

Effects of grid-placed contacts on circuit performance

Jun Wang^{*}, Alfred K. Wong[†]

Dept. of Electrical and Electronic Engineering, Univ. of Hong Kong
Pokfulam Road, Hong Kong

ABSTRACT

The impact of grid-placed contacts on application-specific integrated circuit (ASIC) performance is studied. Although snapping contacts to grid adds restrictions during layout design, smaller circuit area can be achieved by careful selection of the grid pitch, raising the lower limit of transistor width, applying double exposure, and shrinking the minimum contact pitch enabled by more effective application of resolution enhancement technologies. The technique is demonstrated on the contact level of 250-nm standard cells with the minimum contact pitch shrunk by 10%. The area change of 84 cells ranges from -20% to 25% with a median decrease of 5%. The areas of two circuits, a finite-impulse-response (FIR) filter and an add-compare-select (ACS) unit in the Viterbi decoder, decrease by 4% and 2% respectively. Delay and power consumption are also estimated to decrease with area.

Keywords: Resolution enhancement techniques, Grid-placed contacts, Standard cells, Double exposure, Critical dimension, Process-related factor.

1. INTRODUCTION

The critical dimension (CD) of optical lithography—the minimum size that can be defined—is a function of three parameters, as shown by the following expression [1]:

$$CD = k_1 \frac{\lambda}{NA}. \quad (1)$$

The critical dimension is proportional to the wavelength of the exposure light λ and the process-related factor k_1 , and decreases with increasing numerical aperture (NA) of the projection system. Smaller dimensions can be printed by decreasing the wavelength, increasing the numerical aperture and reducing k_1 . The ultimate resolution can only be achieved by all three measures [2].

As the measure of lithography aggressiveness, the k_1 factor is the only parameter that can be controlled by lithographers for a given exposure system. Although the theoretical lower limit of k_1 is 0.25 [2], image quality degrades noticeably when k_1 falls below 0.75. Resolution enhancement techniques (RETs), such as modified illumination, optical proximity correction (OPC), phase-shifting masks (PSMs), are needed to improve image quality for low- k_1 lithography. These RETs are effective in reducing the k_1 factor. Over the last two decades, the k_1 factor has decreased to about 0.5 [3]. But additional improvement requires communications between the technology and design communities.

By considering circuit manufacturability in layout design, it is expected that the k_1 can be further reduced by fabrication-friendly layout. This paper examines the impact of fabrication-friendly layout on chip performance. Test circuits are designed using fabrication-friendly standard cells to study the effects on area, speed, and power consumption.

^{*} junwang@eee.hku.hk; phone 852 2859-2698; fax 852 2559-8738; <http://www.eee.hku.hk>

[†] awong@eee.hku.hk; phone 852 2859-7096; fax 852 2559-8738

2. FABRICATION-FRIENDLY LAYOUT

Modified illumination is essential at low- k_1 imaging for successful lithography. However, image quality depends not only on the size and shape of a pattern, but also on its environment [4]. For example, contacts are usually placed randomly in the layout. However, from the lithography point of view, optimization of the illumination configuration is almost impossible for randomly-placed contacts because of the simultaneous existence of dense and sparse contacts. No illumination scheme allows optimal imaging of both dense and sparse contacts.

By limiting the circuit pattern configuration, fabrication-friendly layout allows lithography optimization. In this study, we examine the effects of placing all contacts on a grid. To do so, the minimum contact pitch can be shrunk by optimization of the illumination configuration [5]. Fig.1 shows the imaging process of the randomly-placed (traditional) and grid-placed (fabrication-friendly) layout. In the randomly-placed layout, the layout (Fig.1.a) translates to a mask with the same features (Fig.1.b), which are imaged onto on the die (Fig.1.c). In the grid-placed layout, contacts are snapped to grid points (Fig.1.d). Assist contacts are placed at grid points that do not have a contact (Fig.1.e). The assist contacts are sized such that they do not print onto the die but nevertheless create a mask spectrum that allows the illumination to be optimized (Fig.1.f).

There is a trade-off between randomly-placed (traditional) and grid-placed (fabrication-friendly) layout. Excessive lithography friendliness may be so restrictive on layout compaction that circuit area increases unacceptably. Although the contacts can be designed smaller and packed closer in grid-placed layout (Fig.1.d), the initial area increase should be small enough such that it can be offset by shrinkage of the minimum contact pitch.

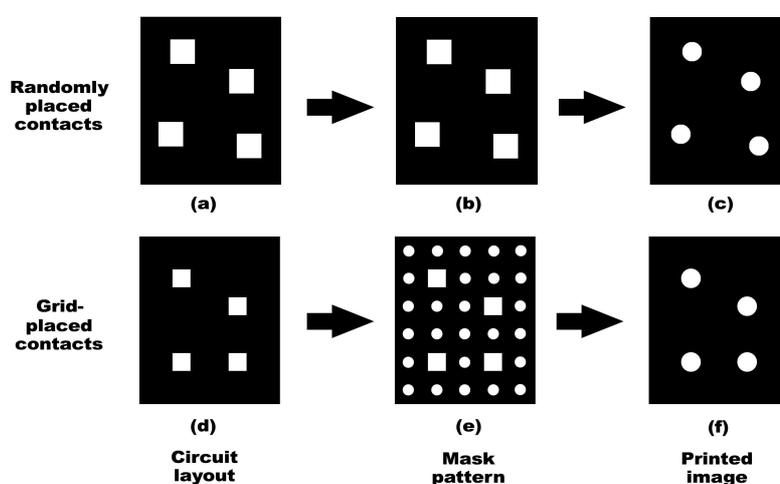


Figure 1: Imaging process of randomly-placed and grid-placed contacts.

3. LAYOUT DESIGN

A 250-nm standard cell library is used to demonstrate the effects of grid-laced contacts on circuit performance. Standard cell is an important structure for ASIC design. A standard-cell based ASIC die typically consists of three types of cells: I/O cells, mega cells and standard cells. I/O cells are cells laid on the periphery of the die as connection points to outside circuitry. Mega cells are typically pre-designed mega-logic structures such as RAMs and ROMs. Standard cells are micro-logic structures providing basic logic functions. Standard cell circuits also exist in designs such as microprocessors and their peripherals. As the elementary and flexible structures in physical layout, standard cells deserve our first investigation.

3.1 Layout structure of standard cells

Fig. 2 shows the structure of a typical standard cell. Each standard cell in a library is rectangular with a fixed height but variable width. During placement in physical design, cells are placed in rows with the power supply paths overlapped. A standard cell is composed of 6 main layers: N-well, N-diffusion, P-diffusion, poly-silicon, contact and metal-1. The N-well, N-diffusion, P-diffusion and poly-silicon layers are primarily used to construct the P-MOSFETs and N-MOSFETs. Intra-cell routing is accomplished by poly-silicon and metal-1. Contacts act as connections between the routing layers and the underlying layers.

For standard cells, the height is determined by the number of metal-1 tracks. A metal-1 track consists of a metal-1 line and space separating two metal-1 lines. The typical height is 10 tracks: 3 for power supply and 7 for intra-cell connection. Because the height is fixed, any area change is caused only by a cell width change

3.2 Grid pitch

First of all, before selecting the grid pitches in the height and width direction, the minimum grid pitch should be found. This lower limit on the pitch is naturally the minimum allowed by the design rules. The minimum contact pitch of the technology under study is 600nm. If we assume that the pitch can be shrunk by 10% because of regular contact placement, 540nm is used as the shrunk minimum contact pitch and the minimum grid pitch in this study.

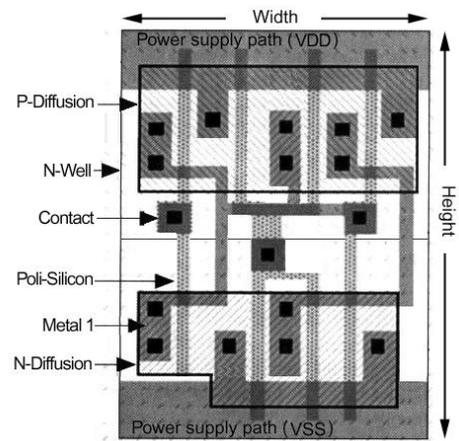


Figure 2: Structure of a standard cell.

The minimum allowable pitch is not necessarily the grid pitch. To minimize the number of affected contact, the grid pitch should be the most common pitch in layout. Distributions of contact pitches of standard cells are collected in both the height (vertical) and width (horizontal) directions. Plotted in Fig.3 are four representative pitch distributions of standard cells. Fig. 3.a and Fig. 3.b show the vertical and horizontal pitch distributions of combinational cells such as NAND gates and multiplexers. Fig. 3.c and Fig. 3.d are the vertical and horizontal pitch distributions of sequential cells such as flip-flops. Because the outputs of combinational cells are

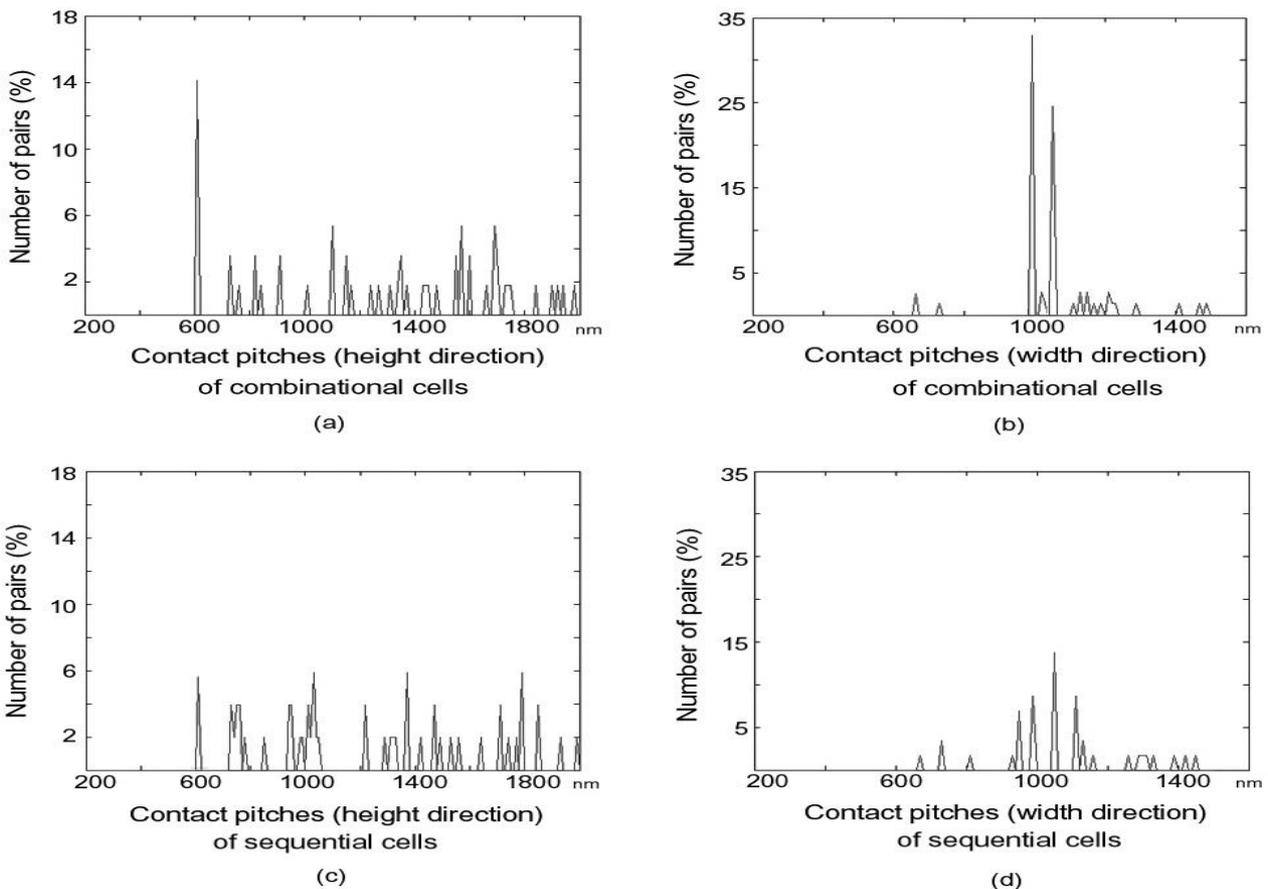


Figure 3: Representative pitch distributions of standard cells. (a and b are for combinational cells, c and d are for sequential cells)

functions of the inputs only, the layouts of combinational cells are typically simpler and more regular than that of sequential cells, which are circuit elements with memory. The consequence is that the pitch distributions of combinational cells (Fig.3.a and Fig.3.b) show sharp peaks while the peaks of sequential cells (Fig.3.d) are not as distinct.

3.2.1 Pitch in the height (vertical) direction

To archive a better connectivity between signal supply paths (metal-1 lines) and the source or drain regions of wide MOSFETs, multi-contact connection is used (Fig.4). This type of multi-contact configuration is used in combinational cells with large load capacitance, in which wide MOSFETs are needed to provide large driving current. Covered by the same metal-1 line, these contacts are separated by the minimum contact pitch (600nm) causing the peak in the pitch distribution at the minimum contact pitch (Fig.3.a). Because MOSFETs are mostly placed vertically (Fig.2) in standard cells, this peak only appears in the vertical pitch distribution.

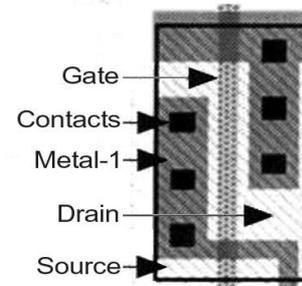


Figure 4: Multi-contact connections for a large MOSFET.

For multi-stage sequential cells, however, MOSFETs in stages other than the output stage are designed smaller to reduce parasitic capacitance. Because there is only one source or drain contact in each narrow MOSFET, most contacts are covered by different metal-1 lines. These contacts can be vias between source or drain to metal-1, poly-silicon or gate to metal-1, and power supply paths to substrate or wells. Since contacts are surrounded by different layers including metal-1, poly-silicon, N-diffusion, and P-diffusion and minimum space of these layers are different, the pitch distribution of contacts in the height direction shows no distinct peak (Fig. 3.c). To accord with the height of standard cells, which is typically integer multiples of the metal-1 track, the pitch of the metal-1 track, 640nm, is chosen as the vertical grid pitch. This pitch can also be used in combinational cells by reducing the number of redundant contacts in multi-contact connections.

Using 640nm as the vertical grid pitch gives us 10 grid points in the height direction because there are 10 metal-1 tracks. This number is often inadequate for regular contact placement because of the differing design rules of the various levels. To be consistent with the metal-1 pitch and to provide more grid points in the height direction, it is desirable to use 320nm—half of 640nm—as the vertical grid pitch. Although 320nm is beyond the resolution limit of the technology (540nm), we propose a method in Sec.3.3 to enable its fabrication.

3.2.2 Pitch in the width (horizontal) direction

The pitch distribution for both combinational and sequential cells show peaks near 1000 nm in the width direction. These peaks arise from the layouts in which the MOSFETs are drawn one by one in the width direction (corresponding to a series connection electrically) with contacts placed in both the source and drain regions (Fig.5). The pitches between these source and drain contacts are about 1000nm (with small difference arising from the different structures of MOSFETs, as discussed later). By shrinking the minimum contact size by 10%, these peaks can be shifted to about 950nm (Because the space separating the contacts from the gate and the size of gate can not be shrunk by the grid contact placement, the percentage of shrinking of pitch between source and drain contact is less than that of contacts). Since 950nm is not an integral multiple of 540nm (the shrunk minimum contact pitch), the use of the shrunk minimum contact pitch as the grid pitch will increase the cell areas significantly because these frequently occurring pitches will be snapped from 950nm to 1080nm (two times of the shrunk minimum contact pitch of 540nm).

The two peaks around 1000 nm in Fig.3.b deserve further discussion. These two peaks, 1080nm and 980nm respectively, originate from two types of MOSFET layout configurations as illustrated: wide and narrow (Fig.6). For MOSFETs with narrow gates

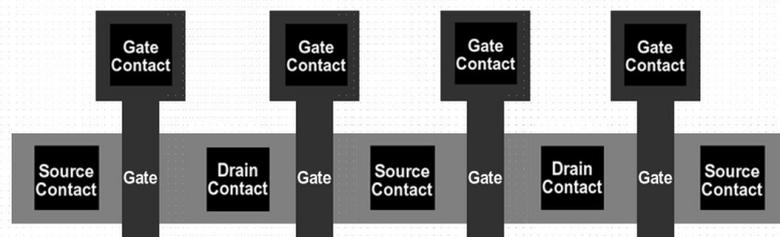


Figure 5: Series connection of MOSFETs.

(Fig.6.a), the widths of source and drain are wider at the contact than the gate. The pitches between the source and drain contacts can be determined by the following expression:

$$\text{Pitch.N} = (\text{CO.E} + \text{PO.C}) \times 2 + \text{PO.W} + \text{CO.W}, \quad (2)$$

where Pitch.N (1080nm) is the pitch between the source and drain contacts of narrow MOSFETs, CO.E is the minimum extension of the source or drain region beyond a contact, PO.C is the minimum clearance from a gate to a source or drain region, PO.W is the minimum gate length, and CO.W is the size of the contact. For wider MOSFETs (Fig. 6.b), the pitch between the source and drain contacts is determined by the following expression:

$$\text{Pitch.W} = \text{CO.C} \times 2 + \text{PO.W} + \text{CO.W}, \quad (3)$$

where Pitch.W (980nm) is the pitch between the source and drain contacts of wider MOSFETs, and CO.C is the minimum clearance from a contact to a gate. Typically, CO.E + PO.C is larger than CO.C. As the result, Pitch.N is larger than Pitch.W.

After shrinking the contact size, and hence CO.W in Eq.2 and Eq.3, by 10%, Pitch.N and Pitch.W can be reduced to 1050nm and 950nm respectively. The horizontal grid pitch should be the most common pitch in the width direction, i.e., either Pitch.N or Pitch.W. If Pitch.W (950nm) is used as the horizontal grid pitch, Pitch.N must be increased from 1050nm to twice of Pitch.W (2x950nm=1900nm). On the other hand, if Pitch.N (1050nm) is used as grid pitch, Pitch.W will increase from 950nm to 1050nm to match the grid pitch. Therefore, although only a small fraction of MOSFETs are narrow, Pitch.N must be used as the grid pitch to avoid a significant area increase. However, the average area increase is still on the order of 5% because all wide MOSFETs need to increase their source and drain pitches. This approach will result in standard cell area increase despite a 10% contact pitch shrink.

3.2.2.1 Minimum width of MOSFETs

A solution to the double peak problem, which leads to significant area increase after layout modification, is to decrease the number of peaks to one. One approach is to increase the minimum MOSFET width such that the peak corresponding to narrow MOSFETs is eliminated. Pitch.W, 950nm after shrinking the contact size by 10%, can then be used as the grid pitch in the width direction. Increasing the transistor width also reduces the on-resistance of MOSFETs. It should be mentioned that there are also the disadvantages of this modification including decreasing of the range of available transistor widths and the increasing of gate capacitance of MOSFETs. The effects of increasing the minimum width of MOSFETs require further study.

3.2.2.2 Gate contacts

In addition to the source and drain contacts, there are also poly-silicon contacts that connects poly-silicon lines or gates to metal-1 lines. Fig.7 shows the configuration of a MOSFET with gate, source, and drain contacts. Gate contacts are often placed in the middle of the source and drain contacts in the width direction. If the pitch between source and drain contacts is used as the grid pitch, the gate contacts have to be aligned with the source or drain contacts.

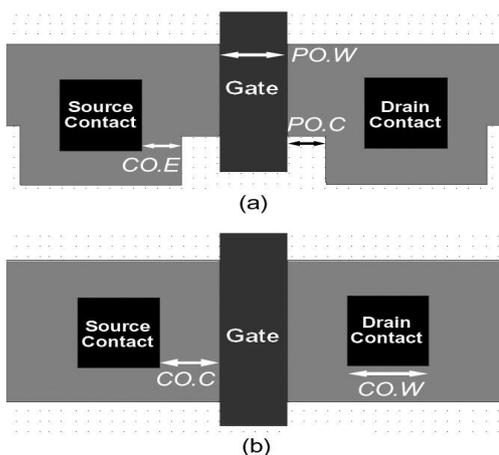


Figure 6: Contact pitch of narrow MOSFETs (a) is larger than that of wide MOSFETs (b).

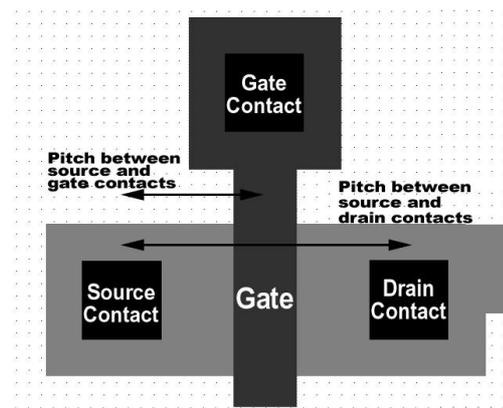


Figure 7: Gate contacts are placed in the middle of source and drain contacts.

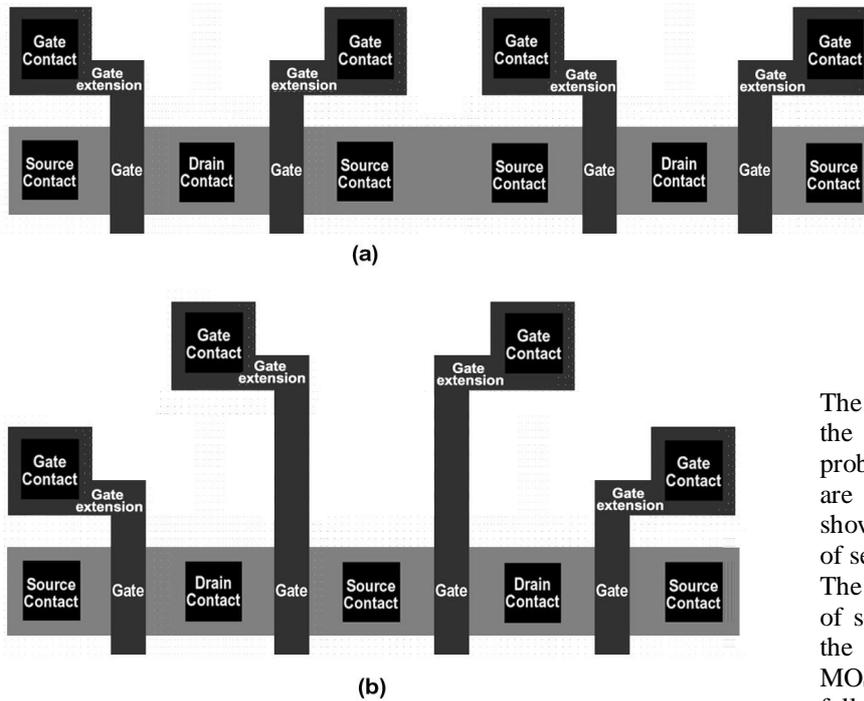


Figure 8: Two scenarios of snapping contacts of series MOSFETs connection to the grid. The grid pitch is assumed to be the pitch of source and drain contacts.

The alignment of the gate contacts with the source and drain contacts creates a problem when more than two MOSFETs are connected in series (Fig.5). Fig.8 shows two scenarios of snapping contacts of series MOSFETs connection to the grid. The grid pitch is assumed to be the pitch of source and drain contacts. Because of the gate extension of the previous MOSFET, the gate contact of the following MOSFET have to be moved

apart to satisfy the minimum spacing requirement of poly-silicon layer. Regardless of displacement in the height direction (Fig.8.a) or in the width direction (Fig.8.b), extra space is needed to snap the gate contacts on grid. Since the case shown in Fig.8 occurs frequently, using the source and drain contact pitch as the grid pitch will increase circuit area significantly.

Similar to the vertical grid pitch, it is desirable to use the pitch between the gate and source contacts (475nm)—half of that between the source and drain contacts—as the horizontal grid pitch. Then configurations such as that shown in Fig.5 can be accommodated easily. Although 475nm is beyond the resolution limit of the technology (540nm), the method we propose below enables its fabrication.

3.3 Double exposure

Two main conclusions can be drawn from the discussion in Sec.3.2.1 for the vertical grid pitch and Sec.3.2.2 for the horizontal grid pitch. First, pitches in the width and the height direction should be chosen separately according to layout configurations in these directions. Second, the desired grid pitches (320nm in the height direction and 475nm in the width direction) to avoid a significant area increase are smaller than the shrunk minimum contact pitch (540nm). Although the desired grid pitches are beyond the resolution limit, this kind of grid is still manufacturable because contacts do not occupy neighboring grid points. The actual spacings of contacts satisfy the design rules. That means the dense grid can be decomposed into several sparser grids, each of which is within the resolution limit. The contacts on these sparser grids are turned into different masks which are sequentially exposed to form the contact level image. Since the original dense grid is decomposed into several sparser ones, it is called the “virtual grid”.

To balance the extra cost for additional masks with the desire to reduce the grid pitches in both the width and height directions, double exposure is the best choice. Fig. 9 illustrates our proposed approach. Suppose we need to image a dense grid consisting of contacts and assist contacts with the horizontal grid pitch P_x and the vertical grid pitch P_y , both beyond the resolution limit (Fig.9.a). We can decompose the dense grid into two sparser grids with their axes rotated, as shown in Fig.9.b and Fig.9.c. The rotated grid pitch of sparse grids, P_{\min} , is determined by the following equation:

$$P_{\min} = \sqrt{(P_x)^2 + (P_y)^2} . \quad (4)$$

That means the rotated grid pitch P_{\min} is about 40% larger than the virtual grid pitches P_x and P_y if $P_x = P_y$. In the technology under consideration, $P_x = 475\text{nm}$ and $P_y = 320\text{nm}$. This gives a P_{\min} of 570nm, larger than the shrunk

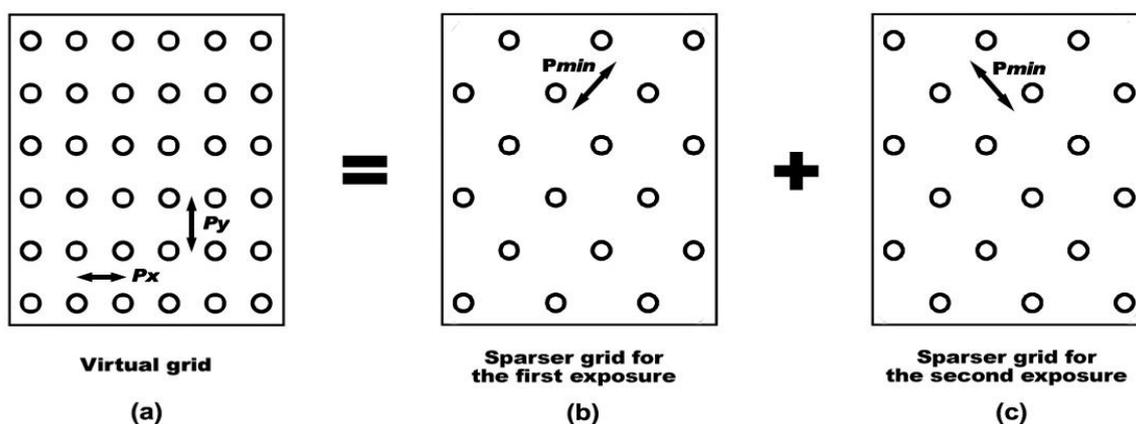


Figure 9: Exposure of the dense virtual grid by combining two exposures of sparser grids.

minimum contact pitch (540nm). A first mask images the contacts on one sparse grid onto the die during the first exposure while a second mask images contacts on the other sparse grid onto the die during the second exposure. The double exposures of the first and second masks produce all of the contacts on the dense grid onto the die. It should be reminded that this method works because no contacts are neighbors on the virtual grid.

In addition, illumination source configuration should be adjusted to the double exposures. To image the grid in Fig.1.f, quadrupole illumination with poles at 45° (Fig.10.a) is optimal. However, for imaging of the grid shown in Fig.9.b and Fig.9.c, the poles should be placed at the vertical and horizontal axes (Fig.10.b) to make an optimized projection.

3.4 Placement in physical design

During placement in physical design, cells are placed in rows with the power supply paths overlapped. To keep a global contact grid across the whole die, the virtual grid of different cell must match. This leads to additional restrictions during standard cells placement which are not considered in this study.

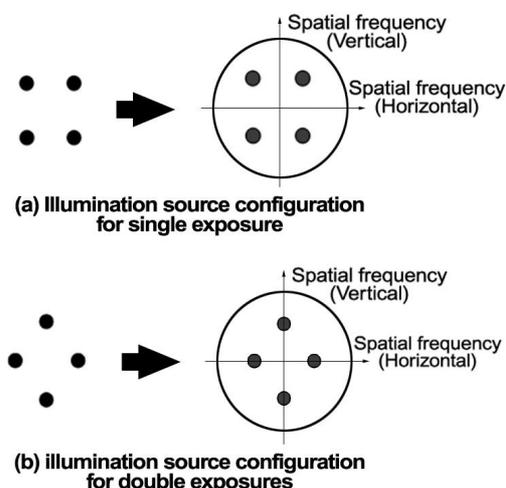


Figure 10: Poles of quadrupole illumination source should be placed at the axes for double exposure.

4 CIRCUIT PERFORMANCE

4.1 Area

To study the effect of grid-placed contacts on standard cell area, 84 cells in a 250nm library are modified by shrinking the minimum contact pitch by 10% and snapping all contacts on the grid with 320nm as the virtual pitch and 475nm as horizontal pitch. The area increase differs from cell to cell. Combinational cells generally have smaller area increase than sequential cells. Because the outputs of combinational cells are the function of the inputs only, the layout of combinational cells are typically more regular than that of sequential cells, whose outputs are functions of both the inputs and the outputs. This result also can be understood from Fig.3.a, which shows that the contact pitches of combinational cells in the width direction have well-defined peaks. Changes of cell area are plotted in Fig. 11, which shows the histogram of percentage area change of 84 standard cells. The percentage changes of standard cell area ranges from -21% for an inverter to +27% for a D-type transparent latch with a median decrease of 5%.

Although the average area change is -5%, it cannot be concluded that the area of circuits will decrease by 5%. Since different circuits use different combination of standard cells, changes in circuit area will also be vary from circuit to circuit. Two circuits, a finite impulse response (FIR) filter and an add-compare-select (ACS) unit in the Viterbi decoder, are designed using the modified standard cells to study the effect on circuit area. The area of FIR circuit decreases by 4% while that of the ACS unit shrinks by 2%. These initial results are encouraging, although more extensive data are needed for a definitive conclusion.

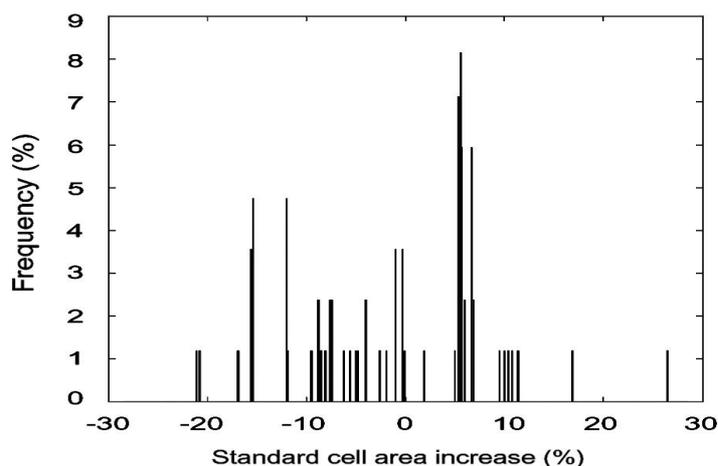


Figure 11: Histogram of percentage area change of 84 standard cells after snapping contacts on grid. (minimum contact pitch is shrunk by 10%)

4.2 Delay and power consumption

Circuit delay and power consumption depends on parasitic capacitance and resistance. Parasitic resistance, arising mainly from MOSFET on-resistance, are virtual unaffected by placing contacts on grid. Widening of narrow MOSFETs actually decrease the on-resistance proportionally. On the other hand, parasitic capacitance depends strongly on the layout. The main sources of parasitic capacitance are the source and drain capacitance and the gate capacitance. They contribute approximately 45% and 50% to the total parasitic capacitance. Although widening of narrow MOSFETs increases the gate capacitor slightly, the increase is so small that the total gate capacitance can be considered unchanged. However, the source and drain capacitance is proportional to the active area dimension. The percentage change of the total parasitic capacitance is therefore roughly half of that of the cell area. This change of capacitance is reflected in the intrinsic delay and power consumption. The intrinsic delay and power consumption of 10 standard cells examined fit this prediction well.

On the other hand, parasitic capacitance depends strongly on the layout. The main sources of parasitic capacitance are the source and drain capacitance and the gate capacitance. They contribute approximately 45% and 50% to the total parasitic capacitance. Although widening of narrow MOSFETs increases the gate capacitor slightly, the increase is so small that the total gate capacitance can be considered unchanged. However, the source and drain capacitance is proportional to the active area dimension. The percentage change of the total parasitic capacitance is therefore roughly half of that of the cell area. This change of capacitance is reflected in the intrinsic delay and power consumption. The intrinsic delay and power consumption of 10 standard cells examined fit this prediction well.

5 CONCLUSIONS

Fabrication-friendly layout does not necessarily mean circuits area increase. Introduction of the virtual grid concept and the use of double exposure make it possible to place contacts onto a dense virtual grid with pitches beyond the resolution limit, allowing more freedom for fabrication-friendly layout designs and lithography optimization. By examining the distribution of contact pitches of standard cells, shrinking minimum contact pitch by 10%, placing contacts on the dense virtual grid and using double exposure, circuit performance can be improved, even with lithography-friendly layout restrictions.

REFERENCES

1. B. J. Lin, "Method to point optical images at low- k_1 factors", V. Pol, ed., vol. 1264, pp. 2-13, Proc. SPIE, 1990.
2. A. K. Wong, *Resolution Enhancement Techniques in Optical Lithography*, SPIE Press, Washington, 2001.
3. T. A. Brunner, "Pushing the limits of lithography for IC production," in *Proceedings of the IEEE Electron Devices Meeting*, pp. 9-13, IEEE, 1997.
4. A. K. Wong "Microlithography: trends, challenges, solutions, and their impact on design," *IEEE Micro*, Mar/Apr. 2003.
5. A. Rosenbluth, S. Bukofsky, C. Fonseca, M. Hibbs, K. Lai, A. Molless, R. Singh and A. Wong, "Optimum mask and source patterns to print a given shape", *Journal of Microlithography, Microfabrication and Microsystems*, vol.1, no.1, pp.13--30, Apr. 2002.