A Pre-Placement Net Length Estimation Technique for Mixed-Size Circuits

Bahareh Fathi
University of Calgary
Calgary, Canada
bfathi@ucalgary.ca

Laleh Behjat
University of Calgary
Calgary, Canada
laleh@ucalgary.ca

Logan M. Rakai
University of Calgary
Calgary, Canada
lmrakai@ucalgary.ca

ABSTRACT

An accurate model for pre-placement wire length estimation can be a useful tool during the physical design of integrated circuits. In this paper, an a priori wire length estimation technique for mixed-size circuits is proposed. The proposed technique is capable of predicting the wire lengths for individual nets, and uses both relevant factors used in previous research as well as new factors that can affect the net lengths in mixed-size designs. The proposed model's main characteristics include reporting individual net lengths, suitability for mixed-size designs, and the power to predict preplacement net lengths before and after clustering. The net lengths estimated by this model are shown to be an average of 10% more correlated to after placement lengths compared to the most elaborated model of literature. The model can be used for a priori individual net length estimation and predicting the possible effects of clustering on lengths of individual nets during the placement stage.

Categories and Subject Descriptors

B.7.2 [**Hardware**]: Integrated Circuits—*Design Aids*; G.1.2 [**Mathematics of Computing**]: Numerical Analysis—*Approximation*

General Terms

Algorithms, Design

Keywords

Wire Length Estimation, Placement, Hypergraph Clustering, Physical Design

1. INTRODUCTION

At today's nanometer technology nodes, the interconnect delay is surpassing the device delay; therefore, there is a need to predict interconnect delay before the layout phase. An ideal wire length estimation technique can predict the lengths of individual wires before placement and routing. The results from such estimation can be used not only to help ensure that a design can meet the timing constraints, but also to guide the placement process so that better placement results can be obtained.

Several a priori net length estimation techniques exist [4, 12, 16, 18, 28]. However in most of these techniques, average lengths of groups of nets [16] or wire length distributions [12, 28] are given. In [18], mutual contractions between nets are studied and the results are used in clustering the circuit. In [4], an individual net length estimation technique is given. In this technique, several characteristics of the nets and cells are used in a linear regression model to obtain pre-placement net length estimates. However, this technique does not consider mixed-size circuits encountered today.

In this paper, a comprehensive pre-placement net length estimation technique that can estimate the individual lengths of nets in mixed-size circuits is developed. The proposed technique is based on a regression model and includes a number of variables which are all available before placement. To make sure that the proposed model covers different parameters that may affect net length, many parameters are either designed or taken from previous literature and their impact on net length is studied. Variables with significant impact on wire length are selected among all the studied variables to form a pre-placement individual net length estimation model.

The accuracy of the model is validated by comparing the estimated net lengths to after placement net lengths. Then the model is applied to find a before placement estimation of net lengths. Finally, the estimated net lengths are used to predict the effects of clustering on individual net lengths, where the lengths of nets are estimated and compared before and after one level of clustering. The results of this comparison can be either used in the next levels of clustering to correct the side effects of clustering or to design more appropriate clustering techniques.

The rest of this paper is organized as follows: In Section 2, previous works on wire length estimation and clustering are reviewed. In Section 3, the proposed net length estimation technique is introduced and its accuracy is evaluated. Applications of the proposed technique in a priori net length estimation and prediction of impacts of clustering are presented in Section 4 and numerical results are given. Conclusions and future work are provided in Section 5.

2. BACKGROUND

2.1 Length Estimation Techniques

Many of the existing wire length estimation techniques rely on the relation known as Rent's rule [22], a power law relation between the number of the pins of a circuit partition and the number of its components. This relation is a consequence of a statistically *homogenous* topology and placement [9].

Donath [11] and Feuer [13] explained that the average wire length can be calculated based on Rent's constant. The results of these

SLIP'09, July 26–27, 2009, San Francisco, California, USA.

studies are theoretically important, however the fact that estimation is sensitive to Rent's constant, which fluctuates in different circuits, makes them difficult to apply in practice [14]. In [12], Rent's constant is exploited for finding the distributions of net lengths. Hamada *et al.* [15] and Pedram *et al.* [28] also introduced models for length estimations. These statistical wire length prediction methods yield average net length and/or distribution of net lengths but do not report the lengths of individual nets.

Heineken and Maly [16] proposed an approach in which nets with similar properties are grouped together and statistical models are developed to estimate net length distribution parameters for each group. Another wire length predictor that is based on predicting the wire lengths for groups of nets is explained in [25].

Bodapati and Najm [4] proposed a model for pre-layout estimation of individual wire lengths of standard cell circuits. This model combines the characteristics of each net, the condition of its neighborhood and a number of global parameters for estimation.

Wire length prediction was utilized for clustering in [18] and [24]. In [18], an estimation of net lengths called mutual contraction was suggested as a metric for scoring potential clusters. In [24], estimated lengths were predicted during the clustering stage and applied as constraints for the simulated annealing refinement stage. Although these techniques introduced innovative ways to consider wire length during clustering, they did not study or evaluate the effects of clustering on net lengths.

In [20], intrinsic shortest path length was introduced as a pre-placement estimate of net lengths and applied on ICCAD04 [1] benchmarks. Although this estimation shows a high correlation to net degree and after placement lengths of degree two nets, it does not cover all the properties of mixed-size circuits.

Besides the pre-placement approaches discussed above, a number of post-placement or on-line (during floor planning or placement) length estimation methods have been presented [5,29,33,34].

2.2 Clustering Algorithms

Clustering is used in many physical design algorithms, e.g. all leading partitioning algorithms since 1997, and almost all placement algorithms in ISPD 2005 and 2006 placement contests [26, 27], used clustering. Algorithms such as edge coarsening, hyperedge coarsening, FirstChoice [21], heavy-edge matching [3] and PinEC [7], form clusters and finalize them without any comparisons with other potential clusters. These algorithms are usually fast and can cluster circuits to very small sizes. However, since no comparison between potential clusters is made, high-quality clusters can be ignored. Best-choice clustering [2] and Net Cluster [23] are examples of score-based clustering techniques in which a score is calculated for each potential cluster. The potential clusters are compared and the clusters with the best scores are finalized. These algorithms can provide a more global view of the circuits at the cost of extra runtime. Other score-based clustering algorithms include edge-separability [10] and fine granularity clustering (FGC) [17].

One of the main drawbacks of current clustering algorithms used in placement is that during clustering there is no consideration for the actual goal of the placement. In the case of placement, using pre-placement wire length estimation during clustering can provide feedback on how clustering is impacting the wire length of nets. One of the main goals of the proposed net length estimation is to be used as a guide during the clustering and placement stages.

3. NET LENGTH ESTIMATION MODEL

Each of the mentioned wire length estimation techniques in Section 2 exhibits favorable features and characteristics. However, to be able to better understand the effects of clustering on individual nets, a mixed-sized design estimation technique is proposed that has the following features: individual net length prediction, suitability for mixed-size designs, high accuracy, coverage of various parameters that affect net length, and the ability to estimate net lengths before and after clustering during the placement stage.

The proposed technique is a polynomial model that includes significant variables which are all available before placement. This model can be expressed in generic form as:

$$EstLength(j) = \sum_{i=1}^{n}\sum_{k=1}^{n} a_{ik}x_i(j)x_k(j) + \sum_{i=1}^{n} b_i x_i(j) + c \quad (1)$$

where $EstLength(j)$ shows the estimated length of net j. Variables $x_i(j)$ and $x_k(j)$ are values of variables of the model for net j, and n is the total number of variables that are included in the model. Coefficients a_{ik}, b_i and c are the coefficients of the polynomial. Different techniques to find these coefficients are discussed in the following sections.

To ensure that the significant variables cover different factors that may affect net lengths, many of the variables stated in (1) are selected from previous models. However, these models do not include all the aspects of new mixed-size circuits; therefore, new variables are also introduced. Experiments show that these new variables prove to yield the best correlation to after placement net lengths compared to other variables.

All the variables that were studied in this paper are presented in Section 3.1. The impact of all parameters on net length is studied in Section 3.2 and the most significant variables are selected to be included in the polynomial model. Three factors are considered for variable selection: (i) strong correlation to the actual net lengths, (ii) coverage of different parameters that may affect net length to have a comprehensive model and (iii) minimization of the overlap between the effects of different variables. Finally, the proposed model is validated by comparing the correlation coefficients to the actual lengths of the nets obtained after placement in Section 3.3.

3.1 Description of Candidate Variables

3.1.1 Existing Variable Definitions

net degree: Net degree is the number of cells connected to a net. In [4, 16], it is suggested that net length is highly correlated to the net degree. However, most existing techniques have used other parameters that indirectly include effects of net degree.

base length: Base length was proposed in [4] as a base for net length. Base length of a net j is calculated as the average of two lengths: The length obtained by placing the cells of the net in a single row adjacent to each other, and the length obtained by placing them on the top of each other in a single column:

$$Lbase(j) = \frac{1}{2}(Deg(j)H_{std} + Deg(j)\frac{W_{avg}}{UtilFac}), \quad (2)$$

where, $Lbase(j)$ is the base length of net j of degree $Deg(j)$. The standard cell height and the average cell width are represented by H_{std} and W_{avg} respectively. $UtilFac$ is the fraction of row width that is used for placing cells. In addition to having a standard cell height, the cells are considered to all have the same width (average cell width) in this definition of base length. Base length is proportional to net degree and constant for all nets of the same degree.

$attint\_nc$: Introduced in [16], $attint\_nc$ of a 2-cell connection is equal to the number of nets attached to either of them excluding the nets that are connected to both of them. A 2-cell connection is a connection between two cells that are attached by a net of degree two or more. $attint\_nc$ is defined only for two cell connections and cannot be used in calculating the length of a whole net.

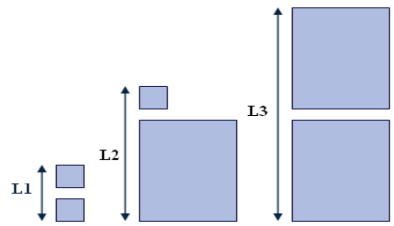

Figure 1: Effect of sizes of cells of a net on its length

degree-two to degree-six congestion metrics: Degree-two to degree-six congestion metrics, proposed in [4], estimate the effect of the nets in the neighborhood of a net on its length. These metrics are based on the number of nets of a certain degree in the neighborhood of a net and the common layouts the nets can take.

N_{2oth}: N_{2oth} [4] is an estimate of how degree-two nets in a design cause congestion in the path of routing of nets and hence increase in their lengths. Degree-two nets are selected since they comprise the largest percentage of the nets.

Mutual Contraction: Mutual contraction is introduced in [18] as a measure of contraction of a group of cells that are connected to a net. This variable is directly proportional to the weight of the connection of each cell to the other cells of the group and inversely proportional to the connection weight of each cell of the group to the cells not belonging to the group. A connection weight is calculated based on the degree of the connecting net, the higher the degree, the less the connection weight. In contrast to other model variables discussed in this section, mutual contraction is expected to be inversely proportional to net lengths. As the proposed model in this paper is a polynomial, the inverse of the mutual contraction variable (called inv. mutual contraction) is used.

3.1.2 Proposed Variables' Definitions

macro base length: Macro base length was defined to include attributes of mixed-size designs. Because of the diverse range of cell sizes in mixed-size designs, the length of a net can be considerably affected by the sizes of its cells. For example, in Figure 1, three pairs of cells are shown that are to be connected with degree 2 nets. Although the degrees of these nets are equal, their lengths would be very different. None of the existing models and variables can predict this.

Macro base length is a measure of base net length calculated using the actual widths and heights of cells instead of a standard height and an average width for all the cells, to distinguish between the nets with a same degree but different cell sizes. This parameter is equal to the average of the height needed for placing all the cells of the net vertically on the top of each other and row width required for placing them horizontally beside each other.

$$Lmacro(j) = \frac{1}{2}\left(\frac{\sum^{l\in Ne(j)} H(l)}{UtilFacH} + \frac{\sum^{l\in Ne(j)} W(l)}{UtilFacW}\right) \quad (3)$$

where $Lmacro(j)$ shows the macro base length of net j , the sum of heights and widths of its cells are shown by $\sum^{l\in Ne(j)} H(l)$ and $\sum^{l\in Ne(j)} W(l)$, respectively, and $Ne(j)$ is the set of cells adjacent to net j. $UtilFacH$ and $UtilFacW$ are the horizontal and vertical utilization factors, respectively, to guarantee sufficient allocation of wiring space.

second level effect: Based on rationale similar to macro base length, sizes of cells within the second level neighborhood of nets are expected to affect nets' lengths. The second level neighborhood of a net contains cells that can be reached from the cells of a net by traversing one connected net. Therefore this variable is proposed to cover the effect of sizes of second level neighbors of nets. Second level effect is the sum of half perimeters of such cells. So if second level effect of net j is shown by $secondLvlEF(j)$, then:

$$secondLvlEF(j) = \sum^{l\in Ne2(j)} (W(l) + H(l)) \quad (4)$$

where $Ne2(j)$ is the set of cells in the second level neighborhood of net j. Width and height of cell l, a second level neighbor of net j, are shown by $W(l)$ and $H(l)$ respectively. The numerical results presented in Section 3.2 shows that second level effect is one of the most significantly correlated variables to net length compared to other variables from literature.

$nettint\_nc$: $nettint\_nc$ is developed by improving $attint\_nc$ and is defined as an attribute of a whole net, not a two cell connection. $nettint\_nc(j)$ of a net j is equal to the number of nets connected to the cells of net j excluding the nets that are common between two or more cells of net j.

3.2 Variable Analysis and Selection

Not all of the twelve variables discussed in the previous section have significant impact on net length. In addition, the effects of some of these variables can overlap heavily with each other. In this section, the correlation of each variable to after placement lengths of individual nets is studied. Then, a set of variables are selected for the final model based on their correlations to the after placement lengths and the effects that they cover.

Two measures are used for finding correlations: correlation coefficient and c_ratio [25].

correlation coefficient : According to the mathematical definitions, $cor(est, aft)$, correlation coefficients of the data sets of estimated net lengths, est, and after placement net lengths, aft, is calculated by $cor(est, aft) = \frac{cov(est,aft)}{\sqrt{cov(est,est)cov(aft,aft)}}$, where $cov(.,.)$ shows the covariance of two data sets.

c_ratio: C_ratio [25] is another measure of correlation of a variable to the after placement net lengths. To obtain the c_ratio for a variable, nets are sorted into twenty bins based on the values of the considered variable. The average of variable values and average net length values are calculated for each bin. Then, pair combinations of the bins are considered. A pair is considered a violating pair if one of its bins has a lower numerical average value of the variable, but higher after average placement length, compared to the other bin. Once all violating pairs are identified, the c_ratio for the variable can be calculated as: $c\_ratio = 1 - \frac{number\ of\ violating\ pairs}{total\ number\ of\ pairs}$. c_ratio is a number between zero and one and higher for more correlated data sets.

The average values of the correlation coefficient, *cor*, and the c_ratios, *c_r*, over the three actual net length data sets for the circuits in the ICCAD04 benchmarks [1] are shown in Table 1. The data of this table is related to the placements performed by the placer Capo 10.3 [6]. Other placers have been tested, such as mPL6 [8] and FastPlace3.0 [32], force directed placers, and Dragon05 [30], a simulated annealing placer. Because of the lack of space the results for these placers are not reported in detail. To minimize the effect of random behavior of each placer, placement is performed three times. All after placement lengths in this paper are calculated by using Batched iterated 1 Steiner (BI1ST) [19].

In columns 2 and 3 of Table 1 the correlations and the c_ratios of the net degree, net deg., and base length, BL, variables are given respectively. In columns 4 to 8 the correlations and the c_ratios of degree-two to degree-six congestion metrics, D2-D6 Con, are

Table 1: correlation coefficients and c_ratio of different variables to after placement net lengths (The numbers are scaled by multiplication to 100). The placement is performed by Capo 10.3.

circuit	Existing Variables									Proposed Variables		
	net deg.	BL	D2 Con	D3 Con	D4 Con	D5 Con	D6 Con	N_{2oth}	Inv. mut. Cont	macro BL	2nd Lvl EF	net nc
	x_1	-	x_2	x_3	x_4	-	-	x_5	x_6	x_7	x_8	x_9
	cor / c_r	cor / c_r	cor/c_r	cor/c_r	cor/c_r	cor/c_r	cor/c_r	cor/c_r	cor/c_r	cor/c_r	cor/c_r	cor/c_r
ibm01	37 / 63	37 / 63	39 / 87	31 / 85	31 / 87	10 / 58	19 / 70	23 / 66	36 / 86	46 / 86	34 / 95	45 / 78
ibm02	48 / 69	48 / 69	48 / 87	46 / 86	50 / 81	10 / 67	41 / 81	44 / 80	35 / 88	48 / 86	47 / 88	61 / 79
ibm03	44 / 66	44 / 66	48 / 81	49 / 81	48 / 80	30 / 83	49 / 82	54 / 85	41 / 87	58 / 84	55 / 87	53 / 78
ibm04	46 / 66	46 / 66	31 / 85	31 / 86	31 / 84	25 / 77	35 / 75	36 / 79	41 / 85	34 / 85	35 / 94	26 / 78
ibm05	64 / 77	64 / 77	47 / 73	49 / 74	49 / 76	23 / 67	38 / 73	49 / 77	55 / 82	62 / 75	47 / 77	55 / 78
ibm06	51 / 72	51 / 72	67 / 77	69 / 77	69 / 80	54 / 81	61 / 78	64 / 75	59 / 84	62 / 85	68 / 74	71 / 76
ibm07	40 / 76	40 / 76	41 / 86	42 / 84	44 / 83	26 / 79	35 / 83	44 / 81	37 / 89	43 / 87	48 / 88	44 / 84
ibm08	50 / 68	50 / 68	45 / 91	45 / 92	46 / 92	38 / 86	53 / 84	52 / 82	49 / 87	47 / 89	47 / 94	35 / 82
ibm09	38 / 74	38 / 74	44 / 89	40 / 86	39 / 86	23 / 81	34 / 84	42 / 83	40 / 88	47 / 86	46 / 87	45 / 81
ibm10	18 / 75	18 / 75	33 / 84	34 / 82	33 / 83	7 / 72	8 / 67	19 / 70	25 / 86	35 / 82	34 / 83	31 / 78
ibm11	34 / 76	34 / 76	37 / 90	35 / 89	36 / 90	16 / 81	30 / 83	38 / 83	38 / 88	43 / 86	39 / 87	39 / 80
ibm12	23 / 68	23 / 68	37 / 84	37 / 85	38 / 84	8 / 68	15 / 62	32 / 70	33 / 88	33 / 86	38 / 80	29 / 79
ibm13	43 / 71	43 / 71	44 / 88	41 / 88	41 / 87	28 / 85	36 / 88	42 / 87	43 / 86	44 / 87	43 / 91	42 / 84
ibm14	43 / 77	43 / 77	27 / 91	26 / 90	27 / 89	20 / 83	31 / 86	34 / 89	33 / 93	41 / 92	34 / 93	20 / 85
ibm15	39 / 78	39 / 78	41 / 94	41 / 92	40 / 92	31 / 89	43 / 89	46 / 91	37 / 91	40 / 89	44 / 94	37 / 86
ibm16	37 / 69	37 / 69	38 / 88	39 / 88	40 / 89	18 / 78	28 / 82	40 / 78	41 / 92	42 / 85	44 / 89	38 / 82
ibm17	39 / 73	39 / 73	34 / 82	36 / 81	36 / 83	17 / 76	27 / 76	34 / 77	29 / 92	38 / 87	34 / 74	34 / 81
ibm18	58 / 76	58 / 76	43 / 89	45 / 91	45 / 91	23 / 79	44 / 86	53 / 81	43 / 88	57 / 84	53 / 91	53 / 80
average	42 / 72	42 / 72	41 / 86	41 / 85	41 / 85	23 / 77	35 / 79	41 / 80	40 / 88	46 / 86	44 / 87	42 / 81

Table 2: Rankings of the candidate variables

	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th
average corr	macro BL	2nd Lvl EF	net nc	net deg, BL	-	D2 Con	N_{2oth}	D3 Con	D4 Con	Inv. Mut. Cont	D5 Con	D6 Con
average c_ratio	Inv. Mut. Cont	2nd Lvl EF	macro BL	D2 Con	D4 Con	D3 Con	net nc	N_{2oth}	D6 Con	D5 Con	net deg, BL	-
corr score	macro BL	2nd Lvl EF	N_{2oth}	net nc	D4 Con	D3 Con	net deg, BL	-	D2 Con	Inv. Mut. Cont	D6 Con	D5 Con

given, respectively. Columns 9 and 10 show the correlation and c_ratios of N_{2oth}, and the inverse of mutual contraction, Inv. Mut. Cont. Finally, the correlations and c_ratios of the three proposed variables, macro base length, Macro BL, second level effect, 2nd Lvl EF, and nettint_nc, net nc, are presented in columns 11 to 13 of the table. Since *attin_nc* is only defined for two cell connections, it is not included in this table.

To be able to better make comparisons between different variables, they are ranked according to the data presented in Table 1. This ranking is presented in Table 2. In the first and second rows of this table, the rankings based on the average correlation and average c_ratio are presented. In the third row, a ranking based on correlation scoring is shown. For correlation scoring, for each circuit, twelve bins, related to twelve variables, are considered, bin number twelve to bin number one. For each circuit, the variable with the highest correlation falls into bin twelve and receives twelve points. The second highly correlated variable falls into the bin eleven and receives eleven points and so on. The points for each variable are counted over the 18 benchmarks. Then, the variables are ordered based on their overall scores over the 18 benchmarks. Correlation scoring takes into account the frequency of a variable having high correlations and avoids assigning a high rank to a variable only because of its extremely high correlation to one or a limited number of circuits.

According to Tables 1 and 2, macro base length and second level effect, variables designed for mixed-size circuits, show the highest average correlation to the actual lengths. Second level effect and macro base length also show the second and third highest average c_ratios. Net degree and base length variables rank in the middle for correlation coefficient and the correlation scoring and poorly in the average c_ratio. On a circuit basis, macro base length has a higher correlation coefficient for most of the tested circuits compared to the base length and net degree variables which are designed for standard cell circuits. For some circuits, net degree and base length show higher correlations compared to macro base length. This is the case for a circuit like ibm05 which contains only standard cells and ibm17 and 18 that do not contain cells with an area larger than 1% of total area of circuit. Macro base length, second level effect and net degree are selected as variables to represent attributes that directly impact the length of a net and base length is discarded because of its overlap with macro base length and net degree variables. By keeping macro base length and net degree in the model and discarding base length, not only the effects of number and sizes of cells of a net are covered but also no effect is emphasized too much because of redundancy.

Figure 2 shows a comparison between base length, macro base length and second-level effect, by plotting average after placement net lengths versus average values of these variables for ibm04. Since there are thousands of nets in each circuit, to be able to clearly compare the variables versus the actual net lengths, the bins used for finding c_ratio are considered and the average net length for each bin is plotted against the average of the variable. The graphs of macro base length and second level effect are very close to linear. However, the plot for base length consists of different areas of increase and decrease in net length with different angles.

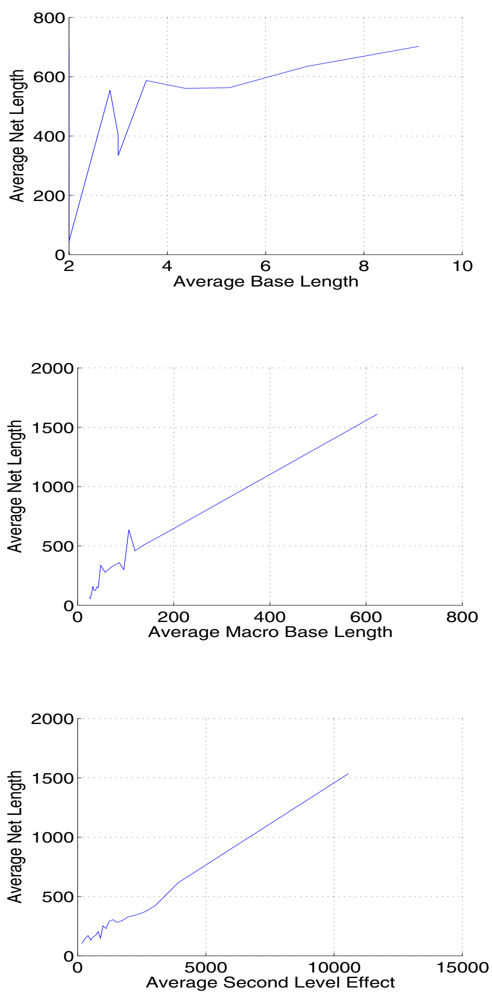

Figure 2: Plots of actual net length versus three variables for ibm04. Net lengths are found by Dragon05.

It should be also mentioned that the difference in range of average base length and average macro base length is according to the fact that for base length average cell height and width are concerned, which tend to be very close to height and width of standard cells (because of high number of such cells). However, macro base length adds up actual heights and widths of cells, which might be very fluctuating and larger compared to standard cell height and width. The difference in ranges of average net length is also due to the fact that long nets are distributed in all bins for base length, but for the other variables, they are collected in the last bins.

The inv. mutual contraction exhibits the highest c_ratio, claiming the top spot; therefore, it is included in the model. However, in average correlation coefficient and correlation scoring, this variable is placed in tenth spot. Variables degree-five and degree-six congestion metrics show the lowest correlation coefficients with a significant gap compared to the third lowest correlated variable (inv. mutual contraction). They also show poor c_ratio values and have the lowest scoring, so they are not included in the model. However, degree two to four congestion metrics show average rankings and are included in the model.

N_{2oth} and nettint_nc variables score highly in average correlation and correlation scoring and their effects are not covered by any other variables, and are included in the model.

Nine of the twelve variables discussed in Table 1 are labeled as significant. These variables and their role in modeling are shown in Table 3. Columns 1 and 2 of Table 3 show the names of the

Table 3: The significant variables of the model and the effect each covers

	Variable	Modeling Role
x_1	net degree	number of cells of a net
x_2 x_3 , x_4	degree 2 to 4 congestion metrics	effect of nets in net neighborhood
x_5	N_2oth	effect of degree 2 nets
x_6	inv. mutual contraction	connectivity between a net and its neighboring nets
x_7	macro base length	sizes of cells of a net
x_8	second level effect	sizes of 2nd level neighbors
x_9	nettint_nc	common and uncommon nets between cells of net

parameters. In Column 3, the role of each parameter is described.

To show that these variables are correlated to lengths found by other placers, Table 4 is presented. The average correlation and c_ratio values for Capo 10.3 are repeated in Column 2 for comparison's sake. The last three columns of this table show the averages of correlations and c_ratios of each parameter over ICCAD04 benchmarks, to the net lengths found by mPL6, Dragon05 and FastPlace 3.0. The first number in each column shows the correlation coefficient and the second number shows the c_ratio, all multiplied by 100. It can be seen that correlations are comparable to the ones reported for Capo 10.3 and for all of the placements, macro base length and second level effect show the best correlations.

Table 4: Correlation coefficients and c_ratios of significant variables of the model to after placement net lengths

variable	Capo10.3 cor/c_r	mPL6 cor/c_r	Dragon05 cor/c_r	FastPlace 3.0 cor/c_r
x_1	42 / 72	39 / 72	40 / 72	35 / 71
x_2, x_3, x_4	41 / 86	40 / 87	42 / 87	43 / 86
x_5	41 / 80	40 / 80	42 / 79	40 / 74
x_6	40 / 88	38 / 89	40 / 88	39 / 88
x_7	46 / 86	45 / 90	47 / 90	52 / 92
x_8	44 / 87	43 / 86	46 / 86	45 / 86
x_9	42 / 81	41 / 84	44 / 84	45 / 84

It should also be noted that the correlation coefficients stated in this paper are lower than what has been stated in [20]. This can be because in [20], all cells are assumed to have unit area and only degree-two nets are considered in the calculations. The net lengths used in that paper are half perimeter wire lengths, but for this research, a more accurate method of finding after placement lengths, Batched iterated 1 Steiner (BI1ST) [19] has been used.

3.3 Model Development

3.3.1 Polynomial Coefficient Calculations

The generic form of the proposed model is presented in (1). To validate this model, the polynomial coefficients a_{ik}, b_i and c should be found by fitting the polynomial to the after placement net lengths.

One of the most common ways of fitting a curve is by using ordinary least square fitting (OLSF), which is based on minimizing the residual squared error. Another curve fitting method, least absolute shrinkage and selection operator (lasso) [31], minimizes the residual sum of squared errors subject to the absolute values of the coefficients being less than a constant (referred to as the Lasso constant). The advantage of this method is that it improves the accuracy by setting some coefficients to zero and determining coefficients for a smaller subset of variables that exhibits the strongest effects.

To find the polynomial coefficients, both lasso and OLSF are used. Figure 3 depicts the proposed algorithm for calculating polynomial coefficients.

Input: $after\ placement\ net\ lengths$
$values\ of\ polynomial\ terms, lasso\ constants$
Output: $polynomial\ coefficients$

1. $best\ correlation$ = -1
2. $polynomial\ coefficients$ = { }
3. for i = 1:$number\ of\ lasso\ constants$
    - 3.1. C_{lasso} = $lasso\ constants(i)$
    - 3.2. Use lasso and C_{lasso} to find effective terms
    - 3.3. Use $OLSF$ to find $current\ polynomial\ coefficients$
    - 3.4. Calculate estimated length of each net
    - 3.5. Calculate $correlation$ of $after\ placement$ and $estimated\ lengths$
    - 3.6. If $correlation$ > $best\ correlation$
        - 3.6.1. $best\ correlation$ = $correlation$
        - 3.6.2. $polynomial\ coefficients$ = $current\ polynomial\ coefficients$
    - 3.7. end
4. end

Figure 3: The algorithm for determining the polynomial coefficients

The inputs of this algorithm are after placement net lengths, values of polynomial terms for each net, and a set of candidate lasso constants. The algorithm returns as output a set of polynomial coefficients, $polynomial\ coefficients$.

In the for loop, lasso with a selected lasso constant is used at each iteration to determine the effective terms. Then the estimated lengths are found by OLSF of the selected terms to the actual net lengths. The correlation coefficients of the estimated net lengths and actual net lengths are calculated. The for loop is repeated for all the different lasso constants in the array $Lasso\_constants$ (20 in this work). The polynomial coefficients corresponding to the best correlation coefficient are given as the output of the algorithm.

3.3.2 Model Validation

After calculating the best coefficients and determining a polynomial model, it needs to be validated and the accuracy of the estimation tested.

The model proposed in this paper is closest to the model proposed in [4] which is also made of a polynomial obtained by curve fitting. Furthermore, many variables in the current model are the same as the model in [4], which is the most elaborated model of literature. Therefore, average correlation coefficients in this model are compared to those obtained from the model constructed in [4]. In estimation techniques, usually 75% of the actual data is used for model building and 25% of the data for testing. The same procedure is applied in this research where 75% of the after placement net lengths are used to develop the model and 25% are used for validating. By this strategy, the validation is determined by nets whose after placement lengths have not been used for making the model. The correlation coefficients to after placement net lengths estimated by the proposed model and the model in [4], are given in Table 5. The after placement lengths are found after placing the nets by Capo 10.3. For all the circuits, the proposed estimation method proves a higher correlation to actual net lengths compared to the model of [4] and on average, correlation coefficient shows 10% improvement.

Table 5: Comparison between experimental results of proposed model and a model of literature

Circuit	Correlation coefficient		Improvement
	Proposed model	Model of [4]	
ibm01	0.59	0.49	10%
ibm02	0.70	0.55	14%
ibm03	0.69	0.58	11%
ibm04	0.61	0.54	7%
ibm05	0.76	0.62	14%
ibm06	0.83	0.63	20%
ibm07	0.57	0.48	9%
ibm08	0.72	0.63	8%
ibm09	0.59	0.52	7%
ibm10	0.45	0.41	4%
ibm11	0.56	0.44	11%
ibm12	0.51	0.40	11%
ibm13	0.59	0.46	12%
ibm14	0.54	0.43	11%
ibm15	0.55	0.47	9%
ibm16	0.56	0.48	8%
ibm17	0.46	0.40	7%
ibm18	0.64	0.58	6%
Average	0.61	0.51	10%

4. PROPOSED MODEL'S APPLICATIONS

In the above procedure, the after placement lengths of a circuit are required for making the model. Therefore, this model on its own is not an a priori estimation model. It is also the case for any model, such as [4], that uses curve fitting. Therefore, this curve fitting is presented only as a validation for the model. In this section, one method of using the model for estimating the nets of a circuit without relying on the after placement lengths is suggested and tested. The model is also applied for finding effects of clustering on net lengths.

4.1 Pre-Placement Net Length Estimation

To have an a priori net length estimation model, fixed polynomial coefficients that do not depend on the tested circuits are required. Once these coefficients are found, they can be saved in a look up table and referred to for estimating the net lengths of different circuits. Polynomial coefficients can be calculated by applying the algorithm of Figure 3 on p nets of a circuit (model making set).

However, there is no guarantee that the calculated polynomial may be successfully applied on the other circuits. Therefore, a set of p nets is sought so that model built on them can accurately predict the lengths of other nets too. An experiment is performed on the ICCAD04 circuits to find an appropriate model making set. From each circuit, 75% of the nets are selected and polynomial coefficients are calculated based on their data. Then, a cross validation task is performed by applying the polynomial of each circuit to estimate the net lengths for other circuits. Correlation coefficients and c_ratios of net lengths estimated in this way and the after placement lengths are shown in Table 6. Each row of this table shows the circuit being tested and each column shows the circuit from which the model making set is taken. It can be seen that many circuits can be estimated well by models of other circuits. However, the model made by ibm08 circuit is more successful compared to other circuits, as indicated by the higher correlation coefficients and c_ratio of lengths estimated by it.

The circuit structure of ibm08 has properties that can explain why it is a good candidate for predicting the lengths of nets for the other benchmarks. ibm08 has nets of 69 different degrees and nearly each set of nets for each degree contains multiple nets. This allows the effect of net degree to be modeled well. It also has an above average amount of cell degrees represented in its circuit. Of those cells, 14 are macro blocks, including one very large block occupying 12% of the total area. This allows its model to capture the effects of different sizes of macros which are seen in the other circuits. ibm08 is also of an average size compared to other circuits of the design. It is proposed that the polynomial coefficients of model of ibm08 can be saved as the fixed polynomial coefficients in a look up table.

4.2 Clustering Effects on Net Lengths

Another application of the a priori individual net lengths can be to evaluate the performance of a clustering technique on individual net lengths. Clustering techniques are shown to be effective in helping to reduce the total half parameter wire length after placement. However, since clustering is performed before placement and accurate individual net length models do not exist, the effect of clustering on individual net lengths is not known.

The proposed net length estimation model is used to evaluate the effect of clustering on net lengths, using best-choice clustering [2]. Best-choice is chosen since it is the state-of-the-art clustering technique, has been successfully used in many placement algorithms and can cluster a circuit to low clustering ratios. Pre-clustering and post-clustering wire lengths when a circuit is clustered to 50% of original size are estimated and compared. A summary of the change in net lengths is presented in Table 7. In this table, the percentage of nets of each circuit with increased, constant and decreased estimated net lengths are shown under inc. , con. and dec. respectively.

Table 7: Percentage of nets with increased, decreased and constant lengths after clustering

circuit	inc.	con.	dec.	circuit	inc.	con.	dec.
ibm01	24	0	76	ibm10	34	1	66
ibm02	55	0	45	ibm11	17	3	80
ibm03	25	1	74	ibm12	42	0	58
ibm04	32	0	68	ibm13	17	1	82
ibm05	45	1	54	ibm14	50	0	50
ibm06	37	0	62	ibm15	34	1	65
ibm07	35	1	64	ibm16	36	0	64
ibm08	46	0	54	ibm17	50	0	49
ibm09	29	1	70	ibm18	40	0	60
				average	36	1	63

As it can be seen from this table that a large percentage of the nets have seen estimated decrease in the length, however, still a substantial percentage of the nets have seen net length increase. Currently, no clustering algorithm considers the nets with increased length in a feedback loop to reduce their sizes. It is proposed to implement a feedback mechanism in a clustering algorithm to try to reduce the percentage of nets whose lengths increase to obtain better clustering results for placement in the future.

5. CONCLUSION

In this paper, a framework for a priori individual net length estimation is given. This model includes variables that attribute to the mixed-size circuits encountered today. Using cross validation, a generic model was formed that is able to predict individual net lengths for different circuits. Future work can include making different generic models for nets with different degrees, and using other models, such as an exponential model as is used in [20].

In addition, the net length estimation is used to study the effects of clustering on individual net lengths. It is shown that after clustering, using best-choice, the lengths of significant percentages of the nets are increased. Future work in this area is to develop a clustering model with a feedback loop that can correct the net length increase.

6. ACKNOWLEDGMENTS

This work was supported by Natural Science and Engineering Research Council of Canada grant number 262463-03.

7. REFERENCES

[1] S. Adya, S. Chaturvedi, J. Roy, D. Papa, and I. Markov. Unification of partitioning, floorplanning and placement. In *Proc. of ICCAD*, pages 550–557, 2004.

[2] C. Alpert, A.Kahng, G. J. Nam, S. Reda, and P. Villarrubia. A semi-persistent clustering technique for VLSI circuit placement. In *Proc. of ISPD*, pages 200–207, 2005.

[3] C. Alpert, D. Huang, and A. Kahng. Multilevel circuit partitioning. *IEEE Trans. on CAD*, 17(8):655–667, 1998.

[4] S. Bodapati and F. Najm. Prelayout estimation of individual wire lengths. *IEEE Trans. on VLSI*, 9(6):943 – 58, 2001.

[5] A. E. Caldwell, A. B. Kahng, S. Mantik, I. L. Markov, and A. Zelikovsky. On wirelength estimations for row-based placement. *IEEE Trans. on CAD*, 18(9):1265 – 1278, 1999.

[6] A. E. Caldwell, A. B. Kahng, and I. L. Markov. Can recursive bisection produce routable placements? In *Proc. of ACM/IEEE DAC*, pages 477–482, 2000.

[7] A. E. Caldwell, A. B. Kahng, and I. L. Markov. Improved algorithms for hypergraph bi-partitioning. In *Proc. of ASPDAC*, pages 661–666, 2000.

[8] T. F. Chan, J. Cong, M. Romesis, J. R. Shinnerl, K. Sze, and M. Xie. mPL6: Enhanced multilevel mixed-size placement. In *Proc. of ISPD*, pages 212–214, 2006.

[9] P. Christie and D. Stroobandt. The interpretation and application of Rent's rule. *IEEE Trans. On VLSI*, 8(6):639–648, 2000.

[10] J. Cong and S. Lim. Edge separability-based circuit clustering with application to multilevel circuit partitioning. *IEEE Trans. on CAD*, 23(3):346–357, 2004.

[11] W. Donath. Placement and average interconnection lengths of computer logic. *IEEE Trans. CAS*, 26(4):272 – 7, 1979.

Table 6: Cross Validation of models generated from a single circuit using Capo 10.3 and BI1st actual lengths. Values pairs are the cross correlation / c_ratio multiplied by 100.

		The ibm benchmark used to determine the polynomial																		
		01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	Avg
The ibm benchmark being estimated	01	59 /97	48 /95	44 /89	31 /84	12 /56	37 /82	35 /88	48 /92	45 /89	44 /96	34 /84	46 /87	41 /91	45 /92	40 /87	42 /90	43 /92	47 /90	41 /88
	02	-7 /70	71 /92	31 /83	-10 /67	30 /69	-17 /71	13 /71	50 /85	21 /86	44 /86	24 /91	-8 /72	24 /78	0 /69	5 /88	-18 /74	-18 /77	1 /87	13 /79
	03	-9 /79	57 /92	69 /84	8 /86	21 /44	48 /84	38 /87	63 /91	34 /80	39 /72	44 /91	31 /88	46 /78	51 /87	46 /88	30 /88	-11 /87	52 /89	37 /83
	04	18 /89	25 /94	20 /87	61 /94	-14 /50	23 /88	15 /84	47 /95	18 /88	-9 /72	20 /88	17 /88	16 /89	6 /86	24 /89	7 /91	-16 /83	-16 /83	15 /85
	05	45 /79	54 /84	55 /76	51 /67	76 /90	49 /70	51 /76	62 /66	54 /78	52 /78	50 /76	56 /81	56 /80	64 /76	50 /66	63 /78	52 /84	66 /83	56 /77
	06	-12 /83	68 /95	27 /87	55 /88	17 /65	83 /92	-21 /87	72 /92	22 /86	64 /89	21 /85	9 /86	45 /83	7 /88	63 /88	68 /94	-19 /83	22 /91	33 /87
	07	24 /91	51 /95	49 /89	47 /89	-12 /58	35 /82	57 /95	50 /94	53 /87	44 /84	51 /85	50 /94	50 /88	44 /93	51 /89	51 /95	44 /93	48 /94	44 /89
	08	31 /83	-27 /87	28 /83	-27 /72	-17 /69	30 /79	-22 /77	72 /96	-18 /84	-24 /74	-10 /87	26 /80	-23 /79	-24 /77	38 /89	-18 /77	-26 /77	-27 /83	-2 /81
	09	3 /82	53 /97	50 /84	36 /88	-25 /49	18 /85	33 /89	54 /98	59 /89	49 /93	52 /85	52 /92	54 /83	10 /85	51 /94	53 /92	24 /89	40 /89	37 /87
	10	1 /84	25 /89	12 /85	-1 /76	-18 /60	17 /82	-21 /83	38 /88	1 /81	45 /95	-3 /84	-6 /87	37 /88	-21 /85	34 /85	38 /91	-19 /84	-16 /83	8 /84
	11	8 /83	47 /96	48 /89	43 /94	-24 /52	21 /86	33 /90	50 /97	53 /89	40 /86	56 /89	49 /94	51 /89	14 /87	49 /96	50 /96	22 /87	29 /87	35 /88
	12	-12 /83	15 /90	27 /94	-22 /78	-24 /52	27 /87	25 /93	43 /92	31 /94	13 /88	30 /95	51 /96	3 /88	-21 /79	36 /92	42 /96	-20 /87	-22 /86	12 /87
	13	20 /88	47 /92	45 /85	20 /85	-16 /45	36 /86	7 /86	51 /95	45 /93	44 /84	49 /92	45 /87	58 /92	21 /89	49 /94	53 /91	23 /84	8 /87	34 /86
	14	5 /89	30 /91	21 /92	31 /90	-5 /63	9 /91	18 /91	47 /93	24 /89	12 /77	18 /88	28 /95	24 /85	52 /96	43 /92	41 /95	4 /90	4 /92	23 /89
	15	17 /92	32 /91	20 /85	28 /85	-14 /57	17 /88	6 /87	49 /96	25 /91	38 /82	35 /89	24 /88	36 /88	7 /89	56 /95	49 /94	3 /90	-6 /86	23 /87
	16	11 /93	22 /92	25 /93	-9 /85	-13 /74	21 /87	-9 /91	51 /94	42 /86	17 /88	34 /91	35 /95	35 /89	-15 /89	33 /88	55 /96	-4 /91	-15 /91	18 /90
	17	22 /94	39 /96	39 /94	34 /85	1 /61	29 /93	33 /95	42 /91	40 /91	41 /87	41 /90	42 /94	40 /92	43 /95	42 /85	42 /95	46 /97	40 /96	36 /91
	18	20 /82	50 /83	44 /83	-1 /80	19 /83	25 /81	31 /87	60 /88	34 /87	46 /83	22 /83	31 /85	39 /81	46 /88	37 /77	41 /89	32 /89	64 /93	36 /84
	Avg	13 /85	39 /92	36 /87	21 /83	0 /61	28 /84	18 /87	53 /91	32 /87	33 /84	32 /87	32 /88	35 /86	18 /86	42 /88	38 /90	9 /87	18 /88	

[12] W. Donath. Wire length distribution for placements of computer logic. *IBM Journal of RD*, 25(3):152 – 5, 1981.

[13] M. Feuer. Connectivity of random logic. *IEEE Transactions on Computers*, C-31(1), 1982.

[14] T. Hamada, C.-K. Cheng, and P. Chau. A wire length estimation technique utilizing neighborhood density equations. In *Proc. of ACM/IEEE DAC*, pages 57 – 61, 1992.

[15] T. Hamada, C.-K. Cheng, and P. Chau. A wire length estimation technique utilizing neighborhood density equations. *IEEE Trans. on CAD*, 15(8):912 – 22, 1996.

[16] H. Heineken and W. Maly. Standard cell interconnect length prediction from structural circuit attributes. In *Proc. of the IEEE CICC*, pages 167 – 70, 1996.

[17] B. Hu and M. Marek-Sadowska. Fine granularity clustering for large scale placement problems. In *Proc. of ISPD*, pages 67–74, 2003.

[18] B. Hu and M. Marek-Sadowska. Wire length prediction based clustering and its application in placement. In *Proc. of DAC*, pages 800 – 5, 2003.

[19] A. B. Kahng, A. Member, G. Robins, and S. Member. A new class of iterative steiner tree heuristics with good performance. *IEEE Trans. on Computer-Aided Design*, 11:893–902, 1992.

[20] A. B. Kahng and S. Reda. Intrinsic shortest path length: a new, accurate a priori wirelength estimator. In *Proc. of ACM/IEEE DAC , Digest of Technical Papers, ICCAD*, volume 2005, pages 173 – 80, 2005.

[21] G. Karypis, R. Aggarwal, V. Kumar, and S. Shekhar. Multilevel hypergraph partitioning: Application in VLSI domain. *IEEE Trans. on VLSI*, 7(1):69–79, 1999.

[22] B. Landman and R. Russo. On a pin versus block relationship for partitions of logic graphs. *IEEE Trans.on Computers*, C20(12):1469 – 79, 1971.

[23] J. Li and L. Behjat. Net cluster: a net-reduction based clustering preprocessing algorithm. In *Proc. of ISPD*, volume 26, pages 200 – 5, 2006.

[24] Q. Liu, B. Hu, and M. Marek-Sadowska. Wire length prediction in constraint driven placement. In *Proc. of SLIP*, pages 99 – 105, 2003.

[25] Q. Liu and M. Marek-Sadowska. Pre-layout wire length and congestion estimation. In *Proc. of DAC*, pages 582 – 587, 2004.

[26] G. Nam. ISPD 2006 placement contest: Benchmark suite and results. In *Proc. of ISPD*, page 167, 2006.

[27] G. Nam, C. J. Alpert, P. Villarrubia, B. Winter, and M. Yildiz. The ISPD2005 placement contest and benchmark suite. In *Proc. of ISPD*, pages 216–219, 2005.

[28] M. Pedram and B. Preas. Interconnection length estimation for optimized standard cell layouts. In *Proc. of ICCAD*, pages 390 – 3, 1989.

[29] T. Taghavi, F. Dabiri, A. Nahapetian, and M. Sarrafzadeh. Tutorial on congestion prediction. In *Proc. of SLIP*, pages 15 – 24, 2007.

[30] T. Taghavi, X. Yang, and M. Sarrafzadeh. Dragon2005: Large-scale mixed sized placement tool. In *ISPD*, pages 245–247, 2005.

[31] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, 58:267–288, 1996.

[32] N. Viswanathan, M. Pan, and C. Chu. Fastplace 3.0: A fast multilevel quadratic placement algorithm with placement congestion control. In *Proc. of ASPDAC*, 2007.

[33] T. Yan, S. Li, Y. Takashima, and H. Murata. A theoretical study on wire length estimation algorithms for placement with opaque blocks. In *Proc. of ASP-DAC*, pages 268 – 273, 2007.

[34] T. Yan and H. Murata. Fast wire length estimation by net bundling for block placement. In *Proc. of DAC, Digest of Technical Papers, ICCAD*, pages 172 – 178, 2006.