CHAPTER 2  Visualizing Device Structure From Layout and Process Flow

Process flow simulation TCAD tools are used in this chapter to show how the planar process converts the layout into the device and interconnect structures. A good test of the working knowledge of a new process engineer or even a circuit designer is to hand them a circuit layout and ask them to sketch the device cross section. This requires an understanding of the ordering and patterning various thin-film layers during the process flow. The simulation of these structures can be carried out in seconds by extruding masks according to layer thicknesses, in minutes by adding topography 2D transition effects and step coverage simulation, or hours by including rigorous simulation of the lateral movement of dopants atoms in the silicon and/or 3D topography effects. The device and interconnect structures associated with a double level metal CMOS process for a logic chip design are used for basic illustration. Examples to explore process technology issues and their assessment with test structures are then considered. The role of the alignment process and the design/preservation of the alignment marks is emphasized. The conventions used for layout specification and typical back-of-the-envelope models used for rapid simulation are also given. Finally the issues in data structures used to describe the device topography and associated impurity doping attributes in communicating with rigorous process simulators are described.

2.1 Role of Process Simulation From the Layout

The microfabrication of integrated circuits typically requires the sequential application of 100 process steps involving over 15 masking patterns. The optimization of the manufacturing process and the resulting circuit performance requires making trade-offs in the complex interrelationships between numerous process parameters for each process step and many layout pattern details at each mask level. Integrated Circuit designers who focus on circuit aspects and process technologists who pursue unit processes need to be able to understand the consequences of their design decisions as these decisions propagate...
throughout the complete IC microfabrication process. The importance of this understanding to industry was dramatically illustrated to the author in 1979 by Milt Gosney [rzz:Gosney78] by a series of 100 SEM micrographs of catastrophic device features from a single chip which were the result of errors in circuit layout. Today Circuit design CAD systems with automatic design rule checking can catch many of these errors. However, the wise circuit designer who understands the technology issues can suggest clever trade-offs in negotiating design rules with process developers. This is no simple task because as devices and features shrink layout feature interactions with their surroundings are becoming much more complex. There are well known proximity effects between patterns due to optical imaging effects, plasma etching effects, and even point defect effects in oxidation and diffusion. Also due to the higher aspect ratios the length of the sequence of process steps over which topography effects can propagate is higher. Finally, systems level integration of optical communication devices and sensors and actuators in bring in entirely new structures and fabrication approaches.

Simulation can aid the product design process in many ways but perhaps its most profound impact is to allow the circuit design and process development to be tightly coupled. The overall CAD environment for integrated circuits is shown in Figure 2.1 [rzz:Lee.vlsi.83]. When a system specification is given, the system is partitioned into several blocks according to their functions. Circuit designers implement function blocks with circuit elements. The circuits are simulated through two loops: at first without layout considerations, and then with layout. An additional CAD effort is necessary in device design which starts from the bottom in Figure pflrslevn. This path determines the physical characteristics of each element fabricated on the chip in the integrated circuit. As a process sequence is given, the process is simulated and modified along with experimental feedback. The devices fabricated from the process are characterized and the resulting models are supplied to the circuit simulator as shown in Figure pflrslevn by means of the dotted line. The final layout and process steps are determined by these two main paths: one which starts from the system specification and the other from the process specification. Tight coupling between circuit and device design is possible through simulation as is indicated in Figure pflrslevn by the solid interconnection of circuit and device simulation through the box labeled SIMPL-2. This allows the same data bases for layout and process flow to be used for communication and in exploring circuit and device consequences. Design rules which were once typically finalized 2 years in advance can now be more dynamic adapted to reap the benefits of making small compromises as the design unfolds.

An example of showing the interrelationship between the layout view of the circuit designer and the process flow view of the technologist is shown in Figure CP.1 (in the color plate section). Here SIMmulation of Profiles from Layout (SIMPL) has been used to generate the cross section of a CMOS device. The designer’s view of the mask set is shown in the upper portion (or view port) of the display. Note that the masks are defined by the color keys on the left of the display. A menu of process steps such as that at the bottom of the display is used to sequentially specify the process steps and associated physical processing parameters. As the process proceeds the effects of layout and process flow on the impurity distribution and the topographical features can be seen in the lower view port. Each material or doping concentration is also defined by the color keys on either the left or right of the display. More detailed information is given in Figure CP.2 which shows a magnified cross section and the doping concentration along a ‘cut-line’ made in the cross section. The doping is positive or p-type in the source/drain region, then negative or n-type in the well, and finally positive again in the substrate.
Simulators for the integration of layout and process flow have been developed throughout the semiconductor industry. In the work began in the early 80’s on SIMPL [rzz:Grimm.IEDM.1983, rzz:Lee.VLSI.1983] and on OYSTER [rzz:Koppleman.83] which is based on solid modeling. (Stanford IEDM 1994, mention extruding). The FABRICS program which emphasizes statistical aspects of processing was also generalized to visualize device cross sections [rzz:FABRICS]. Topography effects from process flow specifications have also been explored by xxx [rzz:Kat87]. [xxx Other work from Japan rzz:Spirit] [xxx East Germany work rzz:East germany] In the late 80’s commercial software began taking on the flavor of simulating process flow sequences in programs such as Depict rzz:Depict TMA and Market rzz:Silvaco. The early 90’s is the era of integrated TCAD environments and collaboration on standards for communicating between rigorous simulators where linking to the layout has become one of the important function components envisioned in most systems. Further information on TCAD environments is given in Chapter 11.

Many of the examples in this chapter and monograph have been generated using the SIMPL program. SIMPL was begun in 1981 following the suggestion of Milt Gosney of MOS-Tech that IC designers badly needed such a tool and with the cooperation of a visiting technologist Yoshi Sakai of Hitachi who provided the inspiration to undertake an N-well CMOS process. Simultaneously Mike Grimm started working on the 1-D extrusion version [rzz:Grimm.IEDM82] and Ken Lee started working on the 2D version [rzz:Lee.VLSI.1985]. The design aids which are the focus of the current discussion were added by Joe Wu [rzz:WU.1987]. Many useful interface features such as invoking aerial images from SPLAT and including doping core samples from SUPREM III were added by Ed Scheckler, Robert Wang, and John Camagna [rzz:VLSI.scheckler]. The ability to bilat-
eraly link with SUPERM IV was added by Robert Wang [rzz.wang.chin]. And major improvements in the data structures, algorithms and features for MEM applications were added by Andre Gabara and Derek Lee [rz.Lee.thesis]. The best source of information on current features in SIMPL is the user guide and the reader may be interested in obtaining a copy of SIMPL to make further explorations [rzz.ILP.software].

2.2 The Device Fabrication Process Sequence

The relationship between layout and device cross-section will first be illustrated using a CMOS example from SIMPL-1 from the work of M. Grimm et al. [rzz.Grimm.IEDM.82]. Figure CP.3 shows the layout for a CMOS inverter and the resulting device cross section. The layout is at the bottom of the figure and the heavy black line is the ‘cut-line’ along which the cross-section is desired. In the layout red is used for polysilicon, pink for N-well, brown for active area, green for p-source/drain implant, black for contacts, and light blue for metal. The red poly mask crossing the brown active area creates the self-aligned gate of the n-channel and p-channel devices. These devices can be seen in the cross-section in the upper view port as the shallow green n-implant interrupted by the red poly gate and the brown p-implant interrupted by the red poly gate. The field is the region with a thick oxide above the silicon. The devices are first insulated by an oxide layer and then openings are made for contacts, and a blanket metal is deposited and patterned to make the electrical circuit. Note that in this inverter the metal interconnects the sources of the two devices as the output and that the metal contact to the drain of the p-channel device which is in the green N-well also makes contact through a bright green n+ region to the well itself.

The process information is typically specified by a series of process steps and associated parameters. An example of the CMOS process flow for SIMPL-1 is given in Figure 2.2. Although a total of 60 process steps is required the entire process can be built from only a few generic commands for deposition, masking, etching and doping are needed. In many cases a series of commands frequently occurs in the same sequence such as deposit-mask-etch. This sequence is often referred to as a unit process such as lithography module. The material type code names such as NTRD, OX, and POLY which refer to nitride, oxide and polysilicon while RST and ERST correspond to resist and exposed resist. Additional parameters in the command give the name and polarity of a mask to be used or specify physical model information such as the thickness of deposition or etching.

The evolution of the device topography during processing can be followed by examining the cross section at various stages of the process flow in Figure CP.4. First in a) a oxide/nitride ((light purple/blue) stack is deposited and patterned so that it remains on the N-well area. The oxide under the nitride is to prevent damage to the silicon surface during oxidation. In b) a LOCUS oxidation (light purple) is then done, the oxide/nitride stack removed, and the N-well implant and drive (green). In SIMPL-1 the doping is shown at its final rather than initial depth to save simulation time. In c) a new oxide/nitride ((light purple/blue) stack is deposited and patterned to remain in the active areas for the p-channel (N-well) and n-channel devices. In d) and additional photoresist (red-pink) is applied, opened outside the well and a boron implant (brown) is made which will form the channel stop in the field in the field region. In e) the second LOCUS is carried out to create the field oxide ((light purple) and the oxide/nitride stack is removed. At this point in f) the device forma-
The Device Fabrication Process Sequence

FIGURE 2.2

Listing of process steps and key parameters for an N-well CMOS process flow using SIMPL-1 commands [rzz.grimm.iedm].

A simple one-to-one correspondence between each device feature and a single mask does not exist for most device technologies. The circuit designer usually make a layout using a mathematical set of mask levels from which the physical mask levels can be derived from Boolean operations. So too the process itself combines mask shapes in Boolean like manner to create device features. In N-well CMOS process just described, the contact to the N-well does not appear explicitly in the layout. However, the contact clearly stands out in the cross-sectional view as the green N-well with an additional highly doped (dark green) region touches the blue metal. This region is formed by the contact and metal patterns acting in concert with n-type source/drain implant (complement of the p-type implant mask). More subtle device technology features such as the threshold implant in the channel region (light yellow under red poly in green N-well) and the difference in height of the p-channel and n-channel devices also appear.

The famous “planar process” is really noting more than the generic operation of combining layout with process flow. An example for a bipolar device technology is shown in Figure CP.5. Here in the layout the buried layer is green, the base is yellow, the emitter and collector contacts are green, the p-type isolation is pink and the metal is light blue. In the cross section the darker green buried layer protrudes below the green collector which is broken by the p-type isolation diffusion. The small green n-type emitter can clearly be
seen inside the shallow yellow p-type base diffusion. The high n-type doping under the metal contacts to the collector can also be seen. A difference in the height of the silicon collector, base, and emitter-contacts also occurs.

In developing processes the technologist is generally most concerned with rapid topography changes and horizontal encroachment of impurities in what might be termed lateral effects. To realistically predict these effects 2D and even 3D simulators are needed. These lateral effects can be seen in Figure CP.1 discussed earlier. Here the spreading of the well doping is apparent. It is well known that the source/drain junction depth must be kept much smaller than the feature size to reduce second order effects in the device operation. Step coverage effects at the contacts are also apparent in Figure CP.1. Simulating these lateral effects requires additional sophistication in the models, algorithms, and even user interfaces which are the principal subjects throughout this nomograph.

2.3 Structures in a Double Level Metal CMOS Process

Meade and Conway created a new paradigm for circuit design by introducing scalable design rules to insulate the IC designer from the details of the process flow. This paradigm shift should not be undone lightly but there are advantages in look carefully and developing design tools which can look on the other side of the fence. For example the circuit designer who’s work is well protected by a design rule checker might be interested in seeing the structures his circuit is creating. A process engineer who is responsible for meeting the design rules should be adept at visualizing what structures a layout with misalignment would create. While the design rules are an contract to divide and conquer made in advance both the circuit designer and process engineer should take an active interest in ways to assist each other and perhaps to amend the design rules. As mentioned in Chapter 1 a fertile area for new Technology CAD and Computer Aided Manufacturing tools is bring together manufacturing, process development and circuit design so that these traditional areas of expertise can work concurrently together.

To illustrate how simulation can help tie together circuit design and process development a circuit layout for a double level metal CMOS process will be considered. This particular process corresponds to the standard process used in the textbook by J.M. Rabaey [rzz.rabaey] on digital circuit design. It is based on the MOSIS CMOS design rules. A generic process flow for this P-well process was first written and SIMPL cross sections were expressly generated by D. Lee for both this monograph and the text by Rabaey. A layout of a logic circuit made for this process is shown in color plate Figure CP.6 and below in black and white in Figure 2.3. In the color (black and white) layout diffusion or active area is green (northeast to southwest crosshatched), polysilicon is red (northwest to southeast crosshatched), first level metal is blue (no stipple pattern but outline), second level metal is magenta (double crosshatched), contacts between first level metal and source/drain or polysilicon are black (square with X), and contacts between second and first level metal are black (square with diamond). The minimum feature size is 2λ where λ is 0.6 μm. The contacts are shown with the 4λ borders but the contact hole itself is 2λ wide. As will be shown later it is important that contacts are not allowed directly between second level metal and source/drains or polysilicon.
A generic process flow for the double level metal CMOS process is shown in Figure 2.4.

Steps 1 through 27 are the front-of-the-line which builds the basic device structure in the
silicon. The N-well is patterned in step 3 by creating an opening in a nitride layer on a thin pad oxide. An n-type phosphorous implant is then made and is driven in while simultaneously growing a 0.3 μm oxide on top of the N-well in a LOCOS process. (This oxide could be used to mask an P-well implant for a twin-well process.) The surface is then stripped down to the silicon and another nitride on oxide LOCOS pad is created and opening the field region. Using the N-well mask the (in the opposite tone) the undoped p-substrate is implanted with boron to raise the doping as a channel stop. The field oxide is then grown by LOCOS. Threshold implants are then carried out separately for the NMOS and PMOS active areas. Polysilicon is deposited and patterned. This is followed by the source and drain implants for the NMOS and PMOS devices. An finally the waver is coated with a deposited oxide layer.

The back-of-the-line processes in steps 28 through 39 provide the interconnection of the devices. Contacts are first opened to the polysilicon and source/drain regions. Metal is then deposited, coated with resist, patterned and etched. A interlayer dielectric is then deposited. This process is repeated again to create the second level of interconnect.

In modern technologies there are many important lateral 2D and even 3D process effects which must be examined. An example of 2D2D process effects for the CMOS inverter is given in Figure CP.7 which is a SIMPL-2 simulation. The mask and material definitions are now displayed at the sides and the vertical scale is now three times larger than the horizontal scale. The oxide (light purple) now has a taper shape or bird’s beak. The lateral diffusion of the N-well which is one of the main causes of the large separation between the p-channel and n-channel devices is now clear. The thinning of metal at steps shows how the topography generated at the opening of the contacts affects step coverage.

A cut-line BB’ in Figure 2.3 can be used to show interconnect features. On the layout this cut-line passes left to right through metal M2, a metal M1 to poly contact (square with diagonal cross), along metal M1, through metal M1 to source/drain contact. In the cross section in Figure CP.8 the metal M2 is the upper most layer and Polysilicon POLY is on top of the field oxide. Metal M1 can be seen first going into a contact to meet the Poly and then traveling on the dielectric D1 to eventually go into the source/drain. Note the two-dimensional effects in filling the contacts. The lateral spread during deposition also encroaches the region of metal M1 and begins to affect its shape.

### 2.4 Design Aid From Process Flow Simulation

This section illustrates the power of design aids at the process flow level. The circuit designer, process technologist and manufacturing engineer must be concerned with many issues in developing a manufacturable product. It is important to understand how well the device topography will work when not every process parameter or alignment is perfect. For example, topography driven design rules for layout which are tolerant to nominal variations must be developed. Process flow simulation is a rich environment for facilitating design for manufacturability with design aids. For example, the technologist who used to have to sketch 40 cross sections by hand to explore misalignment scenarios could automate this process. Looking through a layout for actual occurrences of slightly marginal combinations of masks in a design and generating cross sections could be automated. Also the experienced eye of the technologist in critiquing the quality of features in the resulting cross section for potential problems could be automated.
Several features in the program for SIMulation of Process from the Layout will be used to illustrate features typical of process flow simulators. The menu of features in SIMPL brought by the SIMPL-DIX graphical user interface (GUI) is shown in Figure 2.5. The rectangular boxes indicate that the goals which are typical of any of the TCAD system are to link **rigorous process and device simulation** with **layout and profile editing** in a context where **process flow is managed** and higher level **design tools can be invoked**. SUPREM, CREEP and PISCES represent the process and device simulators. HUNCH, CRITIC and WORST represent design aids for identifying potential problems, critiquing simulated device cross sections and identifying worst case misalignment scenarios.

By invoking various unit process steps, SIMPL can serve as a process flow manager. Figure 2.6 shows how SIMPL serves as a process flow manager for back of the line processes sequences. materials may be deposited in various manners, modified by shrink or reflowed, and then removed by various means. The user is free to select any sequence of processes. Each process simulator communicates from and to a standard profile description to make arbitrary process ordering possible.

An example of the process flow language used to include 2D process effects in SIMPL is shown in Figure 2.7. The advantage of the dialog based approach is that the user need not learn in advance the menu of commands and their formats. Also the user only need get exposed to the level of physical models which they select. With modern interactive graphical capabilities most TCAD interfaces use point and click menus. Typically the technologist would select a predefined process, branch if necessary to change tooling or materials, and branch further if necessary to define the detailed parameters of the equipment or materials.
More detailed information about the doping in the substrate can be generated by asking SIMPL to generate an input file and running for SUPREM3 [rze.scheckler.vlsi]. When the SUPREM command is toggled on the dialog with the user is expanded to include more details of the thermal processes such as time and temperature as required by SUPREM3. This user interview to build the expanded process file only need be generated once and
saved. It can then be used to have SUPREM3 generate the vertical doping profile beneath any point selected by the user in the layout. Typical choices for locations are beneath the gate, in the source/drain, and in the two field regions. SUPREM3 includes the calculation of threshold voltages and sheet resistances which provides important feedback for process development.

The design rules if properly specified permit reasonable misalignment within allowed limits to occur without creating catastrophic errors. This can be illustrated with the double level metal CMOS process as follows. Consider the possibility of a short occurring between the poly and met1 which are separated by only $\lambda$, along the cut-line C-C' in Figure 2.3. The design rules permit con1 to be misaligned to poly by shifting upward by $\lambda$, and met1 to be misaligned to con1 by shifting upward $\lambda$. This means that along the cut-line C-C' in the layout that met1 may not only touch poly but overlap it by a full $\lambda$. It appears that this may cause a short. The WORST function in the SIMPL program was selected, the misalignments introduced, and the corresponding cross section was generated as shown in Figure CP.9. The new layout in the top view port shows the overlap. The cross section in the lower viewport shows that while unwanted topography may be created it does not result in a short due to the presence of the oxide layer between the poly and met1. The real limit to the misalignment of poly, con1 and met1 occurs when the inner core of the contact reaches poly and this can occur only when con1 is misaligned by $2\lambda$ with respect to poly.

Often the technologists have a hunch as to what might lead to a production problem or gotcha and they would like to look through the circuit design layout to see if and where those circumstances occur. The ‘HUNCH’ feature in SIMPL illustrates how these layout dependent problems can be identified. The designer first selects a layout and introduces misalignments if desired using WORST. The ‘HUNCH’ option screen shown in Figure CP.10 allows the user to select sets of masks and Boolean operations on them such as metal crossing poly near a contact. This new mask with an assigned name such as PROB is then highlighted on the layout. All occurrences will thus be displayed in the layout. Figure CP.11 shows outlines of the mask edges with the problem area highlighted in bright yellow. The designer can then specify a ‘cut-line’ through this problematical area for generation of the cross sectional view. Note that the HUNCH definitions are also used to select which of the other mask levels of interest are displayed and in which format such as using outlines.

Technologists usually have calibrated eyeballs for examining SEM cross sections and spotting potential problems. They are generally looking at the topography features and recognizing particular shapes which historically have lead to problems. Process flow simulators can also provide this service. In SIMPL once a cross section has been generated the ‘CRITIC’ feature can be invoked. A menu of specifications to meet then appears from which the user selects. The ‘CRITIC’ is then used to highlight the region in the cross section not meeting a specification such as the thin of metal on the sides of the steps as shown in Figure CP.12.

### 2.5 Process Technology Issues and Test Structures

One challenge for topography simulation is to go ‘from masks to memory’. That is, starting from a set of masks or layout for a memory cell create for the designer a cross sec-
tional view of the device along any ‘cut-line’. The layout for a typical 1 Mbit DRAM cell is shown in Figure CP.13 of the color plate section. There are two poly layers PO1 and PO2 and two metal layers MTL1 and MTL2. A diagonal ‘cut-line’ is specified from a metal-to-polysilicon contact through a metal-to-metal via. Using about 120 process steps the DRAM cross section in Figure CP.13 was simulated in SIMPL-2. This ‘cut-line’ throughout the via shows how a signal on upper metal MTL2 passes through the via to the lower metal layer MTL1 which in turn contacts the upper poly layer PO2 which is the gate of the transfer device.

The process technology used to create the dielectric insulator profoundly affects the quality of the interconnect metal. In the previous example a high temperature reflow of the deposited Phousphosilicate glass was simulated using CREEP to fill in the cracks at steps in the deposited oxide insulator. This results in the acceptable step coverage of MTL1 as can be seen along the horizontal cut-line in Figure CP.14. If the reflow step is omitted sever step coverage problems result as can be seen in Figure CP.15. All abrupt topography edges tend to create cracks through shadowing during the deposition. These cracks continue to create shadowing effects during metal deposition which can result in open circuits. This topography propagation from one unit process step to the next is one of the most serious problems in process integration and may require some modification of several of the process steps in the sequence. Recently a new process planarization step of chemical-mechanical polishing (CMP) has been introduced in an attempt to return the wafer surface to a nearly planar state.

The integration of microelectromechanical (MEM) devices with integrated circuits is of interest for many new applications such as sensors and actuators. TCAD simulation can be very helpful in bring together the process flow for the IC devices with that for the MEM device. A CMOS IC device layout and cross section are shown in Figure CP.16 [rzz.imems.prl]. The layout for an interdigitated MEM capacitor to form an accelerometer on this chip is shown in Figure CP.17. In etching the thick polysilicon to form the fingers polysilicon stringers can result from height variations in the underlying topography as shown in Figure CP.17. Here thick polysilicon must be patterned to separate the moving fingers from the fixed structure. Near the ends of the fingers the underlying features have create trenches which fill in nicely when polysilicon is deposited. The polysilicon which becomes extra thick over these depressions must be over etched to remove all the polysilicon and prevent stringers. The polysilicon will be released from the substrate by a wet etch of the oxide. However, after this release the two polysilicon structures must become electrically independent and the stringer will prevent this.

BiCMOS presents another major process integration challenge in simultaneously forming the bipolar and CMOS devices. An example is shown in Figure CP.18 [rzz.cole and logan]. Here the complete formation of the topographical features on the wafer has been simulated. The bipolar device is on the left, the PMOS in the center and the NMOS on the right. The devices occur at different physical heights and the bird’s beak oxide formations between these active devices are quite apparent. The polycide gate construction and oxide spaces utilized in this advanced technology are also visible. The impurity concentration in the silicon is not shown in complete detail and in the process of extending SIMPL to this technology the ability to display the concentration of the An, As, B and P dopants was added in SIMPL-System-5.
It is very desirable to make assessments from the interconnect and device structure on their potential impact on electrical performance. In the case of topography effects on interconnect this can be carried out by linking the topographical information to an electrical analysis program such as RACPLE in 2D [rzzLee.racple] and FastCap and FastHenry in 3D [rzzWhite.fastcap]. Figure CP.19 shows and example of invoking the RACPLE program in SIMPL. Here the number of equivalent squares in the width to height of the faces of the various topographical features which make up the structure has automatically been generated and displayed. A fast algorithm has been used which approximates the 2D solution to Laplace’s equation to a few percent. A more rigorous analysis is possible even in 3D by linking to FastCap for analysis. In this case, for example, the results of rigorous process simulators such as SAMPLE-3D must first be converted to solids and feed to the analysis program. This process is depicted in Figure 2.8 where the SAMPLE-3D topography is first clipped against cutting planes at material interfaces and simulation boundaries, then thin-triangles are removed, and finally the linking to FastCap is made [rzz.sefler.IEEE]. Without including detailed effects of the IC process on the geometry formation capacitance estimates are typically accurate to only 30%.

Process and device simulation can play a very important role in assessing possible scenarios to develop cost effective IC manufacturing. In developing new products it is often very desirable to reduce the mask and process step count to reduce the cost of manufacturing. In looking for alternatives the process developer must assess many different scenarios often involving differences in oxidation rates of materials, implant compensation and relative diffusion of different dopants. Clever techniques for saving masks are often introduced which go well beyond the self-aligned polysilicon gate. They include ideas like a nonselective $V_T$ implant, combinations of one nonselective with an additional or com-
pensating source/drain implant, dual As and Ph well implants which diffuse differently to form the channel stop and well, rapid oxidation of n⁺ poly compared to the silicon substrate to selectively grow a mask, and out diffusion of dopants from polysilicon depositions [rzz.sze.tech.parillo].

During the process development phase test structures are used to examine in detail every possible aspect of a unit process. This includes effects of material parameters, the processing equipment, and the wafer state or condition on arrival at the unit process step. The structures which are used encompass a variety of points of view. On one hand they attempt to represent the full spectrum mask geometries which might ever be encountered on product wafers. On the other hand, some of the structures are designed to discriminate one source of error from another. In optical projection printing this might be to distinguish among variations in mask making, aerial image changes across the lens field and resist processing variations across a wafer.

To create a scenario similar to the process flow a short loop of several process steps might be used. To deal with the complexity of parameters and the expected statistical fluctuations a statistical metrology using electrical test structures which can be automatically probed are preferred. A more restricted subset of test structures is included on production wafers to monitor device quality and diagnose production problems. These structures are added as a drop-in test structures to the completed die design and often are added in the unused area of the scribe lines. Process flow simulation is useful for understanding the nature of what are often unusual structures used for testing and for analyzing the degree to which they can discriminate one effect from another.

An example of a floor plan of a N-well CMOS test chip is shown in Figure 2.9 [rzz.290]. This floor plan contains structures for gathering information about device, process and manufacturing. The regularity of sets of 2 by 10 probe pads is quite apparent and indicates that the structures are designed to be electrically probed. The basic NMOS and PMOS test structures to determine the I versus V characteristics, V_T as a function of W/L and inverter characteristics are in the upper right area. Also three simple complete operating circuits are included. Much of the upper left of the chip is utilized to determine process related parameters such as sheet resistances of films and implants, linewidth and overlay. Among these are special test structures to assess special effects in lithography and etching. At the sides of the lower portion alignment patterns can be seen. Those on the left are the diagonal crosses used in the automatic alignment system and some of them include a square field to either create a suitable substrate region or protect a patterned mark until it is needed at a later process step. The cross in a box alignment patterns on the right hand side create features on the wafer which can easily be read with visual inspection. The large filled in areas of the chip are generally test structures for assessing yields typically contain contact chains and serpentine metal running over topography steps.

It is interesting to examine a typical electrical test structure in detail. The layout of a test structure for electrically measuring overlay is shown in Figure 2.10. Here the exact placement of the thin cross-hatched layer on top of the shaded layer affects the balance between the resistances of the remaining unblocked shading on the two sides. This unbalance in the resistance of the unblocked shading is a sensitive measure of alignment which can be electrically probed by the connections to the probe pads shown. Note that the misalignment in both the vertical and horizontal directions can be measured with this test structure.
In process development and trouble shooting it is often useful to develop systematic approaches to isolate causes of problems. Figure Figure 2.11 shows a set of layout patterns used to diagnose problems in separating polysilicon fingers in crossing dimple regions in a MEMs process. The dimples are used to create standoffs and prevent the sticking of the polysilicon fingers to the substrate. In etching to separate the fingers material was observed to remain at the bottom of the dimples for fingers at diagonal angles in a wagon wheel pattern. There are many possible causes which include linewidth variation at angles in mask making, increased resist thickness in dimple regions, and interference effects from reflection from polysilicon edges. to attempt to discriminate among these three sources of
error the systematic design of test structures in Figure pflmitntm. Here a variation in the uniformity of the radial lines in the pattern would indicate potential mask making or projection printer resolution changes with orientation. Observing a variation down the column where the dimple width increases would help identify effects of resist filling of the dimple and of interference from reflections of the polysilicon edges at steps. Observations of the printing quality across the field or with a rotated mask could help separate mask effects from exposure tool effects. Similarly changing the dimple depth and resist thickness might help distinguish resist height and reflection effects. An increase in linewidth size across the rows is also included as a mask design variable which might be modified to possibly reduce the sensitivity to the problem. Alternative solutions might be to over expose or over develop the photoresist.

2.6 Process Flow Issues in Designing the Alignment Methodology

2.6.1 Role of Alignment

The alignment methodology is extremely important as it is the dominant factor in determining the packing density of circuits which can be achieved for a given resolution. Determining the best methodology is a very complex problem as there are many mask levels which must be coordinated, there are multiple options as to which prior mask to align to, the intermediate process steps can erode or degrade the alignment marks on the wafer, the signals from these degraded marks may affect the tolerances which the exposure tool can
maintain, and the achievable tolerances in turn affect the best overall strategy of the which sequence of mask alignments is best.

Each mask at its point of use in the process flow contains an alignment key which is aligned to a previously created target mark on the wafer. Since some statistical error occurs in the placement of the features created by each mask it is generally the practice to choose a wafer topography feature created previously by a one particular mask level and align to it. This sequence of alignment steps can be characterized by an alignment diagram [rzzParrilo] such as that shown in Figure pflalgdia for the double level metal process. Certain masks and process steps such as the channel stop and threshold implants do not create new topographical features and thus are not convenient for creating alignment target marks. Except for these cases the remaining masks directly align to the previous mask. Thus the contacts in CON1 directly align to the polysilicon, the first level of metal MET1 aligns to the contacts, etc. Unfortunately in this scenario neither the contacts nor the vias align directly to the ACTV mask which creates the sources and drains. The contacts which align to POLY which in turn aligns to ACTV has an alignment level of indirection N of 2 while the vias in CON2 go through MET1, CON1, POLY, to ACTV has N of 4. If the second level of metal were to be allowed to contact sources and drains directly a different alignment sequence would be desirable.

The design rule agreement between the circuit designer and process development engineer must include an adequate minimum separation $S_{AB\text{MIN}}$ to allow for misalignment. This minimum separation in layout for any two mask levels is a function of the alignment tolerance and the linewidth variation of the features formed by the two mask levels. Assuming uncorrelated standard normal distributions of alignment (with standard deviation $\Delta$ for all alignments) and edge position variation $\delta_i$ for features from the ith mask the standard deviation $\Delta_{AB}$ for the minimum separation is [rzzParillo].

\[
\Delta_{AB} = \sqrt{N_{AB}\Delta^2 + \delta_A^2 + \delta_B^2} \quad \text{(EQ 2.6 -1)}
\]
Here A and B refer to the two mask levels and \( N_{AB} \) is their alignment level of indirection in the alignment sequence. In addition to the statistical variation an edge placement bias may occur for each mask level \( B_A \) and \( B_B \). The net bias \( B_{AB} \) between two mask levels is either the sum or the difference \( B_A \) of \( B_A \) and depending on the polarities and the additive or subtractive nature of the process. The choice of the minimum separation distance can be chosen by allowing sufficient tolerance for the both the net bias and edge position variation \( S_{AB}^{\text{MIN}} \) as

\[
S_{AB}^{\text{MIN}} = B_{AB} + N_\sigma \Delta_{AB} \tag{EQ 2.6 -2}
\]

The factor \( N_\sigma \) is chosen to provide the process capability \( c_{pk} \) desired for manufacturing.

While the above discussion illustrates the basic theory, the practice of developing the alignment methodology and its impact on design rules involves a lot of decisions with important trade-offs between yield and packing density. To illustrate this, suppose for example, that in the double level metal CMOS process described earlier a designer decided to use the second level metal to directly contact a source/drain region. This would of course be flagged by the IC CAD Layout system as a violation but for purposes of illustration we continue the example. Now using the previously specified alignment sequence for this process in Figure 2.12 allowable misalignment vectors are introduced consisting of the maximum drift to the left of one \( \lambda \) for the alignment of con1 to poly, met1 to con1, con2 to met1 and met2 the metal2. The cross section for this process sequence in Figure CP.20 shows that for this allowed alignment scenarios results in MTL2 landing far from the source/drain on the field oxide and failing to make contact. So even if the contact etching process with CON2 were modified to etch through thicker insulator direct contact of MTL2 to source/drain would not be possible without also tightening the alignment requirements. The requirement of a tighter alignment is almost always the case whenever a mask is required to simultaneously overlay features created by separate masks.

2.6.2 Design of Alignment Marks

A process technologist must also consider the generation, preservation and use of the alignment marks. This might be referred to as the care and feeding of alignment marks. This is an important part of the process design in practice because the topography evolution effects during processing can affect the size and style of marks which work best. Symmetrical marks are always used so that edge bias or edge taper on the wafer to first order cancels out.

The alignment methodology consists of designing the creation of the target mark on the wafer, its preservation during intermediate processing and its use with it companion alignment key on the subsequent mask level. This must be carried out in a coordinated manner such that all target marks and mask keys appear on the appropriate mask levels in the mask set. Figure 2.13 depicts the alignment process. A small subarea of the die is dedicated for alignment marks. Within that area a small region is used for each alignment step. As shown the level being currently aligned is a cross which is to be located inside a previously patterned box on the wafer. In this case the first mask (diffusion labeled green) creates a boxes at the alignment A1 and A2 levels. The second mask (thin oxide labeled magenta) has a cross which must fit in the green box at level A1 and it also creates a box a site A3. The third mask (contacts which is black) has a cross which must fit in the green
box at site A2. The fourth mask (metal labeled blue) has a cross which must align to the magenta cross at site A3.

The wafer at various stages of alignment and the patterns on the four mask levels are shown in Figure 2.14. In this scenario the cross is always an opaque chrome (dark) on the mask. Note that a window larger than the target box surrounds the cross. Although process details are not considered here it is necessary to work out with the resist polarity and pattern transfer a way to create the target and preserve it during the intermediate process steps.

Aside from problems in asymmetrical step coverage in creating marks there are also problems in creating sufficient strength of the alignment signal. Sometimes for almost no apparent reason the alignment signal will be extremely weak. This is due to subtle changes in the alignment mark geometry due to within specification changes in process parameters. An example of the lateral filling of trenches by deposition of metal is shown in Figure 2.15. Even the subsurface filling of trenches with polysilicon can be a problem if the filling height after etchback cannot be controlled to prevent interference effects [rzzWong].

### 2.7 Layout Standards

A popular format for exchange of layout data between software tools developed in universities is the Cal Tech Intermediate Form (CIF) defined by Conway and Meade [rzz Conway.mead]. CIF allows four kinds of primitives that specify geometric objects: box, polygon, flash and wire. The conventions for boxes and wires are for example.

\[
B \ W_b \ L_b \ X_c \ Y_c \ D_x \ D_y \tag{EQ 2.7 -1}
\]
FIGURE 2.14 Wafer and mask used at various stages to create, protect and use alignment marks.

FIGURE 2.15 Silicon trench of (a) 1 μm, (b) 2 μm and (c) 4 μm used as alignment marks. Note that the width, depth and even symmetry are degraded by intermediate deposition steps [rzz.flack].

\[ W\, W_b\, X_1\, Y_1\, X_2\, Y_2\, X_3\, Y_3. \]  

(EQ 2.7 -2)
The first character \( \text{B} \) or \( \text{W} \) specifies the primitive type. The following inputs are integers in layout units describing the size and location of the primitive. For a box \( W_b \) and \( L_b \) are the width and length, \( X_c \) and \( Y_c \) are the center, and \( D_x \) and \( D_y \) are (optional direction vectors with a default to +x axis) the direction along the length. For a wire \( W_w \) is the width of the wire and the \( X_n, Y_n \) pairs are the starting, turning and ending points along the wire. A CIF file which describes the layout of a CMOS inverter is given in Figure 2.16. Each mask layer is identified by the starting character \( \text{L} \) followed by a name such as \( \text{NWEL}, \text{PSD}, \text{ACTV}, \text{POLY}, \text{CONT}, \text{METL} \) to identify the associated N-well, p-type source and drain, active area, polysilicon, contacts and metal mask level. The layer definition is followed by a collection of primitive objects such as boxes and wires.

```
L NWEL;
B 4400 2400 -2200 0;
L PSD;
B 2800 1700 -1600 -50;
L ACTV;
B 3500 1000 -2250 0;
B 2500 1000 1000 0;
L POLY;
W 400 -1000 700 -1000 1200 1000 -1200 1000 700;
B 800 400 0 -1500;
L CONT;
B 10000 50 0 0;
B 400 400 0 -1400;
L POLY;
B 399 249 6799 9945;
L 0X;
B 1899 1206 9049 10044;
B 99 600 8049 10019;
B 199 600 7899 10019;
B 199 600 7299 10019;
B 199 600 7099 10019;
B 199 600 7299 10019;
B 199 600 6499 10019;
B 199 600 6299 10019;
```

FIGURE 2.16 Example of a CIF file for a CMOS inverter [rzz.grim.ms]

The shapes defined by \( \text{POLY} \) and \( \text{ACTV} \) layers are sketched in Figure 2.17. Here it is assumed that 100 CIF units is equal to 1 grid or scalable layout unit \( \lambda \). Thus for example, the center of the second ACTV box is located at \((18,0)\) and drawn with a length of 26 and width of 10. The wire in the poly layer starts at \((-18,7)\), goes through turning points \((-18,-12), (18,-12)\) and ends at \((18,7)\). All points within 1/2 of the wire width of 2 \( \lambda \) are included in the masked area.

Industrial grade tools are usually based on a proprietary description called GDSII. This standard differs in that rectangles are specified by [xxx center and width or corners?]. Translation between CIF and GDS is possible and commercially available. Also to incorporate layers on masks of arbitrary transmission and phase a generalized format called the semiconductor mask representation SMR has been developed [rzz.bernard.bacus?].

### 2.8 Elementary Process Models for Lateral Effects

Hand sketching or simulating typical topography effects in processing can be carried out extremely rapidly when only one-dimensional models are used. Here the elementary pro-
cess models used in SIMPL-1 will be used for illustration. Each mask edge creates a new cut in the nonuniform rectangular grid representing the impurity distribution. The doping model is a Gaussian with an adjustable position of peak depth and standard deviation. The doping profile does not move with subsequent processing and for this reason the input parameters must include final depth and standard deviation. A special blocking thickness parameter is used to account for the thickness of intermediate layers which would be sufficient to block the atoms at time of implant. The oxidation thickness input is the thickness which would grow on bare silicon. In cases in which an oxide existed previously the composite oxide thickness is taken as the square root of the sum of squares of the initial and input oxide thicknesses.

Approximate two-dimensional effects can be simulated quite rapidly with elementary “back-of-the-envelope” models as in SIMPL-2. Adjustable parameters reflect the process dependent aspect of lateral topography and doping lithography, etching, doping and deposition. For lithography vertical resist edges are assumed and only a simple bias is allowed. The etching is described by two parameters one is depth and the other is slope of the tapered edge as indicated in Figure 2.18.

The internal LOCOS oxidation model used is shown in Figure 2.19. Only the resulting geometrical structure from the oxidation process is used. This structure is obtained from experiments or rigorous simulation of oxidation. The oxide thickness specified in SIMPL-2 is the thickness of oxide resultant from the oxidation process on bare silicon. If an oxide exists prior to oxidation then \( U_1 \) and \( D_1 \) are approximated by:

\[
U_{1,NEW} = \sqrt{U_{1,OLD}^2 + (\alpha T_{OX})^2}
\]

(EQ 2.8 -1)
where $T_{OX}$ is the given oxide thickness on the bare substrate, $\alpha$ and $\beta$ are the fractions of oxide thickness grown upward and downward for the initial substrate surface, respectively. $X_t$ is the total bird’s beak length and $X_r$ is the length of encroachment. $u_1, u_2, u_3, d_1, d_2,$ and $d_3$ describe the tapering of the profile as indicated in Figure 2.19. There can be previously existing nodes between these new nodes. In this case the existing nodes are moved to points calculated by linear interpolation of the two adjacent nodes.

A simple Gaussian distribution can be used for doping given by
\[ C(z) = \frac{Q}{(2\pi \Delta R_P)^{2}} \exp\left(\frac{(z-R_P)^2}{2\Delta R_P}\right) \]  

(EQ 2.8 -3)

where \( Q \) is the implanted dose, \( R_P \) is the projected range and \( \Delta R_P \) is the projected standard deviation. Since SIMPL-2 does not have a model for thermal annealing, the final profile after any thermal steps should be approximated by a Gaussian distribution. The thickness of oxide thin film covering on silicon at the time of implant that ions cannot fully penetrate is specified by the user. SIMPL-2 assumes a complete “block” by other materials covering the surface. Lateral diffusion is approximated by the same Gaussian distribution decaying radially from the mask edges.

### 2.9 Linking Rigorous Process Simulators

#### 2.9.1 Need for Links to Rigorous process Simulation

The nearly planar interface between the intrinsic device in the silicon and the interconnect above it which was characteristic of IC’s for many years has given way to topographical features such as trenches and capacitor stacks whose formation must be carefully simulated. To provide sufficient accuracy the topography and doping impurity distribution generally require using the best available physical models from various centers of expertise. Thus to facilitate simulating today’s devices the process flow simulators have evolved toward more of a role of that of a process flow manager which invokes the best available rigorous process simulators on a process step by process step basis.

An example of a device requiring combining nonplanar topography and impurity simulation is given in Figure CP.21. Here a bipolar transistor with a polysilicon plug which reaches down to the buried layer has been simulated through using SIMPL to invoke SAMPLE topography simulation and impurity redistribution with SUPREM 4. The red polysilicon under the collector resulted from a opening the trench, a blanket deposition of polysilicon, and an etch back of the polysilicon to clear the planar areas. The dark green (high doping) surrounding the polysilicon indicates that the polysilicon has out diffused to make a good contact. In all cases these simulations are carried out using much more specific information about the actual process. Typically a process description at the level of the single parameter of a simple goal (such as a 0.5 \( \mu \)m oxide growth) now includes process conditions (time and temperature) and parameters for the physical models to be used (pre-exponential, and activation energies for B and B/A). In the case of lithography the goal of a given linewidth is replaced by typically the 25 parameters it takes to describe the exposure system, resist material and underlying thin-film stack.

Communicating to rigorous process simulators is not particularly difficult as the internal topography description and impurity attributes need only be written to a file in a format that the rigorous simulator can read. The rigorous simulators can then be invoked using a shell script. This approach works well but leaves the programmer of the process management system with the rather large task of building a graphical user interfaces to allow the
user to select every possible option allowed in each rigorous process simulator. This problem will hopefully be overcome in future TCAD systems where all simulators will read information from a standard semiconductor process representation and use standard graphical user interface widgets. Bringing back the information from individual simulators to the process manager database is a non-trivial problem when different data representations are used which is almost always the case. This data mapping problem is one of the most difficult tasks which underlies every TCAD system and requires a major research endeavor.

2.9.2 Data Structures Physical Representations

The data structure and physical representation play a big role in the accuracy of the simulation and the ability to communicate between simulators. Data structures must be designed based on the speed and accuracy of algorithms, the communication bandwidth at which data must be moved and the efficient use of memory in adapting to a wide variety of device technology applications. Rapid searching is often facilitated by using additional data structure pointers to presort the data into auxiliary views. Efficient use of memory is made by designing the data structure to be dynamically adaptable to the nature and size of the problem. The use of pointers and dynamic allocation of memory during execution are well suited for these needs. The communication bandwidth can become a problem when a frequently required task involves shipping a copy of detailed topography and doping information to and from a centralized geometry or impurity calculation utility.

This section uses a few of the data structures approaches and issues using examples from SIMPL. Polygons are convenient for representing the arbitrary shapes of topographical features generated in process simulation. In SIMPL-2 the doubly linked polygon list shown schematically in Figure 2.20 was introduced [rzz.lee.phd]. There are three storage locations a, b and c with material name information (in the middle of each node) and their associated pointers to adjacent nodes. The third location is need for branching nodes where three materials meet. This makes it possible to trace the boundary of a polygon of given material type by simply exiting from a starting node via the pointer associated with a given material type. This can be illustrated by following material IV through out Figure 2.20. More recently this data structure has been generalized to an arbitrary number of vertices connecting at a point in an ordered sequence. This was found to allow more efficient management of string collisions through not having to manage false collisions of nearest neighbors whenever more than three materials meet at a point.

An entirely different class of data structures is needed to describe inhomogeneous attributes over the interior of materials such as nonuniform silicon doping or oxide stress throughout a feature. The data structures must be capable of managing both topographical and attribute changes during processing. Fig depicts a bipolar transistor process where a region of the structure is first removed and then backfilled with doped polysilicon. The polysilicon then acts as a diffusion source to form a low resistance path to connect to the buried layer. Figure CP.21 shows the SIMPL simulation of this structure. Note the dark green around the red trench in the substrate. This is the n-type dopant diffusing out of the polysilicon plug to assure a good electrical connection with the buried layer.

Nonuniform rectangular grids and finite elements are frequently used to describe these inhomogeneous attributes. SIMPL-2 is based on the use of a nonuniform rectangular grid
to describe the doping in the silicon substrate. A list of x cuts and y cuts is maintained as well as an array of the values of the impurity types and doping levels associated with each grid point. The grid is generated and deleted automatically as the process evolves. This nonuniform rectangular impurity representation is dynamically managed during the process flow by the Berkeley Topography Utilities. Figure 2.22 shows an example from SIMPL of ‘stitch-back’ of a SAMPLE etched profile into the SIMPL grid used to represent the impurity information [rzz:BTU]. Note that when the topography polygon is clipped against the grid for impurity representation it is tempting to both increase the number of nodes on the polygon and as well as a number of new vertical and horizontal grid lines. To avoid problems of increasing node density and small sliver triangles techniques to regularize the topography and impurity representation which trade-off accuracy for efficiency. As shown in Figure 2.23 SIMPL can both write and read a SUPREM4 save file to
TOPOGRAPHY AND IMPURITY DIFFUSION INTERACTION

**FIGURE 2.22**
Example of issues in combining a rigorous topography description with associated inhomogeneous attributes to maintain accuracy without undesired refinement of either grid.

allow the rigorous simulation of the impurities and electrical performance with PISCES [rzz.vlsi.joint].

**SIMPL-IPX Link To PISCES-IIB**

G. Chin and R. W. Dutton, Stanford University

Extracting SIMPL cross sections for device simulation is now automatic.

**FIGURE 2.23**
Example of the mesh used in SUPREM4 and the data mapping SIMPL [rzz.vlsi.joint].
The general problem of developing common data representations for device cross sections is depicted in Figure 2.22. Arbitrary polygons are needed for many features such as the LOCUS shape and spacers on the gate. Yet other regions like the source/drain and field have planar layers. Similarly for the impurity there are regions in the vicinity of the LOCOS edge and gate edge where a 2D representation is needed amid regions where 1-D representations are adequate. It is desirable to use the best available rigorous simulators in these 2D regions and link their results into those of simpler 1-D simulation. Using common data formats and developing data mapping capabilities for changing formats are the keys to facilitating this communication. The simulation community must agree on matters as simple as coordinate directions, representation for topographical features and descriptions of inhomogeneous attributes over these regions. The early work on a Profile Interchange Format (PIF) [rzz.duvall] set an initial file based standard which has been incorporated for example in VISTA [rzz.vista.pif]. More recently a Semiconductor Wafer Representation (SWR) database query standard in C++ has been defined [rzz.swr]. Unfortunately a full implementation of this top down general approach has to date not been found to be cost effective. Some efforts are being made to integrate and cross compare simulation tools such as for lithography in the TCAD Workbench effort sponsored by Sematech [rzz.TWB] and a more limited set of common formats may emerge.

2.10 Summary

This chapter introduces process flow integration as well as a brief overview of many concepts and issues in technology CAD (TCAD). Given the multi-purpose nature of this chapter it is appropriate to briefly summarize what the purposes were.
• Show the importance of process integration and the powerful role that TCAD can play.
• Walk through a process sequence to see the step-by-step topography time-evolution of the topography and doping.
• Explore the relationship of the device and interconnect structures to layout and process flow for a typical double level metal CMOS process.
• Sample the rich opportunities in TCAD for providing design aids for process development and integration with device and circuit design.
• Illustrate typical technology issues and important role that test masks play in their characterization.
• Highlight the relationship alignment requirements to design rules and line-width variation as well as the importance of designing the creation and preservation of alignment marks.
• Present basic description of the CIF layout format and elementary process models for lateral effects.
• Describe how linking to rigorous process simulation can be accomplished and the issues in mapping topography data and associated inhomogeneous attributes such as impurity doping between simulators.

Many of the topics briefly discussed in this chapter will be considered in greater detail in other chapters as we revisit lateral effects at mask edges. The basic simulation of deposition and etching will follow in Chapter 3 and its characterization in Chapter 4. The lithography material then begins in Chapter 5. TCAD surface representation and system integration issues are treated in greater detail in chapters 11 and 12. The electromagnetic scattering aspects of alignment receive some attention in chapter 10.

2.11 References


