

Research Report

Direct Printability Prediction in VLSI using Features from Orthogonal Transforms

K. Kryszczuk*, P. Hurley*, R. Sayah‡

*IBM Research – Zurich, Switzerland

‡IBM Systems and Technology Group, USA

LIMITED DISTRIBUTION NOTICE

This report has been submitted for publication outside of IBM and will probably be copyrighted if accepted for publication. It has been issued as a Research Report for early dissemination of its contents. In view of the transfer of copyright to the outside publisher, its distribution outside of IBM prior to publication should be limited to peer communications and specific requests. After outside publication, requests should be filled only by reprints or legally obtained copies (e.g., payment of royalties). Some reports are available at <http://domino.watson.ibm.com/library/Cyberdig.nsf/home>.



Research
Almaden • Austin • Beijing • Delhi • Haifa • T.J. Watson • Tokyo • Zurich

Direct printability prediction in VLSI using features from orthogonal transforms

Krzysztof Kryszczuk
IBM Zurich Research Lab
kkkr@zurich.ibm.com

Paul Hurley
IBM Zurich Research Lab
pah@zurich.ibm.com

Robert Sayah
IBM Systems and Technology
sayah@us.ibm.com

Abstract

Full-chip printability simulations for VLSI layouts use analytical and heuristic physical process models, and require an explicit creation of a mask and image. This is a computationally expensive task, often prohibitively so, especially when prototyping new designs. In this paper we show that using orthogonal transform-based fixed-length feature vector representations of 22nm VLSI layouts to perform classification-based rapid printability prediction, can help in avoiding or reducing the number of simulations. Furthermore, in order to overcome the problem of scarcity of training data, we show how re-scaled, abundant 45nm designs can train error prediction models for new, native 22nm designs. Our experiments, run on M1 layer data and line width errors, demonstrate the viability of the proposed approach.

1 Introduction

With the steady reduction of the track width in VLSI designs, printing designed shapes becomes increasingly difficult [10, 8, 1]. A typical VLSI design must undergo a printability simulation in advance of creating physical masks, so as to avoid production yield problems [11, 2]. Printability simulation is a computationally expensive task [3], often prohibitively so for a full-chip simulation of a large layout in new Technology Nodes (TN) beyond 45nm [8].

Efforts towards simulation speedup can take the approach of algorithm optimization and usage of multiple CPUs [8]. These full-chip simulation methods explicitly emulate a mask and simulate a print image, which is the root cause of the excessive computational burden. In many applications, however, it can suffice only to detect if a particular shape is likely to create printing problems, and avoid simulating its corresponding mask and image shapes. Several attempts have been made in this

direction. In [2] a Support Vector Machine (SVM) classifier is deployed to detect pinching and bridging print errors. A dual-graph-based representation for detection of design hotspots is proposed in [1]. In [6], the entire layout can be analyzed in a window-by-window fashion using a predetermined window overlap. For each window, a feature vector is obtained by applying a distance transform and by computing a histogram of the resulting image. Obtained feature vectors are then classified using an SVM classifier.

We show how using two-dimensional orthogonal transforms to represent VLSI layout shapes as fixed-length feature vectors for printability prediction using a trained classifier, full simulation may often be avoided. The idea of using fixed-length feature vector representation for clustering layout constructs was proposed in [5]. In this paper we use the Walsh transform, as in [5], and the two-dimensional Discrete Cosine Transform (DCT), to represent a layout window for printability prediction. We demonstrate that the proposed fixed-length feature vector representation allows more accurate printability prediction for line width errors than the features previously proposed in [6].

The existing 45nm TN will soon be replaced by 32nm and 22nm TNs. A frequent problem arising when working with layouts designed in a new TN is the scarcity of available real design data to train error models. We propose the use of designs from older technology nodes, re-scaled to meet the dimension constraints of the new technology node, to train the printability prediction classifiers. We demonstrate that such a strategy enables construction of a competent classifier for printability prediction for the 22nm TN from abundant existing 45nm design data.

Section 2 describes extraction of the Walsh and DCT features. Section 3 describes the concept of classification-based print error prediction for VLSI, Section 4 gives details and results of our experiments, and Section 5 concludes the paper.

2 Extraction of Walsh and DCT features

The 2D DCT and Walsh transforms are closely related to the Fourier transform. They are known methods of representing images in a fixed-length feature vector format in pattern recognition applications, for instance face recognition [9, 12]. Given image W , a matrix M_F of 2D transform coefficients is computed as follows:

$$M_F = D \cdot W \cdot D^T, \quad (1)$$

where D is the 1D transformation matrix of DCT or Walsh transform respectively. Elements of M_F are consequently selected, usually in a zig-zag pattern, and placed into a feature vector [12]. The resulting feature vector begins with the coefficients representing the low spatial frequencies present in the image, and going towards higher frequencies as the vector length increases. The maximal number of features possible to obtain using Eq. 1 is equal to the number of pixels in W . However, this is usually by far in excess of the optimal number of features. We describe the feature selection procedure used in this paper in Section 4.

3 Classification-based VLSI printability prediction

The proposed scheme of classification-based VLSI printability prediction consists of a training step and a classification step. In the training step, a layout with overlaid printability error markers is used. The location of the error markers on the training layout is determined via prior detailed, computationally expensive full simulations using the Calibre[®] nmOPC[™] tool. The training and testing layouts are divided into overlapping square tiles, whose side length corresponds to 5 track lengths of the layout, similarly as in [4]. In this way we assure that the entire immediate neighborhood of an error marker is included in the tile. The tiles overlap by 50%. All tiles are subsequently reduced to images of 32×32 pixels, with no loss of effective shape resolution.

Next, the tiles are assigned class labels before they can be used for classifier training. For each error marker, one tile containing it is assigned to the error class (A). All tiles that do not contain error markers are labeled as no-error class (B). For each tile a feature vector is extracted, as described in Sec. 1.

Using the two sets of feature vectors, a Bayesian classifier is trained [7], by estimating distributions $p(x|A)$ and $p(x|B)$. For a newly observed tile probability that given tile I will be free of print faults is given

by

$$P(B|x) = \frac{p(x|B)P(B)}{p(x|A)P(A) + p(x|B)P(B)}, \quad (2)$$

where x is an instance of a fixed-length feature representation of the considered layout fragment W . Distributions $p(x|A)$ and $p(x|B)$ are estimated using training data from prior simulations. We assumed non-informative prior probabilities $P(A)=P(B)$.

4 Experimental evaluation

For the experiments, we used a fragment of the M1 layer of a native 22nm design, and a section of the M1 layer of a 45nm design. The 45nm design has been scaled down in order to match its track width with that of a native 22nm design. Both layout fragments subsequently went through a detailed printability simulation. The obtained error markers served as ground-truth class labels for classifier training (45nm), and for evaluating the classification accuracy (45nm and 22nm). Examples of tiles belonging to the error and no-error classes, originating from the re-scaled 45nm layout, are shown in Figure 1.

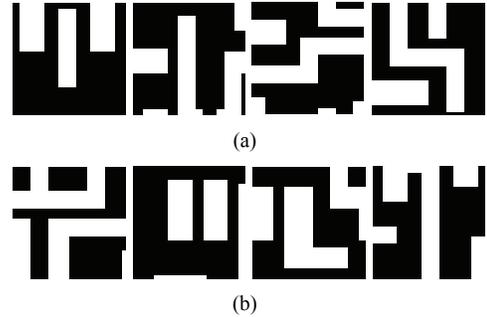
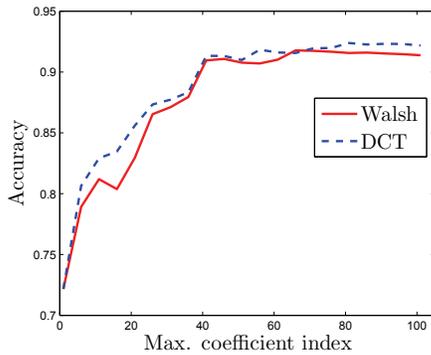
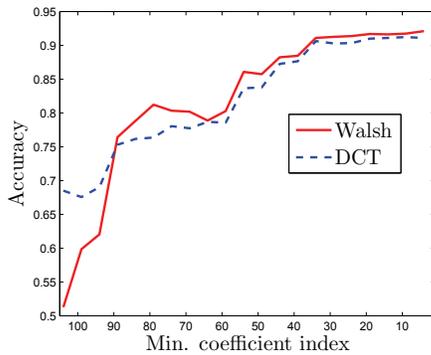


Figure 1. Sample layout fragments containing line width printability errors (a), and without errors (b). White area corresponds to the metal tracks.

The number of the DCT and Walsh features (transform coefficients) was selected empirically using the wrapper approach. A vector of 105 features for both feature extractors were extracted from each tile. The resulting data set was divided into two disjoint sets, training and testing, of equal size per class. The candidate feature sets were created by forming vectors $F_n = [c_1, c_2, \dots, c_n]$ for $n \in \langle 2, 3, \dots, 105 \rangle$, and $F_m = [c_{105}, c_{104}, \dots, c_m]$ for $m \in \langle 104, 103, \dots, 1 \rangle$, to verify the importance of the coefficients representing low (LF)



(a)



(b)

Figure 2. Impact of the presence of coefficients corresponding to low-frequency (a) and high-frequency (b) in the feature vector, for DCT and Walsh transform, accuracy in terms of $(1 - EER)$.

and high (HF) spatial frequencies. For each considered candidate set of features, a classifier was built using the training set, and accuracy on the testing set was recorded in terms of Equal Error Rate (EER). The results of the wrapper feature selection experiments are shown in Fig. 2. It is evident from Fig. 2 that the LF coefficients are of primary importance in achieving high classification accuracy. Based on the results, we used 75 first coefficients of DCT and Walsh transforms as features in two following experiments, E_1 and E_2 .

We compared the accuracy of the DCT- and Walsh-feature based classifiers with the Distance Transform-based approach (DT) from [6], on the task of predicting the line width errors for the 45nm data (E_1), and for a native 22nm design. All error models were trained on re-scaled 45nm data (E_2). The number of layout tiles used for training and testing is given in Table 1. We eliminated duplicate layout tiles, which in the case

	error tiles	no-error tiles
45nm	1543	44508
22nm	118	881

Table 1. Number of layout tiles (feature vectors) available for re-scaled 45nm, and native 22nm layouts.

	$HTER_{min}$	EER	$HTER_{0.5}$
training: 45nm, testing: 22nm			
Walsh	1.87%	2.35%	2.15%
DCT	1.75%	1.95%	1.87%
DT	11.75%	14.00%	14.00%
training: 45nm, testing: 45nm			
Walsh	5.41%	5.49%	5.88%
DCT	5.01%	5.16%	5.43%
DT	17.9%	18.96%	19.65%

Table 2. Classification error rates for experiments E_1 and E_2

of the highly-regular 22nm layout led to much lower number of non-identical feature vectors than in the case of a more diverse 45nm design.

In the E_1 experiment we used the cross-validation strategy; the 45nm data was divided into two disjoint sets of equal sizes. The classifiers were trained using one of the sets, then deployed to classify the other set, and vice-versa. Reported error rates are averaged for both cross-validation runs. In the experiment E_2 , the entire scaled 45nm data volume was used to train the classifiers, and the whole 22nm data set was used for testing. We used Bayesian classifiers, with the class-conditional feature distributions were modeled as Gaussians with full covariance matrices¹. Table 2 shows the obtained error rates in terms of minimal achievable Half-Total Error Rate ($HTER_{min}$), Equal Error Rate (EER), and $HTER$ for a naive Bayes threshold of $\tau = 0.5$ ($HTER_{0.5}$). The DET curves for both experiments are shown in Fig. 3.

The results shown in Fig. 3 and Table 2 clearly show that the classifiers based on the DCT and Walsh features outperformed the classifiers based on the DT features, in both experiments (E_1 and E_2). The error rates obtained by using the DCT and Walsh features are similar. The recorded differences between $HTER_{min}$, EER and in particular low values of $HTER_{0.5}$ show that the classifiers are not easily biased and fairly insensitive to the choice of the decision threshold, which is very important in practical applications. The error rates for E_2

¹We also tried other model types, including GMM-based ones, and discriminative classifiers, with comparable outcomes.

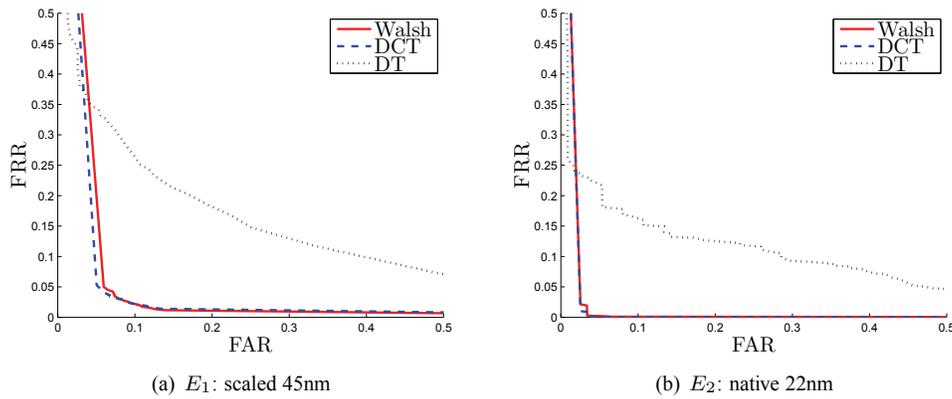


Figure 3. DET curves representing the performance of the printability predicting-classifier, trained using scaled 45nm data, applied to classify (a) E_1 : scaled 45nm data (training and testing sets disjoint), and (b) E_2 native 22nm data.

are considerably lower than those for E_1 , due to the fact that the 22nm layout was more regular than the 45nm layout.

5 Conclusions

In this work we presented an application of DCT and Walsh transforms for effective feature extraction for printability error prediction in VLSI designs, using an example of the line width error. Obtained results suggest that classification-based print error prediction in transformed domain is a viable, computationally inexpensive alternative to full-scale chip simulations involving explicit mask and image creation. We also showed that very good error prediction accuracy can be obtained by training error models on abundant, re-scaled 45nm design data, and deploying the models to predict printing errors in new 22nm designs. This work is an initial proof of concept and we plan to extend it model other error types and different layout layers.

References

- [1] X. X. Andrew B. Kahng, Chul-Hong Park. Fast dual graph based hotspot filtering. *Trans. of Computer-Aided Design of Integrated Circuits and Systems*, (Volume 27, Issue 9):6281–04, 2008.
- [2] A. Borjon, J. Belledent, S. D. Shang, O. Toublan, C. Miramond, K. Patterson, C. Lucas, Kevin; Couderc, Y. Rody, F. Sundermann, J.-C. Urbani, S. Baron, Y. Trouiller, and P. Schiavone. High accuracy 65nm OPC verification: full process window model vs. critical failure ORC. In *Proc. of the SPIE*, volume 5754, pages 1190–1201, 2005.
- [3] Y. Cao, Y.-W. Lu, L. Chen, and J. Ye. Optimized hardware and software for fast, full chip simulation. In *Proc. of SPIE, Symposium on Microlithography*, volume 5754, pages 407–414, 2005.
- [4] V. Dai, L. Capodici, J. Yang, and N. Rodriguez. Developing DRC Plus rules through 2D pattern extraction and clustering techniques. In *Proc. of the SPIE 7275*, 2009.
- [5] D. L. DeMaris, D. Maynard, B. B. Reuter, and S. Zhong. An information retrieval system for the analysis of systematic defects in VLSI. In *Proc. of the ICTAI 2004*, pages 216–223, 2004.
- [6] D. Drmanac, F. Liu, and L.-C. Wang. Predicting variability in nanoscale lithography processes. In *Proc. of the 46th IEEE/ACM Design Automation Conference*, San Francisco, 2009.
- [7] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley Interscience, New York, 2nd edition, 2001.
- [8] J. Ferguson. Reducing physical verification cycle time. Technical publications, Mentor Graphics, 2008.
- [9] X. Jia and M. S. Nixon. Profile feature extraction via the walsh transform for face recognition. In *Proc. Int. Conf. on Intelligent Robots and Visual Communications*, pages 46–52, 1992.
- [10] B. Kruseman, A. Majhi, C. Hora, S. Eichenberger, and J. Meirlevede. Systematic defects in deep sub-micron technologies. In *Proc. of ITC 2004 (International Test Conference)*, 2004.
- [11] E. Sahouria, Y. Granik, N. Cobb, and O. Toublan. Full-chip process simulation for silicon DRC. In *Proc. of the International Conference on Modeling and Simulation of Microsystems (MSM)*, pages 32–35, 2000.
- [12] C. Sanderson and K. K. Paliwal. Features for robust face based identity verification. *Signal Processing*, 83(5):931–940, 2003.