Quality Control From The Silicon Broker's Perspective

Danny Cohen and Vance Tyree
University of Southern California/Information Sciences Institute

The process of converting VLSI digital designs into working parts requires many steps. To ensure the overall quality of the finished parts, the quality of every process and of every interface has to be controlled. This article describes the measures taken by the MOSIS (MOS Implementation Service) system to ensure this overall quality, and also explores several related matters.

The MOSIS system (the "silicon broker" to the Defense Advanced Research Projects Agency or DARPA) supports the separation of design from fabrication. As a broker, MOSIS accepts designs expressed in Caltech Intermediate Form (CIF), and returns packaged parts. These designs are not optimized for a specific fabrication line; instead, they are "general" in the Mead-Conway sense. As used by MOSIS, the Mead-Conway design rules are general enough that most fabrication lines can support them. [The question is not whether a certain fabrication line can support the design rules, but rather how best to fit a given design file to the fabrication line by using a computer to modify the value of lambda automatically.]

The MOSIS brokerage operation involves many interfaces between different organizations. To guarantee the quality of the finished parts, each interface requires explicit quality control. The question a user of the MOSIS system must ask is simply "Do my packaged parts truly correspond to the design that I submitted?" This question applies to both geometric and electrical properties, and this interface is the only one of which the naive user is aware. However, many more interfaces must also be verified thoughout the entire fabrication cycle.

Design Rules

Many different IC manufacturers supply fabrication services to brokerage services such as MOSIS. These multiple sources are necessary to avoid delays due to heavy production loads of specific fabricators, or to fabrication-line failures. It is important that users realize that there are as many different geometric design rules for a given technology as there are fabricators that support it.

It is not desirable for silicon brokers to deliver a large set of design rules to the user community in an attempt to match each fabrication line. One reason is that these design rules are often proprietary. A more important reason, however, is to shield the designer from the idiosyncracies of particular fabrication lines. The MOSIS set of design rules are conservatively designed to match most fabricators' requirements. (The most widely known generalized design rules for nMOS depletion-load technology are the Mead-Conway design rules.)

Generalized geometric design rules are usually created to match as many fabricators as possible without giving up too much circuit area and speed. These rules are developed through study of the design rules of many fabricators, and by scaling generalized rules to the average relative feature size. In this way, any deviation from one fabricator to another can be corrected by a small "bloat" or "shrink" of selected layer features. This bloat or shrink is performed by the silicon broker; it is not apparent to the designer.

Submission of Designs

The first interface is the communication of a design file from the originating system to MOSIS. This transfer is accomplished only via electronic messages (either by "electronic mail" or by "file transfer"). Unfortunately, many of the systems used for VLSI design are not connected directly to a computer network; therefore, a file-migration sequence between computer systems is needed until a network access point can be reached. The migration process is sometimes less than perfect, due to communication errors or to various "favors" provided by friendly operating systems. Therefore, guarantees have been applied to ensure the integrity of the original design file until it finally reaches the MOSIS system. Conventional communication-protection techniques cannot provide the desired level of protection. For end-to-end protection, MOSIS uses a special checksum function (designed to protect CIF files and to ignore sequences that do not affect the CIF interpretation) in addition to "hop-by-hop" protection along each communication link.

Converting CIF to MEBES

When a fabrication run closes, the CIF files are converted into MEBES, the pattern-generation format used by e-beam machines to write masks. Full-wafer masks are used because the many different dies involved in each run make it impractical to step and repeat 10X reticles.

Like any other program of substantial size, the CIF/MEBES conversion is practically impossible to verify mathematically. Instead, it must be "validated" by extensive checking of a variety of test data designed specifically for this purpose. Experience has shown that this test data should include deeply nested geometrical transformations, acute ("sharp") angles, and ill-conditioned polygons.

Even after the code for the CIF/MEBES conversion has been accepted, it will still have to be upgraded, improved, and adapted to possible changes in the operating environment. As this software is modified, it must be validated. Validation is usually a matter of comparing the new versions with the older

(and already verified) ones, to make sure that a given CIF input produces the corresponding MEBES output.

These patterns cannot be compared digitally, because the MEBES representation is not unique. Therefore, optical methods are typically used: either writing the patterns side by side on glass for comparison under a specially built microscope, or writing out the "exclusive OR" of these patterns and examining it with a standard microscope.

Mask-Making

The next step is the transfer of MEBES files to the mask house, together with control information specifying the location of each pattern on the mask, its polarity, and the required "bloat." Unfortunately, of all the steps in fabrication process, this step is the most error prone. Although mask-making technology is relatively advanced, the information-transfer technology for this purpose is quite crude. Most mask houses can control pattern-positioning to within 1/40 of a micron, with spot size of 1/4 micron. At the same time, most mask houses can only accept information from users via 800-bpi magnetic tapes. (These tapes were the industry standard some time ago; today's standard tape densities begin at 1600 bpi.)

The process of bringing the mask-maker and the fabricator to agreement on the actual size of mils and microns is usually referred to as "correlation."

The "standard" software used for this transfer does not provide enough error protection. Our experience has shown that more files are damaged during this transfer than in any other step.

The interface between the mask house and the fabrication line also includes the following details:

- Mask size (*e.g.*, 4" x 4" x 0.060" or 5" x 5" x 0.090")
- Type of glass (*e.g.*, standard or low-expansion)
- Type of chrome (*e.g.*, anti-reflective or low-density)
- Layer polarity (dark or clear)
- Critical dimensions and alignment marks (which may differ from line to line)

Critical-dimension (CD) marks are usually designed to yield the minimum design-feature size on the silicon. For example, if it is known that on a certain fabrication line a 4.0-μm feature on the mask yields silicon features between 3.3 μm and 3.7 μm, then the mask features must be expanded to 4.5 μm so as to achieve 4.0-μm silicon features. Therefore, the CD on the wafers are 4.0 $\pm$ 0.2 μm. This adjustment of feature sizes on masks is called "bloating/shrinking" (or "compensation").

In this case, the CD mark for this level would be 4.5 μm. Both the mask-maker and the fabricator should be notified of this fact, so that they can verify feature sizes on the masks. In addition to the size of the CD in "mils" (milli-inches) or microns, the size of the mils and microns themselves must also be specified. Not only do some industrial companies differ on the size of a mil (some say it's 25.4 μm; others, 25.0 μm), but they also use microns of several sizes (due to empirically derived dimensional standards).

What is MOSIS?

MOSIS* is the fast-turnaround "silicon broker" that serves the requirements of DARPA's VLSI research program that involves participants from several universities and other research and development organizations. To meet the need for fast-turnaround fabrication at reasonable cost, DARPA has established the MOSIS system at USC/ISI. The MOSIS system accepts design files (expressed in CIF) that are submitted via a computer network such as the ARPANET or TELENET.

These designs are fabricated by commercial fabrication lines, on wafers that combine many die types, each of which may contain several different projects. This procedure spreads the fabrication cost over many independent projects.

After verifying the quality of the fabricated wafers by parametric probing, the wafers are sawed into separate dies, which are individually packaged in standard DIPs and sent to the designers. Because of the prototypical nature of the circuits, MOSIS does no functional testing.

The MOSIS system has successfully completed more than 20 fabrication runs by various commercial vendors. The total turnaround time in the MOSIS operation is 4 to 6 weeks, except when unusual fabrication problems occur.

*The MOSIS system was developed at USC/ISI under the sponsorship of the Defense Advanced Research Projects Agency (DARPA) under contract No. MDA903-81-C-0335. The views and conclusions in this article are those of the authors, and should not be interpreted as representing the official opinion or policy of DARPA, the U.S. Government, or any person or agency connected with them.

A micron is theoretically well defined. Unfortunately, this precise definition involves a comparison with one one-millionth of a standard meter, once defined as the length of a certain rod of platinum (kept in Paris), but now defined as *1,650,763.73 wavelengths of radiation in a vacuum, corresponding to the unperturbed transition between levels 2p10 and 5d5 of the atom of Krypton-86 (the orange-red line)*. Both of these definitions are very precise; but they are not very practical when one has to accept masks. Standard mask-acceptance procedures use calibration plates to define the most commonly used sizes. The variation in these standard calibration plates exceeds 0.25 μm for 5-μm features, depending on factors such as the source and the date of the plate, and the feature polarity. The process of bringing the mask-maker and the fabricator to agreement on the actual size of mils and microns is usually referred to as "correlation."

The variation between the actual size of features on the masks and the "drawing size" depends not only on the bloating applied (if any), but also on the actual parameters of the photoprocessing in the mask-making operation. Therefore, the final feature size has to be verified before the mask is delivered to the fabrication line.

In our experience with MOSIS, we had one case in which masks found by the mask-maker to be within a strict $\pm$0.25-μm specification were also found by a fabricator *not* to be within a more relaxed specification of $\pm$0.50 μm, when measured by the same instrument used by the mask-maker. This discrepancy was due to the above-mentioned variation in the size of microns, or, in other words, to the lack of adequate correlation.

Furthermore, because the chrome used for mask-making is

not perfect, defects are always present on the masks. These defects consist of either excessive chrome where it should not be, or lack of chrome where it should be. Depending on the polarity of a mask, these defects may cause extra connectivity or a lack of connectivity, either of which could be fatal to the functionality of the device. Even though some defects (*e.g.*, an extra piece of metal that does not "short" any devices) may cause no harm at all, the removal of as many of these defects as possible is very important. This task is usually carried out by the QA departments of good mask-making operations and fabrication lines.

Defects may be found either by visual ("manual") inspection or by automated means for optical comparison of similar dies. The latter method requires the use of special equipment, and requires that certain geometric constraints pertaining to that equipment be applied to the placement on the mask of the dies to be compared.

Fabrication Lines

The next interface, and probably the most important one, is with the silicon fabrication line. This interface consists of specifications and negotiations leading to masks in and wafers out. The main issue is the preparation of designs that comply with the broker's design rules for optimal fabrication on a given line.

> Suppose that on a certain run, a contact was found to have a yield of 99.99% (i.e., one contact in every 10,000 fails). In other areas, this may be an acceptable yield; but for a 64K memory chip with at least 5 contacts per bit, it is a catastrophe.

This issue is resolved by finding the best bloat/shrink values to be applied to the various layers of the designs, such that the smallest value can be used as the "feature size" (or "lambda").

Unfortunately, design rules are not easy to verify. They do not sharply divide all the circuits that are guaranteed to operate from those that do not work. Design rules represent the best possible compromise between yield and performance. The more conservative the rules are, the more likely it is that a circuit will work; the more aggressive they are, the better a circuit's performance will be, possibly at the expense of yield.

Fabrication quality is usually measured by the quality of the resulting parts. Functioning devices are the ultimate testimony to quality. But a silicon broker, dealing with prototype devices, must have objective quality measurements that do not depend on the quality of the design submitted by the customer.

The broker must be able to defend the quality of the fabricated parts and to tell customers (after taking into account reasonable yield): "If it doesn't work, check your design." A broker who cannot say that, should not accept the wafers. To reach this position, the broker must be sure that designs that comply with the design rules produce working parts with a reasonable yield. And the broker must obtain this assurance for each separate run.

As mentioned above, design rules are hard to verify. For example, consider a simple contact built according to given

design rules. Suppose that on a certain run, a contact was found to have a yield of 99.99% (*i.e.*, one contact in every 10,000 fails). In other areas, this may be an acceptable yield; but for a 64K memory chip with at least 5 contacts per bit, it is a catastrophe.

Suppose, further, that steps were taken to improve the yield to 99.9995%. How can one verify this yield improvement? Surely not by testing a small set of contacts (*e.g.*, like one on every die). Only very extensive random fault structures (RFS), built specially for this purpose, can provide relevant information about the yield level.

Because each wafer has its own unique "misalignment set," the RFS analysis has to be performed on every wafer, on every run, to validate the yield of such a contact. And this is just one example. Doing it for all design rules, both geometrical and electrical, in quantities that provide significant statistics, may not leave enough room for the actual "payload" (the parts for the users); therefore, it is not very practical.

Consequently, wafers have to be accepted on the basis of tests that are not exhaustive, but that are revealing enough to inspire confidence. The design of such a test set is still a research issue. Several MOSIS users are also doing work in this area; in particular, researchers at the University of California at Berkeley, Stanford University, the Jet Propulsion Laboratory, and the National Bureau of Standards are participating in the effort to maximize confidence while minimizing test-structure area.

Prototype runs in R&D are different from production runs because the former include many designs which, along with their testing procedures, have not yet been debugged. Some prototype devices experiment with new circuits and concepts, and some devices "push" the standard design rules slightly to determine their limits. Therefore, it is important to be able to verify the quality of the wafers, so that designers can address design and circuit issues rather than having to direct the entire testing effort toward solving yield problems.

In summary, the broker's tests have to verify that the designs that comply with the broker's design rules have a reasonable yield. The broker also must verify experimentally the electrical parameters of the wafers, so as to be able to tell users the parameters to expect and the values to use in their circuit simulators. To meet these goals, a good brokerage should have test vehicles that can provide the required parameters.

Most fabrication lines have their own process-control monitoring (PCM) devices. (These devices are "dropped in" at several places on every wafer, and hence are often called "drop-ins.") The PCM devices help to "fine-tune" a line by monitoring it constantly; they also form the basis for their own continuing quality-assurance program. Because these structures are unique to each foundry, a common basis is necessary for monitoring MOSIS fabrication runs.

Wafer Testing

When a contract is written for wafer fabrication, the acceptance criteria should be defined. There are many ways to define these criteria; however, most of them state that a certain number (usually M out of N) of tests must be within a given specification. This leads to more questions: Which parameters? What specifications? Which devices? How many? How measured? and others.

For example, the simple specification "V_{th} = 0.8 V" raises

many questions. (Fortunately, the magnitude of volts, unlike microns, does not have to be specified.) Is 0.81 V out of the specification? Obviously not, furthermore, a range must be defined. Is it everywhere on the wafer, or only on certain sites? What if it is met in only 4 out of 5 areas? In only 3? Which devices should be used for the measurements (and by whom)? Should minimum-size devices or large ones be used? What method of measurement should be used?

V_{th} is defined as the voltage at which a transistor starts to conduct. This "starting" point is not well defined. It is best to think of it as the intersection of the maximum-slope tangent of the I_{ds}/V_g graph with the V_g axis. It should be computed by obtaining many points on this curve. Because obtaining points is a time-consuming chore, it is common practice to compute V_{th} by extrapolation from only two points (and sometimes from only one). Different measuring methods yield different results. Therefore, the measurement method, as well as the range of accepted values, must be specified. Furthermore, V_{th} measurements on minimum-size devices yield different results than do measurements on larger devices; therefore, the size of the measurement device must also be specified.

MOSIS uses the following nMOS depletion-load device parameters as a basis for accepting wafers:

- Enhancement and depletion thresholds
- Field threshold
- GAMMA (the SPICE model parameter)
- Junction breakdown
- Feature sizes
- Sheet resistances for diffusion, polysilicon and metal

These PCM test results do not provide MOSIS with the required level of characterization of the fabrication run. Instead, they are used mainly as wafer-acceptance criteria. MOSIS must also give device-performance information to users, both *a priori* and *a posteriori*, to let users predict and analyze the performance of their circuits. Therefore, parametric monitors must be used to characterize each wafer on each fabrication run.

If the spatial distribution of parameters across wafers is guaranteed to be uniform, then just a very few test sites are needed to determine the statistics of these parameters and to verify the uniformity of the spatial distribution. Unfortunately, this distribution is not very uniform; therefore, many parametric measurements must be made to gain an understanding of the parameters. Because each test device uses an area which otherwise could contain user circuits, a compromise must be reached on the allocation of silicon area for test vehicles.

MOSIS compromises by applying a few drop-in dies to every wafer for tests that require a large area, and by placing small parametric test-strips (consisting of many smaller tests) on each die.

Test Dies and Silicon 'Canaries'

The drop-in test dies provide information (including yield) for the generalized design rules (for example, random fault structures). They also include "typical comprehensive user devices" which are expected to represent the circuits on the wafer. These circuits are intended to give "early warning" of fabrication problems. It is hoped that if unexpected problems arise during fabrication (*i.e.*, problems that do not affect, and

therefore are not caught by, any of the parametric measurements), this "canary" circuit will not behave as usual. Subsequent functional testing will reveal the presence of problems, without however indicating their exact nature.

Canaries warn against creeping problems, just as canaries were used in coal mines to warn against poisonous fumes. These silicon canaries are useful for predicting the yield for "typical" circuits submitted by users. The canaries are static and dynamic circuits designed according to generalized geometric design rules so as to test the limits of the published design rules. It is hoped that these circuits represent users' designs. A canary that "sings" perfectly, when none of the user designs works, is clearly too robust.

The MOSIS drop-ins also include patterns and circuits designed to optically and electrically measure the precision of the alignment between various layers—especially for the "invisible" layers. Because the spatial frequency of the parameters is not as high as the die placement, it is possible to mix

Canaries warn against creeping problems, just as canaries were used in coal mines to warn against poisonous fumes. These silicon canaries . . . are static and dynamic circuits designed according to generalized geometric design rules so as to test the limits of the published design rules . . . A canary that "sings" perfectly, when none of the user designs works, is clearly too robust.

several types of test strips on one wafer, thus increasing the number of parametric tests performed on each wafer.

The quality control program for the MOSIS service was developed with the following goals in mind:

- To provide wafer acceptance and selection criteria;
- To correlate vendors' PCM test results with those of MOSIS, to monitor the fabrication quality;
- To determine defect density and yield of devices designed with generalized design rules; and
- To extract transistor-model parameters for circuit simulators (*e.g.*, SPICE).

The test strip placed on each die provides primarily the information needed to meet the first two goals; it also provides some of the information needed to meet the last two. A set of test structures was developed for use on the test strip to provide as much information as possible, while occupying the least area.

The test strip provides the following information:

- Sheet resistance of diffusion, polysilicon, and metal
- Electrical line-width measurements of diffusion, polysilicon, and metal
- Contact resistances of metal-polysilicon, metal-diffusion, and buried contacts. It is also necessary to test the contacts to determine whether the current direction is symmetric.
- Field-oxide transistor-voltage threshold (metal-field, poly-field)

- Gate capacitance (C_{ox})
- Inverter characteristics (including V_{high}, V_{low}, inverter threshold, and gain) for K = 4 and K = 8 inverters. (Conceptually, K is the ratio of pull-up dynamic impedance to pull-down dynamic impedance.) The K = 8 inverters are driven through pass transistors.
- Inverter delay, as derived from ring oscillator performance. The ring oscillator also serves as a small yield-monitoring device.
- Transistor characteristics, including voltage threshold (enhancement and depletion, GAMMA and KP at several substrate bias voltages) and subthreshold leakage current in enhancement transistors
- Source/drain diode breakdown voltage

Our experience with the MOSIS system has shown that packaging is not usually a problem, . . .

While the projects are being packaged, probe data is analyzed to extract transistor model parameters and yield data. Data from the drop-in test dies is used in conjunction with the test-strip data to extract transistor-model parameters. Yield information is obtained from a combination of random fault structures (RFS), and from static and dynamic canaries.

Extraction of SPICE model parameters in the MOSIS environment is somewhat more approximate than one might expect. Because MOSIS is not limited to only one fabricator (or even necessarily to the same production line in a given fabricator's facility), it is impossible to "fine tune" the model parameters to force the simulation results to match the actual measured device characteristics. Therefore, for every run, MOSIS provides the users with *a posteriori* device parameters that represent the observed device characteristics for every wafer and for each individual die. Obviously, for prediction purposes, this is not necessarily as accurate as it would be if the same fabrication line were always used. The goal is to provide users with typical model parameters that track the different fabrication lines. Users requiring more accurate information can be informed in advance which model parameters best match the fabrication line to be used in the next run.

Random fault structures are included in the drop-ins on each wafer run to permit monitoring of fabrication quality. Each RFS is designed to provide information about the relative reliability of a certain feature commonly used in circuits. A typical drop-in test die contains strings of contacts between layers (to test for contact reliability), long runs of conductors running over oxide steps (to test for conductor-step coverage), interdigitated conductor runs (to test for bridging faults), and other features. Larger random fault structures are included on runs used for qualifying new vendors or new fabrication lines, to estimate the relative yields of the various circuits on the devices. These structures include wires in each layer, contacts of all types, wires over oxide steps, etc. These structures provide MOSIS with a means to determine whether the standard design rules are pressing the limits of the fabrication line.

The Final Steps

The next interface is with the device packagers. The packager is given wafers and bonding maps, and returns packaged parts. Due to the relative simplicity of this operation, it is easy to detect packaging errors such as loose contacts, broken bonding wires, and wrong bondings. Our experience with the MOSIS system has shown that packaging is not usually a problem, and that random checks suffice to verify the integrity of this phase.

The MOSIS system places on every die a ring oscillator that is bonded to the leads of the package, unless it conflicts with a user's circuit requirements. The routine check of these oscillators, via the bonded pads, gives MOSIS a large bonding sample, and serves as a (more or less random) packaging check. On one occasion, this routine check revealed a systematic problem with the die attachment. Once this problem was identified it was solved.

The marking of both wafer runs and individual dies has proved to be very important. Individual die markings, consisting of unique numbers assigned to each die on the wafer, let faults be traced to a given position on the wafer. Through human error, dies were once interchanged on a mask, controlling an implant layer that was invisible (naturally, in accordance with Murphy's Law) and hard to find. On another occasion, during fabrication, two masks of the same layer were interchanged between two different MOSIS runs. Due to the repetitive positioning of both the MOSIS test-strips and the PCM devices belonging to that vendor, neither of these interchanges would cause any test to fail; but they would destroy most (if not all) of the user's devices. (What a robust canary, to keep singing among all the poisoned miners!) In both cases, the markings revealed these interchanges, and in both cases the mistakes were caught immediately.

One of the last interfaces occurs when the devices are sent to the designers. This is the last interface for MOSIS, but only the first (or second) step in the designer's long journey from concept to a debugged working system.

About the Authors

Danny Cohen received the BS degree in mathematics from the Technion (Israel Institute of Technology) in 1963 and a PhD degree in computer science from Harvard University in 1969. Danny is a project leader at USC/ISI. His research interests include aircraft cockpit information systems, interactive real-time systems, flight simulation, voice communications, computer architecture, VLSI, computer networking, and graphics. He has been on the computer science faculties of Harvard University, the Technion and Caltech.

Vance C. Tyree earned a BSEE in 1965 and an MSEE in 1966 at the University of California, Berkeley. He is currently a member of the research staff at USC/Information Sciences Institute and is an Adjunct Assistant Professor in the UCLA Computer Science Department and the Electrical Engineering Department. He is developing the quality assurance program for the MOSIS IC brokerage service. His research interests include fault-tolerant VLSI system architectures and concurrent architectures for VLSI systems.