

Wire-length Prediction using Statistical Techniques

Jennifer L. Wong[‡], Azadeh Davoodi[†], Vishal Khandelwal[†],
Ankur Srivastava[†], Miodrag Potkonjak[‡]

[‡]University of California, Los Angeles, CA 90095, {jwong,miodrag}@cs.ucla.edu

[†]University of Maryland, College Park, MD 20742, {azade,vishalk,ankurs}@glue.umd.edu

Abstract

We address the classic wire-length estimation problem and propose a new statistical wire-length estimation approach that captures the probability distribution function of net lengths after placement and before routing. The wire-length prediction model was developed using a combination of parametric and non-parametric statistical techniques. The model predicts not only the length of the net using input parameters extracted from the floorplan of a design, but also probability distributions that a net with given characteristics obtained after placement will have a particular length. The model is validated using both learn-and-test and resubstitution techniques. The model can be used for a variety of purposes, including the generation of a large number of statistically sound and therefore realistic instances of designs. We applied the net models to the probabilistic buffer insertion problem and obtained substantial improvement in net delay after routing.

INTRODUCTION

Wire-length has become one of the most critical metrics in physical design primarily due to the rise of the deep sub-micron era. There is a large number of different parameters and constraints, such as the bounding box of the net, number of routing grids and the grid capacity, total number of nets routed in the vicinity of the pertinent net, that are all potentially relevant, but are typically very hard to capture into consistent wire-length model. Hence, estimating an exact value for the wire-length is a very hard problem.

We have developed a new wirelength model that uses data that can be extracted once the placement of the designs is completed. In order to build the model we used a combination of parametric and non-parametric techniques [2, 3]. Statistical models and prediction methodology can be used in many ways. For example, one can use the prediction information to evaluate the suitability of a particular floorplan for obtaining final routing where nets satisfy a particular user specified condition. For instance, the goal can be to determine which among a number of competing floorplans is most likely to result in a final design with few long nets or overall small sum of wirelengths. They are also a natural component of the overall probabilistic design automation methodology. One such probabilistic algorithm is [5] which performs buffer insertion assuming wire-lengths which are estimated as distributions. We used our models in the probabilistic buffer insertion approach of [5] and obtained massive improvements in net delay ($\sim 40\%$) after routing when compared with traditional bounding box strategies [1].

STATISTICAL MODELLING

We present a statistical approach for predicting the length of a given net on a given chip that is characterized using a set of features that can be rapidly obtained after floorplanning. Our primary objective is to predict the length of each net given a set of features that can be rapidly extracted from the floorplan of a chip. The goal is not only to predict the length, but also to quantify the probability that the net will have a particular length after routing. Furthermore, the operational constraint is to use only features that can be extracted with low computational effort. The final major objective is to statistically validate all obtained results and to establish intervals of confidence on all deduced models and their parameters.

The starting point for model development was the definition of relevant features of nets that are available after placement. We used two types of features: atomic and composite. Atomic features are ones that are directly extracted from the design. Composite features were created by combining atomic features using simple rules. Most often the composite rules were ratios of two atomic features. We used a state of the art commercial placement and routing tool (Cadence) to collect data that is used to build our statistical models. We use the post placement information as input parameters for building the model for each net. The objective is to identify metrics that influence the post routing wire-length for each net. The basic intuition lies in the fact that the net length is inversely proportional to the amount of routing area available and directly proportional to the routing hardness. Furthermore, a net is hard to route if its available routing area is being claimed by other neighboring nets. We have considered the following post placement properties of the nets. A detailed discussion of all properties can be found in [6].

- Number of Net Terminals.
- Bounding Box (BBOX) for net i .
- Minimal Spanning Tree (MST).
- Convex Hull (CHULL) of net i 's terminals.
- Unique terminals in the bounding rectangle of Net_i .
- Space Utilization Factor (SUF) for the net i .
- White Space of Net_i .
- Resource Competition Metric (RCM) for Net_i .
- Total number of overlapping neighbors.
- RCM for overlapping neighbors of Net_i .
- Sum of RCMs for neighbors of overlapping neighbors.
- Total overlapping area of the net for all neighbors.
- Num. common terminals of neighboring nets to Net_i .
- Neighbor utilization factor (NUF).
- Neighbor hardness factor (NHF).

Figure 1 summarizes the flow of the developed statistical modeling technique for prediction of the wire-lengths of the

1. Feature Definition;
2. Feature Extraction;
3. Preliminary Data Exploration;
4. Features Evaluation and Normalization and Compound Feature Selection;
5. Net_Characterization {
6. Nets Categorization;
7. Preliminary Linear Regression on percentiles;
8. Outliers Detection;
9. Outliers Modelling;
10. Final Linear Regression on percentiles; }
11. CDF and PDF model generation;
12. Chip characterization;
13. Development of Mapping Function to New Designs;
14. Evaluation and Validation;

Figure 1. Modeling Approach Overall Flow.

nets. The first step is the identification of relevant net properties. Two types of net properties are employed. The first group consists of properties related to the net itself. The second group consist of metrics that aim at predicting encountered congestion during routing of a given net due the routing requirement of neighboring nets. The second step was data collection. All designs were routed using the Cadence placement and routing tool. Once the data was available, we started with a randomly selected design and built a number of models. We used only 60% of all nets for this propose and preserved the rest of data for conducting statistical test and validation procedures. It was immediately apparent that each of the following three features (bounding box, minimum spanning tree, and convex hull) predict the length of a majority of the nets remarkably well. Each of them had R^2 value above 0.8 individually. The R^2 value is square of residuals, i.e. difference between the predicted variable and its predicted value using an individual property. The statistical t-test indicates that the probability that this correlation is accidental is less than 10^{-16} . While independently each of the measures (bounding box, minimum spanning tree, and convex hull) are strong predictors, their combination results in only marginally better prediction. Therefore, we decided to use bounding box as the basis of our model because of its low computational cost.

Closer examination of the data, indicated that short nets (ones with a bounding box value less than 6,000) and long nets (bounding boxes larger than 6,000) had very different properties. Most importantly, the first group, short nets, had significantly better statistical fits and essentially no outliers. Therefore, we decided to treat these two sets of nets separately. Statistical t-test indicates that correlation is significantly higher for the separated sets than for the overall set. Once the data was divided into two sets, we conducted a linear regression-based procedure for fitting data for different percentiles. For each percentile (in range 10% to 90%) a separate fit is obtained and validated using the t-test. After that, to further enhance the accuracy of our model, we conduct an outliers detection procedure that identified a small subset of data that required specialized models. For this purpose we have developed a CART model [2]. The next step was to repeat linear regressions on the data after outliers were

Bench	# layers	# nets	Area	Net Area	Total Term
ibm01	8	11507	5.89E+09	1.95E-06	44266
ibm02	10	18429	7.65E+09	2.41E-06	78171
ibm07	10	44394	1.63E+10	2.73E-06	164369
ibm08	10	47944	1.76E+10	2.73E-06	198180
ibm10	10	64227	2.97E+10	2.16E-06	269000
ibm11	10	67016	2.31E+10	2.90E-06	231819
ibm12	10	67739	3.44E+10	1.97E-06	284398

Table 1. Chip level characteristics for ibm designs obtained using Cadence routing and placement tool.

Bench	BBOX	MST	CHULL
ibm01a	9.01E-07	1.09E-06	1.12E-06
ibm02a	6.44E-07	1.39E-06	2.03E-06
ibm07a	5.23E-07	6.06E-07	6.38E-07
ibm08a	4.78E-07	6.05E-07	5.98E-07
ibm10a	3.33E-07	3.91E-07	4.13E-07
ibm11a	3.35E-07	3.79E-07	3.98E-07
ibm12a	4.06E-07	4.74E-07	5.06E-07

Table 2. Congestion metrics for ibm designs.

removed. The next two steps were dedicated to the development of a PDF and CDF for wire-length prediction and to interchip prediction. The goal of interchip prediction is to use global parameters of the chip in order to predict how features, such as global congestion and the number of nets and terminals, impact the PDF for wire-length distribution.

Finally, we conducted extensive model evaluation using learn-and-test and the resubstitution procedure in order to verify that the developed model is sound and no overfitting was done. In the rest of this Section, we elaborate on several key steps of the procedure. The final statistical model used the following features: number of terminals, RCM of the net, RCM of overlapping neighbors, total number of overlapping neighbors, and the number of common terminals for a given net. A complete description of all steps can be found in [6].

The PDF is built using the following procedure. First a subset of k nets are randomly selected for short nets. In our experimentation, we used value $k = 50\%$. The data is separated in bins that are dictated by BBOX values. The size of bin was determined in such a way that all bins contain the same number of points. The total number of bins was 10. The randomly selected subset of data is used to establish new percentile points for each bin containing data. All percentile points are normalized against the bounding box with shortest nets. The normalization is done in such a way that the average discrepancy between the values that correspond to the identical percentile is minimized. The data is fit using polynomials of low degree (three and four in our experimentation). The procedure is repeated a large number of times, the average value for each of percentile is calculated and fit using a least linear squares approach. This process was terminated once the percentile validation method indicated that we achieved user specified intervals of confidence for the PDF model. The same procedure is repeated for long nets.

For interdesign modeling, we considered the following atomic chip properties: (i) the area of the chip; (ii) the number of nets; (iii) the average and median of bounding box areas,

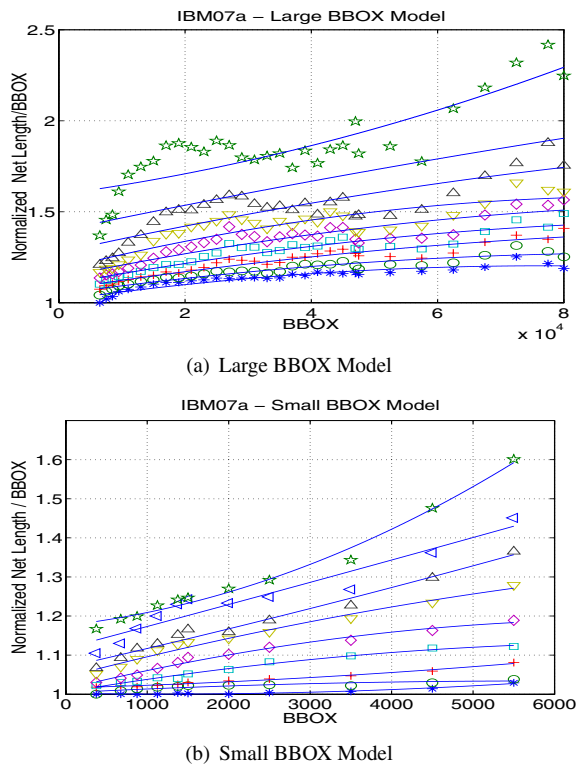


Figure 2. Linear Regression Model for ibm07 design using Cadence Router.

MST, and convex hulls for all nets (iv) the average number of terminals per net; (v) the percentage of the number of nets with a small number of terminals (two, three, or four). The composite chip metrics included ratios of all atomic chip properties and their simple statistical measures such as moments of low orders. Table 1 shows the chip level characteristic of the designs. The first column denotes the name of the benchmark, followed by the number of chip layers and the number of nets in the benchmark. The fourth column denotes the total area of the chip. The overall congestion of the design is denoted in the fifth column by the total number of nets over the area of the design. The final column specifies the total number of terminals in the benchmark. Table 2 denotes the normalized average size of the bounding box, MST and CHULL for each net for each design. The statistics are normalized against the area of the chip. We denote by c_i and c_j the overall congestion of designs i and j measured by the normalized sum of convex hull area for each design divided by the total area of the design. Furthermore, we denote by NL_i and NL_j the number of layers used in designs i and j . Our model indicates that the length of the net in design i (L_i) can be calculated using the length of the net with the same BBOX in design j (L_j) using the following formula $L_i = L_j \frac{NL_j}{NL_i} * (\frac{C_i}{C_j})^{0.48}$. This model is built using least squares data fitting approach [7]. We build this model using a randomly selected subset of four designs.

The model was validated using learn-and-test resubstitution

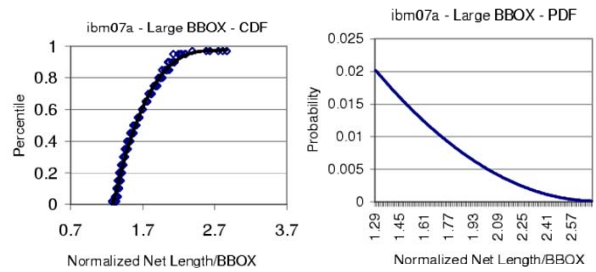


Figure 3. CDF and PDF for Large BBOX nets in ibm07 design for Cadence Router.

procedures [3, 6]. For example, we have applied the learn-and-test validation technique to both trend modeling and outlier identification. In both cases, for single chip models, we obtained predictions with 3% of accuracy for more than 96% of instances.

We now present the obtained statistical wire length model. The prediction abilities of the model are illustrated in Figures 2(a). The demonstration example used for the development of the model is imb07. It is important to emphasize that the model was actually developed using only 60% of randomly selected nets. Figures 2(a) and 2(b) show the normalized net length with respect to BBOX for different sizes of BBOX. The continuous lines in these two figures indicate the prediction models for small and large BBOX respectively. The bottom line correspond to 10% percentile and the top line to 90% percentile value. All other lines indicate the value of expected length for percentiles that differ by 10% increments. Table 3 presents the parameters of the models and the obtained R^2 values. Note that the square of residuals is consistently high. The t-test indicates that for both sets the probability of accidental coincidence is less than 10^{-18} . Therefore, it is clear that the model is both theoretically and practically sound.

As can be seen from the table, the variability of the net lengths is well captured as indicated by the high value of the R^2 coefficient, in particularly for the small BBOX model. There are two main reasons why it is much easier to accurately predict short nets. The first one is that there are significantly more short nets than long nets and, therefore, the statistical model can be developed using a much larger number of samples. The second reason is that short nets usually have significantly fewer terminals, simple structure, and can leverage on relatively small areas of white space in their vicinity.

Figure 3 shows a cumulative distribution function (CDF) and a probability distribution function (PDF) for short nets. The x-axis indicates the normalized discrepancy against the most likely values. Again, the continuous line indicates the prediction provided by the model and each plot point corresponds to the length of the nets in a particular bounding box bin selected by the resubstitution procedure. From the PDF figure we can conclude that the majority of nets are routed using a wire-length that is close to theoretical minimum and that longer nets are statistically rare. A similar figure was obtained for long nets [6]. We evaluated the accuracy and

ibm07a - Small BBOX Linear Regression Models				
Percentile	a	b	c	R^2
90	9E-09	2E-05	1.1758	0.9876
80	9E-10	5E-05	1.0686	0.9762
70	2E-10	6E-05	1.1175	0.9184
60	-2E-09	5E-05	1.0439	0.9804
50	-4E-09	5E-05	1.0131	0.9849
40	-2E-09	3E-05	1.0055	0.9876
30	1E-09	6E-06	1.0160	0.9854
20	-9E-10	1E-05	1.0038	0.8049
10	1E-09	-3E-06	1.0023	0.9702

ibm07a - Large BBOX Linear Regression Models				
Percentile	a	b	c	R^2
90	5E-11	5E-06	1.5944	0.7185
80	-1E-11	7E-06	1.3948	0.6268
70	-2E-11	8E-06	1.2767	0.6890
60	-5E-11	9E-06	1.1828	0.7460
50	-3E-11	7E-06	1.1383	0.8111
40	-3E-11	6E-06	1.0981	0.8655
30	-2E-11	5E-06	1.0720	0.9109
20	-3E-11	5E-06	1.0476	0.9033
10	-3E-11	5E-06	1.0135	0.8862

Table 3. Linear Regression Fit Parameters and R^2 for Small & Large BBOX of ibm07 design. Coefficients a , b , and c are used for the quadratic model ($ax^2 + bx + c$).

consistency of the PDF and CDF using the resubstitution procedure. We generated 100 different subsets that contain 60% of initial data and build a PDF and CDF of the wire length model. For a hundred randomly selected points their PDF and CDF values were recorded for each of the resubstitution models. The non-parametric interval of confidence was calculated for each point and for the overall probability and cumulative distribution functions. The analysis indicates that with a probability larger than 96% the model is accurate within $\pm 7\%$.

PROBABILISTIC BUFFER INSERTION

We illustrate the effectiveness of the wire-length prediction model using the buffer insertion problem. The problem can be formally stated in the following way. Given the fan-out wiring tree with parasitic resistances and capacitances, wire-lengths, potential buffer locations, sink required times, sink capacitive loads and a delay constraint at the driving gate, the problem is to place buffers into the tree such that the required arrival time at the input of the driving gate is maximum. We also consider the optimization of the number of buffers used to satisfy the delay constraint.

The buffer insertion problem was formalized by [8] and models the fan-out wiring tree as a set of distributed RC sections. The Elmore Delay model [4] is used to compute the delay of such a wiring tree. In order to estimate the parasitics for each wire-segments we need to determine the exact wire-lengths. Now let us suppose that this optimization is being performed during the in-place mode during which the exact wire-length is not available. The only available information is about the bounding box of the nets. Using the placement information we can generate the probability distributions of individual wire segments of the wiring tree and perform buffer inser-

	Prob.		BBox	
	Delay	# Buf	Delay	# Buf
Net1	8.31	14	9.59	9
Net2	6.0	20	8.74	17
Net3	6.4	22	8.51	17

Table 4. Post Routing Comparison: Prob. vs BBox based Buffer Insertion on ibm08 design.

tion probabilistically. The work in [5] proposed such a probabilistic approach to buffer insertion. For brevity, we omit the details of that algorithm. We ran probabilistic buffer insertion on a placed net (placed using Cadence Qplace) and also traditional buffer insertion [8] assuming bounding box as the net length estimate. After buffer insertion, the entire circuit was routed and the net delay was computed using real wire delay values.

The following table compares the post routing net delays from probabilistic and traditional buffer insertion. Table 4 reports the comparison. It can be seen that post routing, the probabilistic approach produces significantly better results than BBOX based approach indicating the effectiveness of our models and the superiority of a probabilistic approach.

CONCLUSION

We have built a compact statistical model that predicts the probability that a given net will have a particular wire-length. The model is characterized using a small set of parameters that are easily extracted from the design's floorplan. The runtime of the model is less than one second even for the largest designs. The model was validated using both learn-and-test and resubstitution evaluation techniques. We demonstrated the effectiveness of our model through extensive experimentation with state of the art commercial and academic tools.

REFERENCES

- [1] C. J. Alpert and A. Devgan. Wire segmenting for improved buffer insertion. In *ACM/IEEE Design Automation Conference*, pages 588–593, 1997.
- [2] L. Breiman, et al. *Classification and Regression Trees*. Chapman and Hall, 1984.
- [3] B. Efron and R. Tibshirani. *An introduction to the bootstrap*. Chapman & Hall, 1993.
- [4] W.C. Elmore. The transient analysis of damped linear networks with particular regard to wideband amplifiers. In *Journal of Applied Physics*, volume 19 of 1, 1948.
- [5] V. Khandelwal, et al. A probabilistic approach to buffer insertion. In *IEEE ICCAD*, pages 560–567, Nov 2002.
- [6] J. L. Wong, et al. Wire-length prediction using statistical and probabilistic techniques. UCLA Technical Report #040027, pages 1–8, July 2004.
- [7] William H. Press, et al. *Numerical recipes in C: the art of scientific computing*. Cambridge Univ. Press, 1992.
- [8] L.P.P. van Ginneken. Buffer placement in distributed rc-tree networks for minimal elmore delay. In *Int'l Symposium on Circuits and Systems*, pages 865–868, December 1990.