6-1-2014

Layout regularity metric as a fast indicator of process variations

Esraa AbdelAzim AbdelHamid Swillam

The American University in Cairo

School of Sciences and Engineering

Layout Regularity Metric as a Fast Indicator of Process Variations

A Thesis Submitted to

Department of Electronics Engineering

In partial fulfillment of the requirements for
the degree of Master of Science

by Esraa Swillam

under the supervision of Dr. Mohab Anis
March, 2014

The American University in Cairo

School of Sciences and Engineering

Layout Regularity Metric as a Fast Indicator of Process Variations

A Thesis Submitted by

Esraa AbdelAzim AbdelHamid Swillam

To the Department of Electronics Engineering

March/2014

In partial fulfillment of the requirements for

The degree of **Master of Science**

Has been approved by

Thesis Supervisor

Affiliation :

Thesis Internal Examiner

Affiliation :

Thesis Internal Examiner

Affiliation :

Department Chair/Director:

Date:

Dean:

Date:

Abstract

Integrated circuits design faces increasing challenge as we scale down due to the increase of the effect of sensitivity to process variations. Systematic variations induced by different steps in the lithography process affect both parametric and functional yields of the designs. These variations are known, themselves, to be affected by layout topologies. Design for Manufacturability (DFM) aims at defining techniques that mitigate variations and improve yield. Layout regularity is one of the trending techniques suggested by DFM to mitigate process variations effect. There are several solutions to create regular designs, like restricted design rules and regular fabrics. These regular solutions raised the need for a regularity metric. Metrics in literature are insufficient for different reasons; either because they are qualitative or computationally intensive. Furthermore, there is no study relating either lithography or electrical variations to layout regularity. In this work, layout regularity is studied in details and a new geometrical-based layout regularity metric is derived. This metric is verified against lithographic simulations and shows good correlation. Calculation of the metric takes only few minutes on 1mm x 1mm design, which is considered fast compared to the time taken by simulations. This makes it a good candidate for pre-processing the layout data and selecting certain areas of interest for lithographic simulations for faster throughput. The layout regularity metric is also compared against a model that measures electrical variations due to systematic lithographic variations. The validity of using the regularity metric to flag circuits that have high variability using the developed electrical variations model is shown. The regularity metric results compared to the electrical variability model results show matching percentage that can reach 80%, which means that this metric can be used as a fast indicator of designs more susceptible to lithography and hence electrical variations.

Acknowledgements

I would like to thank my supervisor, Dr. Mohab, for being helpful, flexible and supportive through the hard times I had during my graduate studies. I want to thank him for giving me the freedom to explore different research ideas and to choose what I loved. I, also, want to thank him, together with Dr. Ayman Elezabi in helping me in admission and guiding me to get a partial fellowship from the AUC which helped me a lot.

I wouldn't have been here without the love and support of my family, specially my mother and my husband who bore a lot with me and were always there to encourage and push me when I was down. My father and brother helped me a lot with their academic background. My sister and best friend love and support helped me through all the hard times.

I would like also to thank my company, Mentor Graphics for supporting part of the research and thank my manager, Omar Kamal, for his continuous encouragement and support. I want also to thank my technical leads, Andres Torres and Kareem Madkour and my colleague Sarah Mohamed for guiding me technically through different parts of this thesis.

Before all, I thank Allah for giving me the inspiration, power and persistence to do this work.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Integrated Circuits design and manufacturing face increasing challenges as technology scales down to follow Moore's law. The main cause of these challenges is that the wavelength of light used in lithography (193nm) is much larger than the target features to be printed. Extreme Ultraviolet (EUV) was thought to be the enabler for 22nm and beyond, however its high cost and some technical challenges for volume production pushed the industry to find other solutions for getting smaller dimensions using the 193 nm light source [10].

This chapter gives a brief introduction about the efforts done to enable scaling in the form of resolution enhancement techniques. It will discuss one of the main challenges which is process variations, its sources and impact. Then, it will talk about the efforts in mitigating process variations and design for manufacturability techniques (DFM) as some of these efforts. The following part will discuss layout regularity and how designers headed towards more regular solutions to mitigate the effect of variations and how this raised the need for a regularity metric. Section 1.2 will show the contribution of this research and section 1.3 will brief how this thesis is structured.

1.1.1 Technology Scaling Down

The resolution limit of optical lithography is often discussed in the context of the famous Rayleigh criterion defined as

$$CD = k_1 \frac{\lambda}{NA} \tag{1.1}$$

where CD is the critical dimension, λ is the wavelength, and NA is the numerical aperture of optical system. The parameter k_1 depends on the process specifications. Scaling down means smaller CD, to get smaller CD; NA should be increased, k_1 and λ should be decreased. To increase NA a powerful system of lenses is being used together with the idea of immersion lithography. As can be noticed from (1.2), NA can be increased not only by increasing the size of the lens to increase $\sin(\theta)$ but also by choosing a medium whose refractive index n is higher than that of air which is 1. This medium should have the following properties; low optical absorption at 193nm, compatible with photoresist and the lens material and non-contaminating. All these properties exist in pure water whose refractive index is 1.44.

$$NA = n \sin(\theta) \tag{1.2}$$

Another way to improve resolution for the same λ is through decreasing k_1, this is done by several resolution enhancement techniques (RET), examples of these are listed below:

- Optical Proximity Correction (OPC)
- Sub-Resolution Assist Features (SRAF)
- Phase Shifting Mask (PSM)
- Off Axis Illumination (OAI)
- Double Patterning

Each and every one of these techniques added more restrictions to design which made design rules very complex. Furthermore, abiding to design rules alone, still don't guarantee yield because of the non-linear effects induced by sub-wavelength fabrication and due to process variations which will be discussed in the following section.

1.1.2 Process Variations

Variation is the deviation from intended or designed values for a structure or circuit parameter of concern. Process variations effect can be severe enough to affect the functional yield so that after printing the circuit doesn't work or it can affect the parametric yield so that the manufactured circuit works but with performance different from that of simulations. As minimum feature size decreases with scaling down, the effect of process variations becomes more and more dramatic as the tolerance becomes tighter.

There are two main categories for the sources of process variations; environmental and physical. This thesis concentrates more on the physical sources which arise mainly from different manufacturing steps as lithography, etching, ion implantation and chemical mechanical polishing.

1.1.3 Solutions for process variations

There are several techniques to mitigate the effect of process variations. Some are process-oriented and some are design oriented. Design oriented solutions includes design rules, layout modifications or robust circuit and system design. From the design oriented techniques to mitigate variations and improve yield comes the design for manufacturability branch of techniques.

Design for Manufacturing (DFM) are techniques that aim at improving the printability and yield through better communication between the design step and the manufacturing step [11]. DFM encourages ideas that provide better designs prior to manufacturing to decrease process variations effects and thus give better yield. DFM can be accomplished through a combination of minimizing process variations, and minimizing

device and circuit sensitivities to process variations. Many of the rules advocated in the DFM method include measures to mitigate pattern dependent variation, such as reduction of geometric primitives, layout regularity, redundancy techniques, pattern density uniformity, and optimal process feature sizing. Other areas of consideration in DFM are design for robustness and design for test[12]. This thesis concentrates on layout regularity as one of the trending techniques in mitigating process variations effect.

It was noticed that some designs and design styles are less sensitive to process variations, that's why SRAMs and FPGAs are the earliest products to migrate to a new, smaller, technology node. These design styles are able to reach high yield production level faster than other designs because they contain a very high degree of regularity. The repetitive bit-cells in SRAMs and Configurable Logic Blocks (CLBs) in FPGAs result in a limited number of unique geometry shapes in the layouts. One can afford to perform detailed analyses and simulations to employ strong-RETs on these shapes to improve manufacturability. The repetition of fundamental building blocks also enables the pre-qualification of the building blocks in silicon for known layout neighborhoods.

There are two types of regularity, the design regularity in which all building blocks have predictable neighborhoods is considered macro-regularity, and this is the type of regularity observed in SRAM and FPGAs. As feature sizes shrink, another type of regularity is indispensable which is micro-regularity . Micro-regularity is achieved by minimizing the number of unique geometry patterns used in a layout from a bottom-up approach by restricting the flexibility to draw arbitrary shapes. Constraints such as single width, single pitch, and single orientation of critical dimensions are imposed to enhance the printability and characterization predictability of critical layers. These constraints are imposed to enable and to simplify strong-RETs analyses and applications, thereby improving the printability of designs that adhere to micro-regularity [13]

1.2 Contribution

The contribution of this work is defining the required regularity for better robustness against process variations and coming out with a metric that can assess this regularity.

The main advantages of the regularity metric are that it is very fast; it takes a few minutes to run on a 1mm x 1mm design, it can be used to compare regularity of patterns within the same layout and it is closely correlated to the variations induced in lithography process so it can be used as a fast indicator of the patterns with high lithography variations.

This metric has the following possible usage scenarios:

- It can be used in pre-layout phase to help the designer create correct by construct patterns from the beginning.
- It may be used post layout in design verification phase to give something similar to critical area analysis where it highlights the problematic patterns that are more susceptible to process variations. The best practices deduced from the model will be used to provide hints to the designer to help him solve these problematic areas.
- Due to its simplicity and high speed, it can be used prior to lithographic simulations to select the areas of interest, which are areas that are more probable to have variations, then run lithographic simulations on only these areas to detect hotspots.

1.3 Structure of the thesis

This thesis is structured as follows; Chapter 2 gives the background needed to understand the basic manufacturing steps and their contribution as sources of variations, section 2.2 shows how variation impacts transistor parameters, while section 2.3 shows some variation mitigation methodologies on different design levels. Chapter 3 is dedicated to regularity, it starts with showing the effects of topology on electrical parameters in section 3.1 which clarifies the reason behind adopting regularity, then, in section 3.2, it shows the regular design solutions and how they evolved, section 3.3 shows the need for regularity metric and the existing metrics in literature. Chapter 4 shows the work done in deriving a regularity metric, it starts, in section 4.1, with a first trial using

image processing and Fourier transform then it shows the problems encountered in this methodology and how we moved to a geometry-based metric, then, in section 4.2, it shows the steps of deriving the geometry based metric and how it was validated against a variability metric that is a modified version of one found in literature, then it shows the work done to use the regularity metric as a fast indicator of electrical variability where it shows the comparison of regularity metric results to the results of a selected electrical variability method in literature.

Chapter 2

Background: Process Variations; Sources, Impact and Mitigation

Process variations are increasingly becoming the limiting factor in both IC design and manufacturing, so that, to a great extent, they dictated the style and progression of design in integrated circuits. Understanding the sources of process variations is critical to developing better design rules for circuits, ensuring accurate tests for robustness, and controlling manufacturing conditions for optimal design yield.

Process variations sources can be classified mainly into two categories *environmental* sources, those arise during the operation of a circuit, and include variations in power supply, switching activity, and temperature of the chip or across the chip. Variations resulting from environmental sources depend primarily on architectural and operating decisions such as power grid design and component placement. Time-varying variation in these environment parameters can be a significant design concern. The second category of variations, which is our main concern in this thesis, is due to *physical* factors. Examples of these physical factors are the fluctuations that appear during manufacturing and result in structural device and interconnect variations, which are essentially permanent. Physical variations can also be classified into device and interconnects and also into geometry, material, and electrical categories as shown in Figure 2.1.

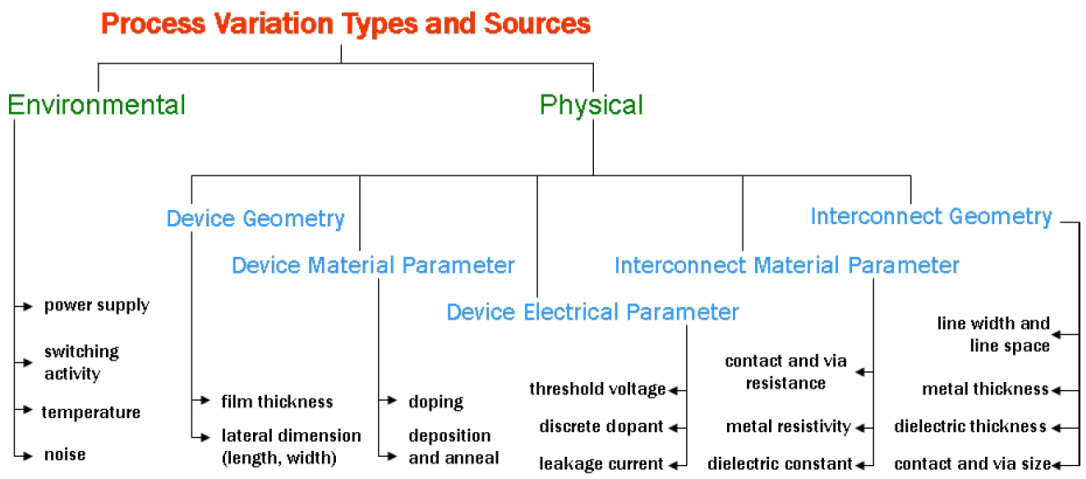


FIGURE 2.1: Process variation types and sources [1]

As can be noticed from Figure 2.1, almost every step in the manufacturing process introduces variations in the end device which affect the functional and parametric yield of the whole design.

To understand more about physical process variations sources, this chapter provides a basic introduction to semiconductor processing. It gives a brief overview of some of the main manufacturing process steps and it places emphasis on variations introduced by each manufacturing step. The impact of variations on different transistor parameters is discussed in section 2.2.

Because the manufacturing process may never be deterministic, designers always try to use design techniques and methodologies that are known to reduce process variation. Accordingly, the last section shows the different approaches followed to mitigate variation effects in each design level.

2.1 Sources of variation

There are several steps in integrated circuits fabrication. Theses steps can be classified into front-end processes and back-end processes. The front-end ones are those related to fabrication of devices like MOS transistors. They involve oxidation, doping and annealing with some patterning and etching steps. On the other hand, back-end

processes are those involved in creating interconnect wires, they include metallization, deposition and chemical mechanical polishing [14].

Some of these process steps are considered to be major sources of variation, these are: 1) photolithography, 2) etching, 3) ion implantation and rapid thermal annealing, and 4) chemical-mechanical polishing (CMP). The effect of variations of each step depends on the feature being fabricated, so that each process step affects subsequent transistor and interconnect parameters. For example, variations in lithography and etch affect the physical dimensions of transistors. CMP variations affect wires and vias that form the interconnect between transistors. On the other hand, variations in ion implantation and rapid thermal annealing directly impact the doping of transistors. One more thing to be considered is that the the impact of variations in a certain process step also depends on the application in which the fabricated transistor is used. For example, variation in the size of the source/drain area of a transistor has much greater impact on the overall performance when that transistor is used in an analog versus digital application [4]. The following sections will show in details some of the process steps and how they contribute as sources of variations.

2.1.1 Photolithography

Photolithography is the first step in IC manufacturing. It consists of several steps that start with photoresist application followed by exposure, baking and development. Photolithography is also one of the main contributors to physical variations. The subsections below are going to discuss the different steps in photolithography, the challenges each one has and also the variations they introduce.

The current industry method is projection printing. The exposure system consists of three main components. First, a source of light at a particular wavelength is needed (193 nm). Second, a reticle or mask that has the desired pattern drawn in the form of transparent and opaque areas. The light is passed through the mask to transfer that pattern onto the photoresist-coated wafer. The pattern on the mask is greater than the desired one on the wafer so the mask pattern is demagnified as it is transferred to the

wafer. To accomplish this image reduction, the third component of the exposure system, a lens consisting of several optical elements (called projection optics), is required to project a reduced image of the mask onto the target wafer. The mask is exposed to light together with the wafer that has been coated by photoresist.

Projection optics give rise to diffraction effects, which in turn limit the minimum printable feature size of the lithography system as expressed by the Equation 1.1. In order to improve the resolution of a lithography system, the most obvious solution is to reduce the illumination wavelength. The industry has shifted the illumination wavelength from the mercury e-line (577 nm) to the g-line (436 nm) and iline (365 nm), and on to deep ultraviolet (DUV) wavelengths 248 nm (KrF excimer laser illumination), and, presently, to 193 nm (ArF excimer laser). At the 248-nm illumination wavelength node, lithographers began printing feature sizes that were smaller than the illuminating wavelength. This improvement has been possible due to the other two parameters in Equation 1.1, namely NA and $k1$. The numerical aperture is given by Equation 1.2. The step-wise exposure of single, small portions of the wafer (called fields) allows systems to be built with very high NAs, since the acceptance angle is greatly increased. "Dry" systems (where the medium between the wafer and the lens is air with n = 1) used to have NAs over 0.8, compared to an NA around 0.25 in the 1970s, allowing for a three-fold improvement in resolution at a given wavelength. As lens size increases, imperfections in the lens, or aberrations, become more difficult to avoid during lens manufacturing. To combat this difficulty, lithography systems limit the portion of the lens that is actually used during imaging. This is accomplished by using a slit centered over the lens that allows only a portion of the mask pattern to be imaged onto the wafer at one time. To print the entire field, the mask is then translated in one direction while the wafer is translated in the opposite direction in a scanning fashion as shown in Figure 2.2.

There are limits to the potential continued increase in NA. First, when imaging in air, NA cannot exceed 1 , since the index of refraction for air is approximately 1 and the $sin(\theta)$ term cannot exceed 1. State-of-the art immersion lithography systems increase NA to values greater than 1 (1.25) through the use of a medium of higher refractive index, water, between the lens and wafer.

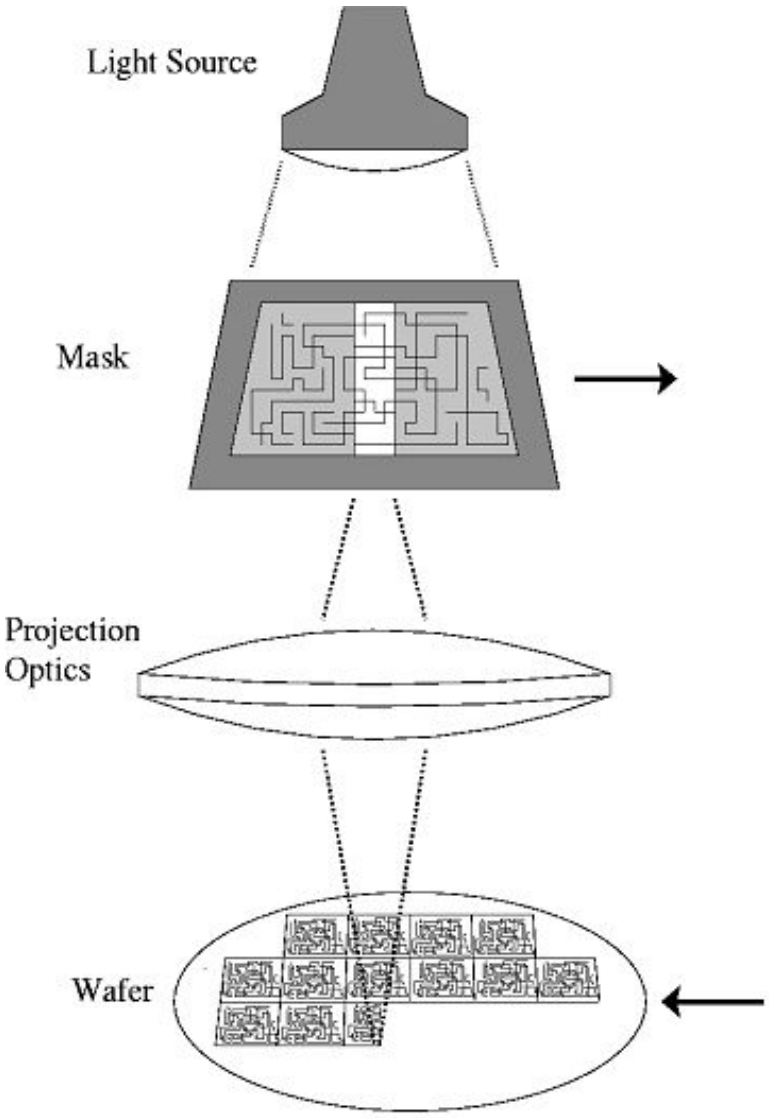


FIGURE 2.2: Illustration of the projection scanner exposure system. To print each lithographic field, the wafer and mask are translated in opposite directions in a scanning fashion [2].

However, the increase of NA has limitations, since the depth of focus (DOF), defined to be the range over which the wafer can be moved along the optical axis such that the image stays in focus, is inversely proportional to the square of NA as shown in 2.1.

$$DOF = k_2 \frac{\lambda}{NA^2} \tag{2.1}$$

Photoresist

A photoresist is a material that changes its chemical properties upon exposure to light. Photoresist has three main components: a resin (base material), a photoactive compound (PAC), and a solvent that controls the viscosity of the substance. There are two types of photoresists; positive photoresist and negative photoresist. A positive photoresist becomes more resistant to a certain solvent called developer when exposed to light, while the negative photoresist becomes less resistant to the developer. Figure 2.3 shows the two types of photoresist. In positive resists, the PAC behaves as an inhibitor, nominally preventing the resist from dissolving in developer solution. While, in case of negative photoresist, exposure to light triggers a chemical reaction which changes the PAC into a sensitizer, a compound which enhances the dissolution rate in developer by breaking the resin structure material down.

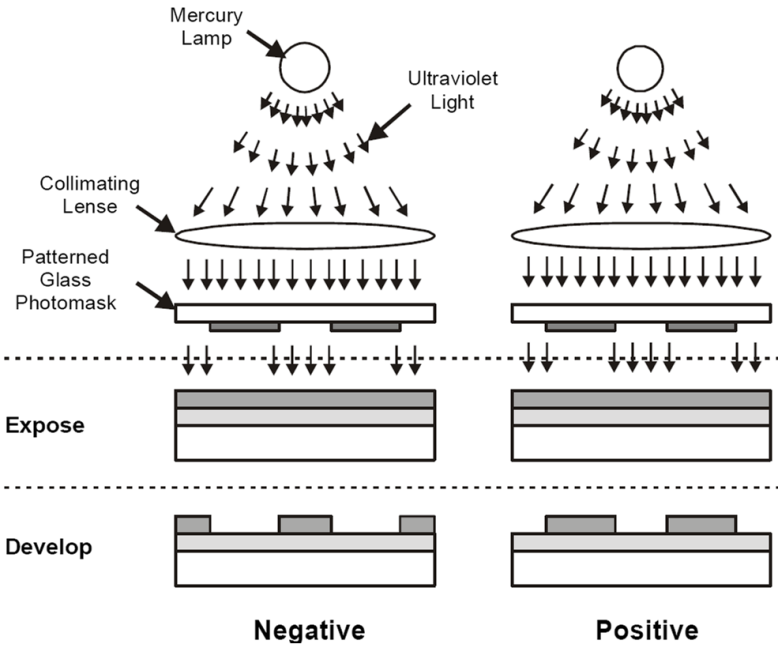


FIGURE 2.3: Types of photoresists

Photoresist is applied to the wafer with a spin-coating process as shown in Figure 2.4, which yields an exceptionally uniform film across the wafer. However, from wafer to wafer and lot to lot, variability in spin speed, resist viscosity, and adhesive properties between the resist and substrate lead to film thickness variation. This variation will in turn lead to variation in the printed CD [2].

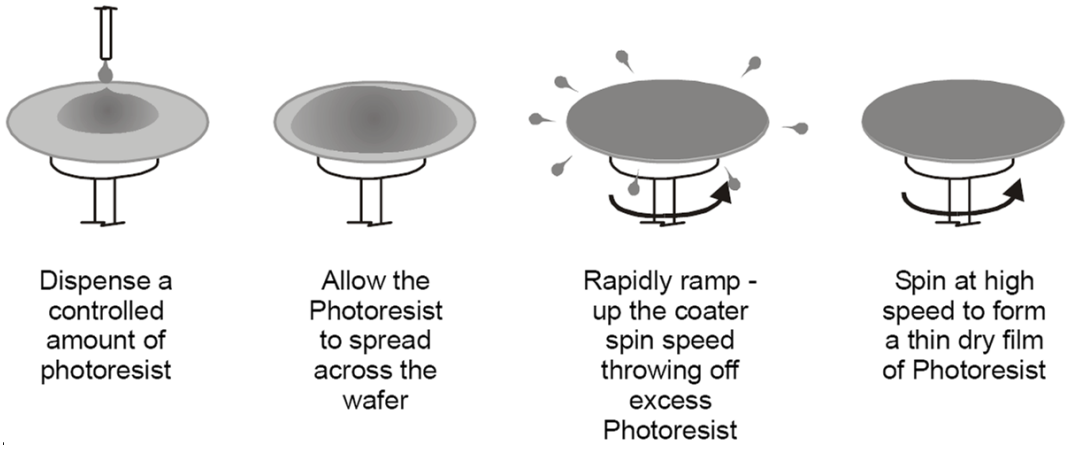


FIGURE 2.4: Photoresist spin coating process

The resist used is called chemically amplified resist. Chemical amplification is achieved by an agent in the resist which greatly increases the potency of the photochemical process, so that an incident photon will create a cascade of chemical reactions. The resist sensitivity relies heavily on thermal activation, which is carried out in a post-exposure bake (PEB) step. The resulting image (prior to development) in the resist depends strongly on the thermal dose, so both PEB temperature and duration are key factors. PEB bake stations must be designed to place the wafer perfectly level on the heated surface so that the same thermal dose is received at all points on the wafer; also, the nominal temperature must be carefully maintained. At a nominal temperature of roughly 100C, deviations of a single degree will lead to 5 nm deviations in resulting feature width. There are two other bake steps in a standard lithography flow: the post-application bake (PAB) and an optional hard bake. The PAB step serves to drive off some of the solvent in the film following spin-coat, serving to create a firmer film prior to exposure. The hard bake step follows development, and is intended to drive off any remaining solvent in the remaining resist to strengthen the masking features prior to the etch process. Since neither step has a strong effect on the resulting pattern, they generally do not receive much attention from a process control standpoint.

Now the wafer is immersed in the developer solution so that the pattern intended to be transferred remains in the form of the photoresist. Finally, the underlying regions which are not protected by photoresist are etched away; when the remaining photoresist is stripped away, the layer has been patterned as desired.

2.1.2 Exposure

The most critical, difficult, and expensive step in the general lithography flow as shown in Figure 2.5 is the exposure step.

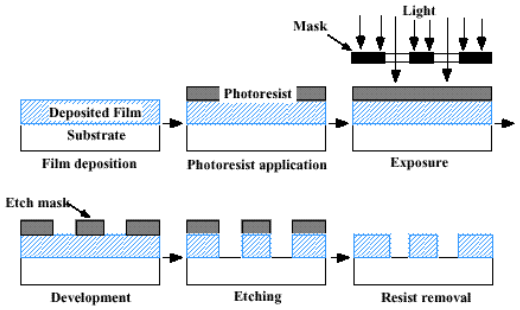


FIGURE 2.5: Photolithography and etching

As mentioned before, scaling of the wavelength of light used for patterning ceased at 193nm due to increased cost of lithography technology, materials, and equipment development and deployment. The resulting lithographic defocus causes both systematic and random line-width variations. Resolution Enhancement Techniques (RET) helped in mitigating these variations effect. RET include Optimal Proximity Correction (OPC), Sub-Resolution Assist Features (SRAF) and phase-shifted mask lithography explained briefly below.

- Optical Proximity Correction (OPC)

  Proximity effect refers to features with the same CD printing differently due to environment variation, which results in across-chip line-width variation (ACLV). Line shortening, another form of image distortion, occurs mainly due to diffraction at low k_1 imaging. As the CD decreases, line shortening increases dramatically. This behavior is very challenging because of its effect on overlay control and circuit density. The third type of error is corner rounding where sharp corners are filtered out by pupil this can cause the effective channel width to degrade due to rounding the elbow corner, corner rounding can cause functional and parametric errors [3].

OPC is the technique of predistorting mask patterns so that the printed patterns are as close to the desired shapes as possible. Examples of OPC techniques are: (a) selective line biasing; which increases the width of lines based on their nominal dimensions and environment (b) Serifs and hammer heads to sharpen corners as shown in Figure2.6.

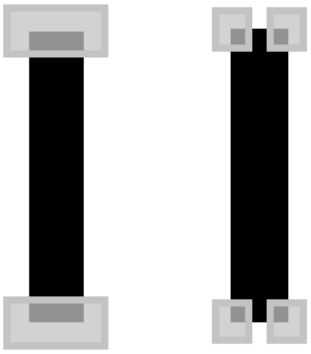


FIGURE 2.6: Hammer head (on the left) and Serif (on the right)

There are two major OPC strategies: rule-based strategy and model-based strategy [3]. In the rule-based method, the design is corrected based on rules extracted previously from simulations, experiments, and etc. Also, the amount of correction applied to a feature or an edge of a feature is determined by predefined tables. Although rule-based methods are efficient for small number of parameters, they cannot be efficiently applied as parameters and rules increase. Therefore, rule-based methods are not suitable for a process with long optical interaction range or assist-feature insertion, where many parameters are effective and many solutions exist.

Rather than applying rules, model-based OPC methods use mathematical models of the fabrication process to determine the correction. In addition to optical imaging, the process model can incorporate other effects such as flare and photoresist diffusion [3].

- Sub-Resolution Assist Features (SRAF)

  SRAFs are extra features that can be inserted in both sides of a line to create a dense environment. These assist-features are smaller than the main line such that they are not printed onto the wafer. Therefore, while the line appears dense

to the projection system, it is printed as a sparse line. However, assist-features should be optimized carefully in terms of number, size, and placement of features.

- Phase Shifting Mask (PSM)

  Image resolution is improved by taking the advantage of the interference generated by phase differences on phase shifting masks. Alternating PSM [15] and attenuated PSM [16] are the most popular PSM methods. In alternating PSM, the target line is bordered by transmitting regions with $180°$ phase difference. As shown in Figure2.7, this phase shift leads to destructive interference, resulting in a sharp dark image. The reliance on destructive interference between two bright regions to create a dark image makes alternating PSM applicable for imaging only small dark areas. In a phase shift mask, the regular process is applied to print regions of $0°$. Then, $180°$ phase-shifted regions are etched into the substrate followed by a post-processing for intensity balancing. Although alternating PSM improves imaging quality, it causes CD and placement errors due to imbalanced intensity between regions of $0°$ and $180°$ phases. Moreover, in addition to lens aberration, alternating PSM is very susceptible to mask defects not only due to the additional mask processing steps, but also due to the enhanced printing of small defects.

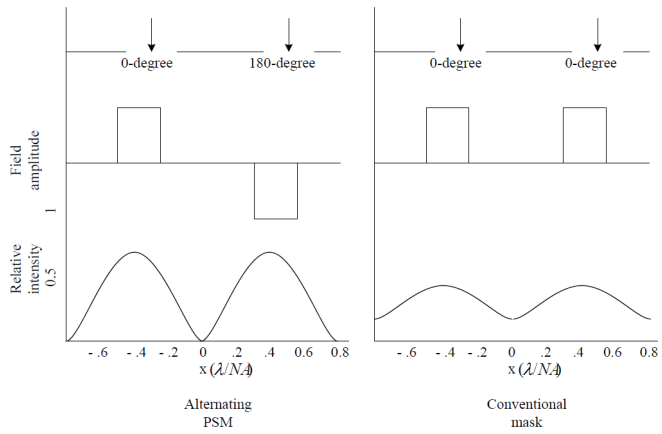


FIGURE 2.7: Mask schematics and images of a line printed with alternating PSM and conventional mask [3]

The opaque regions of an attenuated phase shift mask transmit the projected light partially with a $180°$ phase shift relative to the bright regions. The intensity

transmission of attenuated PSMs, usually around 7%, does not allow any image formation form the background. However, the destructive interference between the clear area and partially transmitting background enhances the image contrast of the bright region. Compared to alternating PSM, attenuated PSM is suitable to print sparse spaces, e.g. an isolated contact, because of its large DOF and exposure latitude. However, this method cannot compete with alternating PSM in printing narrow dark regions.

- Off Axis Illumination (OAI)

  When the light strikes in a photo-mask with small pitch, it is diffracted aggressively. If the incident light is parallel to the axis of the optical system, the zero-th diffracted order remains parallel to the axis while the other orders are diffracted sideways. This deviation increases as the pitch and coherency factor σ of the light decrease. For sufficiently small pitches, only the zero-th diffraction order passes through the projection lens while others are lost and no pattern is created on the wafer consequently. Therefore, for imaging at $k_1 <= 0.75$, the light source should be adjusted for a successful lithography. OAI is an optical system set up in which the exposure light strikes the photo-mask at an oblique angle rather than perpendicularly [3]. By making the illumination off-axis, all diffraction orders are titled, which makes it more likely that higher diffraction orders pass through the projection lens and contribute to image formation on the wafer.

- Double Patterning

  Double-patterning (DP) is a method for breaking up a layout so that sub-resolution configurations are separated between two distinct masks. These masks are exposed and processed sequentially to obtain the original design pattern by composing the layout features from the independent patterning steps.

2.1.3 Etch

Once the desired pattern has been formed in the photoresist layer, the etch process is used to transfer the photoresist pattern into the layer of underlying material. Etch

processes fall into two broad categories; wet etching and dry etching. Wet etch processes, where the entire wafer is immersed in a solution that reacts with and thereby removes the exposed parts of the underlying material, tend to be used only for non critical processes due to the inherent difficulty in controlling the properties of the etch. For example, wet etches are usually isotropic, the etch rate is uniform in all directions, so features undercut the regions defined by the photoresist mask. In addition, wet etch processes introduce a high risk of defects, making these processes unsuitable for defining small features. Dry etching processes, on the other hand, are relatively clean and can be tuned to generate the desired selectivity, etch rate, and anisotropy. The dry etch is performed using a plasma in a controlled, vacuum environment called a plasma chamber. The plasma is created using an electric field to break down an inert gas into electrons and ions; the ions are accelerated toward and bombard the cathode to generate secondary electrons, which are in turn accelerated back toward the anode and thus collide with the gas to create more ions.

To begin the etching process, a feeder gas is then introduced into the chamber and is broken down by the plasma into ions, radicals, electrons, and neutrals. Etching then occurs via two mechanisms: first, the chemical reactant species diffuse to the surface of the wafer and react with the exposed material. Byproducts may then be desorbed, diffused away from the wafer, and transported out of the plasma chamber. Second, physical bombardment by ions (driven into the surface of the wafer by the electric field used to set up the plasma) causes damage to the surface of the exposed film, both promoting the chemical reaction and the removal of byproducts. In some plasma etch processes (referred to as ion milling processes), material removal almost completely depends on the mechanical energy of ion bombardment, but for front-end pattern transfer process, a balance between chemical and mechanical etch is carefully struck.

Since the plasma may be easily started and stopped, it is much easier to start and stop the dry etching process than a simple wet etch process. In addition, the vertically oriented bombardment of ion species (arising from engineering the plasma environment to have a mean free path large in comparison to the dimensions of electrode separation) gives rise to high anisotropy, or nearly vertical sidewalls in the defined features. Finally,

since there are many fewer particles existing in the plasma environment than there are in an immersion environment, the defect problem is strongly reduced. However, since the etch rate will depend on local concentrations of the reactant species as well as the removal rate of the reaction byproducts, IC layout dependent microloading effects can arise during the etch process [17].

Etching is considered a source of variation; it also affects several features; in the case of transistor gates, or an insulator, such as silicon dioxide or a low-k dielectric, in the case of interconnect. Variations in process conditions such as chamber temperature and pressure, RF power, electrode spacing, and gas flows often result in variations if not properly controlled using statistical process control.

2.1.4 Ion Implantation and Annealing

Creation of transistors involves doping them with ions to define the type of transistor (PMOS or NMOS). The substrate, as well as other components of the transistor such as the highly-doped source/drain regions, are doped with different ion species (e.g. B, As, P). These ions are accelerated at high energy into the wafer during the ion implantation step and are then ”activated” by heating the wafer (annealing) in order to ensure the implanted ions are properly substituted within the existing crystal structure of the underlying silicon substrate. Once again, local and global process conditions such as implant energy and dose, tilt angle and temperature profiles all result in variations of the implanted ions. Layout features and proximity effects, such as distance to well edge, can also affect uniformity of ion implantation. In advanced process technologies, device volumes are so small that only several tens to low hundreds of dopant atoms are needed within the channel area, directly underneath the gate, for the required doping concentrations. Due to the small numbers, variation in the dopant counts and even the placement of the atoms within the transistor body is of significant concern, as regions of a single transistor will experience different local doping concentrations. This variation mechanism is known as Random Dopant Fluctuation (RDF) and was brought to light as early as 1975 [5]. It is easy to see that by changing the number or even the placement of the atoms the electrical performance of the transistor can be greatly impacted. As

transistor volumes continue to shrink, without individual placement of dopant atoms, RDF is unavoidable due to the decreasing absolute number of dopants required. As RDF and other variation mechanisms arising from ion implantation and annealing become increasingly significant in modern processes. The most common approach to mitigating this type of variation is to increase device size which reduces relative variation as the number of dopant atoms necessarily increases with device size. Such a solution is fundamentally incompatible with further transistor scaling, making it unsustainable in the long term and requiring solutions either at the process or design levels. Other solutions include: improved process modules for the doping step, new device structures not requiring doping and/or circuit design that is robust to device variation.

2.1.5 Chemical Mechanical Polishing (CMP)

CMP is used to achieve smooth and planar surfaces from which subsequent layers are able to be fabricated. Decreasing depth-of-focus in modern lithography systems increases the need for exquisite planarity and without such planarity, features to be patterned may be out of focus due to surface height fluctuations. This results in subsequent lithographic variations. However, CMP is not a variation-free process itself; it is a significant source of systematic variation resulting from both process conditions, including variations in down force, rotational speed, pad conditioning, and temperature as well as designed feature sizes and pattern dependencies. The primary effects of CMP variation are shown in Figure2.8, where copper lines can be ”dished” and inter-layer dielectrics ”eroded”, causing variation in copper line thicknesses.

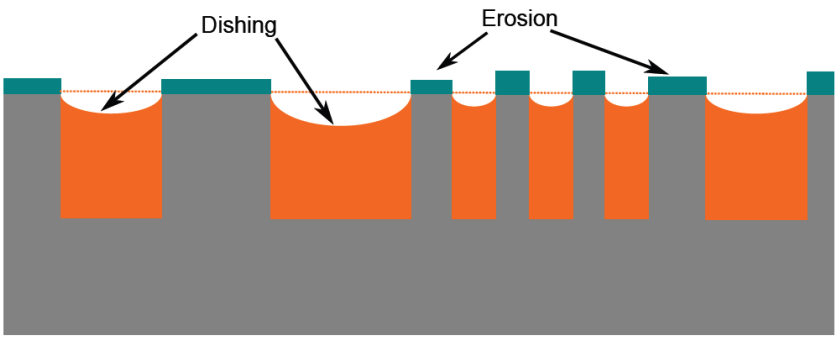


FIGURE 2.8: Dishing of metal and erosion of inter-layer dielectrics in CMP. [4]

Mitigation strategies for variation resulting from the CMP process module began with improved process control by using feedback from the process itself to guide when the polishing should end. More recently, mitigation strategies have focused on improved modeling and design modification: since variations arising from the CMP process tend to be limited to pattern dependencies and features sizes, appropriate modeling of the process and product design can reveal areas of particular susceptibility to the types of variation depicted in Figure2.8. With this information, automated mitigation strategies have been devised to enforce or adjust pattern densities (e.g., by using design rule constraints or automated "dummy fill" insertion) to dramatically reduce a design's susceptibility to CMP-caused variation [18].

2.2 Impact of Variations on Transistor Parameters

Each of the variation sources highlighted above impacts the electrical properties of transistors and interconnect. These effects are best understood in the context of transistor performance. In a typical digital integrated circuit, a transistor either charges or discharges a capacitive load, and the time required to do so determines the performance of the transistor. This time is a function of the capacitance being driven, the voltage to which it must be driven and the current used to drive it, as shown in (2.2). For simplicity, we use the ideal I-V equation for a MOSFET in the saturation regime as shown in Eq2.3, where μ is the mobility of a charge carrier through the device, C_{ox} is the gate oxide capacitance, W and L are respectively the width and length of the transistor, V_T is the device threshold voltage and V_{GS} is the bias between gate and source. Though this equation is idealized and neglects important details in modern transistors, it is sufficient to illustrate the impacts that the variation sources mentioned above have on a transistor.

$$t_d = \frac{C_{load} V_{DD}}{I} \tag{2.2}$$

$$I_D = \frac{1}{2}\mu C_{ox}\frac{W}{L}(V_{GS} - V_T)^{\alpha} \tag{2.3}$$

$$t_d = \frac{C_{load} V_{DD}}{\frac{1}{2}\mu C_{ox}\frac{W}{L}(V_{GS} - V_T)^{\alpha}} \tag{2.4}$$

MOSFET Parameter	Relevant Process Module(s)
μ	Ion implantation, annealing, diffusion
C_{ox}	Gate oxidation
W, L	Lithography, etch
V_T	Ion implantation, annealing, gate oxidation, (lithography, etch)

TABLE 2.1: Process steps effect on transistor parameters[4]

Table 2.1 shows the MOSFET parameters and relevant process modules that directly affect each of those parameters. It is clear that a single process module can affect multiple transistor parameters, and thus decoupling the effects of one variation source from another are difficult. The following sections explore variations from the perspective of the device as listed in the table.

2.2.1 Mobility (μ)

Mobility expresses how easy can charge carriers (electrons or holes) travel through the channel of a MOSFET when applying an electric field. It can be expressed mathematically as in (2.5), where q is the electronic charge, τ_c is the mean free time between carrier collisions, and $m_{n,p}$ is the effective mass of either an electron (n) or hole (p).

$$\mu_{n,p} = \frac{q\tau_c}{2m_{n,p}} \tag{2.5}$$

However, mobility is also affected by doping concentration. Figure2.9 shows mobility as a function of the doping concentration. This effect is because the doping concentration determines the mean free time between collisions. Mobility is also affected by stress effect of nitride liners and silicon germanium source/drains, in the form of either stretching or compressing the silicon lattice [5].

Any process step which affects doping concentration or stress will necessarily affects transistor mobility. Therefore, ion implantation and annealing directly affect mobility as these process steps primarily determine doping concentrations. However, as seen in Fig. 2.9, since doping concentration is on a log scale and typically does not vary

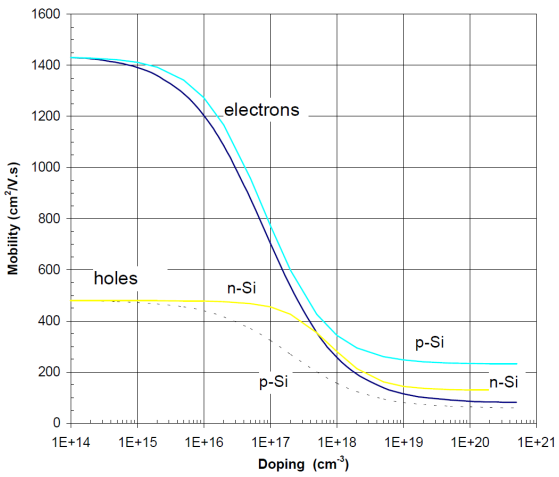


FIGURE 2.9: Mobility as function of doping in Si [5]

by orders of magnitude from one transistor to another, the impact that ion implantation and annealing have on mobility is relatively small.

Intentional and unintentional stresses, whether by stress engineering or proximity to STI, can have large impacts on transistor mobility. Mobility can have great improvements with the application of strain engineering. Even unintentional stresses due to STI proximity can cause within-die mobility variations on the order of a few percent depending on transistor distance to the STI edge. Recent characterization of mobility in advanced processes indicates relatively large variations due to fluctuations in the intentional stresses introduced in these processes [19].

2.2.2 Oxide Capacitance (C_{ox})

Gate oxide capacitance is the capacitance between the gate stack (polysilicon and silicon dioxide) and the inverted channel of the MOSFET. It is shown in (2.6) that the oxide capacitance is a function only of the oxide thickness (t_{ox}) and the dielectric constant of silicon dioxide or other gate insulator.

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} \tag{2.6}$$

Gate oxidation is, somehow, a well controlled process step. However, with gate oxide thicknesses on the order of five atomic layers ($1nm$), even small variations of one atomic layer have the potential to greatly impact, not only oxide capacitance, but also threshold voltage and mobility. With SiO_2 gate oxide, variations of a single monolayer (approximately $0.2nm$) are typical and result in 20% shifts in oxide thickness. Furthermore, though not shown in (2.6), oxide thickness exponentially impacts gate currents due to tunneling. As a result, variations in oxide thickness can have significant effects on idle leakage power. To improve this situation, Intel used a ”high-K” gate dielectric, hafnium dioxide (HfO_2), to allow for thicker gate oxides while maintaining oxide capacitance and gate control over the channel, but reducing gate leakage by three orders of magnitude. Moving to a new gate oxide material not only reduces gate leakage currents but reduces the impact of variability on C_{ox} due to the much larger physical oxide thickness. Nevertheless, variations in the inter-facial oxide of ”high-K” stacks do still occur and can affect performance.

2.2.3 Transistor Dimensions (W, L)

The previous section highlighted the importance of one of the smallest dimensions of a transistor, the gate oxide thickness. (2.3) shows that the width (W) and length (L) of a transistor are critical in determining the current through that transistor. W must be increased or L decreased to increase current and thus performance. Since decreasing L also reduces load capacitances and increases transistor density, scaling has continually reduced L in the pursuit of increased performance, making it the most critical dimension in a transistor today.

Lithographic patterning and etching define both of these dimensions. However, since W is always larger than L, only variation in channel length is typically of concern. (2.4) shows that the delay of a transistor is directly proportional to the channel length, so any variation in channel length will be directly reflected in transistor delay. As transistor lengths have decreased well below the wavelength of light patterning them, relative variation in channel length has increased which corresponds to an increase in performance variability. Another effect of variation in channel length is variation in threshold voltage

variations due to the Drain-Induced Barrier Lowering (DIBL) phenomenon, which have a more dramatic effect as channel lengths scale down.

2.2.4 Threshold Voltage (V_T)

The threshold voltage of a MOSFET is the gate-to-source bias (V_{GS}) that results in a channel forming just under the gate, allowing current conduction from source to drain of the transistor. In an ideal long-channel MOSFET, the threshold voltage is determined by only the doping concentration (N_{ch}) and the oxide capacitance (C_{ox}) as shown in (2.7), where V_{FB}, the flatband voltage, and ϕ_{F_p} , the Fermi Potential of the substrate, are dependent only on the doping concentration, and γ is dependent on both the doping concentration and oxide capacitance. In short-channel devices, effects such as DIBL result in V_T being additionally dependent on the channel length (L), source/drain junction depths (x_j), and stresses, meaning that a large fraction of process steps can potentially affect the value of V_T.

$$V_{To} = V_{FB} + 2\phi_{F_p} + \gamma\sqrt{2\phi_{F_p} + V_{SB}} \tag{2.7}$$

Owing to this factors and the intrinsic random variability of RDF, V_T is one of the least controlled transistor parameters. As scaling has led to smaller and smaller device dimensions, control of V_T has gotten progressively worse.

Looking at (2.4), the impact that variation in V_T has on performance can be substantial due to the quadratic dependence on V_T. In many cases, the gate-overdrive, $V_{GS} - V_T$, is large enough that V_T variations can be mitigated, but in low-power designs where V_{DD}, and thus V_{GS}, are not much greater than V_T , the impact is large. Moreover, device leakage currents are exponentially dependent on V_T , so V_T variation has considerable impact on idle power.

2.3 Variation Mitigation

Mitigation of process variation can be divided into process-oriented versus design-oriented approaches. Targeting the process itself involves altering process modules and/or flows or device design, directly impacting variations source. Design-oriented strategies like design rules, layout modifications, or robust circuit and system design are an indirect route to variation mitigation. The following section details some of the significant mitigation strategies and related work in those areas for each level of the hierarchy.

2.3.1 Architecture level

Architectures and systems are often too large and complicated to apply global techniques to mitigate variation. Often the most susceptible building blocks are the ones to which mitigation techniques are applied. For example, the susceptibility of SRAM caches to variation, has led to the development of ”variation-aware” caches which dynamically resize as faulty bitcells are detected or where whole cache lines are replaced with redundant lines [20].

Variation-aware synthesis techniques have also enabled mitigation in on-chip bus architectures for systems-on-a-chip, resulting in decreased power consumption as well[21]. Process variation has, however, reached the point at which system architects can no longer avoid considering its impact. Some researchers considered including process variability in the context of heterogeneous blocks in an embedded application that together must meet some latency constraint. Each block is a voltage/frequency island and the optimal voltages and frequencies for islands are solved for [22]. Others evaluated Adaptive-Voltage Supply (AVS) and Adaptive Body-Biasing (ABB) as potential circuit-level solutions in the context of core-to-core frequency variations in multi-core processors, but concluded that both of these techniques involve tradeoffs between static and dynamic power, making them less desirable potential solutions [23].

Similarly, an adaptive FPGA architecture utilizing ABB is proposed in leading to 3.45X reduction in timing variability and 3X reduction in leakage power. In [24] core-to-core variation in frequency is explored, and allowing the system to turn off cores if the additional power consumed by the core is higher than a certain metric is proposed.

2.3.2 Circuit level

Digital circuits and systems, due to both their bi-stable nature and high gain in the switching region which results in relatively large noise margins, are inherently more robust to variation. Analog circuits, however, are considerably more sensitive to process variation and have employed both layout and design-oriented mitigation strategies, including common-centroid layout, use of larger than minimum-size devices and feedback for many years to cope with process variation. Among the first circuits to employ design-oriented mitigation was the PLL, which made use of feedback to decrease sensitivity to component values.

PLLs are not limited to analog systems though; they are critical components of high-performance digital systems, providing the clock(s) necessary for timing of digital circuits. Being among the most sensitive and critical signals in digital systems, clocks and clocking networks were subject to variation mitigation strategies long before other portions of digital systems. Attempts to reduce clock skew and jitter have produced a myriad number of mitigation schemes such as H-Tree, X-Tree and grid clock distribution, active deskewing [25], and even some proposed exotic solutions such as use of optical clock networks . Also a subject of much research in variation mitigation are SRAM circuits in cache memories, which are particularly susceptible to variation due to their small physical size, leading to SRAM cell and sense amplifier redesign as well as other so-called ”peripheral assists” to mitigate the impacts of variation [26].

With relative process tolerances continuing to degrade, digital paths have become the focus of multiple variation mitigation techniques. Some of these include replica critical path monitor circuits around which adaptive feedback schemes can be built, such

as Adaptive Body Biasing (ABB), Adaptive Voltage Scaling, and detection and correction of timing errors within microprocessor pipeline stages [27]. Shown in Figure2.10 a shadow latch controlled by a delayed clock, augments each flip-flop in the design. In a given clock cycle, if the combinational logic, stage L1, meets the setup time for the main flip-flop for the clock's rising edge, then both the main flip-flop and the shadow latch will latch the correct data. In this case, the error signal at the XOR gate's output remains low, leaving the pipeline's operation unaltered. If combinational logic L1 doesn't complete its computation in time, the main flip-flop will latch an incorrect value, while the shadow latch will latch the late-arriving correct value. The error signal would then go high, prompting restoration of the correct value from the shadow latch into the main flip-flop, and the correct value becomes available to stage L2 [6].

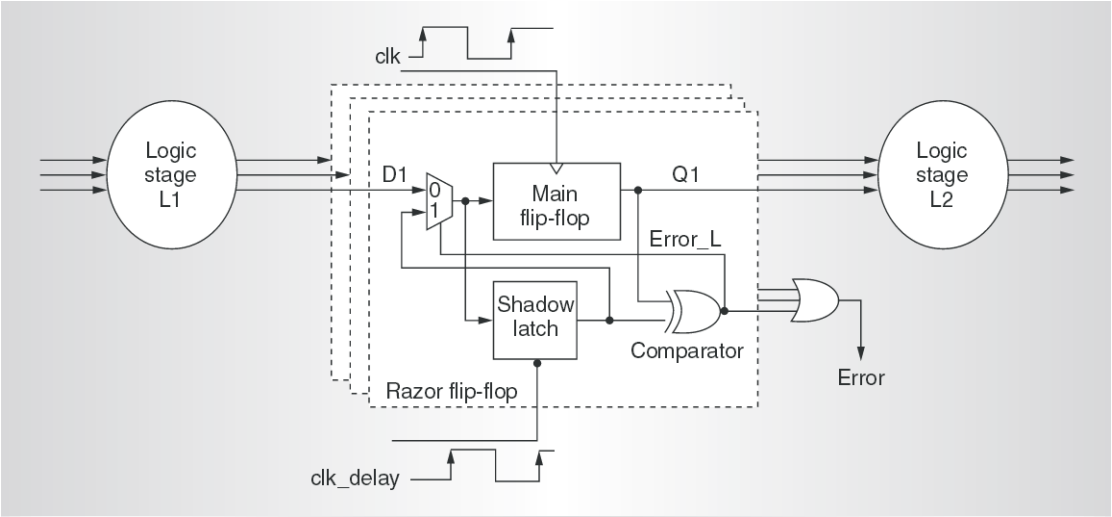


FIGURE 2.10: Pipeline stage augmented with Razor latches and control lines[6]

Portable electronics and sensors have motivated the need for low energy operation in the sub-threshold operating regime, in which transistor leakage currents are exponentially affected by variations in VT and which poses unique, but surmountable challenges to circuit design. With appropriate gate library modeling and design and variation-aware timing methodologies, sub-threshold circuits can be designed to be as robust as their above-threshold counterparts [28]. Understanding of spatial components of variation have also guided some mitigation work. Such understanding can provide improvement in digital circuit delay variability if appropriate process control

mechanisms are put in place to take advantage of spatial correlation data. Furthermore, statistical static timing, in which timing of digital paths is described by statistical distributions, has garnered attention in aiding designers assess the susceptibility of critical paths to variation and make appropriate changes to ensure high yield.

2.3.3 Device level

At the device level, it is more relevant to discuss process-oriented variation reduction approaches. Statistical process control aims to reduce ”back-end” testing and assembly costs in IC fabrication lines by increasing yields in the ”front-end” steps. Measurements coupled with inferential signals from individual process modules are utilized to dynamically change process conditions and recipes to ultimately benefit yield. Statistical process control has pervaded nearly every process module, including oxidation [29], photoresist application [30], etch [31], and CMP processes [18].

In addition to feed-back/forward process control, continuous improvements have been made to process steps and in the use of materials to improve both device performance and variability. The resolution enhancement techniques described before, immersion lithography, other patterning improvements and introduction of high-k/metal-gates to high-volume CMOS manufacturing [7], all provide examples of techniques meant to improve manufacturability of high-performance transistors as shown in Figure2.11.

Common to modern semiconductor manufacturing are combination ”process-design” mitigation techniques, where process and design teams actively engage each other to trade-off ease-of design, manufacturability and performance. This process is best implemented where design and manufacturing occur in the same organization, but Design For Manufacturability (DFM) kits are increasingly made available by foundries as well.

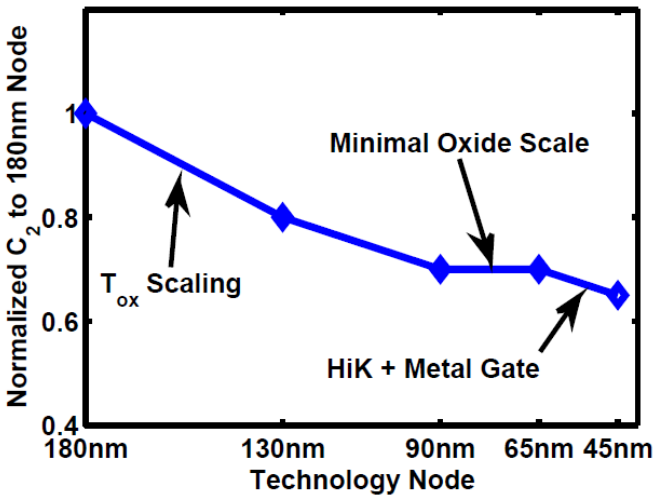


FIGURE 2.11: Improvement in process due to oxide scaling and new materials [7]

2.3.4 Design For Manufacturability

DFM main goal is to bridge the gap between design techniques and manufacturability needs. This can be accomplished by both minimizing process variations, and minimizing device and circuit sensitivities to process variations. Many of the rules advocated in the DFM methods include measures to mitigate pattern dependent variation, such as reduction of geometric primitives, layout regularity, redundancy techniques, pattern density uniformity, and optimal process feature sizing. Other areas of consideration in DFM are design for robustness and design for test. These suggested approaches are detailed below.

Reduction of Geometric Primitives

A geometric primitive refers to the category of shape in a design. The rectangle is the most basic primitive in circuit layout. With each new geometric primitive comes a new set of uncertainty in processing. Primitives process differently, whether in lithography, etch, plating, or through other steps in the process cycle. As shown in Figure2.12 Resolution Enhancement Techniques (RET), must be applied to compensate for the distortions that new primitives may introduce. These techniques are costly and often

insufficient for arbitrary random layout patterns. Reducing the number of these primitive shapes in a design ensures higher certainty during the manufacturing stages.

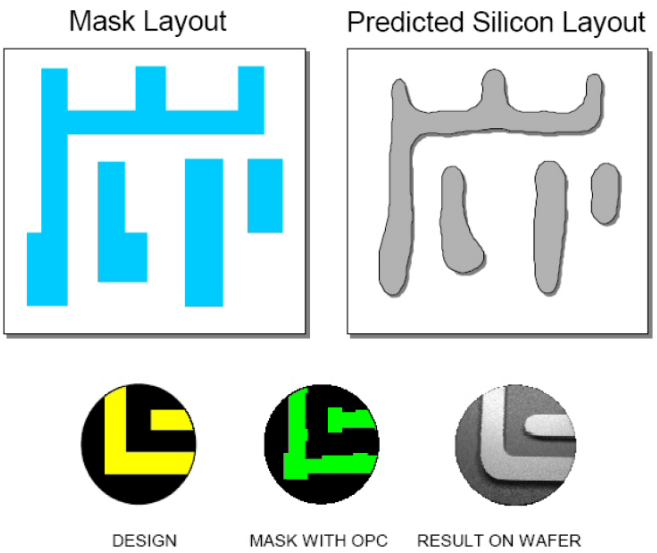


FIGURE 2.12: Increasing geometrical primitives increases uncertainty in final design [8]

Redundancy to Improve Manufacturability

To improve planarity of polysilicon or metal layers in a design, metal "dummy" fill is inserted throughout each layer. Dummy fill consists of redundant pieces of metal placed in empty areas of a design with automated CAD tools to increase planarity. At the cost of increased capacitance, this fill narrows the layer's range of pattern density and significantly reduces CMP variation. Inserting dummy cells and dummy transistors are also methods for reducing edge effects for transistors, memories, and other blocks in a design. Dummy transistors insertion in analog layout is shown in Figure2.13. Finally, redundant via placement may improve manufacturability by decreasing the likelihood that a single via causes logical failure in a design due to bad processing [32]. Combined, these techniques are a wise combination to reduce the variation effects of manufacturing and improve overall yield in an IC process.

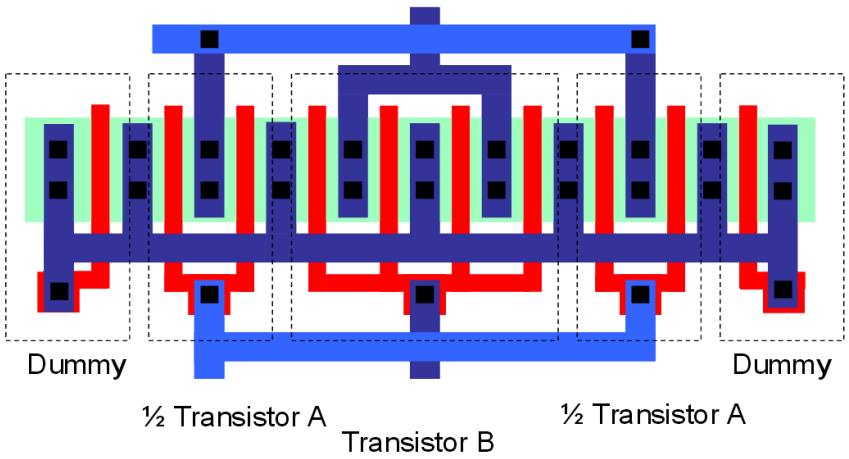


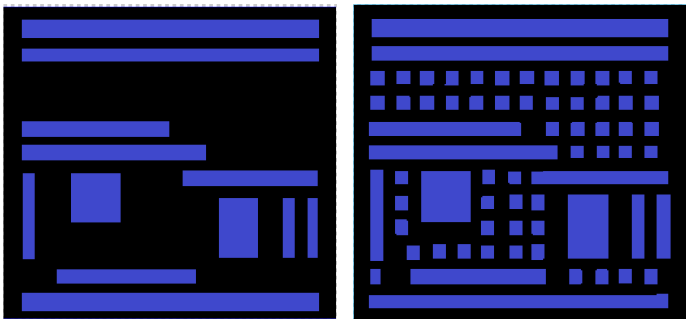
FIGURE 2.13: Inter-digitized transistors with dummy transistor insertion at sides

Pattern Density Uniformity

One main source of systematic process variability mentioned before in section2.1.5 is the irregularity in surface topography due to an unevenly distributed pattern density [33]. As shown before in Figure2.8, the high pattern density region on the rigth will cause successive steps to have higher topography, while the low density region to the left will cause successive steps to be lower.

Fluctuations in local metal thickness can affect the resistance and capacitance of the region, adding unexpected loads to wires, for example. Clock skew resulting from mismatch in the affected interconnect is one important effect of non-uniform pattern density. Electromigration also becomes an increased concern with sub-optimal nonregular metal thicknesses [34]. In the worst case, logical failure from the improper plating and polish of the design.

Fortunately, dummy fill algorithms have the ability to compensate for the nonuniformity of most designs as shown in Figure2.14. Although dummy fill can decrease macro-level fluctuations in pattern density uniformity, it is not taht preferable becuase it introduces significant capacitance to a design, which may be a high cost for high performance circuits. Therefore, a designer must still consider the most regular approach to design and layout with consideration of density in mind from the beginning.



(a) (b)

FIGURE 2.14: Metal layer before and after dummy fill insertion

Layout Regularity

Design performance is affected by the methodology and technique used during transistor layout [35]. Examples of the layout techniques that may affect performance in a design are number of fingers per transistor, transistor orientation, proximity spacing, local and global effect of polysilicon density, interconnect orientation, choice of metal layer, and coupling capacitance across metal layers.

2.4 Summary and Conclusion

This chapter covered the basic manufacturing steps and their contribution to variations, then it showed the impact of variations on transistor parameters and in the last section it covered some of the mitigation techniques starting from architecture level to device level and talked about design for manufacturability efforts to mitigate variations and improve yield.

As shown in the last section layout regularity is one of the suggested DFM techniques to mitigate variations. Different layout techniques have their effects on different phenomena like delay, leakage, electromigration and stress. Industry started heading towards regular solutions for design to mitigate variations and enable scaling for the next technology nodes. Since measuring the layout regularity is the main aim of this research, the next chapter will detail the effect of layout topology on different phenomena and what were the solutions suggested, then it will show how researchers started

recommending using more regular solutions for designs and how this raised the need for a regularity metric.

Chapter 3

State of the Art Layout Regularity Techniques

The fact that variations sensitivity is pattern dependent, which means that some layout patterns exhibit less sensitivity to variations than others, raised the interest in studying the relation between the layout topology and variations sensitivity in different process steps and for different purposes.

3.1 Effect of Layout Topologies

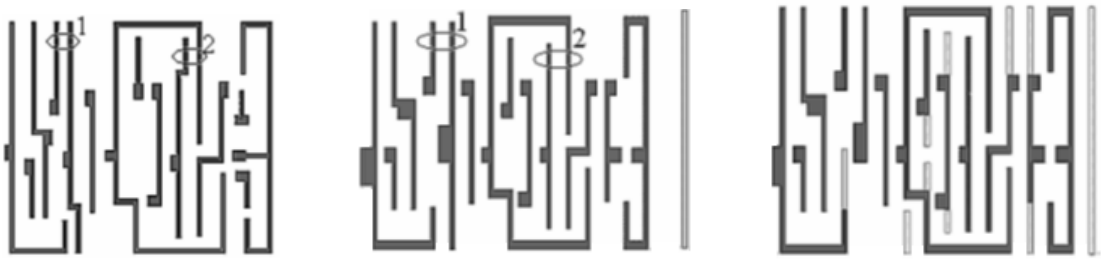
Layout topology refers to how the geometries in the layout of different layers look like and how they are arranged and connected. It was found that different layout topologies result in different effects on delay, leakage, electromigration and stress. This section will show how researchers were able to find certain topologies that can decrease variations and how these topologies led, later, to the adoption of regular layouts.

3.1.1 Delay

Delay is one of the most important electrical parameters that is taken into consideration at each design step. This part is going to show how researchers came out with certain recommended shapes in poly, diffusion, interconnects and contacts that can either decrease delay or decrease variations in delay.

3.1.1.1 Poly Pitch and Orientation

Garg et al. [36] studied gate length variability due to using high poly pitch and large range of poly pitches in the layout. The main concern was about delay spread so a the effect of parameters causing process variability in CMOS devices (gate insulator thickness, gate length, gate width and threshold voltage) on delay spread were compared using Monte Carlo simulation and it was found that gate length is the most dominant. The solution suggested was using Contacted gate pitch, dummy poly insertion, using one orientation for poly gates and multiple orientations for poly interconnects and making interconnects wider than minimum width.



(a) (b) (c)

FIGURE 3.1: Poly layouts of master slave flip-flop. a) Conventional layout. b) Litho-driven layout, Poly lines restricted at contacted gate pitch. c) Layout b with dummy Poly lines inserted at the missing CGP grid lines.

3.1.1.2 Poly and Diffusion Density

Rapid Thermal Annealing (RTA) makes a significant contribution to manufacturing process variations, degrading the parametric yield [37]. A lot of research was carried out

to measure the effect of this variability on delay. Joshi et al. [38] made experimental measures on 45nm experimental test chip that showed the effect of anneal temperature variation on device performance.

In attempt to find what affects RTA variations, people used models to measure the reflectivity of wafer surface with different densities for poly and diffusion, results showed that doped poly-silicon R (and hence inverter delay) correlated well with the pattern density of the exposed STI layer [39], [40].

Solutions proposed in this field were mainly targeting evening out the density of shallow trench isolation (STI) and insertion of dummy fills, Wei et al. [37] solved a floorplanning problem to reduce the RTA variations by evening out the STI density distribution, they also inserted dummy polysilicon fills to enhance the uniformity.

3.1.1.3 Contacts and Interconnects Shapes

Ban et al. [32] proposed a new equivalent contact resistance model which accurately calculates contact resistances from contact area, contact position, and contact shape. Based on the impact of contact resistance on the saturation current, they performed robust S/D contact layout optimization by minimizing the lithography variation as well as by maximizing the saturation current without leakage penalty. The results on 32nm node standard cells show delay improvement under nominal process condition and reduction in the delay variations between the fastest and slowest process corners.

Plasma etching of interlayer dielectrics (ILD) for metal contacts, vias, and lines contributes to circuit performance variations. These performance variations can appear as varying resistance and capacitance and this may lead to circuit signal delays and skews. In [17] the sources of etch rate non-uniformity were investigated, and a methodology for effectively capturing etch rate variation at the chip scale based on layout pattern dependencies was proposed.

Chemical-mechanical planarization CMP is used to planarize bare wafers, in FEOL to remove and planarize overburden oxide, and in BEOL to remove excess copper and barrier, and to planarize inter-level dielectric. While several advancements have been

made in CMP technology, imperfections remain and have always been a concern due to rapidly shrinking topography variation tolerances. CMP is known to suffer from pattern-dependent problems known as dishing and erosion [169]. These two effects arise because of the existence of multiple materials of different softness that get polished simultaneously. Dishing quantifies the height difference seen in one material, while erosion captures the height loss of the harder material while polishing. Two methods to reduce pattern-dependent effects are filling and slotting In fill insertion, non-functional or dummy geometries are added to increase the density of a material. A common objective is to make the material density over the chip uniform by adding fill to regions that have less material. Slotting works in the opposite way by removing material from large features without compromising their electrical functionality. CMP imperfections manifest themselves into electrical variations in several ways. In FEOL, oxide dishing in STI wells and nitride erosion can cause poor isolation between devices and increase inter-device parasitics. Excessive nitride erosion into the underlying silicon, and failure to completely remove oxide from over the nitride can cause device failure. In BEOL, copper dishing and dielectric erosion affect the interconnect resistance and capacitance, and consequently the interconnect delay. Poor planarity also poses difficulty in patterning the layers above and can cause large defocus during exposure. Planarization non-idealities also compound for higher metal layers due to the non-planarity of the underlying layer [2].

3.1.2 Leakage

Due to lithography process, the gate length of the device is distorted at the gate edge and the end of the gate which is referred to as non-rectilinear gate (NRG), as illustrated in Figure 3.2 . The NRG may increase sub-threshold leakage by more than 15X from that of an ideal layout. The shortening of poly line end may even cause device failure. Regular Layout that follows single pitch and single orientation improves design predictability. Limiting the layout to regular pattern effectively enhances yield and minimizes the variations of gate length [41].

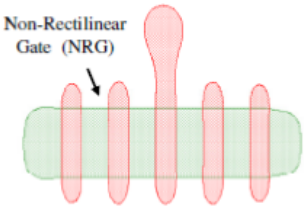


FIGURE 3.2: Post lithography aerial image of poly gate and diffusion showing NRG and line end shortening

The amount of heat transferred, and hence the local anneal temperature, is affected by the layout pattern dependence of optical properties in a region. This variation in local anneal temperature causes a variation in performance and leakage across the chip by affecting the threshold voltage (Vth) and extrinsic transistor resistance (Rext). Experimental results based on a 45nm experimental test chip show anneal temperature variation of up to 10.5^oC results in 2.45X variation in device leakage across the chip [38].

3.1.3 Electromigration

Electromigration is increasingly relevant to the physical design of electronic circuits. It is caused by excessive current density stress in interconnects. The ongoing reduction of circuit feature sizes has aggravated this problem. It is therefore an important reliability issue to consider electromigration-related design parameters during physical design. A simple efficient methodology to predict the probability of narrow interconnects in any layout given some basic information such as interconnect width, spacing, and density is proposed in [34]. Subsequently, the probability of chip's failure due to narrowing defects can be predicted. In [42] Lienig presented various physical design constraints that affect electromigration, then he introduced components of an electromigration-aware physical design flow. In addition to the regular design steps, this flow contains three current-density-driven design and verification tools which allow an effective consideration of electromigration-related constraints during physical design.

Narrow interconnects are susceptible to electromigration that may lead to interconnect open or short defects and consequently a chip failure. It is expected that narrowing defects will present a serious challenge for IC reliability with decreasing feature size.

Intentional and unintentional stresses, whether by stress engineering or proximity to STI, can have large impacts on transistor mobility. Mobility improvements greater than 10% over unstrained silicon have been reported as strain engineering has matured. Even unintentional stresses due to STI proximity can cause within-die mobility variations on the order of a few percent depending on transistor distance to the STI edge [4].

3.1.4 Stress

A compact stress model that physically captures stress distribution in the channel is essential to bridge the process technology with design optimization. In [43] starting from the first principle, a new layout-dependent stress model is proposed as a function of layout, temperature, and other device parameters. Furthermore, a method of layout decomposition is developed to partition the layout into a set of simple patterns for efficient model extraction.

3.2 Regular Solutions

As clear from previous sections, using certain topologies , which are actually more regular shapes in terms of density, pitch and orientation, will result in better designs in terms of delay, leakage, electromigration and stress. Furthermore, more regular layouts with higher degree of repetition will save some of the time spent in applying RET techniques and post-layout verification and decrease the un-expected non-linear effects due to un-deterministic neighborhoods. The following part will show how technology headed towards a more regular solution for layouts for the aforementioned reasons and what are the challenges faced by these solutions.

3.2.1 Restricted Design Rules

The computation cost and complexity of RETs techniques are becoming bottlenecks in the Design to Silicon flow. This has motivated the recent calls for restrictive design rules such as fixed width/pitch/orientation of gate-forming polysilicon features. Lavin et al. in [44] explored how design might take advantage of these restrictions, and presented some preliminary ideas for how to reduce the computational cost throughout the back end of the design flow through the post-tapeout data processes while improving quality of results: the reliability of OPC/RET algorithms and the accuracy of models of manufactured products. This methodology was called L3GO (Layout using Gridded Glyph Geometry Objects). Their approach has been to extend the design restrictions on polysilicon features mentioned above to all layers of a design, and then tailor the layout flow and tools to take advantage of these restrictions. L3GO layouts have conventional overall structures (cells and layers) shapes are defined by a coarse grid, typically a simple fraction of the minimum manufacturable feature pitch, and three types of glyphs, shown in Figure 3.3

- Point glyphs: 0-dimensional points lying at grid points, typically used for vertical interconnections (contacts and vias).
- Stick glyphs: 1-dimensional line segments drawn between two grid points, typically used for FET gates or for interconnections.
- Rectangle glyphs: 2-dimensional, axis aligned rectangles whose vertices are on grid points, typically used for diffusion regions.

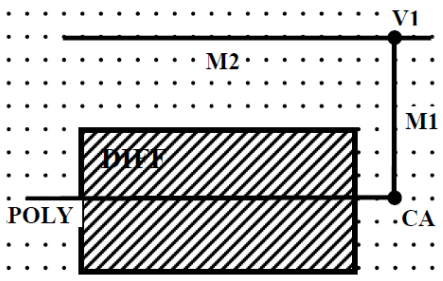


FIGURE 3.3: L3GO Grid and Glyphs

The advantage of glyph is that they have grid-base nature and they can also have attributes to describe circuit identity.

Layout creation restricts the inputs to lie on layout grid. For verification, L3GO layouts are subject to a few dozen design rules compared to hundreds of design rules in conventional layout. These rules can be done directly on the L3GO data model to take advantage of the fact that the L3GO rules can be easily specified and checked in terms of allowed and forbidden patterns.

3.2.2 Regular Fabrics

Enforcing shape-level regularity in ordinary logic has been challenging since generic logic tends to be irregular, and the added constraints to produce physical regularity can result in substantial area-delay penalty.

Several regular layout fabrics have been proposed with varying degrees of performance overhead and flexibility. FPGAs are very flexible but suffer from huge area, power and performance overheads. Another example of regular fabrics is Via Patterned Gate array; VPGA [45] which is formed by regular geometry, logic and routing layer structures that are customized for an application using via layers. Via programmable gate arrays offer programmability of logic as well as interconnect using vias and contacts. A recent example of VPGA is Via Configurable Transistor Array (VCTA) which is a regular fabric that is based on a single basic cell which is repeated along the circuit as shown in Fig 3.4. The basic cell contains a transistor array and an interconnection grid already in place to maximize layout regularity. The transistor array is composed of a number T of PMOS and T NMOS transistors. The T transistors in each case are connected in series by default. However, parallel connections can be implemented by properly setting up vias [46]. VPGA offer performance, power and area closer to ASICs as they do not have complex SRAM based programmable logic as FPGAs.

Generally, homogeneous logic elements that are configurable to perform various logic functions tend to be underutilized [47], while logic elements in standard cells libraries are too fine-grained to be efficient building blocks of regular ICs. Somewhat

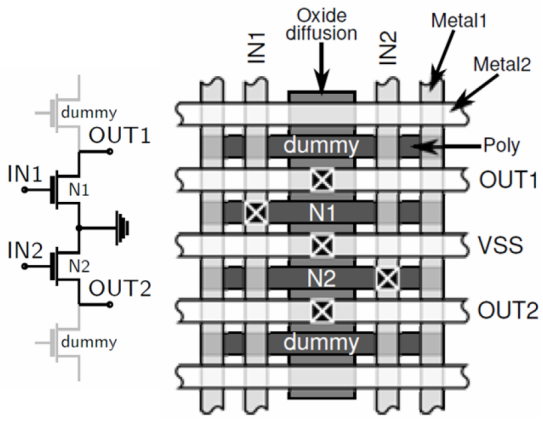


FIGURE 3.4: VCTA schematic and layout for the NMOS transistor array with T=2

less restrictive regularity can be achieved by more manufacturable cell libraries with regular structures which will require supportive placement techniques.

In [48] a method to generate coarse-grained logic building blocks was investigated. An important concern was that these building blocks should produce efficient implementations while facilitating various manufacturability benefits afforded by regular ICs. What is called a ”brick generator” was introduced. Its function was that it produces regular logic bricks to be used instead of standard cells was presented. The design flow as shown in Figure 3.5 directly maps a RTL netlist into logic bricks to effectively utilize the efficient circuit level implementation and the layout compactness of logic bricks.

The boundaries of logic bricks contain similar logic primitives to provide a common geometrical interface between any two bricks. In terms of logic granularity, bricks are more coarse grained than simple standard cells as the logic function performed by each brick is more complex than a standard cell.

Using logic bricks, robust IC designs that are based on a small number of RET-friendly regular geometry patterns can be constructed.

3.3 Regularity Metrics

There are two types of regularity; micro regularity and macro regularity. Micro-regularity identifies a small set of easy-to-print patterns, and the macro-regularity provides known

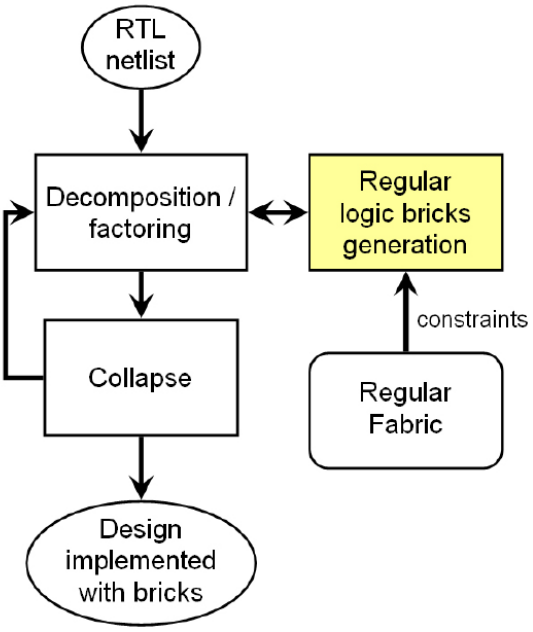


FIGURE 3.5: Overall design flow for Regular Logic Bricks methodology

neighborhoods for effective OPC applications and predictable performance characterization [13]. To evaluate and compare the degree of regularity of designs, researchers started developing methods to quantify regularity and come out with a metric that can help designers to compare between different designs, as the work done in [49] and [50] which will be discussed in details below.

3.3.1 Spatial Fourier Transform

In [49] the method was to use two-dimensional Fourier transform to compare between three different layout styles. The comparison was based on analyzing the number of dominant frequency components in the Fourier transform of each layout style. The regular layout that uses a small number of layout patterns placed at a fixed pitch is expected to have a high degree of repetition and thus, a finite number of dominant frequency components. The results of Fourier transform analysis on a 50 μm x 50 μm slice of poly-silicon and contact layers of an SRAM array , random logic implemented using standard cells, and logic implemented using bricks on a regular design fabric are shown in Figures 3.6, 3.7 and 3.8. Modern SRAM bitcell contains unidirectional

poly on a fixed pitch, so it is expected to find a single dominating frequency, the plot in Figure3.6 shows a dominating frequency component. The other peaks, seen at a multiple of the dominating frequency component are just harmonics that result from using Fourier analysis. Also observable from the plot are the nonzero frequencies in the perpendicular orientation, which show the periodicity of the bitcell in the SRAM array. While the Fourier transform of the contacts in the SRAM shows a much greater number of frequency components due to nontypical contact shapes in the bitcell layout, it is still possible to identify the few dominating frequencies that are present. Similar analysis of a standard cell design in Figure 3.7 shows that the number of patterns and their placements are not limited, resulting in a large number of frequency components in the Fourier plot. However, the Fourier transform plots of layout implemented using logic bricks on a regular design fabric shown in Figure 3.8 seem to have few dominant frequency components like that of the SRAM.

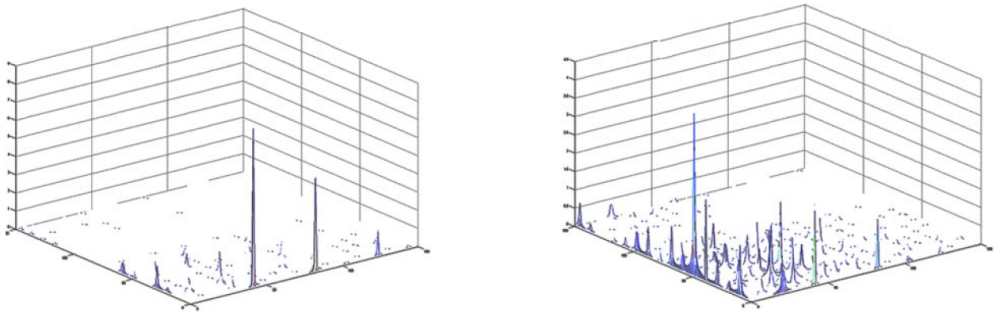


FIGURE 3.6: SRAM Fourier transform of poly (on the left) and contacts (on the rigth)

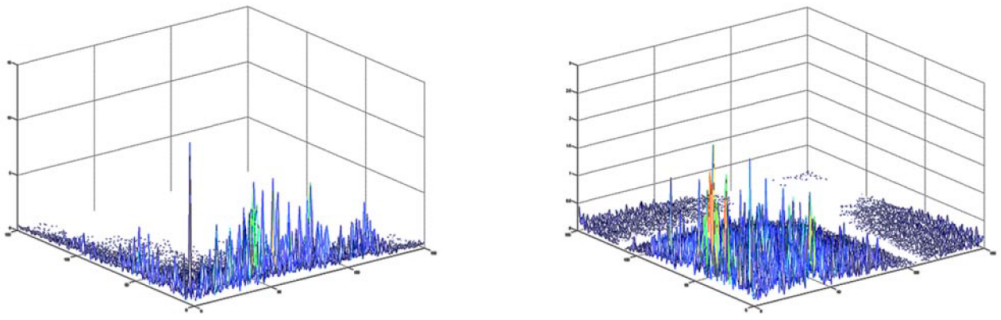


FIGURE 3.7: Standard Cell Fourier transform of poly (on the left) and contacts (on the rigth)

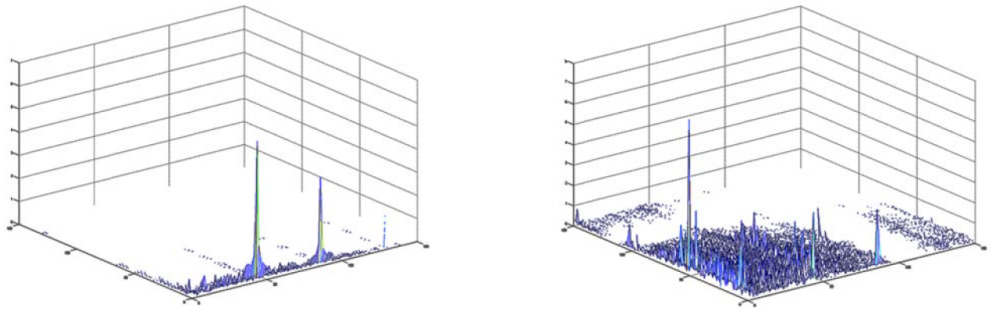


FIGURE 3.8: Regular fabrics Fourier transform of poly (on the left) and contacts (on the rigth)

3.3.2 Fixed Origin Corner Square Inspection (FOCSI)

More recent work related to regularity metric was done in [50] on 65 nm tecnology node. Layout regularity was defined, for a certain layout layer, as the ability to generate this layer by a reduced number of ”layout areas”. Layout areas (LAs) are squares of a given shape and size (e.g., squares of 160 nm x 160 nm)for 65 nm. Different types of LAs in the layer were called layout generators. Therefore, the lower the number of generators found, the higher the regularity is. According to this definition, the maximum regularity is achieved when a single generator can be used to generate the whole layer. On the other hand, the minimum regularity occurs when all LAs are unique, and therefore, there is no repetition at all, i.e: each LAs is generators. To create layout generator for certain layer a first approach to divide the complete layout in contiguous squares and then to compare each one to all the others noting the number of different ones. This option was found to be inadequate because, as shown in Figure 3.9 , even if the layer was composed by only one layout pattern, low repetition and a high number of generators were observed, and thus very low regularity.

The Fixed Origin Corner Square Inspection (FOCSI) the layout layer in order to detect all upper left pattern corners and then considered these corners as the origins of the square LAs to be compared. This way FOCSI improved how regularity was captured Figure 3.9 shows how FOCSI worked. Black crosses indicate all the corners considered. In this case, types 1 and 2 generators can be detected. Note that in this figure different LA sizes are also illustrated with red and blue squares. Once the

corners are fixed, various sizings can be applied to squares in order to evaluate different granularities of regularity. Implementation was done by exporting the layout layer as an image and treating each pixel as a sample and codifying each sample. Codification assigned a 0 value to the sample when the layout is empty and a 1 value to represent the layer material. The whole layout layer is codified as a matrix of 0's and 1's. Then, all upper left pattern corners are identified and LAs were created. Finally, LAs were compared sample by sample against each other in order to calculate the number of different generators. The result of this step of FOCSI metric is the number of different generators of the layout layer under study refered to as Rlayer. The lower Rlayer is the higher regularity is for this layer.

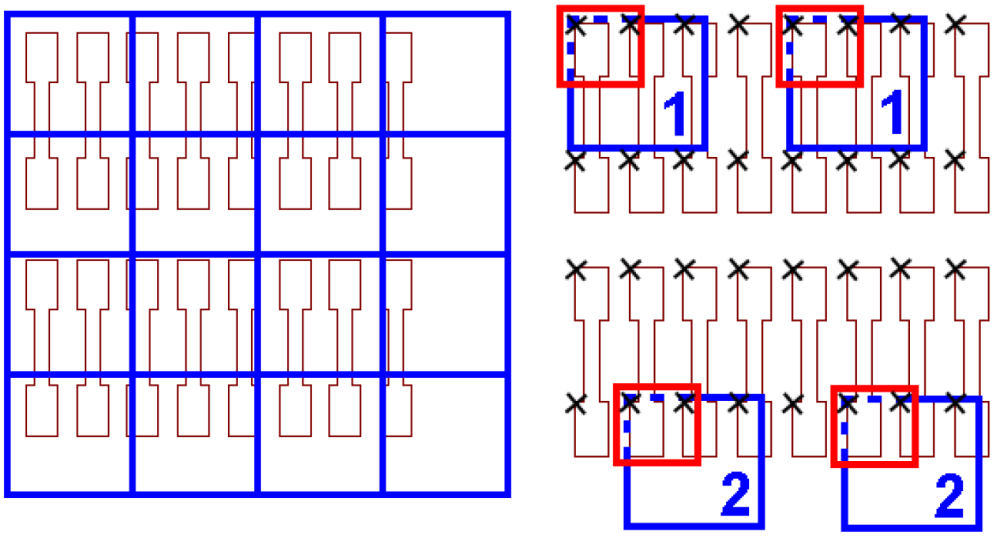


FIGURE 3.9: Creating layout areas: contiguous squares (on the left) and FOCSI (on the rigth)

3.4 Summary and Conclusion

This chapter started with showing different efforts in relating layout topology to delay, leakage, electromigration and stress and how this motivated heading toward more regular design solutions. Then it showed two of the main regular design solutions in literature. The existence of regular design solutions raised the need to a regularity metric, the metrics existing in literature were shown.

In conclusion, the pros and cons of the metrics existing in literature can be summarized as follows. The two-dimensional Fourier transform can provide a visual comparison of regularity but it does not give enough information to be able to compare accurately two layouts of somehow similar regularity. Fourier transform does not quantify regularity. It is a graphical representation giving an intuitive and qualitative measure of regularity. It can be used to compare regular versus non-regular layouts but it is difficult to use it to compare similar layouts in terms of regularity, like, for instance, two layouts developed with regular design techniques. Also, it can be noticed that Fourier transform is somehow computationally intensive and expected to have long runtime so will not be suitable for a production tool. FOCSI method seems to be better than Fourier Transform in the sense that it can provide more quantitative measure and can compare the regularity of two layouts created using the same technique. However, converting the layout to image and comparing it pixel by pixel is definitely compute intensive and will require long runtime too. Furthermore, the existing metrics are more concerned with macro-regularity so their output can't help in spotting certain irregular areas of a given layer that need improvement. The conclusion is that a metric that is more concerned about micro-regularity is needed. Such a metric will be important during both layout design and post tape-out verification to identify areas that needs further modification to improve their regularity. The work done to derive such a metric will be shown in Chapter 4.

Chapter 4

Layout Regularity Metric

Previous chapters showed the effects of variations and methods of mitigation of these effects. They also showed the work done towards creating more regular design and how this raised the need of a regularity metric.

This chapter will show the work done in deducing a new fast layout regularity metric. It will, first, show a preliminary metric that was inspired from image processing and Fourier transform and how this metric was found to have several challenges and problems that directed the decision to look for a more geometrical based metric.

4.1 Preliminary Exploration for Developing a Metric

At this part of research, the target was to look for a methodology that can differentiate between regular and irregular patterns in a quantitative way. One of the main problems of the two-dimensional Fourier transform method mentioned before, was that it was more of a qualitative measure. This part uses the same basic concepts of the Fourier transform method with some modifications to give a quantitative metric. The idea of the method was from [9], a paper from a slightly different domain that is concerned with regularity of texture and recognition of defects. The decision was to check the applicability of this method in VLSI layout domain. Below is a brief of the steps of the used algorithm.

- Convert the layout patterns into monochrome images and get their intensity function I(x,y).
- Calculate the autocorrelation function of I in both x and y directions and normalize it.
- Calculate the $Interaction\_map$ where:

$$Interaction\_map = 1 - Autocorrelation(I) \tag{4.1}$$

- Calculate the $Contrast\_curve$ in the direction in which regularity is to be calculated. Where the $Contrast\_curve$ in x direction is the $Interaction\_map$ for certain y, examples of $Contrast\_curves$ are shown in Figure 4.1

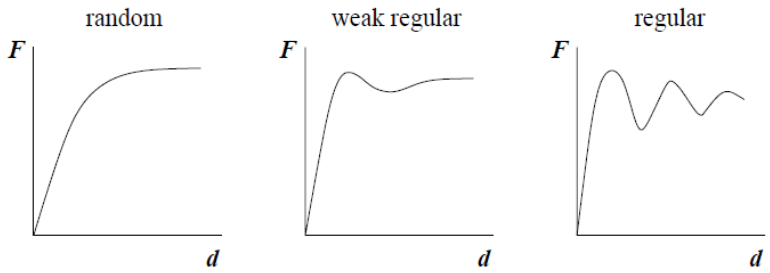


FIGURE 4.1: Contrast curves of a random, a weak regular and a regular pattern [9]

- Use the periodicity in the $Contrast\_curve$ as shown in Figure 4.2 to calculate intensity regularity (R_{int}) and positional regularity (R_{pos}) and directional regularity (R) Where d_1 and d_2 are the two lowest minima, F_1 is the value of contrast curve

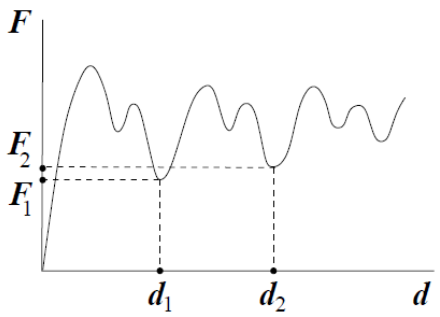


FIGURE 4.2: Computing regularity for a certain contrast curve [9]

at lowest minima (d_1).

$$R_{int} = 1 - F_1 \tag{4.2}$$

$$R_{pos} = 1 - \frac{|d_2 - 2d_1|}{d_2} \tag{4.3}$$

$$R = \sqrt{(R_{int}.R_{pos})^2} \tag{4.4}$$

To test the validity of this approach, about 30 patterns were created with different degrees of visual regularity. Shown below in Table 4.1 and in Figure4.3 some selected patterns patterns and their regularity values.

Patterns	Pattern 1	Pattern 2	Pattern 3
Regularity value	0.928	0.851	0.652

TABLE 4.1: Regularity values for some patterns

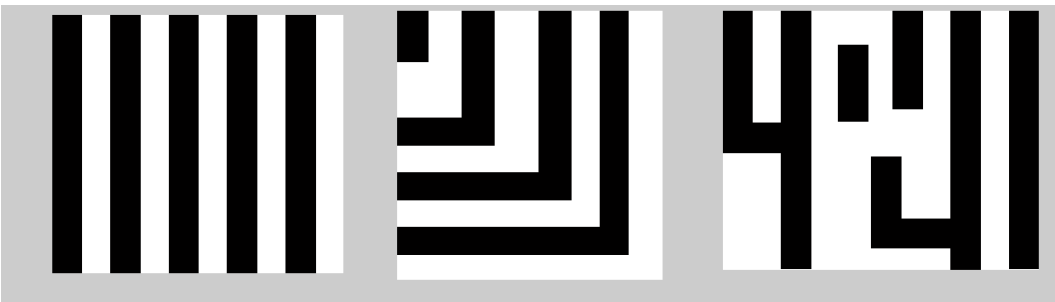


FIGURE 4.3: Patterns 1, 2 and 3 respectively from left to right

As seen from the values in table 4.1, the results of this metric seem reasonable for some cases. However it shows strange results in other cases, like in Figure 4.4. Pattern 4 is a simple deviation of pattern 1, only by changing ending line of the pattern, this pattern has regularity value of 6.217, which is much less than pattern 1. This large

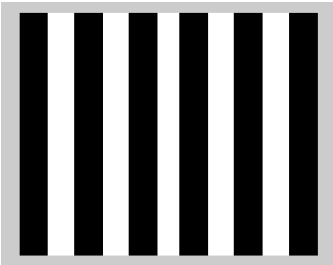


FIGURE 4.4: Pattern 4

dependence on the start and end of pattern is due to the use of Fourier transform and its well known aliasing problem. A trial was done to multiply the image by a $raisedcosine$ before processing, but it distorted the image and didn't improve the results a lot. Another major issue with this method is that it is computationally intensive as well and will not be suitable for large layouts and for production use. That's why the research headed towards seeking another method of deriving the metric so that it could be faster and more accurate.

4.2 Geometrical Based Metric

At this part of research, a simpler idea, based on the geometrical properties of patterns, like area, parameter, etc. was adopted. The following sections will show in details the work done. They start from showing how the metric was derived, followed by verifying the regularity metric against a variability metric and end up by choosing an electrical model from literature and comparing the results of regularity metric against the results of the electrical model to prove the validity of using the regularity metric as a fast indicator of litho-induced electrical variations.

4.2.1 Proposed Methodology

This section shows the methodology used in deriving and verifying the metric. It starts with defining the required regularity and showing the importance of having this regularity definition related to process variations and then, shows how the metric was derived from simple geometrical properties. Verification of the metric against a variability metric is shown as well to prove that it can be used as an indicator of process variations.

4.2.1.1 Derivation of Regularity Metric

Since this research is more oriented towards micro-regularity, the regularity mentioned here, means the regularity of certain pattern or cell or part of the layout and not the regularity of the whole layer as done in the research mentioned in the background section. The idea was to use a geometrical based technique for simplicity and less computational time. The first phase was to figure out the required regularity by each process step including lithography, etching, rapid thermal annealing and chemical mechanical polishing [17], [39], [38] and [40] . The second phase was trying to find a common definition for regularity that would decrease variations induced in each of the aforementioned process steps. Adding to that the well-known fact that a pattern consisting of parallel lines with equal widths separated by equal spaces as that shown in Figure 4.5 is considered the most regular pattern and it has the least variability. The layout used

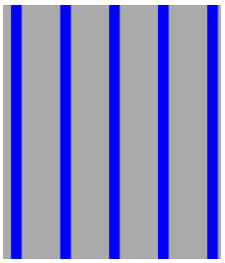


FIGURE 4.5: Most regular and least variable pattern

in the experiment of deriving the metric was an industrial 40nm node layout with size of $50\mu m$ x $50\mu m$. Layers used were poly, metal1 and metal2. The layout was divided into windows of equal sizes to create patterns. The size of the windows was chosen to be equal to the optical interaction length (1μ m). This resulted in about 2500 pattern/layer. It was observed that the results obtained from runs on poly layer were the most indicative because they contained patterns that are diverse in their degrees of regularity. Different geometrical properties of patterns in the layout were compared to those of the most regular pattern in Figure 4.5. A metric was derived using a simple equation shown in 4.5. This equation contains geometrical properties in a way such that the metric has a maximum value when the pattern resembles the most regular pattern. This means that the regular pattern has: (a) single orientation, (b) regular density, and (c) regular pitch. The metric value decreases as the pattern has line ends, jogs,

corners and shapes of different orientations and different densities. Figure 4.6 shows two of the most irregular patterns in the poly layer of the layout under experimentation according to our metric.

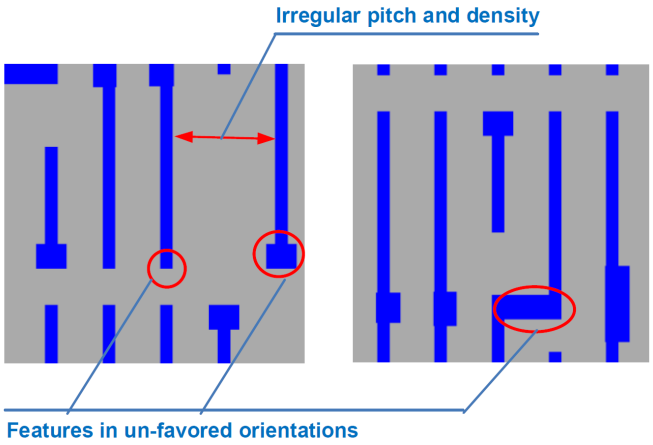


FIGURE 4.6: Two of the most irregular patterns for poly layer

$$RM\alpha \frac{\sum lengths\ of\ edges\ in\ favored\ orientation}{\sum lengths\ of\ edges\ in\ unfavored\ orientation} \frac{\sum perimeter\ (shapes\ of\ layer)}{\sum area\ (shapes\ of\ layer)} \frac{\sum perimeter\ (shapes\ of\ derived\ layer)}{\sum area\ (shapes\ of\ derived\ layer)} \tag{4.5}$$

Where:

RM is the regularity metric.

$derivedlayer$ is a layer created between the edges of projecting shapes within certain distance specified by the minimum spacing for each layer, example of derived layer for a certain pattern is shown in 4.7.

The regularity metric consists of three terms; the first term accounts for single orientation, the second term accounts for regular density and the third term accounts for regular pitch.

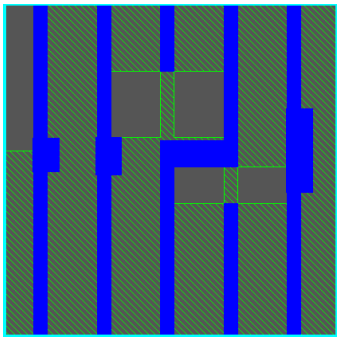


FIGURE 4.7: Derived Layer in hashes(green)

To verify the relation between the regularity as defined by our metric and variability, a variability metric was needed, the following section will show the basics behind this variability metric and how it was derived.

4.2.1.2 Variability Metric

To verify the relation between the regularity as defined by our metric and variability, lithography simulations were used at different process conditions of dose and defocus to model the variations. Dose is the variation across the wafer in the intensity of the light used in the manufacturing process. Focus is the variation in the alignment of the wafer in the Z-axis during the manufacturing process.

Process Window

Process window is the range of normalized exposure and focus variation within which a feature will print within acceptable tolerance. As shown in Figure4.8, the gray area is the process window within which the critical dimension will be within acceptable range.

The lithographic process window is not static, so that the process window for one feature may not be the same as the process window for an adjacent feature. This can be due to several factors like the context of the feature and the wafer topology. Figure 4.9 shows how different process conditions can affect the printability of a feature.

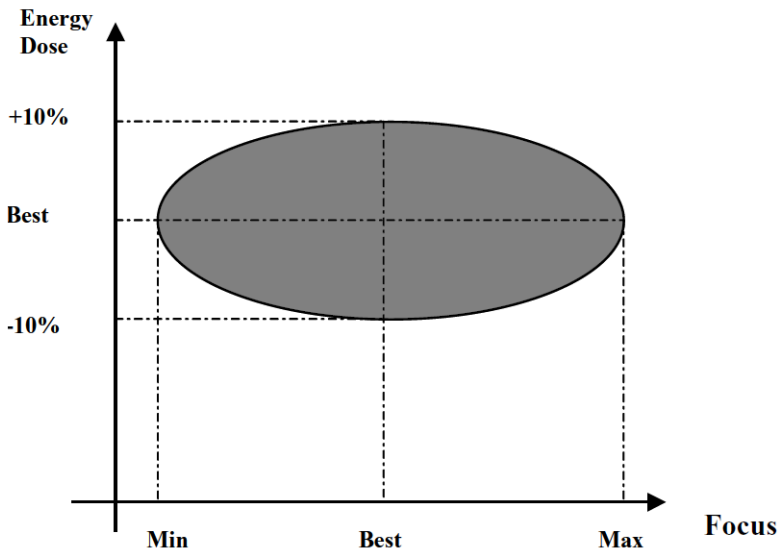


FIGURE 4.8: Process Window

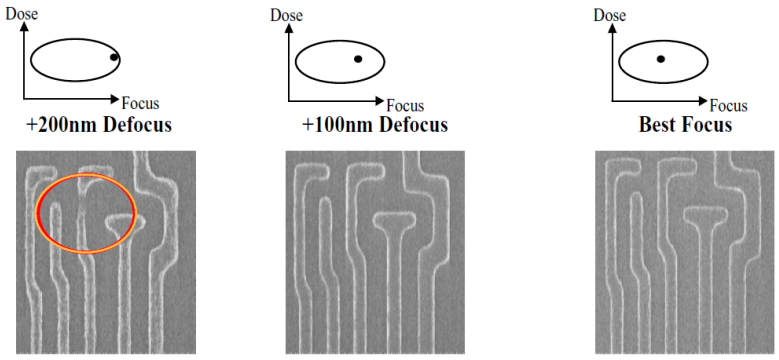


FIGURE 4.9: Impact of process variations on printability

Process variability band

Process Variability band or PVband is the band created from printing contours of different process conditions as shown in Figure 4.10. It represents the area within which a feature will print as the process conditions vary [51]. PV-bands are created by simulating the printed image at the various process conditions and combining the resulting printed images. PV-bands have widths due to process variation [52].

There are two types of PV-bands:

- Regular PV-bands Regular PV-bands represent the maximum and minimum edge displacement as in figure 4.11

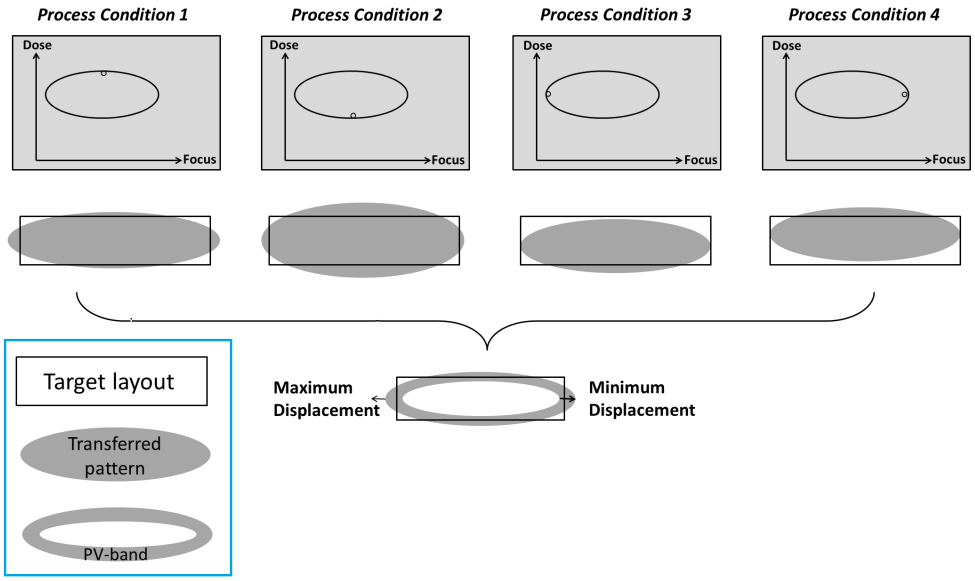


FIGURE 4.10: PV-Band representing a set of simulated print images

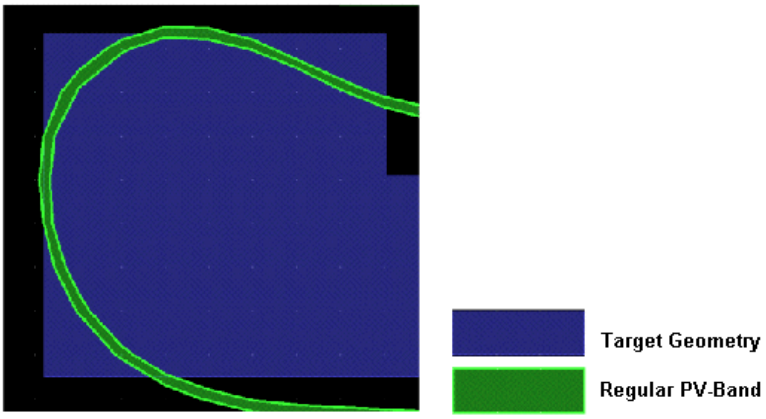


FIGURE 4.11: Regular PVband

- Absolute PV-bands Absolute PV-bands represent the maximum and minimum deviation from the target layout as shown in figure 4.12. Whereas regular PV-bands let you see how much feature shape varies from one process condition to another, absolute PV-bands let you see how much the feature shape will deviate from the drawn shape. If the outer PV-band edge is within the target layer, the outer PV-band will snap to the design edge. Similarly, if the inner PV-band edge is outside the layout target, the inner PV-band edge will snap to the design edge. Regular PV-bands are always within the absolute PV-bands [52].

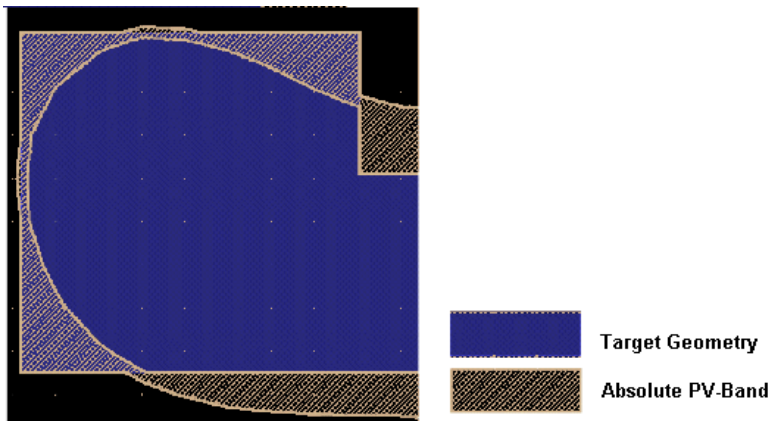


FIGURE 4.12: Absolute PVband

To verify the relation between the regularity as defined by our metric and variability, a variability metric to indicate the variability for each pattern was needed. This metric used the idea of process variability band (PV-band) [51]. The variability metric is defined as the area of absolute PVband divided by the perimeter of the drawn layer 4.6. By this definition, the variability metric represents the average width of the PV-band which is an indication of the average variations.

$$VM = \frac{Area\ (PVbands)}{Perimeter\ (drawn\ layer)} \tag{4.6}$$

4.2.1.3 Verification

A change in dose of $+/-3\%$ and defocus of $100nm$ were used in generating PVbands. Lithography models were calibrated to best match the silicon results. Calibre Litho-Friendly Design was used in these simulations.

Plotting the regularity metric value of each pattern versus its variability metric value, a correlation can be observed, such that patterns with high regularity had low variability and vice versa as shown in Figure 4.13.

It has been noticed that the idea of dividing the layout into equal windows irrespective of the cell placements is not realistic. This way of layout division resulted in patterns of low regularity as highlighted in Figure 4.13. These patterns were found to be created around cell boundaries as that shown in 4.14.

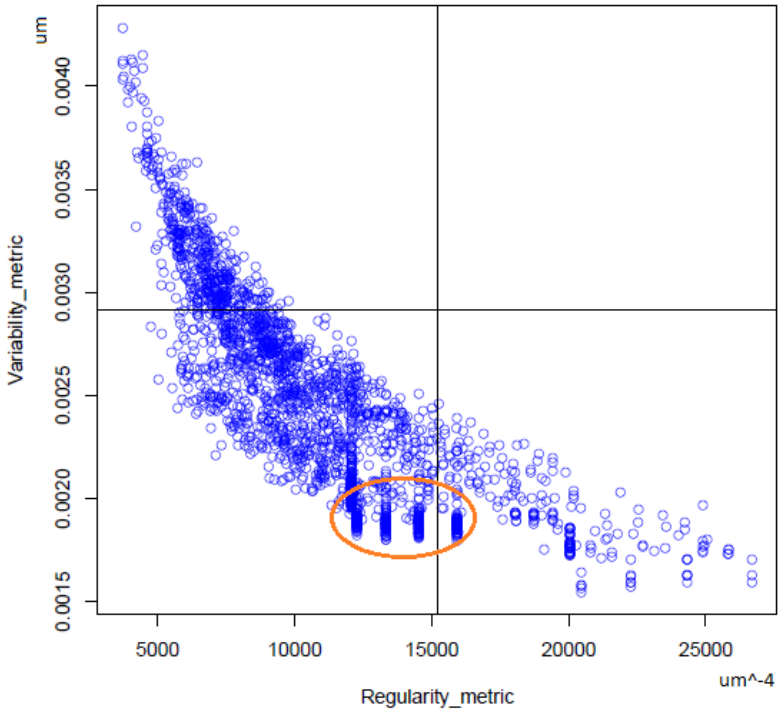


FIGURE 4.13: Regularity metric versus process variability metric for each pattern.

FIGURE 4.14: Pattern taken at the edge of two cells

Motivated by the above results the decision was to explore the possibility of using the regularity metric on logic cell placements in layout instead of arbitrary patterns. This way, the regularity metric can help in detecting electrical variations induced by systematic lithographic variations. The electrical modeling method used for that purpose is shown in the following section.

4.2.2 Electrical Variability Modeling

Lithography simulation enables estimation of critical dimension (CD) variations at different process points. The challenge is to transform shapes generated by lithography simulation to a form that preserves the electrical properties. Various recent works address the problem of non-rectangular gate (NRG) transistor modeling. The method as in [53] was used in this paper, where gate slicing method is used to compute the current of the NRG device as shown in Figure 4.15. Considering the non-linear effect of narrow width; the gate slicing method was modified to include an extra term which compensates for such non-linear effects [53]. When compared to TCAD, this method showed an average current error of only 1.6% and had a correlation of 0.98 with measured transistors current [54].

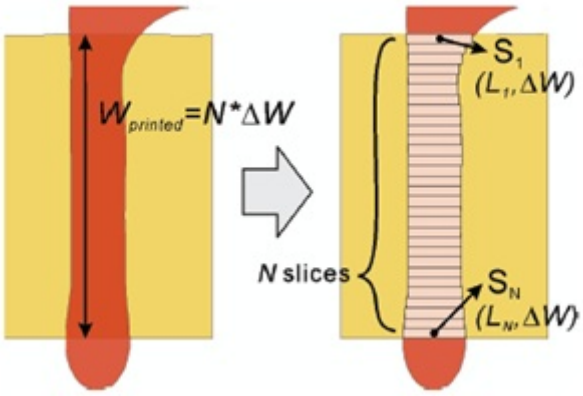


FIGURE 4.15: The NRG device contour broken into parallel slices

The method adopted here didn't need to compute the equivalent gate dimension. It can directly calculate current from non-rectangular gates, thus providing a quick and easy electrical performance analysis without SPICE simulations. The current is calculated from (4.7).

$$I = \sum_{i=1}^{n} I_{per_{\mu m}}(L_i) \times W_i + \frac{1}{2}[I_{offset}(L_1) + I_{offset}(L_n)] \tag{4.7}$$

Where $I_{per_{um}}$ and I_{offset} are obtained from transistor characterization shown in Figure 4.16. The lithography variations were modeled by simulating across dose and

defocus variations as in section4.2.1.3. Changes in dose of $+/-3\%$ and defocus of 100nm were used. The current calculation for each transistor was performed on each process condition: nominal, minimum and maximum. The "ON" current of each transistor was calculated. Transistors with process-induced "ON" current variations larger than ten percent relative to nominal current are considered as "electrically variable" transistors [55].

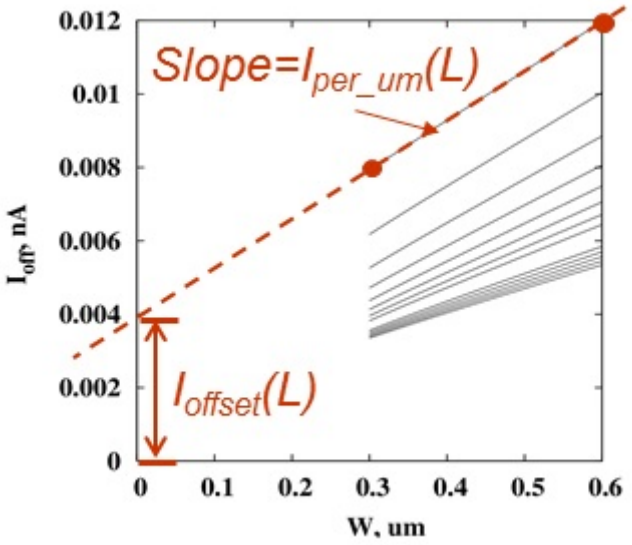


FIGURE 4.16: Transistor current for different gates width (W) and length (L)

4.2.3 Results and Discussion

4.2.3.1 Setup

In this section the electrical model and the regularity metric defined before were used to characterize the cells in the critical path for two different designs. The target here is to compare the cells that are found to be electrically variable to those found to be irregular and investigate the validity of using the regularity metric as a fast indicator of highly variable cells. To define how the regularity part is obtained, the regularity metric value is calculated for the poly layer of all cells in the critical path, then cells are arranged in ten bins according to their regularity metric value and those cells in the least two bins are considered the most irregular patterns.

4.2.3.2 Results

To be able to compare the results of the regularity versus that of the electrical variability metric in [55] we used the same design benchmarks. b22 design of the ITS'99 benchmark circuit is composed of 7266 cells, 104 cells are on the critical path and of the 104, 49 were found electrically variable. S13207 design of the ISCAS'89 benchmark designs has 2419 cells , 21 cells where on the critical path and out of the 21 cells , 16 were found to be electrically variable. Table 4.2 shows the number of cells on the critical path with electrically variable transistors according to the definition in previous section. It also shows the number of most irregular cells in the critical path. Then shows the number of "matches" which are the variable cells that the regularity metric was able to detect, the number of "misses" which are the cells that were found to be variable but the regularity metric didn't consider them highly irregular. The "extras" are the cells detected by the regularity metric as irregular while they were not found to be variable.

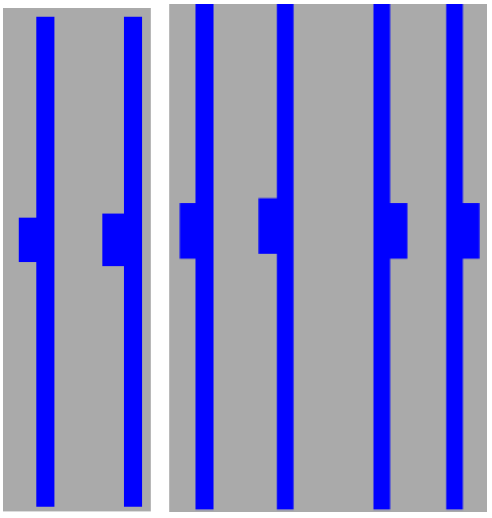
	Design 1	Design 2
Design details	b22 design of ISCAS'89 benchmark	s13207 of ITS'99 benchmark
Cells on critical path	104	21
Electrically variable cells	49	16
Irregular cells	33	14
Matches	31	12
Misses	18	4
Extras	2	2

TABLE 4.2: Results of electrical variability and regularity of two different designs

4.2.3.3 Discussion

In this part a detailed analysis was done to understand the misses and the extras in each of the experiments. They were found to be few numbers of cells but with different placements. The variability results showed that some of these placements were found variable and others were not, which means that these cells were affected by their neighborhood. A detailed study below clarifies this conclusion. The 49 electrically variable cells in design 1 were found to consist of 11 unique cells; the regularity metric detected

10 out of the 11 as irregular. In design 2, the 16 electrically variable cells were found to consist of 14 unique cells; the regularity metric detected 10 out of the 14 as irregular. Misses: All 18 misses in design 1 were found to be different placements of one cell shown in Figure 4.17 (a). According to the regularity metric, this cell has medium regularity. This cell had 58 placements in the critical path; 18 placements were found electrically variable while 40 were not. This can be attributed to the small size of the cell which makes it highly affected by neighboring cells. For design 2, one of the four misses was the same cell as in Figure 4.17 (a). The other three misses in design 2 were different mirror images of the cell in Figure 4.17 (a). Figure 4.17 (b) shows an example of design 2 misses.



(a) (b)

FIGURE 4.17: Examples of misses.

Extras: The extras in design 1 were two cells; one of them (Figure 4.18 (a)) had 24 placements in the critical path; 23 placements were found variable while only one was not. According to the regularity metric this was irregular cell. The other cell (Figure 4.18 (b)) didn't have any other placements in the critical path but it was noticed that the gate region is far from the irregularities in the poly which may be the reason why it was not found electrically variable. The same cell in Figure 4.18 (b) was one of the extras in design 2 while the other one was a cell that looked a multiple of that in Figure 4.18(a).

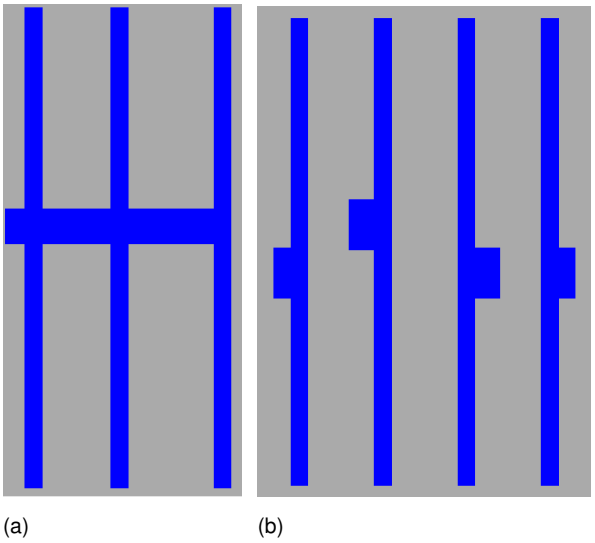


FIGURE 4.18: Examples of extras.

4.3 Summary and Conclusion

This chapter showed the work done to derive and verify a layout regularity metric. An auto-correlation based metric was first explored and found not suitable for its low speed and accuracy. A geometrical based metric was then derived using a definition of regularity that suits different process steps. The regularity metric showed good correlation with a variability metric that was derived using the concept of process variability bands. As an application, a method that models litho-induced electrical variations was selected from literature and its results were compared to that of the regularity metric. The results showed good matching percentage between the elctrical model and the regularity metric which proves the validity of using the regularity metric as a fast indicator of litho-induced electrical variations.

4.3.1 Publications Resulting From This Thesis

- E. Swillam, K. Madkour and M. Anis, ”Layout Regularity Metric as a Fast Indicator of High Variability Circuits”, IEEE System On Chip Conference,Erlangen, Germany,September 04-06, 2013 (In press).

Chapter 5

Conclusions And Future Work

This work was based on the fact that regular designs induce less process variations and are more robust to variations in general. This research has studied the impact of layout topology on different sources of variations and came out with a definition regularity that is correlated to variability.

5.1 Conclusions

In this thesis, a fast geometrical based layout regularity metric was derived and verified by lithography simulations. This metric can be used to compare different patterns or cells in the same layout to each other. The metric has shown good correlation with process variability metric. An electrical variability model has been selected and used to characterize cells in the critical path of two benchmark designs. When the regularity metric results are compared against electrical model results, it is seen that the metric has high matching percentage; which proves that the metric can be used as a fast indicator of highly variable cells.

5.2 Future Work

As future work, the metric can have several applications. During design, it can be run on the standard cells of a certain library to evaluate each cell independently then the values from the regularity metric can be used to predict the effects of neighborhood on certain cells, this information can be fed to a placer to do regularity aware placement. It can also be used in post-layout verification to highlight areas of high probability of variations for further litho-simulations so this will decrease the area of litho-simulations and thus the turn around time. Some modifications can be applied so that after running the metric, the user can get hints of how to improve problematic areas to be more regular and, thus, less sensitive to variations.

Another suggestion is to try to integrate the regularity metric in regular fabrics design flow so that it can run on each brick as as soon as it's generated and evaluate its regularity to make sure that the generated brick are ”regular by construct”, it can also be used during placement and routing of regular fabrics.

An important point to notice is that for 32nm and beyond, poly layers are becoming more regular and the concern of regularity will be more towards higher metal layers. This metric can be easily migrated to other technology nodes by changing the parameters that are related to technology.

Appendix A

Auto-correlation based metric

```
%***here lists all the images to process
img_available=['regularx.bmp' ; 'regular5.bmp' ; 'regular2.bmp' ; 'regular3.bmp' ; 'r
%***loop on the images
Reg_vec=[];
for i=1:size(img_available,1)
  img=img_available(i,:);
  %***now read the images
  img_read=imread(img);
  figure;
  imshow (img_read)
  xlabel 'x';
  ylabel 'y';
  %***calculate the FFT of the image (ftt2 for 2 dim fft)
  img_read_fft=fft2(img_read);
  figure;
  imshow(log(abs(fftshift(img_read_fft)) + 1), [])
  %***calculate the conjugate of the fft
  img_read_fft2=conj(img_read_fft);
  %***autocorr=IFFT [FFT[I]* X FFT(I)]
  img_autocorr_tmp=ifft2(img_read_fft2.*img_read_fft);
```

```
  %***using ifftshift to bring zero frequencies to the middle
   img_autocorr=ifftshift(img_autocorr_tmp);
  %***calculating the max of autocorr to use it in normalization
  max_img_autocorr=max (max(img_autocorr,[],1));
  %***normalizing autocorr
  img_autocorr_norm=img_autocorr/max_img_autocorr;
%   figure;
%   surf (img_autocorr_norm);
%   xlabel 'x';
%   ylabel 'y';
%   title 'img autocorr';
  %***calculate the interaction map
  interaction_map=1-img_autocorr_norm;
  %interaction_map=interaction_map_tmp';
  figure;
  surf (interaction_map);
  xlabel 'x';
  ylabel 'y';
  title 'interaction_map';
  %x_dimension
  %[peaks,loc]=findpeaks(img_autocorr_norm(1,:),'0',MPH)
  %figure
  %plot (img_autocorr_norm(1,:));
  %interaction_map_min=min(interaction_map);
  %interaction_map=1-img_autocorr;
  %surf (interaction_map);
  %plot (img_autocorr (1,:))
  %surf (img_autocorr);

  %***finding local minima and maxima of the interaction map
```

```
  for j=1:size(interaction_map,1)

    [maxtab, mintab] = peakdet(interaction_map(j,:) ,0.01);
    if size(mintab,1)< 2
        Reg=0;
    else
        [d1,d2]=find_2_lowest_minima(mintab);
        d1
        d2



        Reg1=1-((abs(d2-2*d1))/d2);
        %to avoid negative values reg2 is set to 0 when d2 falls far from
        %3d1
        if d2<1.5*d1
            Reg2=0;
        else
            Reg2=1-((abs(d2-3*d1))/d2);
        end
        Reg=max(Reg1,Reg2);
    end

    Reg_vec=[Reg_vec;Reg];
    mean_reg_vec=mean(Reg_vec);

%     figure;
%     plot (interaction_map(j,:))
%     hold on; plot(mintab(:,1), mintab(:,2), 'g*')
%     plot(maxtab(:,1), maxtab(:,2), 'r*')
  end
```

```
  %f=abs(fftshift(img_read_fft));
  %figure;
  %surf (f);
  %title 'fft';
  %autocorr (img_read (1,:),450)
    figure;
    k=1:size(Reg_vec);
    plot ((Reg_vec))
end
```

Bibliography

[1] D. Boning and S. Nassif, *Design for High Performance Microprocessor Circuits*, ch. Models of Process Variations in Device and Interconnect. IEEE Press, 2000.

[2] P. Sharma, *Manufacturing-Aware Physical Design Techniques*. PhD thesis, University of California, San Diego, 2007.

[3] A. K. Wong, *Resolution Enhancement Techniques in Optical Lithography*. SPIE Press, 2001.

[4] N. A. Drego, *Characterization and Mitigation of Process Variation in Digital Circuits and Systems*. PhD thesis, Massachusetts Institute of Technology, June 2009.

[5] R. W. Keyes, "The effect of randomness in the distribution of impurity atoms on fet thresholds," *Applied Physics*, vol. 8, 1975.

[6] D. E. et al., "Razor: circuit-level correction of timing errors for low-power operation," *IEEE Micro*, vol. 24, pp. 10–20, Nov 2004.

[7] K. Kuhn, "Reducing variation in advanced logic technologies: Approaches to process and design for manufacturability of nanoscale cmos," in *IEEE International Electron Devices Meeting*, pp. 471–474, 2007.

[8] M. R. et al., "Layout design methodologies for sub-wavelength manufacturing," in *ACM/IEEE Design Automation Conference*, June 2001.

[9] D. Chetverikov, "Pattern regularity as a visual key," *Image and Vision Computing*, vol. 18, p. 2000, 1998.

[10] C. W. Gwyn and P. J. Silverman, "Euvl: transition from research to commercialization," in *Proc. SPIE*, vol. 5130, 2003.

[11] P. Gupta and A. B. Kahng, "Manufacturing-aware physical design," in *Proc. International Conference on Computer Aided Design*, pp. 681–687, 2003.

[12] M. Gazor, "Design for manufacturability with regular fabrics in digital integrated circuits," Master's thesis, Massachusetts Institute of Technology, May 2005.

[13] K. Y. Tong, *Logic Design Regularity for Robust Integrated Circuits*. PhD thesis, Carnegie Melon University, 2006.

[14] P. D. Friedberg, *Spatial Modeling of Gate Length Variation for Process-Design Co-Optimization*. PhD thesis, Univerity of California, Berkeley, 2007.

[15] L. L. et al., "Alternating phase-shifted mask for logic gate levels, design and mask manufacturing," *SPIE Optical Microlithography*, vol. 3679, 1999.

[16] B. J. Lin, "The attenuated phase-shifting mask," *Solid State Technology*, January 1992.

[17] K. O. A. et al., "Pattern based prediction for plasma etch," *IEEE Transactions on Semiconductor Manufacturing,*, vol. 20, no. 2, pp. 77–86, 2007.

[18] X. Xie, *Physical Understanding and Modeling of Chemical Mechanical Planarization in Dielectric Materials*. PhD thesis, Massachusetts Institute of Technology, June 2007.

[19] S. D. et al., "High mobility si/sige strained channel mos transistors with hfo2/tin gate stack," in *IEEE International Electron Devices Meeting*, 2008.

[20] A. A. et al., "Process variation in embedded memories: Failure analysis and variation aware architecture," *IEEE Journal of Solid-State Circuits*, vol. 4, pp. 1804–1814, September 2005.

[21] S. P. et al., "Process variations aware robust on-chip bus architecture synthesis for mpsocs," in *IEEE International Symposium on Circuits and Systems*, 2008.

[22] D. Marculescu and S. Garg, "Process-driven variability analysis of single and multiple voltagefrequency island latency-constrained systems," *IEEE Transactions on Computer Aided Design*, vol. 27, pp. 893–905, May 2008.

[23] E. H. et al., "Impact of process variations on multicore performance symmetry," in *Design, Automation and Test in Europe Conference and Exhibition*, 2007.

[24] J. Donald and M. Martonosi, "Power efficiency for variation-tolerant multicore processors," in *International Symposium on Low-Power Electronic Design*, 2006.

[25] N. Drego, "Low-jitter receiver circuit for on-chip optical clock distribution," Master's thesis, Massachusetts Institute of Technology, June 2003.

[26] B. Calhoun and A. Chandrakasan, "A 256kb sub-threshold sram in 65nm cmos," in *IEEE International Solid-State Circuits Conference*, February 2006.

[27] J. T. et al., "Adaptive body bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage," *IEEE Journal of Solid-State Circuits*, vol. 37, November 2002.

[28] N. V. et al., "Nanometer mosfet variation in minimum energy subthreshold circuits," *IEEE Transactions on Electron Devices*, vol. 55, pp. 163–174, January 2008.

[29] R. G. et al., "Process optimization tweaking tool (pott) and its application in controlling oxidation thickness," *IEEE Transactions on Semiconductor Manufacturing*, vol. 2, May 1989.

[30] S. Leang and C. Spanos, "Statistically based feedback control of photoresist application," in *IEEE/SEMI Advanced Semiconductor Manufacturing Conference and Workshop*, 1991.

[31] K. S. W. et al., "On endpoint detection of plasma etching via optical emission interferometry," in *Electrochemical Society Meeting*, May 1996.

[32] Y. B. et al., "Compact modeling and robust layout optimization for contacts in deep sub-wavelength lithography," *Design Automation Conference*, pp. 408–411, 2010.

[33] T. Park, *Characterization and Modeling of Pattern Dependencies in Copper Interconnects for Integrated Circuits*. PhD thesis, Massachusetts Institute of Technology, May 2002.

[34] R. S. Ghaida and P. Zarkesh-Ha, "Estimation of electromigration-aggravating narrow interconnects using a layout sensitivity model," in *IEEE International Symposium on Defect and Fault-Tolerance in VLSI Systems*, pp. 59 –67, September 2007.

[35] K. M. Gonzalez-Valentin, "Extraction of variation sources due to layout practices," Master's thesis, Massachusetts Institute of Technology, May 2002.

[36] M. G. et al., "Litho-driven layouts for reducing performance variability," in *IEEE International Symposium on Circuits and Systems ISCAS 2005*, pp. 3551–3554, 2005.

[37] Y. W. et al., "Physical design techniques for optimizing rta-induced variations," *Proc. Asia and South Pacific Design Automation Conference*, pp. 745–750, 2010.

[38] V. J. et al., "Analyzing electrical effects of rta-driven local anneal temperature variation," *Proc. ASP-DAC*, pp. 739–744, 2010.

[39] I. A. et al., "Rta-driven intra-die variations in stage delay, and parametric sensitivities for 65nm technology," *Tech. Dig. Symp. VLSI Technology*, pp. 170–171, 2006.

[40] J. C. S. et al., "Reduction of rta-driven intra-die variation via model-based layout optimization," in *Proc. Symp. VLSI Technology,*, pp. 152 –153, june 2009.

[41] A. R. S. et al., "Design rule optimization of regular layout for leakage reduction in nanoscale design," in *Proc. ASP-DAC 2008*, pp. 474–479, 2008.

[42] J. Lienig, "Introduction to electromigration-aware physical design," in *Proc. International Symposium on Physical design*, ISPD '06, (New York, NY, USA), pp. 39–46, ACM, 2006.

[43] C. W. et al., "Modeling of layout-dependent stress effect in cmos design," in *Tech. Dig. IEEE/ACM Int. Conf. on Computer-Aided Design ICCAD*, pp. 513 –520, nov. 2009.

[44] M. L. et al., “Backend cad flows for restrictive design rules,” in *Proc. Int. Conf. Computer-Aided Design*, vol. 0, (Los Alamitos, CA, USA), pp. 739–746, IEEE Computer Society, 2004.

[45] L. P. et al., “Exploring regular fabrics to optimize the performance-cost trade-off,” in *Proc. 40th annual Design Automation Conference*, DAC ’03, (New York, NY, USA), pp. 782–787, ACM, 2003.

[46] M. P. et al., “Design of complex circuits using the via-configurable transistor array regular layout fabric,” in *IEEE International SOC Conference (SOCC)*, pp. 166–169, Sept 2011.

[47] A. K. et al., “Heterogeneous programmable logic block architectures,” in *Proc. Design, Automation and Test in Europe Conference and Exhibition*, pp. 1118–1119, 2003.

[48] K. Y. T. et al., “Design methodology of regular logic bricks for robust integrated circuits,” *Proc. Int. Conf. Computer Design ICCD*, pp. 162–167, 2006.

[49] T. J. et al., “Maximization of layout printability/manufacturability by extreme layout regularity,” *Journal of Micro/Nanolithography, MEMS, and MOEMS*, vol. 6, no. 3, pp. 031011–031011–15, 2007.

[50] M. P. et al., “Fixed origin corner square inspection layout regularity metric,” in *Design, Automation Test in Europe Conference Exhibition (DATE)*, 2012.

[51] J. A. Torres and N. Cobb, “Study towards model-based drc verification,” in *SPIE 5992, 25th Annual BACUS Symposium on Photomask Technology*, vol. 59923L, 2005.

[52] Mentor Graphics, *Calibre Litho-Friendly Design User’s Manual*, 2010.3 ed., 2010.

[53] J. W. et al., “Fast and simple modeling of non-rectangular transistors,” in *Proc. SPIE*, vol. 7122, pp. 71223S–71223S–10, 2008.

[54] E. S. et al., “Full flow for transistor simulation based on edge-contour extraction and advanced spice simulation,” in *SPIE Optical Microlithography*, vol. 7275, pp. 727519–727519–9, 2009.

[55] K. M. et al., "Work smarter not harder – how to get more results with less modeling," in *SPIE Optical Microlithography*, 2014 (In press).