents the two-step optimization using Taguchi's approach and finally Section 5 concludes the paper.

## II. Design of Experiment

The Design of Experiment (DOE) method is used to obtain the regression analysis curve that shows the impact of each component attribute on the system performance. This study will be extremely useful for system designers by providing them with suggestions for selecting the component parameters that can maximize system performance. For this purpose we will first perform regression analysis on the basis of the least-squares best fitted model and then secondly, we will apply the Taguchi two step optimization technique. The two analyses are described in the following subsections.

### *A. Regression Analysis*

Regression is a mathematical measure of the average relationship between two or more variables. Multiple regression models seek to predict an outcome from several variables. The method of least squares, which is the way of finding the line that best fits the data, is usually used, thus it will be applied in our experiment. One measure of adequacy of a model is the sum of the squared differences (between actual and predicted values). The squared differences (or residual values) provide a gauge of how well a particular line fits the data: if the squared difference is large the line is not a good representation of the data; if the squared difference is small the line is a good representation. The Analysis of Variance (ANOVA) model enlightens whether the model overall results in a significantly good degree of prediction of the outcome variable or not. The t-statistic measures how extreme a statistical estimate is. These analyses will be used to measure the adequacy of our model.

### *B. Taguchi 2-Step Optimization Technique*

Taguchi technique pioneered by the work of Dr. G. Taguchi is based on the assumption that input-output relationships are linear and there are no interactions between the input parameters. Experiments are designed using a special set of Latin squares "Orthogonal Arrays" (OA) to treat the design process such that the system performance is controlled early during the design stage. In OA, the columns represent the experimental factors to be studied and the numbers in the rows indicates the factor levels. The OA facilitates the experimental design process by assigning factors to the appropriate columns. The contribution of the individual factors that influence system performance is the deciding key in system design. ANOVA is used to analyze the results of the OA experiment and determine how much variation each performance influencing factor has contributed. By studying the main effects of each factor, the general trends of the influence factors towards system performance can be characterized and best results can be predicted.

Orthogonal arrays are used to study many parameters simultaneously with a minimum of time and resources to produce an overall picture for more detailed design and decision making. The signal-to-noise ratio (S/N) is employed to measure performance. S/N analysis determines the optimum set of components from variations of the performance influencing factors within the results. The signals are those factors which are invariant. Noise factors are any uncontrolled variables. Control factors are the variables that are set by the designers that will characterize the performance.

Taguchi's OA analysis is used to produce the best parameters for the optimum system performance, with the least number of experiments. The OA method treats performance influencing factors at discrete levels and often saves time and indirectly reduces the product development cost.

## III. Experimental Results

To explain the experimental results we have taken the problem of designing a NOC model. NOC architectures, discussed in detail in [14], have three main functional components: resources, buffers, and switches. Each component is characterized by different attributes. Resources, which are data producers, are characterized by packet injection rate (that is, the rate at which they produce packets to be delivered through the NOC). We assume 10 possible values for the injection rate, from 0.1, 0.2 … 0.9, 1.0. Buffers are used for storing data packets and are characterize by size and scheduling criteria. Switches are also characterized by scheduling criteria. Buffer sizes can have different values, such as: 1, 2, 3, 4 …. and scheduling criteria can have 4 possible values: PB (Priority Based), PBRR (Priority Based Round Robin), RR (Round Robin) and FCFS (First Come First Served). Thus, there are many possible combinations for compiling a NOC architecture and depending on the attributes of the components integrated, the resulting system will exhibit different performance.

The main performance measure for NOC architectures is network latency, which is the time taken by a data packet to travel from one resource to another. We assume there are three types of data packets injected into the network: high, medium and low priority packets, and there are different latency requirements for each type of packet. Using system simulation software (such as, MLDesigner) we can model the NOC architecture at an abstract level and measure the performance characteristics of a limited number of combinations of components. It is however unfeasible to examine all possible combinations of components in a timely manner. Therefore, it would be extremely helpful to understand the impact of each component attribute on the system performance, for example, the impact of the scheduling criteria on the network latency. The purpose of regression analysis is to extract these relationships.

The three response variables in our experiment are: latency for high priority (R1), mid priority (R2), and low priority (R3) data packets. These response variables are dependent on three input variables: injection rate (I), buffer size (B), and scheduling criteria (S). In this experiment, we assume we have measurements of the response variables for the following levels of the input variables: injection rate at three levels: 0.1, 0.5 and

1.0; buffer sizes at three levels 2, 5, and 10; scheduling criteria at four levels: FCFS-level 1, RR-level 2, PB-level 3 and PBRR-level 4. Our measurements obtained through the system simulation software are however different than the results of a real experiment. This is because simulation always follows a pattern, while real experiment results will deviate from expected value due to noise in the environment. Although this noise can be minimized, it can never be eliminated. To account for it we have introduced a noise function in our simulation model, uniformly distributed with 5% allowable error. We have then taken the simulation results twice to replicate a real scenario. A snapshot of the input data for the experiment is represented by the first 6 columns of Table 1.

The objective of performing regression analysis is to determine the best fitted regression model on the data provided for the NOC architecture. Excel's Data Analysis ToolPak, Design Expert and SPSS are the statistical packages used for the analysis. This analysis will:

1. Identify the best regression model for each of the 3 response variables (latency for high priority, mid priority and low priority packets), based on the least-squares method (using Excel's Analysis ToolPak)
2. Measure model adequacy using the sum of squared differences and ANOVA (in Design Expert); and check that independent variables contribution in predicting the estimates of the outcome using coefficient β analysis (in SPSS)

TABLE I. SNAPSHOT OF INPUT DATA AND R1 ESTIMATION AND RESIDUALS SQUARED FOR EACH OF THE FOUR MODELS

| Input data | | | | | | Model 1 | | Model 2 | | Model 3 | | Model 4 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| S | B | I | R1 | R2 | R3 | P R1 | $RS^2$ | P R1 | $RS^2$ | P R1 | $RS^2$ | P R1 | $RS^2$ |
| 1 | 2 | 0.1 | 24 | 24 | 24 | 33.62 | 92.48 | 24.64 | 0.41 | 50.47 | 700.87 | 41.50 | 306.17 |
| 1 | 2 | 0.1 | 24.9 | 24.9 | 24.9 | 33.62 | 75.98 | 24.64 | 0.07 | 50.47 | 654.03 | 41.50 | 275.48 |
| 2 | 2 | 0.1 | 23.1 | 22.87 | 24 | 14.40 | 75.76 | 23.99 | 0.79 | 19.95 | 9.93 | 29.54 | 41.46 |
| 2 | 2 | 0.1 | 23.7 | 23.3 | 24.6 | 14.40 | 86.56 | 23.99 | 0.08 | 19.95 | 14.07 | 29.54 | 34.09 |
| 3 | 2 | 0.1 | 22.02 | 24 | 24 | 13.74 | 68.52 | 23.33 | 1.72 | 7.99 | 196.85 | 17.58 | 19.72 |
| 3 | 2 | 0.1 | 22 | 24 | 24.5 | 13.74 | 68.19 | 23.33 | 1.77 | 7.99 | 196.29 | 17.58 | 19.54 |

Since the true functional relationships between the response variables and input variables are unknown, four different models are analyzed and the one with the least error is then selected and further discussed. These combinations are:

(1) Quadratic model, includes main effects + interaction + square terms: S, B, I, S*B, B*I, S*I, $S^2$, $B^2$, $I^2$;
(2) Model with main effects + interactions: S, B, I, S*B, B*I, S*I;
(3) Model with main effects + squared terms: S, B, I, $S^2$, $B^2$, $I^2$;
(4) Model with main effects only: S, B, I.

### A. Comparison of the Four Models

Table 1 shows the results for response variable R1 (latency for high priority packets). Columns 7, 9, 11 and 13 show the predicted values for R1 for each of the four models, and columns 8, 10, 12 and 14 show the corresponding residuals squared. The errors calculated for each of the four models are: 6210.197, 12419.96, 15218.29, and 21428.0544 respectively. This indicates that the quadratic model has the lowest error as expected and therefore, is the best. The regression equation for the response variable R1 is:

$$\mathbf{R1 = 49.639 - 40.96*S + 10.529*B + 8.30*I - 3.03*S*B + 0.06*B*I - 0.438*S*I + 9.28*S^2 - 0.03*B^2 - 1.63*I^2.}$$

### B. Measuring the Adequacy of the Quadratic Model

The ANOVA table, coefficient estimates in terms of coded factors and diagnostics results for the quadratic model, are presented in Tables 2 and 3. The model's F-value of 37.44 implies the model is significant. There is only a 0.01% chance that a value this large could occur due to noise. The values of "Prob > F" which are less than 0.0500 indicate which model terms are significant. In our case S, B, S*B, $S^2$ are significant model terms.

TABLE II. ANOVA TABLE FOR THE QUADRATIC MODEL [PARTIAL SUM OF SQUARES - TYPE III]

| Source | Sum of Squares | df | Mean Square | F Value | p-value Prob > F | |
|---|---|---|---|---|---|---|
| Model | 33749.92 | 9 | 3749.99 | **37.44** | < 0.0001 | significant |
| *S* | *14972.34* | *1* | *14972.34* | *149.48* | *< 0.0001* | |
| *B* | *5260.91* | *1* | *5260.91* | *52.52* | *< 0.0001* | |
| *I* | *322.37* | *1* | *322.37* | *3.22* | *0.0777* | |
| *SB* | *9005.32* | *1* | *9005.32* | *89.91* | *< 0.0001* | |
| *SI* | *2.35* | *1* | *2.35* | *0.023* | *0.8788* | |
| *BI* | *0.42* | *1* | *0.42* | *4.203E-003* | *0.9485* | |
| $S^2$ | *6204.79* | *1* | *6204.79* | *61.95* | *< 0.0001* | |
| $B^2$ | *3.28* | *1* | *3.28* | *0.033* | *0.8571* | |
| $I^2$ | *1.69* | *1* | *1.69* | *0.017* | *0.8969* | |
| Residual | 6210.20 | 62 | 100.16 | | | |
| *Lack of Fit* | *6197.49* | *26* | *238.37* | *675.45* | *< 0.0001* | *significant* |
| *Pure Error* | *12.70* | *36* | *0.35* | | | |
| Cor Total | 39960.12 | 71 | | | | |

TABLE III. COEFFICIENT ESTIMATES FOR THE QUADRATIC MODEL

| Factor | Coefficient Estimate | df | Standard Error | 95% CI | | VIF |
|---|---|---|---|---|---|---|
| | | | | Low | High | |
| Intercept | 25.54 | 1 | 3.16 | 19.22 | 31.86 | |
| S | -19.47 | 1 | 1.59 | -22.65 | -16.28 | 1.01 |
| B | 10.48 | 1 | 1.45 | 7.59 | 13.37 | 1.02 |
| I | 2.60 | 1 | 1.45 | -0.30 | 5.51 | 1.01 |
| SB | -18.19 | 1 | 1.92 | -22.02 | -14.35 | 1.01 |
| SI | -0.30 | 1 | 1.93 | -4.16 | 3.57 | 1.00 |
| BI | 0.11 | 1 | 1.75 | -3.38 | 3.61 | 1.01 |
| $S^2$ | 20.89 | 1 | 2.65 | 15.58 | 26.19 | 1.00 |
| $B^2$ | -0.49 | 1 | 2.70 | -5.88 | 4.90 | 1.02 |
| $I^2$ | -0.33 | 1 | 2.54 | -5.40 | 4.74 | 1.00 |

The normal probability plot of the residuals (Figure 1) resembles a straight line indicating error distribution is normal. It follows the usual normality assumptions. There is no evidence of possible outliers. Figure 2 shows a plot of the residuals versus the ascending predicted response values. It tests the assumption of constant variance. The plot should be a random scatter (constant range of residuals across the graph). Expanding variance ("megaphone pattern <") in this plot indicates the need for a transformation. The cube plot from Figure 3 indicates the treatment combinations. The experiment layout and treatment combinations notation is represented for the $4^1 3^2$ factorial design. Also, the contour plot (Figure 4) and the 3-D response surface (Figure 5), indicate that the there are interaction and square term effects in the model.

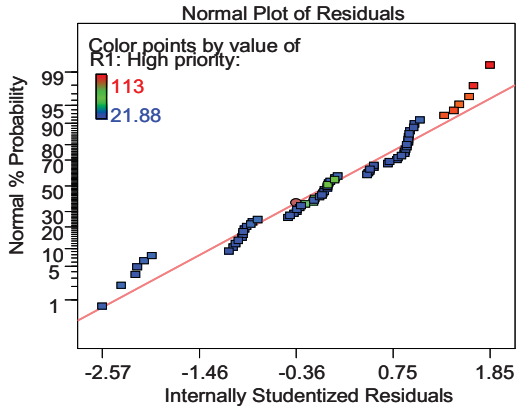


Figure 1. Normal probability of residuals

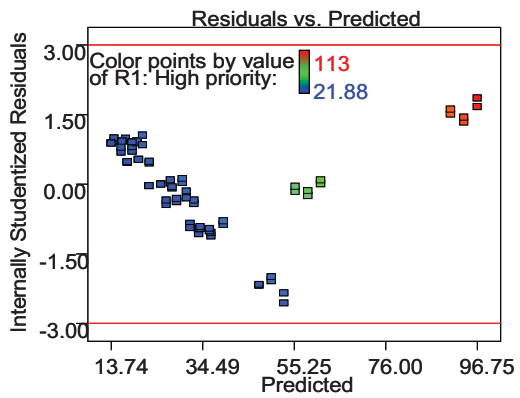


Figure 2. Predicted vs. residuals

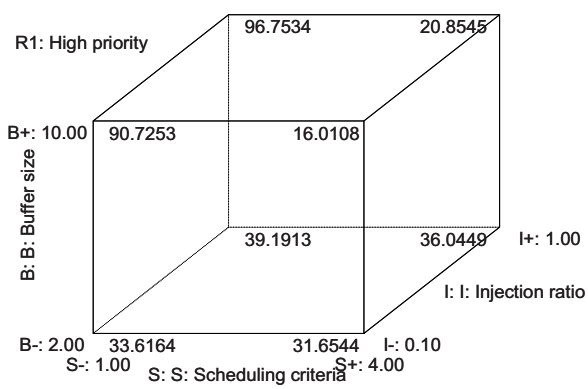


Figure 3. The cube plot

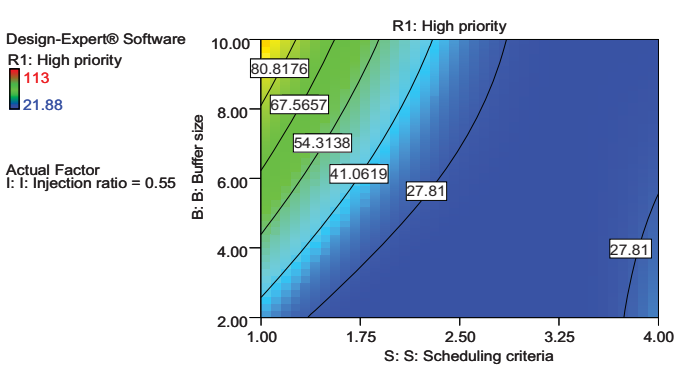


Figure 4. Contour plot

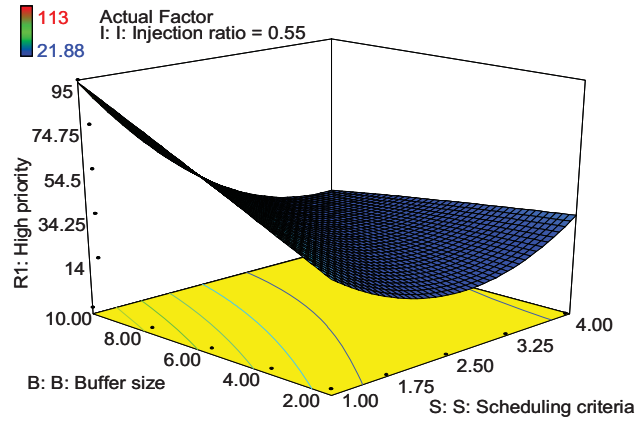


Figure 5. 3-D response methodology

## C. *Measuring Terms Contribution in the Quadratic Model*

The stepwise linear regression is conducted for the selected quadratic model using SPSS. Independent variables with the highest t-statistics are added first and this continues until there are none left with t-statistics less than 0.05 significance level. Each time a new independent variable is added to the equation, a removal test is made of the least useful variable. Thus, the regression equation is constantly reassessed to see whether any redundant variable can be removed. The overall model is summarized in Table 4. R2 is 0.836 for model d, which means independent variables can explain the 83.6% of variation in the response variable.

TABLE IV. MODEL SUMMARY

| Model | Predictors | R | $R^2$ | Adjusted $R^2$ | Std. Error of the Estimate |
|---|---|---|---|---|---|
| a | (Constant), S | 0.568 | 0.322 | 0.312 | 19.67174 |
| b | (Constant), S, $S^2$ | 0.691 | 0.477 | 0.462 | 17.39716 |
| c | (Constant), S, $S^2$, B | 0.782 | 0.611 | 0.594 | 15.11956 |
| d | (Constant), S, $S^2$, B, S*B | 0.915 | **0.836** | 0.827 | 9.87954 |

Table 5 shows that the model is significantly better at predicting the outcome than using the mean as the best guess. The F-value is quite high and the significance value is zero and hence the model is adequate. Therefore, the null hypothesis that the model is not adequate is rejected in this case.

TABLE V. ANALYSIS OF VARIANCE (ANOVA)

| Model | Predictors | | Sum of Squares | df | Mean Square | F-value | Sig. |
|---|---|---|---|---|---|---|---|
| a | (Constant), S | Regression<br>Residual<br>Total | 12871.711<br>27088.406<br>39960.117 | 1<br>70<br>71 | 12871.711<br>386.977 | 33.262 | 0 |
| b | (Constant), S, S2 | Regression<br>Residual<br>Total | 19,076.51<br>20883.612<br>39960.117 | 2<br>69<br>71 | 9538.253<br>302.661 | 31.515 | 0 |
| c | (Constant), S, S2, B | Regression<br>Residual<br>Total | 24415.243<br>33:24.9<br>39960.117 | 3<br>68<br>71 | 8138.414<br>228.601 | 35.601 | 0 |
| d | (Constant), S, S2, B, S*B | Regression<br>Residual<br>Total | 33420.566<br>6539.55<br>39960.117 | 4<br>67<br>71 | 8355.142<br>97.605 | 85.601 | 0 |

As can be seen in Table 6, the $\beta$ value which tells the relationship between the response and each independent

variable indicates that there is a negative relationship between S, S*B and R1 whereas there is a positive relationship between $S^2$, B and R1. Looking at the t-values one can see that the slopes are valid for all the models. They are significant values and hence the null hypothesis is rejected. Table 8 shows the variables excluded from the model, and Table 7 shows the residual statistics of the model.

TABLE VI. MODEL COEFFICIENTS

| Model | Predictors | Unstandardized Coefficients B | Unstandardized Coefficients Std. Err. | Standardized Coefficients Beta | t-value | Significance |
|---|---|---|---|---|---|---|
| a | (Constant) | 65.512 | 5.679 | | 11.536 | 0.000 |
| | S | -11.959 | 2.074 | -0.568 | -5.767 | 0.000 |
| b | (Constant) | 111.928 | 11.415 | | 9.805 | 0.000 |
| | S | -58.375 | 10.414 | -2.770 | -5.605 | 0.000 |
| | $S^2$ | 9.283 | 2.05 | 2.238 | 4.528 | 0.000 |
| c | (Constant) | 97.141 | 10.382 | | 9.357 | 0.000 |
| | S | -58.375 | 9.051 | -2.77 | -6.450 | 0.000 |
| | $S^2$ | 9.283 | 1.782 | 2.238 | 5.210 | 0.000 |
| | B | 2.61 | 0.540 | 0.366 | 4.833 | 0.000 |
| d | (Constant) | 54.197 | 8.125 | | 6.671 | 0.000 |
| | S | -41.197 | 6.178 | -1.955 | -6.668 | 0.000 |
| | $S^2$ | 9.283 | 1.164 | 2.238 | 7.973 | 0.000 |
| | B | 10.188 | 0.864 | 1.427 | 11.788 | 0.000 |
| | S*B | -3.031 | 0.316 | -1.42 | -9.605 | 0.000 |

TABLE VII. RESIDUAL STATISTICS

| | Minimum | Maximum | Mean | Std. Dev. | N |
|---|---|---|---|---|---|
| Predicted Value | 16.3414 | 93.8484 | 35.6149 | 21.6959 | 72 |
| Residual | -23.36706 | 19.15162 | 0.00000 | 9.59720 | 72 |
| Std. Predicted Value | -0.888 | 2.684 | 0.000 | 1.0000 | 72 |
| Std. Residual | -2.365 | 1.939 | 0.000 | 0.971 | 72 |

TABLE VIII. EXCLUDED VARIABLES

| Model | Excluded variables | Beta In | t | Sig. | Partial Correlation | Collinearity Statistics Tolerance |
|---|---|---|---|---|---|---|
| a | B | 0.366 | 4.115 | 0.000 | 0.444 | 1.000 |
| | I | 0.09 | 0.911 | 0.366 | 0.109 | 1.000 |
| | $S^2$ | 2.238 | 4.528 | 0.000 | 0.479 | 0.031 |
| | $B^2$ | 0.359 | 4.020 | 0.000 | 0.436 | 1.000 |
| | $I^2$ | 0.086 | 0.872 | 0.386 | 0.104 | 1.000 |
| | S*B | 0.171 | 1.432 | 0.157 | 0.170 | 0.670 |
| | B*I | 0.279 | 2.995 | 0.004 | 0.339 | 1.000 |
| | S*I | 0.092 | 0.798 | 0.427 | 0.096 | 0.741 |
| b | B | 0.366 | 4.833 | 0.000 | 0.506 | 1.000 |
| | I | 0.09 | 1.031 | 0.306 | 0.124 | 1.000 |
| | $B^2$ | 0.359 | 4.712 | 0.000 | 0.496 | 1.000 |
| | $I^2$ | 0.086 | 0.988 | 0.327 | 0.119 | 1.000 |
| | S*B | 0.171 | 1.626 | 0.109 | 0.193 | 0.670 |
| | B*I | 0.279 | 3.454 | 0.001 | 0.386 | 1.000 |
| | S*I | 0.092 | 0.904 | 0.369 | 0.109 | 0.741 |
| c | I | 0.09 | 1.19 | 0.238 | 0.144 | 1.000 |
| | $B^2$ | -.053 | -0.119 | 0.906 | -0.015 | 0.029 |
| | $I^2$ | 0.086 | 1.139 | 0.259 | 0.138 | 1.000 |
| | S*B | -1.420 | -9.605 | 0.000 | -0.761 | 0.112 |
| | B*I | 0.098 | 1.047 | 0.299 | 0.127 | 0.653 |
| | S*I | 0.092 | 1.042 | 0.301 | 0.126 | 0.741 |
| d | I | 0.09 | 1.848 | 0.069 | 0.222 | 1.000 |
| | $B^2$ | -.053 | -0.182 | 0.856 | -0.022 | 0.029 |
| | $I^2$ | 0.086 | 1.767 | 0.082 | 0.213 | 1.000 |
| | B*I | 0.098 | 1.621 | 0.110 | 0.196 | 0.653 |
| | S*I | 0.092 | 1.613 | 0.112 | 0.195 | 0.741 |

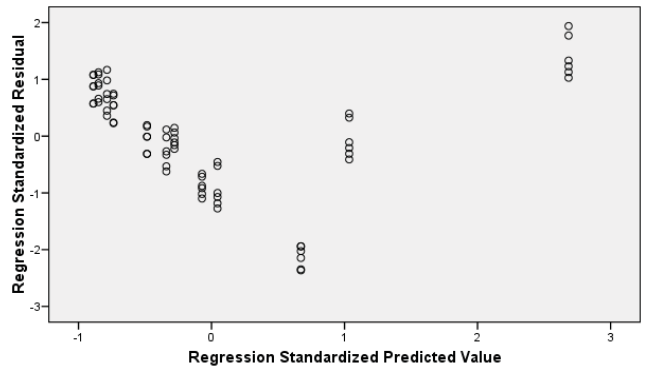


Figure 6. Scatter plot of standardized residuals vs. standardized predicted values

Figure 6 shows the relationship between the standardized predicted values of the response variable and the standardized residuals. From the plot it is seen that the assumptions of random errors are met.

Similar analyses are performed for the second and third response variables. The details of the analyses are omitted due to space limitations. The errors calculated for the four models indicate that again the quadratic model is the best for both response variables. The regression equations for these responses are given in Section 5.

## IV. TWO-STEP OPTIMIZATION USING TAGUCHI APPROACH

Taguchi two–step optimization technique is used to obtain a smaller design in terms of number of experiments. There are 3 factors (S at 4 levels, and B and I at 3 levels) and two replicates in our design. In total there are 3*3*4*2 = 72 runs. In Taguchi approach, the design is optimized by reducing the runs from 72 to 12*2 using $L_{12}$ orthogonal array, which is similar to fractional design (it reduces the full factorial to smaller number of experiments). Table 9 shows the $L_{12}$ orthogonal array for this study. The specific objectives of the study are: (1) to estimate the contribution of individual influencing factors; (2) to select components attributes such that system performance requirements are satisfied; (3) to approximate system performance for selected components. Taguchi's technique is based on the mean and variability around the mean (S/N ratio), also shown in Table 9.

TABLE IX. L12 ORTHOGONAL ARRAY & S/N CALCULATIONS

| S | B | I | R1 | | R2 | | R3 | | Avg | S/N ratio |
|---|---|---|---|---|---|---|---|---|---|---|
| level 1 | level 1 | level 1 | 24 | 24.9 | 24 | 24.9 | 24 | 24.9 | 24.45 | 49.47 |
| level 1 | level 2 | level 2 | 56 | 57 | 56 | 57 | 56 | 57 | 56.5 | 55.83 |
| level 1 | level 3 | level 3 | 111.3 | 113 | 111.51 | 113.8 | 111.48 | 113.9 | 112.49 | 54.9 |
| level 2 | level 1 | level 2 | 26 | 26.4 | 26 | 26.5 | 25 | 25.7 | 25.93 | 49.14 |
| level 2 | level 2 | level 3 | 28.46 | 29.2 | 31.75 | 32.6 | 26.7 | 27.3 | 29.33 | 37.37 |
| level 2 | level 3 | level 1 | 26.82 | 27 | 26.57 | 26 | 22.02 | 22.7 | 25.18 | 36.64 |
| level 3 | level 1 | level 3 | 27 | 27 | 27.84 | 27.5 | 32.98 | 33.2 | 29.25 | 35.37 |
| level 3 | level 2 | level 1 | 21.88 | 22 | 21.88 | 22 | 63.73 | 63.73 | 35.87 | 19.96 |
| level 3 | level 3 | level 2 | 25 | 25 | 25 | 25 | 139.69 | 140.22 | 63.31 | 16.1 |
| level 4 | level 1 | level 1 | 24 | 23.2 | 24 | 23.7 | 23.01 | 24.5 | 23.7 | 48.18 |
| level 4 | level 2 | level 2 | 25 | 25.6 | 28.7 | 29.3 | 66.76 | 67.4 | 40.46 | 21.38 |
| level 4 | level 3 | level 3 | 28.26 | 30. 1 | 28.83 | 30.2 | 146.87 | 146.8 | 68.51 | 16.6 |

Then the response tables (Tables 10 and 11) are generated for both mean and S/N ratio. Here Δ is calculated by subtracting the lowest value from the highest for each factor. The factor having the highest Δ value is given the rank one. In Table 10, buffer size is given rank 1, and scheduling criteria is given rank 2. In Table 11, factor S is given rank 1 and factor B is given rank 2. Both tables agree that factors S and B are the most important factors for the design. Factor I is the least important both from the response table of mean and S/N ratio.

TABLE X. RESPONSE TABLE FOR MEAN

| Level | FACTORS | | |
|---|---|---|---|
| | S | B | I |
| 1 | 64.483 | 25.843 | 36.413 |
| 2 | 26.818 | 40.541 | 46.553 |
| 3 | 42.814 | 67.378 | 59.899 |
| 4 | 44.235 | N/A | N/A |
| Δ | 37.665 | 41.535 | 23.486 |
| **Rank** | **2** | **1** | **3** |

TABLE XI. RESPONSE TABLE FOR S/N RATIO

| Level | FACTORS | | |
|---|---|---|---|
| | S | B | I |
| 1 | 53.404 | 45.545 | 38.568 |
| 2 | 41.054 | 33.639 | 35.617 |
| 3 | 23.816 | 31.063 | 36.064 |
| 4 | 28.725 | N/A | N/A |
| Δ | 29.588 | 14.482 | 2.951 |
| **Rank** | **1** | **2** | **3** |

Taguchi's technique is a 2-step optimization process and the two steps are indicated in the Table 12. In step two, the level at which a factor is most effective is identified. In terms of S/N ratio, it is the highest value of the selected factor in the response table of S/N ratio which is considered. In terms of the mean it is the lowest value of the selected factors that needs to be considered. Therefore, the optimal components for the model will be S1, S2, B1, I1,2,3.

TABLE XII. TAGUCHI 2–STEP OPTIMIZATION

| STEP 1 | | | STEP 2 | | | |
|---|---|---|---|---|---|---|
| **Factors** | **Affect S/N** | **Affect mean** | **Affect both S/N & mean** | **Affect S/N** | **Affect mean** | **Affect neither** |
| **S** | * | * | S1 & S2 | | | |
| **B** | * | * | B1 in both | | | |
| **I** | | | | | | I1,2,3 |

The above experiment indicates that factors S and B are the most important variables since they affect both the variability and the mean. Also, it can been observed from the optimization process that factor S should be at level 1 or 2 which is FCFS or RR and factor B should be at the lowest level, i.e. 2. Factor I, affects neither the mean nor variability at all the 3 levels.

## V. CONCLUSION

In this paper we used regression analysis to estimate the relationships between system performance and component attributes, to help an architect in the component selection process. Specifically, for NOC, our application domain, system performance is measured in terms of latency for high, mid and low priority packets. Our results show that quadratic model was the best regression model for all these three response variables. The regression equations are:

$$R1 = 49.639 - 40.96*S + 10.529*B + 8.30*I - 3.03*S*B + 0.06*B*I - 0.438*S*I + 9.28*S^2 - 0.03*B^2 - 1.63*I^2$$
for response 1

$$R2 = 49.06 - 41.434*S + 10.58*B + 10.43*I - 2.97*S*B + 0.004*B*I - 1.24*S*I + 9.49*S^2 - 0.03*B^2 - 0.86*I^2$$
for response 2

$$R3 = 68.82 - 54.23*S + 2.29*B - 1.77*I + 2.65*S*B + 0.11*B*I + 1.49*S*I + 9.63*S^2 + 0.04*B^2 + 5.66*I^2$$
for response 3

Taguchi's OA analysis was then used to identify optimal components for best system performance, with the least number of experiments. From the 2-step optimization, the factors S and B are identified as the most important variables since they affect both the variability and mean. The optimal components are S1, S2, B1, I1, I2, I3, indicating that scheduling criteria chosen should be FCFB or RR and buffer size should be at level 1 (i.e., 2) for the optimum results (best system performance). Our analysis enhances system design productivity by reducing system development cost and time.

## REFERENCES


[1] Desoli, G., Filippi E.: An outlook on the evolution of mobile terminals: from monolithic to modular multi-radio, multi-application platforms. IEEE Magazine on Circuits and Systems, vol. 6, issue 2, pp. 17-29. (2006)

[2] Rhines, W.C.: Sociology of design and EDA. IEEE Transaction of Design and Test, vol. 23, issue 4, pp. 304-310. (2006)

[3] Lee, A.E., Xiong Y.: System level types for component-based design. Workshop on Embedded Software, California. (2001)

[4] Bennett, T.L., Wennberg, P.W.: Eliminating Embedded Software Defects Prior to Integration Test. CrossTalk the Journal of Defense Software Engineering (2005)

[5] Agarwal A., Pandya A.S., Lho, Y.U.: NOC Architecture Design Methodology. Int'l Journal KMIC, vol.1, pp. 57-64. (2005)

[6] Shankar, R., Kalva, H., Agarwal, A., Jain, A.: Annotation methods and Application Abstraction. IEEE Int'l Conf. on Portable Information Device, Orlando, FL, pp. 1-5. (2007)

[7] Sangiovanni-Vincentelli, A.: Quo Vadis SLD: Reasoning about Trends and Challenges of System-Level Design. Proceedings of the IEEE, vol. 95, no. 3 pp. 467-506 (2007)

[8] Agarwal A., Hamza-Lup, G., Shankar, R.: An Integrated Methodology for QoS Driven Component Design and Component Selection. 1st Annual IEEE Systems Conference, Hawaii, pp. 1-7. (2007)

[9] Hamza-Lup, G., Agarwal, A., Shankar, R., Iskander, C.: Component Selection Strategies Based on System Requirements & Dependencies on Component Attributes. 2nd Annual IEEE Systems Conf., Montreal, Canada. (2008)

[10] Bartholet, R.G., Brogan, D.C., Reynolds Jr., P.F.: The Computational Complexity of Component Selection in Simulation Reuse. Winter Simulation Conf., pp. 2472-2481. (2005)

[11] Fox, M.R., Brogan, D.C., Reynolds, P.F.: Approximating Component Selection. Winter Simulation Conf., vol.1 , pp. 429-434. (2004)

[12] Haghpana, N., Moaven, S., Habibi, J., Kargar, M., Yeganeh, S.H.: Approximation Algorithms for Software Component Selection Problem. Asia-Pacific Software Engineering Conf, pp. 159-166 (2007)

[13] Potena, P.: Composition and Tradeoff of Non-Functional Attributes in Software Systems: Research Directions. 6th Joint Meeting on European Software Engineering Conf. & the ACM SIGSOFT Symp. on the Foundations of Software Engineering, pp. 583-586. (2007)

[14] Benini, L., Bertozzi, D.: Network on chip architecture and design methods. IEEE proceeding Computation Digital Technology, vol. 152, no. 2, 261-271. (2005)