GENERALIZED IC LAYOUT RULES AND LAYOUT REPRESENTATIONS

Carlo H. Séquin

Computer Science Division
Electrical Engineering and Computer Sciences
University of California, Berkeley, CA 94720

1. LAMBDA - BASED LAYOUT RULES

The evolution of very large scale integration (VLSI) is accompanied by a proliferation of ever more sophisticated and thus often more complicated processes. At the same time the cost of the design and layout of a VLSI circuit has become so prohibitive that one can simply not afford to hand-code the layout of a particular circuit for every process variant. A simpler and portable set of layout representations is thus highly desirable. Lyon (1981) has recently given an extensive introduction and description of a set of simplified design rules for the prevailing silicon-gate NMOS process. These simple rules, introduced by Mead and Conway (1980) and extensively used in the context of the multi-project-chip efforts (Conway *et al.* 1980), are based on a single parameter *Lambda* (λ) which makes the rules scalable and thus potentially gives them a much longer lifetime. As long as technology makes nearly uniform advances on all fronts, i.e. worst case misalignment and the smallest feasible features in all mask levels scale down at roughly the same pace, these rules and the associated layouts maintain their validity for all but the most stringent demands.

While longevity is clearly an important attribute of these rules, there are other advantages too: These rules are very simple and thus much more suitable for novice designers. They can be readily expressed on a single color plate, whereas a typical set of industrial design rules comprises more than a hundred separate rules and easily spans a dozen pages of instructions and sketches. But even for the experienced designer there are advantages. By using these simple layout rules and placing all features on a fixed *Lambda grid*, one can work more effectively, uncluttered by the nitty-gritty details of the ultimate fabrication rules, and focus on the more important higher level aspects of the layout. Even if the layout is partially done by a computer, the simpler rules have the advantage that they can reduce the computational task in circuit verification and layout rule checking.

2. EXTENSION TO OTHER PROCESSES

This paper presents a simple set of generalized rules that are formalized independently of a particular fabrication sequence, emphasizing those tolerances that are common to many different processes. The set is simple, regular, and easy to remember. It can be used as a starting point for the more sophisticated set of layout rules of a specific fabrication process by adding a - hopefully rather short - list of well justified exceptions.

The simplicity of the generalized set of rules has advantages for the designer as well as for the developer of efficient design tools. This approach using fixed rules with added exception lists has the further didactic advantage that it explicitly points to the differences between various processes and thus highlights potential trouble spots.

All dimensional rules are based again on a single parameter *Lambda* (λ) that characterizes the linear feature "resolution" of the complete wafer implementation process and permits first order scaling. These layout rules are subdivided into three groups of less than ten rules each, concerning respectively: Mask Feature Sizes, Overlaps and Separations, and Macroscopic Rules associated with the chip periphery and bonding pads.

3. MEANING OF A LAYOUT - CONCEPTUAL LAYOUT LEVELS

The more sophisticated CMOS and bipolar processes are often so complicated that it becomes inefficient for the designer to think in terms of *all* the mask levels actually used in the fabrication process, many of which may only be artifacts of a particular implementation sequence. The design process can be abstracted to a few conceptual layout levels that represent the relevant physical features to be found in the final silicon wafer. The actual mask levels required for implementation, which may vary for different fabrication lines, can then be computer generated from these conceptual design levels.

For example, rather than dealing with one or two thin-oxide masks, channel stop masks, n- and p-type implant masks, as may be used in a typical silicon-gate bulk CMOS process, the designer should think in terms of the n+ and p+ regions that form the operational source and drain electrodes of the transistors of both polarities. Thus the designer may draw only a subset of the masks required for actual manufacturing, or perhaps a composite of two or more layers. An example is the "green" areas in the silicon-gate NMOS process used by Mead and Conway (1980). While these areas are referred to as "diffusion", they really describe the thin oxide areas; the actual diffusion or implant, which is applied uniformly across the whole wafer, is being properly masked by the thick oxide and by the polysilicon features present at that time.

The introduction of such conceptual levels raises the important question of what the layout actually represents:

a) Is it a scaled picture of the mask geometries ?
b) Does it correspond to the final features on the wafer ?
c) Is it an idealized geometrical representation of the intent of the designer in terms of certain device parameters ?

We advocate that, for the kind of design environment and methodology that such a set of generalized layout rules reflects, the most appropriate interpretation lies somewhere between options b) and c). While it is desirable to have the designer think in terms of the envisioned results rather than the means of achieving them, a direct specification of a layout in terms of the final device parameters is beyond the state of the art of current implementation technology. Interpretation c) would understand the "channel width" of a MOS transistor as a scalable parameter which is strictly proportional to the drain current of the device. Since such an ideal parameterization is possible only in a few simple cases, this approach is not practical in general. Nevertheless, the designers should be shielded from the detailed steps of the implementation process and should be able to think in terms of the features they will find on the finished chips that are returned to them. The computer can be used to generate the necessary transformations and combinations of these desired geometrical device features in order to derive the required mask geometries for their fabrication.

3.1. A Comprehensive Set of Levels

At a high enough conceptual level all MOS processes use some of these same basic features: A maximum of two different substrate areas of opposite polarity; strongly doped areas of both polarities forming the source/drain regions of both types of transistors and the contacts or guard rings to the substrate or to the wells; some implants to adjust the threshold of groups of select transistors; gate electrodes; interconnection runs; and contact windows between different conducting levels. Thus a relatively small set of conceptual layout levels depicting these basic features should be sufficient to outline the geometry of any MOS circuit. The following set of conceptual mask levels is believed to be sufficient for all silicon-gate MOS and CMOS processes.

CIF name	*representing*
ANWL,APWL	Well area in the bulk or type of island doping in CMOS SOS
ACAP	Heavily doped area for capacitor electrodes
AND	N-doped area for n-channel source/drains
APD	P-doped area for p-channel source/drains
AIIN,AIIP	N or P-type implant to adjust FET thresholds
AII2,AII3	even more implant levels ...
ASI	Poly-Si electrodes
AME	Metal interconnects and pads
ACC	Contact cut
ABC	Buried contact
AOC	Overglass cut

To illustrate the difference between these abstract levels and the actual fabrication mask geometries, we will discuss the simple case of ANWL, the outline of the final n-type well in the p-type substrate. We assume that during wafer processing the dopants forming this well diffuse laterally to a distance of 2 λ. Thus, while a minimum diffusion mask window of 2 λ can readily be implemented, the minimum feasible width of a strip of n-well diffusion will be

6 λ; and if the diffusion mask windows are separated by less than 6 λ, neighboring strips may accidentally merge. The designer working with the conceptual levels will thus be told that minimum feature dimension for this particular layer is 6 λ, and that neighboring strips must be separated by at least 2 λ to guarantee geometrical separation. The computer will then shrink by 2 λ, the amount of the lateral outdiffusion, the wide features laid out by the designer.

3.2. Usage of Conceptual Levels

These conceptual CIF levels are an idealized representation of the salient features to be found in the final silicon wafer. Some levels are specifically related to a particular process class: bulk processes have to deal with the well of opposite polarity containing one type of transistors; metal gate processes must use a separate mask to define the thin-oxide areas forming the active gate regions. Obviously a single geometrical layout cannot be used for all possible CMOS processes, but it may be sufficient for most processes in a particular class such as Si-gate bulk CMOS or CMOS-SOS. Every mask required in a particular processing step is then derived from the set of conceptual levels. It might even be possible to use the same layout for both classes of silicon-gate bulk CMOS processes; the role of well and substrate would simply be reversed. Of course the electrical characteristics of the transistors generated by the two complementary processes may be quite different, and such portability can thus be achieved only for uncritical digital circuits.

The use of these abstract levels also has advantages for the design tools; working with fewer levels will increase their efficiency. The simple description in terms of a few conceptual levels is a suitable intermediate form between a layout done at the sticks level and the final, compacted mask geometry. A two-step compaction from sticks to final layout may be more efficient. In the first steps that yield the dominant amount of compaction, one should consider only the conceptual levels since this means less computational work for the processor. The geometrical layout at the conceptual level also permits verification and simulation of the circuit. Transistor ratios, the resistance of interconnections, and all other salient circuit parameters can readily be extracted from the features in these design levels. These simple layouts can either be used to generate relatively relaxed implementations of a particular circuit by extending them "in place" to the full mask set, or they can be used in conjunction with a circuit compaction program to generate a dense layout with a sophisticated set of design rules.

4. META RULES AND GENERALIZED LAYOUT RULES

At the highest conceptual level, a single design rule can be formalized from which all specific design rules can be derived:

> *Under worst case misalignment and maximum edge shift of any feature no serious performance loss should occur.*

However this rule is too general to be of much help, and in order to derive more specific rules, a lot must be known about the processing details. So, two assumptions will be made, to permit a derivation of rules of a more specific nature.

1. Assume: worst case misalignment of any two levels is λ.
2. Assume: maximum edge shift due to processing is ½ λ.

With these two assumptions, a set of rather general, parameterized rules can be derived. They have been grouped together by the nature of one or two related issues that they address.

4.1. Mask Feature Sizes

This section addresses dimensional rules that are concerned primarily with a single mask or feature level and specify minimum internal and external separation of edges. Everything follows form one simple metarule:

MI. *All dimensions must be at least 2 λ.*

This rule can readily be expanded into the following more specific rules which are also illustrated in Figure 1:

1. Minimum feature size in any level is 2 λ.
2. Minimum separation between features is 2 λ.
3. Minimum contact window size is 2 λ × 2 λ.
4. Minimum contact cut separation is 2 λ.
5. Minimum conductor width is 2 λ.

Rules 1 through 4 follow directly from metarule MI. Rule 5 also follows in the same way if it is understood to refer to a conductor such as a polysilicon path defined by a single mask level; however it can be generalized to mean that *any* conducting path should be at least 2 λ wide, even if it is defined jointly by two mask levels, as is the case in a diffused path running along a polysilicon gate.

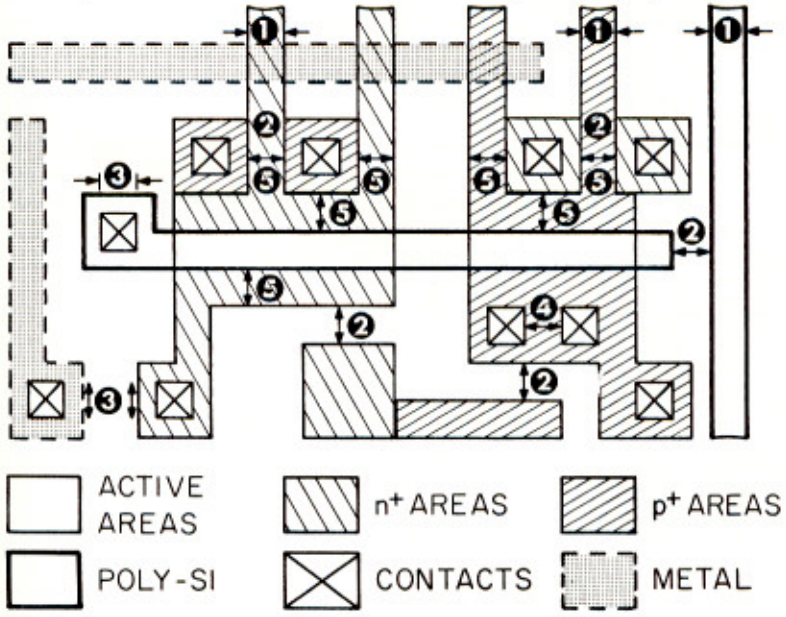

Fig. 1. Rules on mask feature sizes.

4.2. Overlaps and Separations

The underlying meta-rule that governs the overlaps and separation of features on two different masks is:

MII. *Features that should not touch must be separated:*
- *by two λ - if touching has a catastrophic effect,*
- *by one λ - if touching has NO catastrophic effect.*

From this the following rules can be derived (see Figure 2):

1. Frames around contact cuts must be at least 1 λ wide.
2. Keep contact cuts above an active layer 2 λ from edge.
3. Keep contact cuts 2 λ away from poly-Si gates.
4. Separate poly-Si and diffused conductors by 1 λ.
5. FET-gates must extend 2 λ beyond the transistor channel.
6. Doping of a given area should be extended by 1.5 λ.
7. Keep doping 1.5 λ away from undesired areas.
8. Buried contact window frames must be 1.5 λ.
9. Keep buried contacts 1.5 λ away from undesired areas.

Rules 8 and 9 are the simplest way to deal with buried contacts. They correspond in spirit to rules 6 and 7 dealing with threshold adjusting implants. A more detailed analysis is given by Lyon (1981). However, because of the large variety of possible layout configurations of buried contacts, a complete catalog of all cases is beyond the scope of this list.

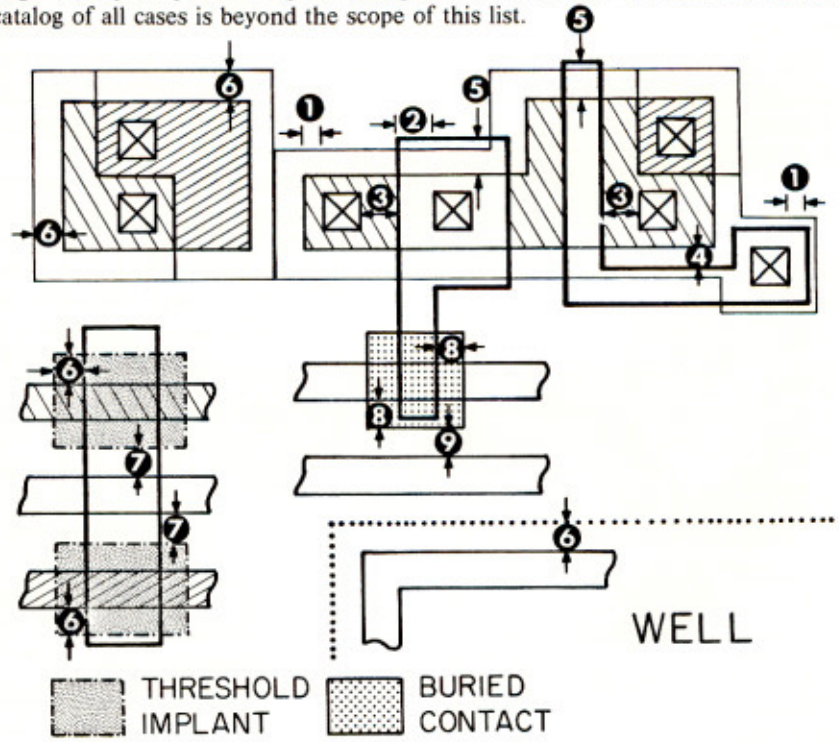

Fig. 2. Rules on overlaps and separations.

4.3. Macroscopic Features

Some features on an IC chip are not normally scaled down proportionally to λ. Features such as the scribe lines between individual chips on the wafer, or size and separation of bonding pads, are really determined by the dimension of the mechanical tools used in dicing and interconnecting these chips. Since the technological progress in this area is slow compared to the advances in photolithography and chip fabrication, the associated rules are expressed in absolute quantities:

1. Total scribe line width is 100 μm.
2. Minimum feature distance from scribe line is 50 μm.
3. Minimum bonding pad size is 120 μm.
4. Minimum overglass cut to bonding pad is 100 μm.
5. Minimum feature distance from bonding pad is 40 μm.
6. Minimum bonding pad separation is 80 μm.
7. Minimum bonding pad pitch is 200 μm.

Rules 3 and 6 result in a bonding pad pitch of 200 μm. This is desirable so that the chips can be wire-bonded easily and with good yield. A few additional constraints sound rather obvious but are often overlooked by the novice designer:

8. The final chip must fit into the cavity of the package.
9. The length of any bonding wire should not exceed 5 mm.
10. Bonding wires must not cross.

Rules 9 and 10 imply that the pads on the chip should roughly line up with the pads in the package and should be spread evenly around the chip perimeter. Thus for use in a 40-pin package one should not put more than a dozen pads along any one edge of the chip.

4.4. Derivation of NMOS Rules

To illustrate the concept of *exceptions* that produce a usable set of rules for a real process, we present the necessary additional rules for the Si-gate NMOS process referred to by Mead and Conway (1980). Only two rules need to be added to obtain their set of layout rules:

a. Transitions from thin gate-oxide to thick field-oxide in the local oxidation processes use up space. The minimum separation of thin-oxide areas thus is 3 λ. In other words, minimum feature size of thick oxide is 3 λ.

b. Metal lines on non-uniform surfaces have poorer edge definition and should therefore have coarser features; a minimum width and minimum separation of 3 λ are often specified.

In addition there are some electrical design rules that apply to most processes, such as:

1. Current density in metal must be less than $10^5 A/cm^2$.

Space constraints do not permit us to discuss them here in more detail.

5. APPLICATION TO CMOS

The above rules and unified layout representations have been generated after review of two Si-gate CMOS processes, a metal-gate CMOS process, two NMOS processes, and a CMOS SOS process. Compatibility with the basic silicon gate NMOS process used by Mead and Conway (1980) and with the CMOS SOS design rules proposed in the Caltech/JPL standards project (Griswold 1980) has been maintained as much as possible.

The design approach using the above proposed conceptual levels will now be illustrated for the inverted bulk CMOS process, chosen for its generality and significance for the future of VLSI. It is a direct and compatible enhancement of the Si-gate NMOS process that permits to mix high-density NMOS circuits and low-power CMOS circuits on the same chip. The process holds the promise of being scalable to channel lengths of only 0.5 μ. The fabrication sequence outlined here follows a process recently developed at Berkeley (Choi 1981). This process can also provide good MOS capacitors by introducing special n+ regions that act as bottom plates prior to the deposition of the polysilicon top plates.

5.1. Process Outline for Inverted Bulk CMOS

The first mask, MNWL, defines the phosphorus doped n-well. An optional mask, MNCA, then defines n+ capacitor plates in the p-type substrate. The gate oxide is grown, and the wafer is covered with silicon nitride. Mask MAAN delineates the thin-oxide, active areas above the p-substrate but leaves all n-well areas covered. The nitride layer is selectively removed where thick field oxide is desired, and a boron implant is introduced to act as a self-aligned p-type channel stop. After a first field oxidation, a separate mask MAAP then removes more of the same nitride layer, this time above the n-well. Photoresist protects the previously defined nitride islands on the p-substrate during the phosphorus implant that forms the n-type channel stop in the n-well. A long local oxidation step produces the thick field oxides above the p-substrate and the n-well. All nitride can now be removed.

After a threshold adjusting implant, heavily n-doped polysilicon is deposited over the whole wafer. Mask MSIN is used to form the gate electrodes of the n-channel transistors. An arsenic implant will then form n+ regions in all thin-oxide regions no longer covered by poly-Si. During this step the future PMOS devices are shielded by polysilicon. Similarly, a second masking step (MSIP) on the same polysilicon layer defines holes for the p+ regions, i.e. the source and drain regions for the p-type transistors in the n-well and the contact points to the substrate. The photoresist layer of this masking step shields the n+ regions during the Boron implant.

After implant drive-in and annealing, a passivation layer of polyimide is applied. Contact holes defined by mask MCC are cut through polyimide and oxide. Aluminum is deposited and patterned by mask MME. If a protective layer is employed, mask MOC is required to provide access to the bonding pads.

5.2. Mask Level Summary for Inverted Bulk CMOS

The masks that need to be generated for this process and their relationship to the CIF levels specified by the designer are:

Fabrication		*Design*
MNWL	Well definition	ANWL, shrunk for outdiffusion
MPCA	Capacitor areas	ACAP
MAAN	NMOS active area	AND ∪ APD ∪ ANWL
MAAP	PMOS active area	APD ∪ AND ∪ ¬ANWL
MSIN	NMOS poly-Si gates	ASI ∪ (ANWL ∩ ¬AND)
MSIP	PMOS poly-Si gates	ASI ∪ (¬ANWL ∩ ¬APD)
MCC	Contact windows	ACC
MME	Metallization	AME
MOC	Overglass windows	AOC

The major differences between the set of design and fabrication levels is that the designer deals only with one layer of polysilicon (ASI); the separate masks MSIN and MSIP required for fabrication can be derived from logical combinations of the well geometry and the n+ or p+ areas. This results in a simpler and more meaningful layout.

5.3. Additional Design Rules for Inverted Bulk CMOS

Starting from the generalized design rules outlined in section 4 of this paper, the following additional rules are necessary to obtain a complete set for the above process.

1. As in all processes using local oxidation, the transitions from thick to thin oxide uses up some space; these areas should thus be separated by 3 λ. At the conceptual level this refers to the placement of separate n+ or p+ areas.
2. Well depth and lateral diffusion can vary anywhere from 1 to 3 λ in different processes. Because of this variation, tolerances for this feature must be rather relaxed. Minimum distances to the edge of this layer are typically specified as 4 λ. Thus the following additional rules result:
   - -a. Minimum feature size for level ANWL is 6 λ.
   - -b. Separation between unrelated wells is 4 λ.
   - -c. Distance of unrelated doped areas from well edge is 4 λ.
   - -d. Guard ring is 4 λ wide; 1 λ outside, 3 λ inside ANWL.
3. The PMOS poly gate should be 3 λ, because of the larger boron outdiffusion.
4. Rules concerning the alignment of the shield masks used in the local oxidation steps of the well and substrate areas can be ignored by the designers who work with the conceptual levels; they are contained implicitly in the rules that govern the minimum distance of features from the well boundaries.

5 There are electrical rules, designed to prevent latch-up of the CMOS circuitry. Latch-up is typically a problem only in the peripheral circuits, which should be designed by an experienced designer and placed into a cell library. The user of these cells need not know these rules.

Note in particular, that in this process, because of the smooth surface of the passivation layer, metal lines need not be wider than the minimum width specified by the generalized design rules, i.e. 2 λ.

6. LAYOUT REPRESENTATION

With the widespread acceptance of CIF2.0 (Sproull and Lyon 1980, Hon and Séquin 1980) as a de facto standard for the description of IC layouts, it is important to standardize not only the syntax of this interchange format but also its semantics.

Known mask names		*features*
used in fabrication	*used in design*	*appearing in Si-gate MOS processes*
"MPWL"	"APWL"	P-Well mask / actual extension
"MNWL"	"ANWL"	N-Well mask / actual extension
"MNCA"	"ANCA"	N-type Capacitor Area
"MPCA"	"APCA"	P-type Capacitor Area
---	"ACAP"	Capacitor of either polarity
"MAA"	---	Active Area for any device
"MAAG"	"AAAG"	Active Area under metal gates
"MAAN", "ND"	---	Active Area for N-channel devices
"MAAP"	---	Active Area for P-channel devices
"MND"	---	N-type Doping
"MPD"	---	P-type Doping
---	"AND"	N-type Doped active areas
---	"APD"	P-type Doped active areas
"MIIN", "NI"	"AIIN"	Ion-Implant, N-type
"MIIP"	"AIIP"	Ion-Implant, P-type
"MSI",	"ASI"	Si-gate for any device
"MSIN", "NP"	---	SI-gate for N-channel devices
"MSIP"	---	SI-gate for P-channel devices
"MSI2"	"ASI2"	second level poly-SI for the future
"MSI3"	"ASI3"	third level poly-SI for the future
"MBC", "NB"	"ABC"	Buried Contact
"MCC", "NC"	"ACC"	Contact Cut
"MCC2"	---	higher level or oversize Cut
"MOC", "NG"	"AOC"	Overglass Cut mask
"MME", "NM"	"AME"	MEtal mask
"MME2"	"AME2"	MEtal mask, second level
"MME3"	"AME3"	MEtal mask, third level

6.1. Mask Level Names

In addition to the already established seven mask layer names for NMOS, we propose here an additional two set of names. One set, in which all layer names start with "M", is a comprehensive and expandable set of mask levels used for the actual device fabrication. The other set, containing names starting with "A", denotes the abstract design levels discussed earlier in this paper.

6.2. Layout Colors and Stipple Patterns

The introduction of some standard colors by Mead and Conway (1980) and the acceptance of these colors by a large University community has simplified dramatically the communication between the designers adhering to this particular representation. It is desirable to develop such a shared culture for a much wider set of processes. Unfortunately, already entrenched conventions and the different limitations of various output devices (screen or plotter) may render this a futile dream. Nevertheless we have worked out a usable set of colors spanning all classes of MOS processes, and we have also done work in developing a readable set of black and white stipple patterns. Restrictions on space do not permit us to present this work here. Interested readers should ask for a copy of the extended version of this paper.

ACKNOWLEDGMENTS

This work was sponsored in part by the Defense Advance Research Projects Agency and performed in part under a consulting contract with Xerox Palo Alto Research Center. I would like to acknowledge many fruitful discussions with R.F. Lyon.

REFERENCES

Choi, T. (1981). "The inverted CMOS process ...," *in preparation.*

Conway, L.A., Bell, A., and Newell, M.E. (1980). "MPC79: A Large-Scale Demonstration of a New Way to Create Systems in Silicon," *Lambda,* Vol 1, No 2, pp 10-19.

Griswold, T.W. (1980). "CMOS SOS Design Rules," CALTECH/JPL Standards Project.

Hon, B. and Séquin, C.H. (1980). *A Guide to LSI Implementation,* Second Edition, Xerox PARC, Palo Alto, CA, Jan 1980.

Lyon, R.F. (1981). "Simplified Design Rules for VLSI Layouts," *Lambda,* Vol 2, No 1, pp 54-59.

Mead, C.A. and Conway, L.A. (1980). *Introduction to VLSI Systems,* Addison Wesley, 1980.

Sproull, R.F. and Lyon, R.F. (1980). "The Caltech Intermediate Form for LSI Layout Description," in: *Introduction to VLSI Systems,* by Mead, C.A. and Conway, L.A., Addison Wesley, pp 115-127.