3

# Microstructures Engineering for Microelectronics



1981 Southcon Professional Program

Georgia World Congress Center, January 13-15, 1981

# MICROSTRUCTURES ENGINEERING FOR MICROELECTRONICS

## ORGANIZERS

Dr. E. R. Chenette
University of Florida
Gainesville, FL

Dr. Jay H. Harris
San Diego State University
San Diego, CA

Dr. Gordon R. Harrison
Georgia Institute of Technology
Atlanta, GA

## CHAIRMAN

Dr. Horst R. Wittman
U.S. Army Research Office
Research Triangle Park, NC

# SIMPLIFIED DESIGN RULES FOR VLSI LAYOUTS

Richard F. Lyon
Member of Research Staff
VLSI Systems Design Area
Xerox Palo Alto Research Center
3333 Coyote Hill Rd.
Palo Alto, California 94304

## INTRODUCTION

The complexities of detailed layout design rules that change over the life of an evolving LSI/VLSI technology have forced a reconsideration of layout design methodology. Even during the design time of a single chip, a target process may scale, making the design inefficient or obsolete before it is produced. The single most important change in any generic process type is the resolution, or feature size, which scales down steadily with time. Therefore, rules parameterized by a single variable (feature size, scale, or resolution) can provide longevity of designs, which will remain workable and efficient as long as the process does not change in a radical way. Such simplified single-parameter design rules for generic processes are the subject of this article.

An interesting effect of simplified design rules is that they enable a restructuring of design, data translation and management, maskmaking, and wafer fabrication jobs in a way that allows much more independence. Clever people who are not associated with any semiconductor company are already becoming independent IC designers, and are using simplified design rules similar to the ones described below.

The *implementation system* referred to in this article should be interpreted as a service capability much like the *MPC79* system described by Conway, Bell, and Newell [1]. Its purpose is to provide a simple interface between the designer and the various other services needed (maskmaking, wafer fab, dicing and bonding, etc.).

## WHAT ARE DESIGN RULES?

Design rules specify to the designer certain geometric constraints (dimensional inequalities) on his layout artwork (design file) such that the patterns on the processed wafer will preserve his design intent in terms of topology and geometry (with high probability), so that functional systems can be built. This means that connected regions will stay connected, disjoint regions will stay disjoint, and electrical length/width ratios of field-effect transistors (for example) will stay reasonably close to the as-drawn values. These crucial properties can not be guaranteed if the design rules are violated.

Design rules should always be specified in terms of the layout artwork (i.e. as rules for the designer), rather than in terms of some intermediate artifact (*e.g.* pattern generator tape, reticle, or mask). The implementation process (data handling, mask making, and wafer fabrication) should always strive to produce final patterns as close as possible to the drawn patterns, in terms of effective electrical widths and other functional properties (not apparent optical width). Some IC fab lines have not taken this approach in formulating their rules, so their rules may be harder to interpret.

For "analog" circuit design, which requires better control of parasitics and electrical parameter tolerances, stricter rules and additional statistical information will be needed. Access to such information often requires a captive and cooperative fab facility, so the independent digital designer should try to avoid needing that much detail.

## STRATEGY FOR SIMPLIFIED RULES

A single parameter, $\lambda$ (the Greek letter *lambda*), has been chosen by Mead and Conway [2] to represent the "resolution" of a typical lithographic step in wafer processing. Design rules can be formulated in terms of $\lambda$, using simple rules for the interpretation of "resolution".

Basically, $\lambda$ may be thought of as a bound on the width deviation of a feature on the final processed wafer from its ideal as-drawn size; it may also be thought of as a bound on the misalignment of any feature (*feature-center*) on one mask level from its ideal position relative to a feature on any other mask level. The combination of these effects, in the worst case, results in *feature-edges* from different mask levels deviating as much as $2\lambda$ from their ideal relationship.

*Typical* feature-width and alignment control for a process should be much better than the worst case implied by the value of λ. Line widths, which may be critical to high-performance devices, will be much better controlled for certain mask levels than this discussion would imply.

Design rules are usually stated as minimum distances between specified feature-edges. Most rules for the required separation of feature-edges may be derived from a pair of simple "meta-rules", stated here:

Fatality meta-rule: *Relative movement of feature-edges by amounts less than 2λ should not be obviously fatal.*

Degradation meta-rule: *Relative movement of feature-edges by amounts less than 1λ should not cause significant degradation in performance.*

Thus, to preserve topology of features on one layer, we have the immediate rules of 2λ minimum width (separation of opposite edges) for any part of any feature, and 2λ minimum space between separate features. However, these "2λ" rules may not be strict enough for all layers, for various reasons relating to the process. In processes with many layers, stricter rules (larger minimum sizes and clearances) typically apply to later steps (upper layers) and to relations between layers that are "far apart" vertically.

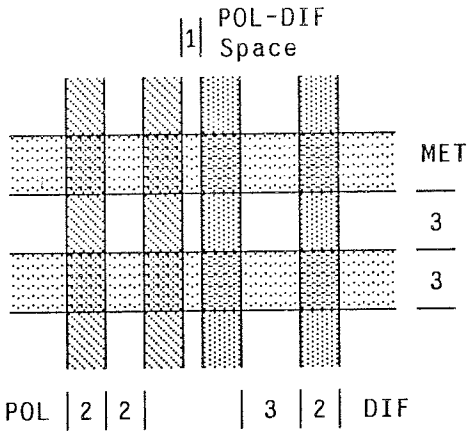
It is possible to draw features on any layer which are not topologically significant and have edges closer than 2λ to each other (narrow stubs and notches); these should be avoided, whether or not they are considered to be violations of the rules, since they can not be counted on to be resolved on the finished wafer. Consideration of such features, and other subtle geometric problems, presents a stumbling block to strict formalizations of the rules. For example, since sharp corners can not be resolved reliably, should the rules include a minimum radius for rounded corners of features? What about internal angles (corners of complemented features)? Accepted practice is to allow corners no sharper than right angles, and to not rely on sharp corners for functional or electrical properties.

## EXAMPLE PROCESS: SILICON-GATE NMOS

In the simplest silicon-gate *n*MOS process, there are three mask levels for conductors (termed POL, DIF, and MET, where POL over DIF also makes enhancement-mode field-effect transistors with POL as the gate) and one mask level for opening contact windows through the insulating oxide between the layers (CUT). See Mead and Conway [2] for a description of the process (including implants for depletion-mode transistors, described below). Design rules for this basic process and its depletion-load variation are discussed in detail below.

### Width and spacing rules for conducting layers

The POL layer (polysilicon, usually the best-controlled layer) should obey the basic "2λ" rules for widths and spaces. The other two conducting layers were found by Mead and Conway to require more width and/or space than the "2λ" rules. See figure 1 for width and spacing rules for the conducting layers.




1. NMOS Conductor Rules
(dimensions in units of lambda)

On the DIF layer (source and drain and interconnect diffusions and channel regions), extra spacing is required (on most processes) to prevent inadvertant conduction through depletion regions, which at high voltage can extend significantly beyond the normal edge of a diffused region; therefore, 3λ is the DIF spacing rule.

The MET layer (aluminum metal interconnect) is the last circuit layer to be patterned, and must reliably cover the non-planar terrain left by structures of the previous layers; therefore, to be consistent with a typical level of conservativeness in the industry, Mead and Conway have chosen 3λ for both minimum width and minimum spacing of MET regions.

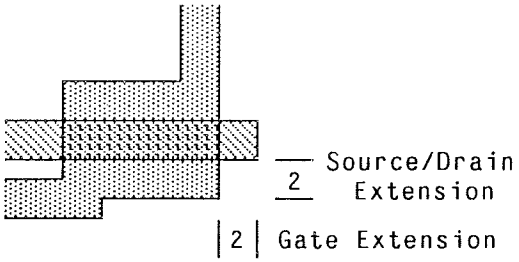
POL and DIF lines which are not purposely coming together to form a contact or a transistor should be kept separated by 1λ. If they should shift between 1λ and 2λ, and begin to overlap one another, the only effect is a narrowing of the diffused width and a (possibly large) mutual capacitance through the gate

oxide between them where they overlap. The MET layer does not interact with DIF and POL except at contacts (and via parasitic capacitance).

Rules for transistors

Where POL crosses DIF, the source and drain diffusion is masked by the gate oxide remaining under the POL region; the source, drain, and channel are thereby *self-aligned* to the gate. The resulting MOSFET has minimum length and width determined by the minimum width of POL and DIF lines, respectively.

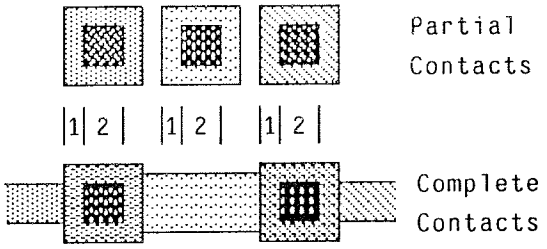
POL and DIF are constrained in two other important ways where they form a transistor; see figure 2. First, POL must absolutely cross DIF completely, or the transistor will be shorted by a diffused path from source to drain; thus 2λ of POL overlap beyond the edges of the DIF region is required. Similarly, DIF must extend beyond the POL gate so that diffused regions definitely exist to carry charge into and out of the gate region; thus 2λ of DIF extension is needed to preserve the source and drain regions.




2. NMOS Transistor Rules

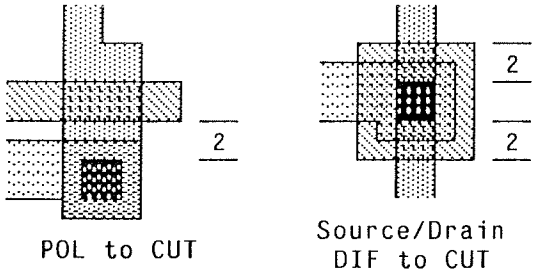
Rules for contacts between layers

The only other interactions between layers involve contacts; there are several rules to assure adequacy of contact area between layers and safe separation of contact cuts from uninvolved conductors. Features on the CUT mask specify windows in the intermediate oxide to expose POL if it exists, or DIF if there is DIF and not POL, so that subsequently placed MET will contact the exposed silicon area. Contacts are made using a minimum of 2λ square of CUT, with a big enough conductor region (POL or DIF) under it to assure a reasonable area of coverage, as illustrated in figure 3; thus 1λ of DIF or POL surrounding a CUT is the rule to prevent the contact area from decreasing with position shifts of less than 1λ, and to prevent etching through to the substrate by the edge of the conductor. The same overlap rule applies to the MET region that overlies the CUT window, to assure adequate contact area and periphery. The aluminum step coverage at the contact window edge is a potential source of problems such as metal migration and breakage; the overlap and large periphery length helps assure low resistance and high current-carrying capacity.




3. Contact Size Rules

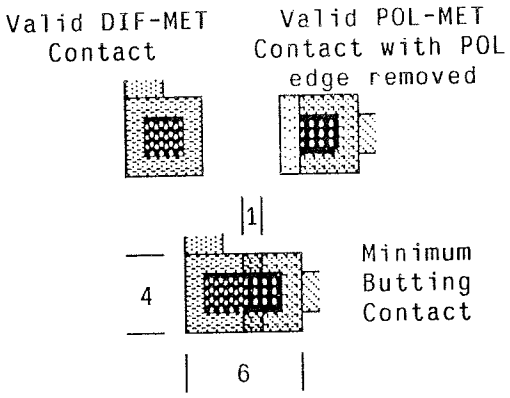
To prevent a CUT window from accidentally allowing contact to an uninvolved conductor, 2λ clearance is required as illustrated in figure 4. Mead and Conway [2] state this rule for POL to CUT clearance, as it relates to placement of a transistor gate near a contact to the source or drain diffusion; DIF to CUT clearance should also be observed in the unusual case of a contact to POL over a channel region, which must be separated adequately from the source or drain diffusion. Some processes suggest no use of CUT over a channel region, as it may cause a significant threshold shift due to the different work function of the resulting silicon-aluminum alloy gate material; when there is no use of CUT over channel, sufficient DIF to CUT clearance will always be assured by combinations of other rules.




4. Contact Clearance Rules

The final complication of the rules for layer interactions involves a structure known as a "butting contact", which allows a compact path between DIF and POL through a small link of MET, as illustrated in figure 5. It is made by removing one edge of POL overlap from a valid POL-CUT-MET contact, and then

overlapping that with a valid DIF-CUT-MET contact, abutting the CUT regions. Keep in mind that the gate oxide under the POL layer prevents contacting DIF below it (there is no conductor there anyway, just channel); thus, this contact has the property that either the DIF-MET contact region or the POL-MET contact region might be diminished even for misalignments less than 1λ, so the contact resistance can be about a factor of two higher, in the λ-shift case, than in a series connection of the simpler DIF-MET and POL-MET contacts. Thus, this contact does not satisfy the "degradation meta-rule" in uses where its increased resistance causes performance degradation; where contact resistance is critical, such as in a DC power distribution network or in series with a low-impedance output driver, more conservative contact structures should be used.

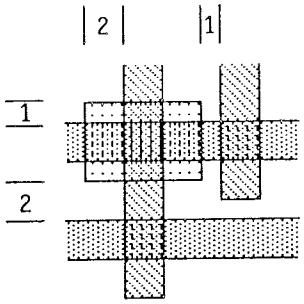



5. Butting Contact Rules

Rules for depletion-mode implants

In order to acheive better speed-power performance, and to allow inclusion of effective clock drivers on-chip without a separate supply voltage, most nMOS processes have added another mask level (IMP, a donor implant) and associated process steps to allow depletion-mode transistors (*i.e.* transistors with negative threshold). Features on the IMP mask define regions that will be implanted with the donor impurity.

See figure 6 for the rules relating to IMP overlap and clearance. It is important to cover the whole area of the POL-DIF intersection that defines a transistor; moving the edge of the IMP feature relative to the transistor edges can result in parasitic enhancement devices in series or in parallel with the desired depletion device. The series device would be fatal (the combined threshold could be positive), so 2λ of overlap beyond the source and drain edges is required. A parallel device would simply increase the effective resistance a bit by narrowing the negative-threshold region, so only 1λ of overlap is required in the direction of gate width.




6. Implant Rules

It is equally important to observe clearances of IMP features from desired enhancement devices, especially to avoid a parallel parasitic depletion device which would be an effective short. Therefore, the clearance rules are also asymmetric, being 1λ from an enhancement device source or drain edge, and 2λ from the side of the enhancement device channel.

To avoid the complications of having to remember and understand these asymmetric rules, Mead and Conway formulated a simpler set of compromise rules, which are 1.5λ of overlap and clearance relative to all device edges. Unfortunately, these rules are not as uniformly conservative as the rest of their rules; and since IMP rules are not critical to density in commonly used gate logic circuits, fab lines have not worked much on improving the resolution of the associated steps. Therefore, the conservative designer will use "2λ everywhere" IMP rules for all new designs that will use modern *n*MOS processes.

Rules for buried contacts

Some processes have added another mask level and processing steps to create DIF-POL contacts without using the MET layer, for increased density. These "buried contacts" (BUR) have a rather extensive set of rules associated with them, which are explained in the expanded version of this paper [3].

## IMPACT OF PROCESS VARIATIONS ON DESIGN RULES

Process options such as buried contacts, extra threshold values, second layer polysilicon (either gate or interconnect), or second layer metal (either interconnect or light shield for imagers) may provide

dramatic improvements in density or performance for a given process resolution, in applications that need or can use those features. Design rules can be formulated, following the techniques illustrated above, for any of these features once the processing is understood.

Some industry experts have expressed the opinion that as processes move into the VLSI domain, simplicity and planarity will become increasingly important to yield, and that therefore buried contacts and additional layers of conductors are not likely to survive (on the other hand, some think that integration will move into multi-layer three dimensional structures!).

The designer's choice of whether to use any optional features will depend on many factors, including process availability, design and maskmaking cost sensitivity, desired design longevity, the density advantage for the particular design, etc.

Other more subtle process variations are commonly being tried out by various fabrication and research groups. The goal is usually to increase the layout density of a particular class of circuits (especially memory and PLA structures) without scaling down the basic processing resolution, or lambda. Thus, the detailed design rules for any fab line will reflect their current state of process variations. Some variations will cause *increases* in some of the minimum distance rules, so designs based on the simple rules stated here may actually become *larger*, since a larger value of λ will have to be used.

## CIRCUIT DESIGN RULES

Some "design rules" are really electrical circuit constraints and suggestions, such as how to construct input protection "lightning arrestors" or output drivers, current limitations for various metal line widths, etc. These are generally outside the scope of this article, but we will discuss an example here because it can be viewed as a topology constraint.

Some processing lines advise against running interconnect over dynamic storage nodes, thereby preventing capacitive coupling of transients into these high-impedance nodes. This restriction could be very detrimental to layout efficiency. We suggest that it is better to use conservative circuit design rules, such as pullup/pulldown ratios appropriate for good noise immunity, rather than to translate circuit parasitic constraints into overly restrictive layout constraints.

This rule exemplifies the *censoring* problem: a designer in a circuit design group associated with a vertically integrated IC facility might have his suggested layout topology "censored" by the layout group to meet "rules" imposed due to past problems between system and fabrication levels of the company. Such interactions are unneeded, and will not occur when clean interfaces are made between the designer, the implementer, and the fabrication services. Clean interfaces are only now evolving, and will come to include simple mechanisms for specifying design rules and verifying that the process "works" for those rules, based on accepted standard test patterns; some evolving standards and interfaces are described in [4].

## AUXILIARY RULES FOR MACROSCOPIC FEATURES

Processing lines often pose auxiliary geometric constraints beyond those needed to insure topological correctness. These rules do not always scale with the other rules; macroscopic rules relating to bonding pads are a prime example. Large MET regions are used as pads to bond wires to, in order to connect the chip to the package pins. Typical rules for these regions are 100 micron (0.1 mm or 4 mils) minimum length and width and 100 micron minimum spacing, with 25 micron clearance to active circuitry (conductors not connected to the bonding pad).

An additional mask layer (PAD) is usually used to pattern windows in a passivating layer of oxide (overglass) or nitride, to allow the bonding wire to contact the bonding pad. The features that specify such windows should stay at least 10 microns inside the MET region, to avoid etching other oxides. This clearance would only need to be 2λ if the masking and etching steps were done carefully enough.

Other "rules" are often posed to describe how to construct scribe lines, alignment marks, identifying numbers, etc. These non-rules (or rules for non-design items in the "starting frame") are really descriptions of additional artifacts to be included in the final mask layout, along with the actual system design. Such details need not be known to the independent designer who has access to a good implementation service that will take care of such things. For designers operating without an implementation system, or for someone who wants to provide such a system, the information in Hon and Sequin's *Guide to LSI Implementation* [5] is invaluable.

## DESIGNING DESIGN RULES

Single-parameter design rule sets, based on the same concept of λ and the same meta-rules, are being developed for other processes. Rules for CMOS/SOS have been developed by Tom Griswold at JPL, and have benefitted from discussions with numerous SOS

fabricators; rules for several bulk CMOS processes are being developed at Hewlett-Packard and at UC Berkeley. Rules for $I^2L$ layouts are being developed by Dick Oettel of Boeing. Rules and meta-rules for a class of junction-isolated bipolar digital/linear processes (with more complicated structures than our "planar" view presented above will easily cover) are being developed at Xerox PARC, guided by the detailed design rule development presented by Glaser and Subak-Sharpe [6].

## CONCLUSIONS

This discussion of design rules has pointed to several ways in which the mere *formulation* of the rules can influence the methodology of designing and building integrated systems. The "simple rules" paradigm discussed here is in conflict, in several ways, with the traditional "vertically integrated" paradigm of the semiconductor industry. We attempt here to summarize the basis for the "simple rules" paradigm.

Rules are simple geometric constraints that assure preservation of topological intent on any of a family of "generic" processes; *i.e.*, the rules are *fab-line independent.*

Single-parameter rules *scale* trivially, and remain relatively efficient over the lifetime of a generic process; thus chip and subsystem designs remain useful for long enough to make the "design library" concept useful.

To take advantage of the longevity of designs possible with these rules, designers should *avoid complicated process options* that may be temporary aberrations from the generic process.

The designer can deal with a *clean interface* if an implementation system is available to deal with other services; the implementation system will only offer process capabilities that are truly generic, in terms of multi-source availability, compatible with one set of simple rules.

The designer is *in control* of all levels of his design, from architecture to geometry, and need not be restrained by arbitrary extra rules, for example on what circuits he can try.

Beyond actual circuitry, the designer needs to know certain other constraints such as minimum bonding pad sizes, which the implementation system will specify if the system is to do the wire bonding; otherwise, the designer can do as he wishes, *independent of the usual practice at the fab line* that will be used.

Simple interfaces and explicit statement of the design constraints, separated into layout design rules and other categories, are already enabling a dramatic increase in the number of practicing integrated system designers.

## REFERENCES


[1] Lynn Conway, Alan Bell, and Martin E. Newell, "MPC79: A large-scale demonstration of a new way to create systems in silicon", *LAMBDA--The Magazine of VLSI Design*, Second Quarter 1980.

[2] Carver Mead and Lynn Conway, *Introduction to VLSI Systems*, Reading, Massachusetts: Addison-Wesley, 1980.

[3] Richard F. Lyon, "Simplified Design Rules for VLSI Layouts", *LAMBDA--The Magazine of VLSI Design*, First Quarter 1981.

[4] Alan G. Bell, "The Role of VLSI Implementation Systems in Interfacing the Designer and Fabricator of VLSI Circuits", *Proceedings of the International Telecommunications Conference*, November, 1980.

[5] Robert W. Hon and Carlo H. Sequin, *A Guide to LSI Implementation*, Xerox PARC Technical Report SSL79-7, Palo Alto, California, 1980.

[6] Arthur B. Glaser and Gerald E. Subak-Sharpe, *Integrated Circuit Engineering*, Reading, Massachusetts: Addison-Wesley, 1977.



Portions of this paper were published in LAMBA Magazine, First Quarter, 1981 Edition