3. An Overview of IC Implementation

3.1 Mask Generation

MOS circuits are constructed as a sequence of patterned layers on the surface of a silicon wafer. Each layer requires a different *working plate*, or *mask*, to provide the pattern; for Si gate NMOS there are typically six working plates in a set. A working plate is a sheet of glass about 100mm square covered with patterned opaque material on one side. This material is usually one of three materials -- photographic emulsion, iron oxide, or chromium. Emulsion is the least expensive but relatively easily damaged in the contact photolithography process (see section 3.2 The Basic Fabrication Process). Chromium and iron oxide give better line resolution and are very hard, but they are more expensive.

The first step in creating the working plates is plotting the files provided by the designer on a photosensitized glass plate. This first plate, called a *reticle*, differs from a working plate in that it contains only one copy of the relevant chip layer and is plotted at 10x the size of the chip on the wafer. The plotting process takes place in a *pattern generator*; the Mann 3000 is typical of such machines. The 3000 projects (*flashes*) the image of a variable size rectangle on the reticle. The size of the rectangle, or *aperture*, the x and y coordinates of the center and the angle with respect to the x axis are specified by the designer. The nature of the reticle making process has a number of important implications for the designer. All shapes on the masks must be decomposed into simple rectangles. The pattern generation process involves complex mechanical motion; proper sorting of the individual rectangles of which the chip is composed can speed up the pattern generation process considerably -- and thus lower the price. For example a Mann 3000 PG machine is fastest at moving in the x direction, followed by aperture change, y direction, and finally, angle change.

Special features have to be included on the reticle which are used in mask making and in IC processing. *Critical dimensions* (CD's) are simple lines or crosses of a fixed size appearing on each layer; they are used by the mask house to adjust exposure and developing time to insure that these marks and hence other features on the mask are the correct size. It should be noted that exposure time has a definite effect on the feature sizes on the reticle, in particular overexposed areas tend to "grow" slightly; for this reason the designer should avoid substantial overlap between flashes. A *parity mark*, consisting of an arrow or triangle, is sometimes included on each mask layer to help the operator orient the mask. The mark is placed outside of the boundary of the chip pattern. *Fiducials* are small crosses which also appear on each layer outside of the boundaries of the chip. These are used in the *step and repeat* process (see below). Often the parity marks and fiducials are

provided by the mask house thus making it unnecessary and undesirable for the designer to supply them. Parity marks and fiducials appear only on the reticles and not on the finished working plates.

Because the interactions between layers are all important in MOS integrated circuits each layer (with the exception of the first) must be critically aligned with a previous layer so that features overlap in the proper way. The fabrication line operators use a set of special patterns on the working plates called *alignment marks* (further described in chapter 4 and section 7.1) to accomplish this.

The reticles are used to make a set of *master plates* in a step and repeat machine which projects an image of the reticle (reduced 10x) onto a photosensitized plate. By precisely stepping the image across the master a matrix of images of the reticle is created. The fiducials are used to control the distance between exposures and to align the reticle images relative to each other.

The working plates are made directly from the masters by contact printing. In cases where a large number of working plates are needed the mask house may make several sets of *submasters* and print the working plates from them.

To facilitate checking of the various layers for mistakes the designer may request color enlargements of the reticles for checking the various layers for mistakes; these *blowbacks* are typically about 100x-150x actual (chip) size. *Black and clear* transparencies (usually 8½" x 11") may also be made at the same time. They are sometimes used in the interaction between the operators on the fab line and the designer to indicate features on the mask such as alignment marks.

3.2 The Basic Fabrication Process

The process discussed here is the standard Si gate n-channel MOS process. The reader need not be concerned with learning all of the details of the process, indeed most of the decisions concerning processing are made by the fabrication line. The designer may not know which particular techniques the fab line uses, and may not care as long as his standard circuits exhibit normal performance. This section is presented to provide background for those interested in what really happens behind the clean room doors.

Integrated circuits are built in layers, some of which are patterned by a *photolithographic* process. Such layers are created in a sequence of steps beginning with the deposition (or growth) of some material on the surface of the wafer. It is then coated with a thin layer of photosensitive chemicals, called *photoresist*, and exposed to ultraviolet light through the proper working plate. The exposure can take place with the working plate pressed against the wafer (*contact photolithography*) or by projecting an image of the working plate onto the wafer (*projection photolithography*). (Projection techniques are becoming more widely used in spite of the extra equipment and maintenance needed since the masks are subject to less wear and contamination than contact masks. Consequently masks last longer and it is easier to control certain kinds of defects incurred in the photolithography steps.) If *negative resist* was used those areas of resist which were exposed to light will be hardened while *positive resist* is softened in the exposed areas. The resist is *developed* by immersing it in a solvent which dissolves the unexposed (for negative resist) or exposed (for positive resist) portions, leaving the desired pattern.

After the developed resist is hardened by baking at a low temperature the material in the uncovered areas is removed by *etching*. Two techniques are widely used today. In the older *wet etching* process the wafer is immersed in a bath of chemical etchant under controlled temperature conditions for a specific amount of time. Wet etching depends on the availability of an etchant which will dissolve the layer beneath the photoresist, yet not significantly attack the resist. The wet etchants for some materials, for example silicon nitride, dissolve photoresist as well as the desired material. Such materials require an intermediate pattern to be formed in another layer which serves as the actual etching mask. *Dry etching* techniques such as *plasma* etching use a stream of ions and electrons to blast away material. The plasma etching technique gives better results for fine geometries and also permits the direct use of resist as an etching mask. Following the etching step the remaining resist is removed, leaving a pattern in the underlying material.

This sequence is repeated for the various layers of the circuit. About six photolithography/etching

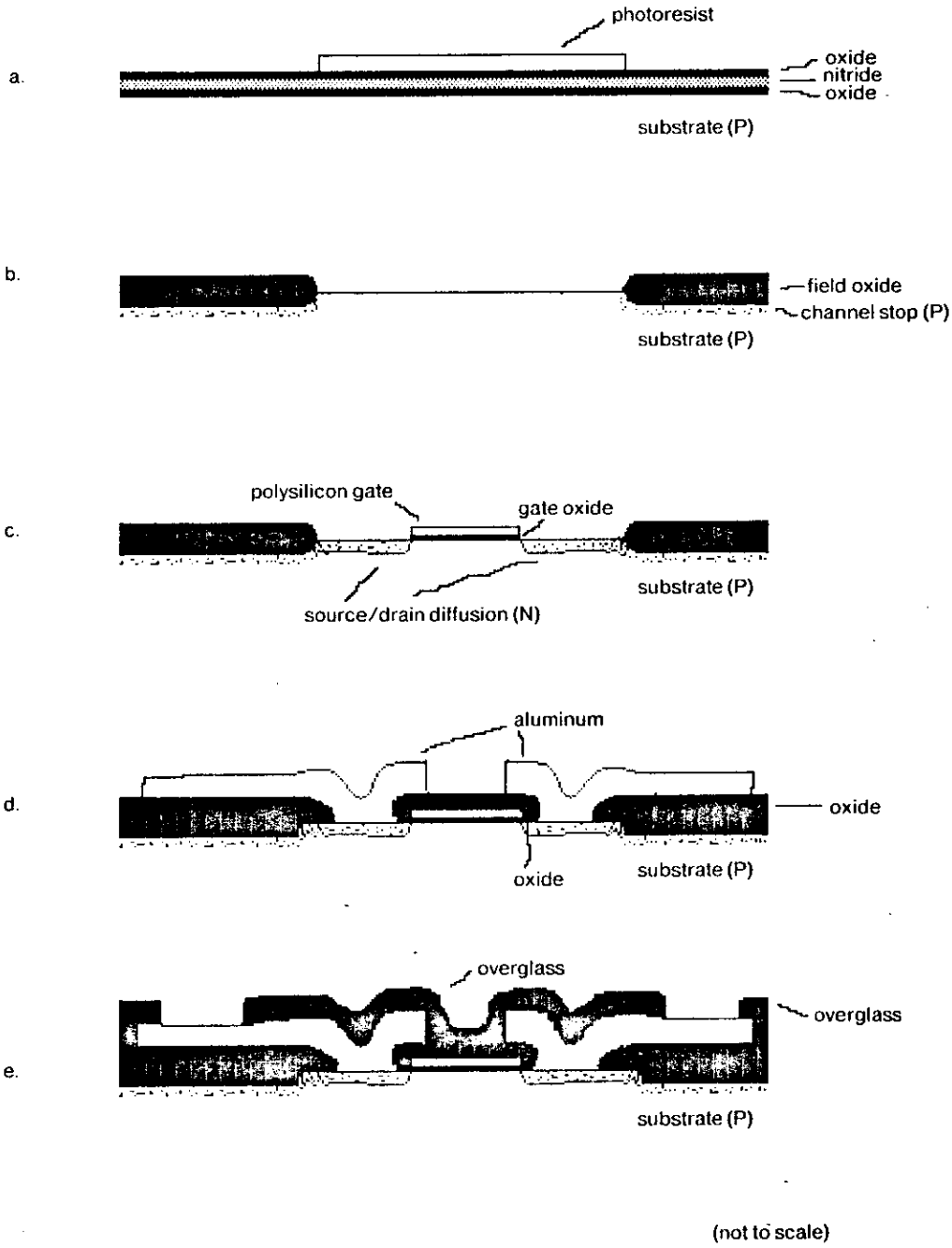
cycles are required to build up a typical Si gate NMOS circuit. The entire process entails over forty individual steps, outlined here.

The wafer of *p type* (*100 crystal orientation* for lowest interface state density) silicon is scrubbed and a thin layer of silicon dioxide (hereafter called "oxide") is thermally grown on the surface. This layer serves as a mechanical buffer zone for the silicon nitride (Si_3N_4) that follows. The buffer zone is needed to relieve stress caused by differences in the coefficients of thermal expansion of silicon and silicon nitride. A layer of Si_3N_4 is deposited by *chemical vapor deposition*, then another layer of oxide is grown. Photoresist is applied over the entire surface and the wafer is exposed to ultraviolet light through the *diffusion layer* mask. The resist is developed, leaving open areas over the *field* region (figure 3.2.1a). The top layer of oxide is etched away wherever there is no photoresist using a hydrofluoric acid solution. After the resist is removed this top layer of oxide is used as a mask for patterning the nitride since the photoresist alone will not stand up to the chemicals used in the wet etching of silicon nitride. A third etching step is used to remove the bottom layer of oxide. *Ion implantation* is used to place the *channel stop* region and a thick *field oxide* is grown over those areas. The field oxide and the channel stop are *self-aligned* with respect to the source/drain diffused areas (the nitride covers the source/drain areas during channel stop implant and prevents oxidation of the underlying silicon during field oxide growth). The remaining nitride and the thin oxide under it are removed resulting in the profile shown in figure 3.2.1b.

Next a layer of photoresist is applied and the wafer is exposed through the *depletion mode implant* mask. The resist is developed, leaving open spaces in the gate regions of the depletion load transistors. Another ion implantation step occurs here (using the resist as a mask) to alter the threshold voltages of the depletion load transistors. The resist is removed and a thin layer of *gate oxide* is grown. If there are *buried contacts* used in the IC design more photoresist is applied, the wafer is exposed through the buried contact mask, the resist is developed, the gate oxide is etched away in the contact areas, and the resist is removed. This allows the *polysilicon gate* material to contact the substrate in selected areas.

A layer of polysilicon is deposited from a chemical vapor and a thin layer of oxide is grown on top of that to provide a surface that photoresist will adhere to. Resist is applied and the wafer is exposed through the polysilicon layer mask. The development of the resist leaves the gates of the transistors covered; the uncovered areas of oxide and polysilicon are etched away (a little field oxide is also removed). After the resist is removed the *source* and *drain* regions are doped (figure 3.2.1c) in a phosphine gas atmosphere. Since the edges of the polysilicon gates define where the source/drain regions begin these features are also self-aligned. Here self-alignment results in a

Figure 3.2.1 Si Gate NMOS Processing Steps



(not to scale)

significant reduction in parasitic capacitance due to the near zero gate to source/drain overlap. A thick layer of oxide containing P_2O_5 is deposited over the surface of the wafer. This layer is reflowed for better coverage of the steps in the surface and a layer of photoresist is applied. *Contact hole* areas are defined using the contact cut mask and the oxide is etched away where the metal layer will contact the underlying features. After the resist is removed the source and drain are doped again (this is to prevent a phenomenon called *spike-through* -- essentially shorting of the aluminum contacts and the substrate through the shallow source and drain regions). A layer of aluminum is evaporated onto the surface of the wafer, followed by the application of more photoresist. Exposure (and subsequent development) through the metal layer mask leaves resist protecting the metal runs and contacts. The uncovered aluminum is etched away and the resist is removed (figure 3.2.1d). The wafer is then *annealed* (heated at a low temperature) to remove radiation damage resulting from the electron beam which is used to heat the aluminum during the evaporation process.

A thick layer of oxide is deposited on the entire surface of the wafer to provide physical protection. Windows to the bonding pads are etched through this layer in another photolithography step using the *overglass layer* mask. At this point (figure 3.2.1e) the wafer is finished, ready to be broken apart, bonded and tested.