Chemical Mechanical Polishing (CMP)

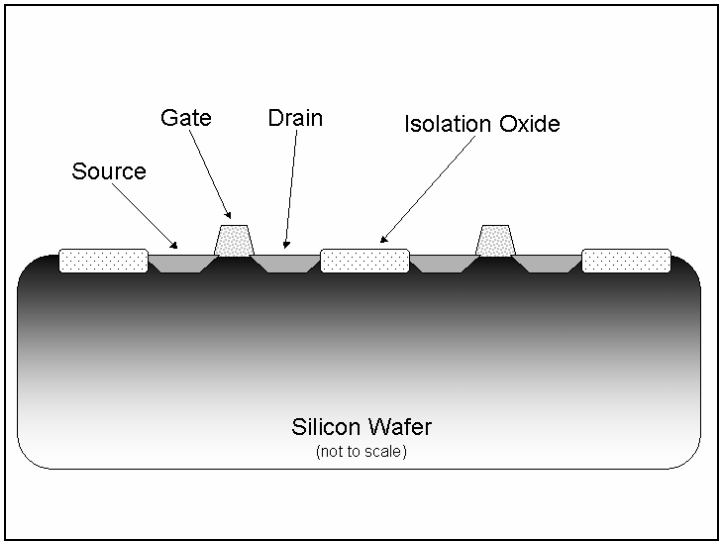
By Dan Woodie

Cornell University
www.cnf.cornell.edu

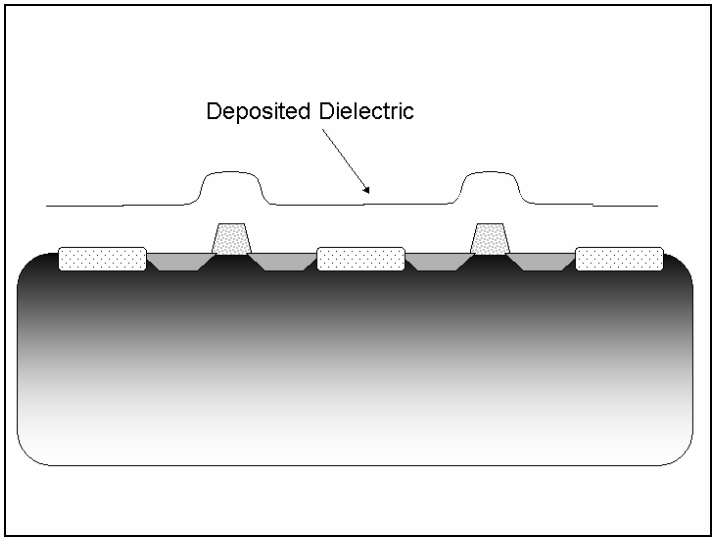
This document is intended to introduce the techniques of Chemical Mechanical Polishing (CMP) to the new user so that they may be better prepared to integrate CMP into their process flow. CMP can be used for many different purposes but it is limited in what it can do and what substrates it can handle. If after reading through this document you still have questions regarding what can be done here at the CNF, please contact the staff member in charge of the tool.

History

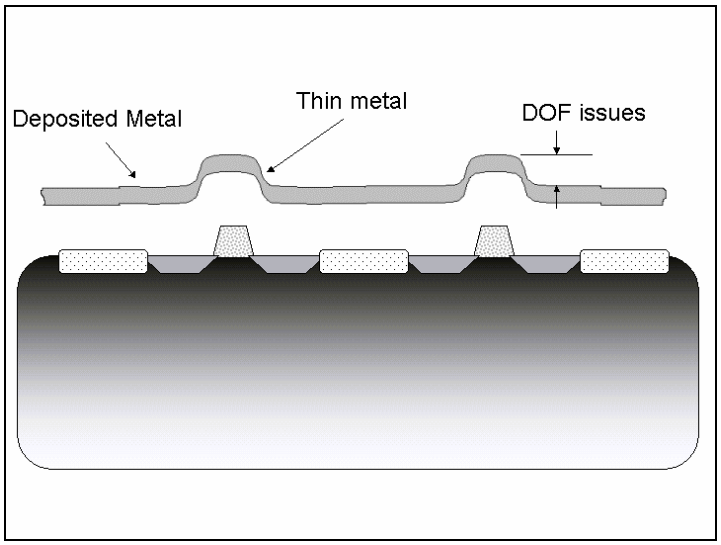
IBM invented CMP in the late 80's to allow for more metal layers in the integrated circuits (IC) that they produced. Originally it was called Chemical Mechanical Planarization (CMP) since that was the purpose for which it was created. A typical transistor wiring process flow of the time is shown.



After creating the transistors in the silicon, a dielectric (typically silicon oxide) was deposited. The deposited material replicates the step height of the underlying surface and in some cases can actually increase the topology. When the metal is deposited to form the first wiring level, the metal thickness can significant thin over the edges of the feature. This causes a reduction in the wire cross-section and a subsequent increase in the wire resistivity.

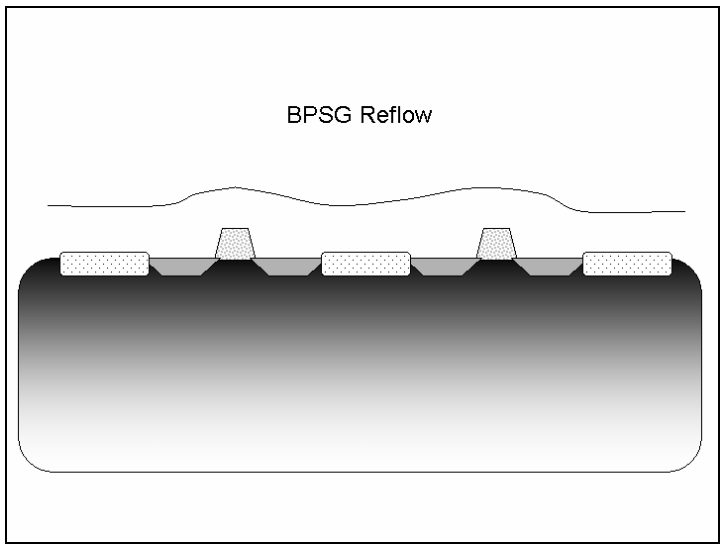


Additionally, the step height causes problems when trying to do high-resolution lithography. Pushing optical lithography tools to print ever-smaller features requires moving toward high numerical aperture (NA) tools. These tools can print smaller features at the expense of a smaller depth-of-focus (DOF) window. This requires that the surface height of the film they are patterning to be within a narrow range for the image to print accurately. Any topology in the surface makes it difficult to focus the image on both the high and low areas.

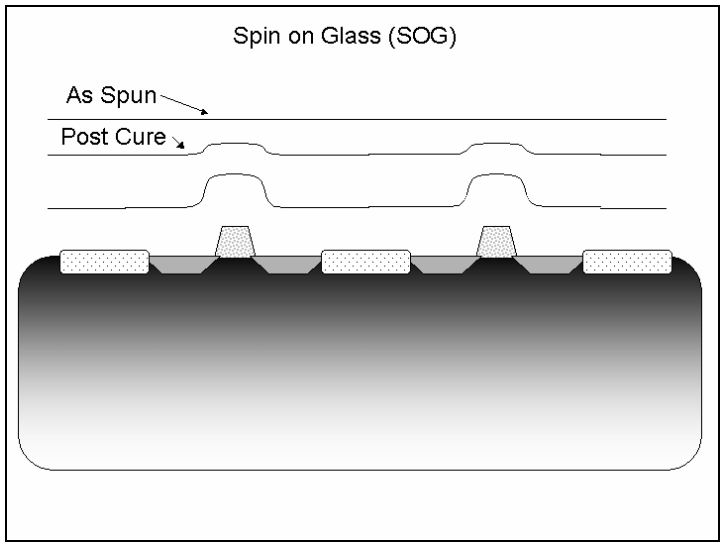


Other Planarization Techniques

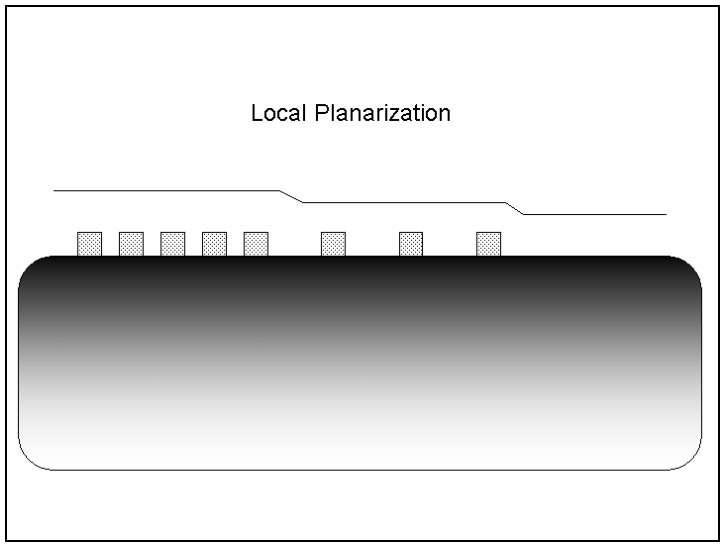
To remove step heights in the dielectric, IC companies used a variety of techniques prior to CMP. One technique was to deposit a silicon oxide layer heavily doped with boron and phosphorus, (Boron-Doped Phosphosilicate Glass-BPSG). This material has a lower melting point than undoped silicon oxide. A high temperature anneal was performed and the material would reflow slightly and smooth out the step heights.



An alternate strategy was to use a Spin on Glass (SOG) material. This is a liquid silicon oxide organic precursor that is spun on the wafer in a manner similar to photoresist. Being liquid, the material planarizes the surface before the solvent is baked off. After it is spun on, the material undergoes a high temperature curing process. During this cure, most of the organic constituents are driven off and the material shrinks to form a type of silicon oxide dielectric. The main issues with this type of process is that the quality of the oxide is very poor compared to a thermal oxide, and it does not completely remove the step height due to differences in the total film shrinkage between thick and thin areas.



Many other techniques were utilized as well but all of them suffered from various drawbacks. The main problem with even the best techniques was that they only achieved local planarization. There was still a height variation between areas of the chips that had different pattern densities. This caused depth of focus problems with the lithography steps.

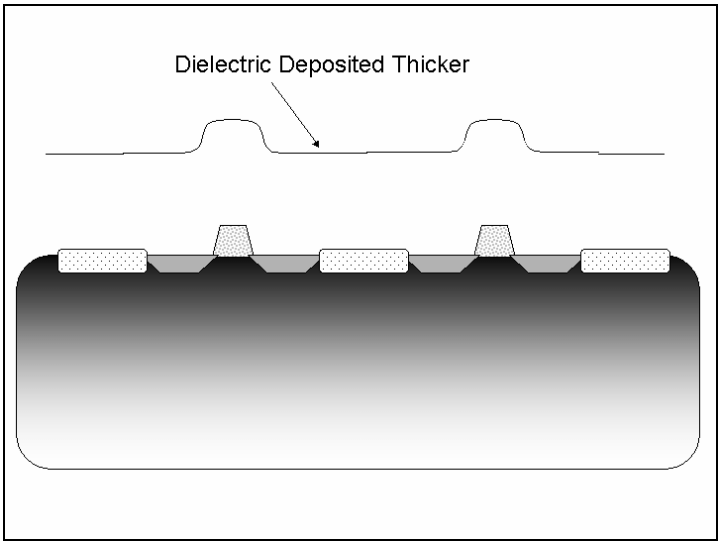


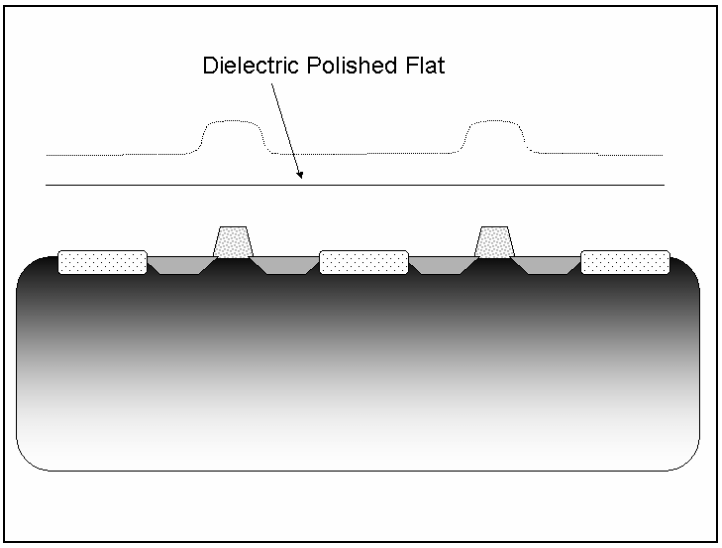
CMP Planarization

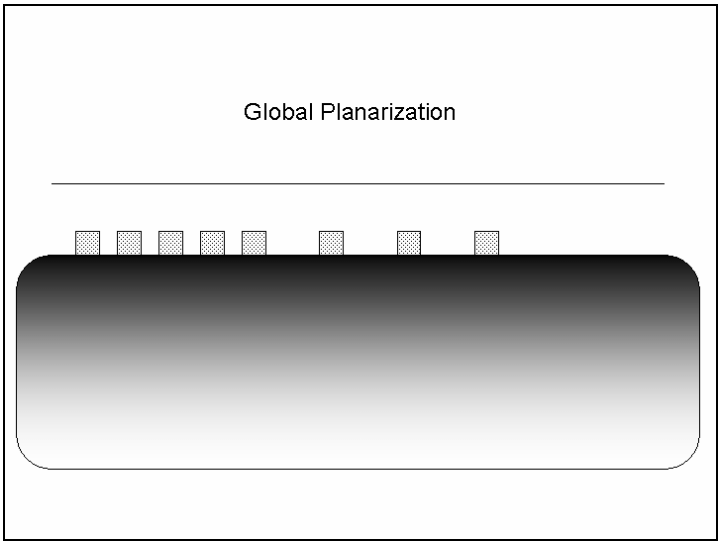
CMP improved on the alternate planarization techniques in many ways. The basic process is to deposit the silicon oxide thicker than the final thickness you want and polish the material back until the step heights are removed. This gives you a good flat surface for the next level. In addition, the process can be repeated for every level of wiring that is added.

CMP is the only technique that performs global planarization of the wafer. This is absolutely required to increase the number of wiring levels in the integrated circuits. Prior to CMP, DOF issues due to global planarization problems limited the total number of IC wiring levels to 3 – 4. With CMP, current state of the art IC production is able to achieve 7 – 8 wiring levels.

These achievements did not come without any cost. Many companies were hesitant to integrate CMP for several different reasons. One is that the process suffers from defect issues due to scratching of the wafer surface and from problems removing the abrasive particles when the polishing is finished. In addition, early CMP suffered from being a bit more of an art than a science. The polishing process was not well understood and variations in the material used to perform the polishing caused process shifts that were hard to correct. As the process has matured, many of those issues have been resolved and CMP is now viewed as a more accepted IC processing technology.



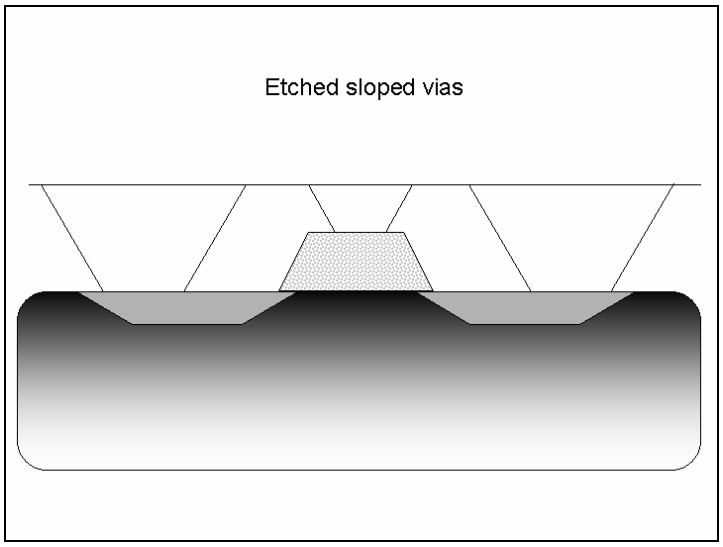


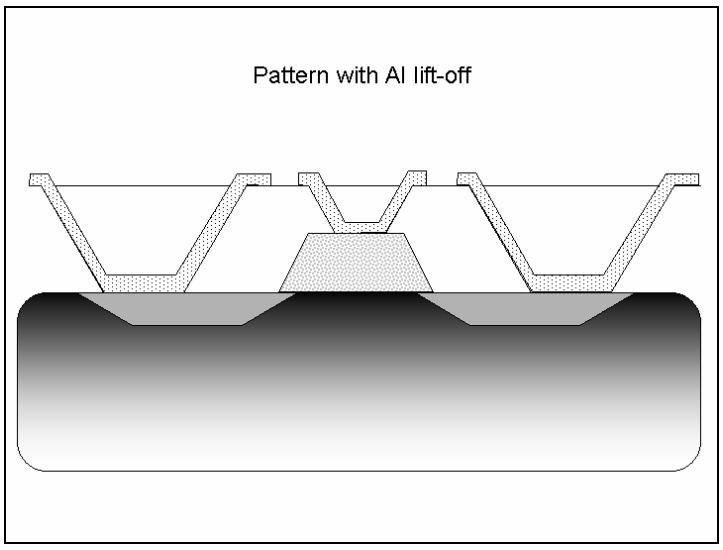


Damascene CMP

An alternate use for the CMP process was for creating inlaid metal patterns on the wafer for the wiring levels. This is called a damascene process. It was used to replace the traditional method of making electrical contacts between the IC wiring levels.

Traditionally to contact the source, drain, and gates of the transistors, large sloped holes or vias were etched into the dielectric, and the wiring metal was patterned over the hole and allowed to contact the lower level directly. The main detractor with this process is the amount of lateral space each contact took up, preventing high-density packing of the transistors.

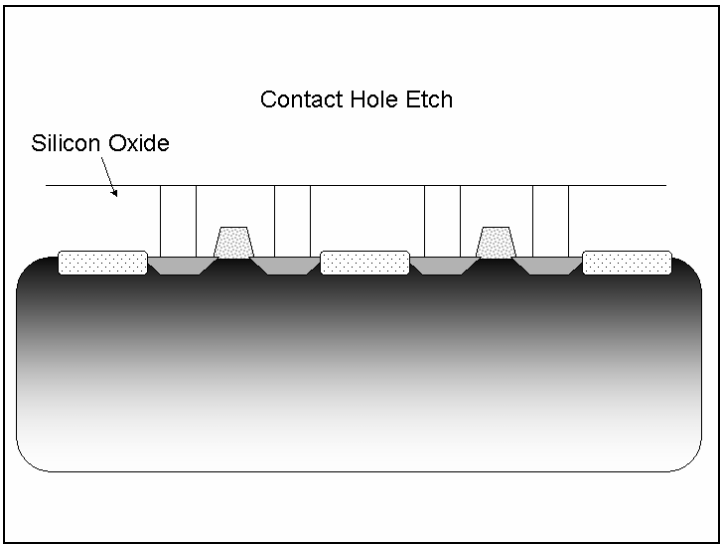


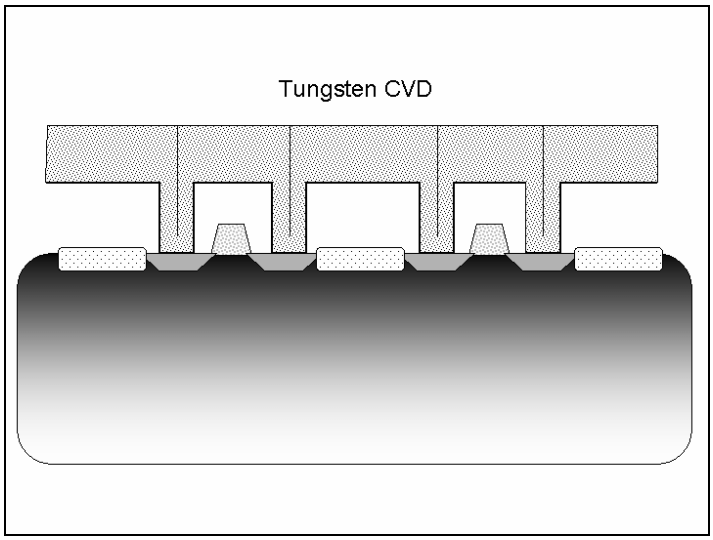


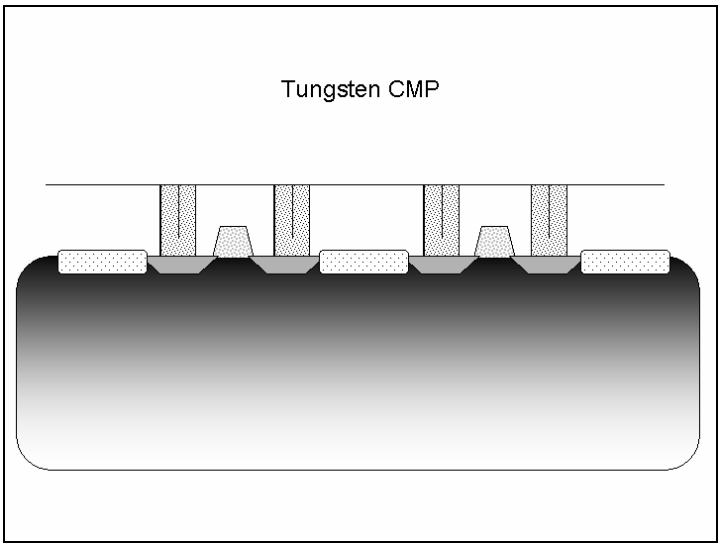
The tungsten damascene process starts with a fully planarized dielectric surface that is patterned with vertical contact holes. These holes can be made much smaller and spaced tighter than the sloped vias of the previous process. Tungsten (W) is then deposited using a chemical vapor deposition process to produce a uniform coating thickness on all exposed parts of the wafer. In this two phase process, the tungsten precursor (WF_6) migrates to the wafer surface where is decomposes into solid tungsten and a volatile by-product. The CVD process 'grows' a crystalline tungsten film that fills the holes from all sides, producing a hole that is completely filled with metal, leaving only a very narrow seam down the middle of the contact hole. Usually a barrier / adhesion layer is put down first (not shown) to reduce electrical resistance to the underlying metal and protect it from the corrosive W CVD chemistry. A CMP process is then

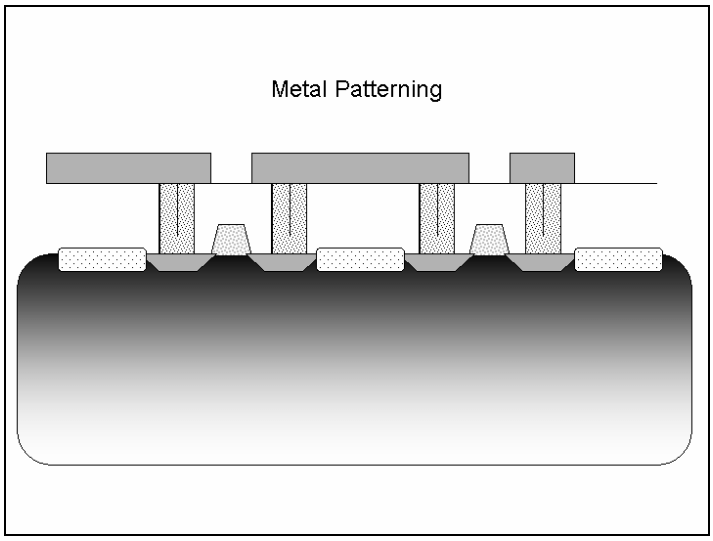
employed to remove the surface tungsten, leaving behind the filled contact holes. This polishing process is designed to be highly selective in removing the tungsten versus the underlying dielectric. This allows the process to use the dielectric as a stopping layer, improving the process latitude. Finally a metal layer is patterned on top of the filled contacts to complete the circuit. This process is repeated with the oxide planarization step to add each wiring level to an IC.

Instead of just being used to pattern vias for connecting two wiring levels, the damascene process can be used with trenches patterned in the dielectric to form the wiring themselves. In this process, a shallow trench is etched in the dielectric in the shape of the desired wire, the metal is deposited on the wafer, and the CMP process selectively removes the material to leave the trench filled. This process is one of the key technologies that has enabled the integration of copper into IC wiring levels. Prior to this, there was no way to easily pattern small copper features since copper cannot be plasma etched. The damascene process is also utilized in the Shallow Trench Isolation (STI) scheme to further permit tighter transistor packing.



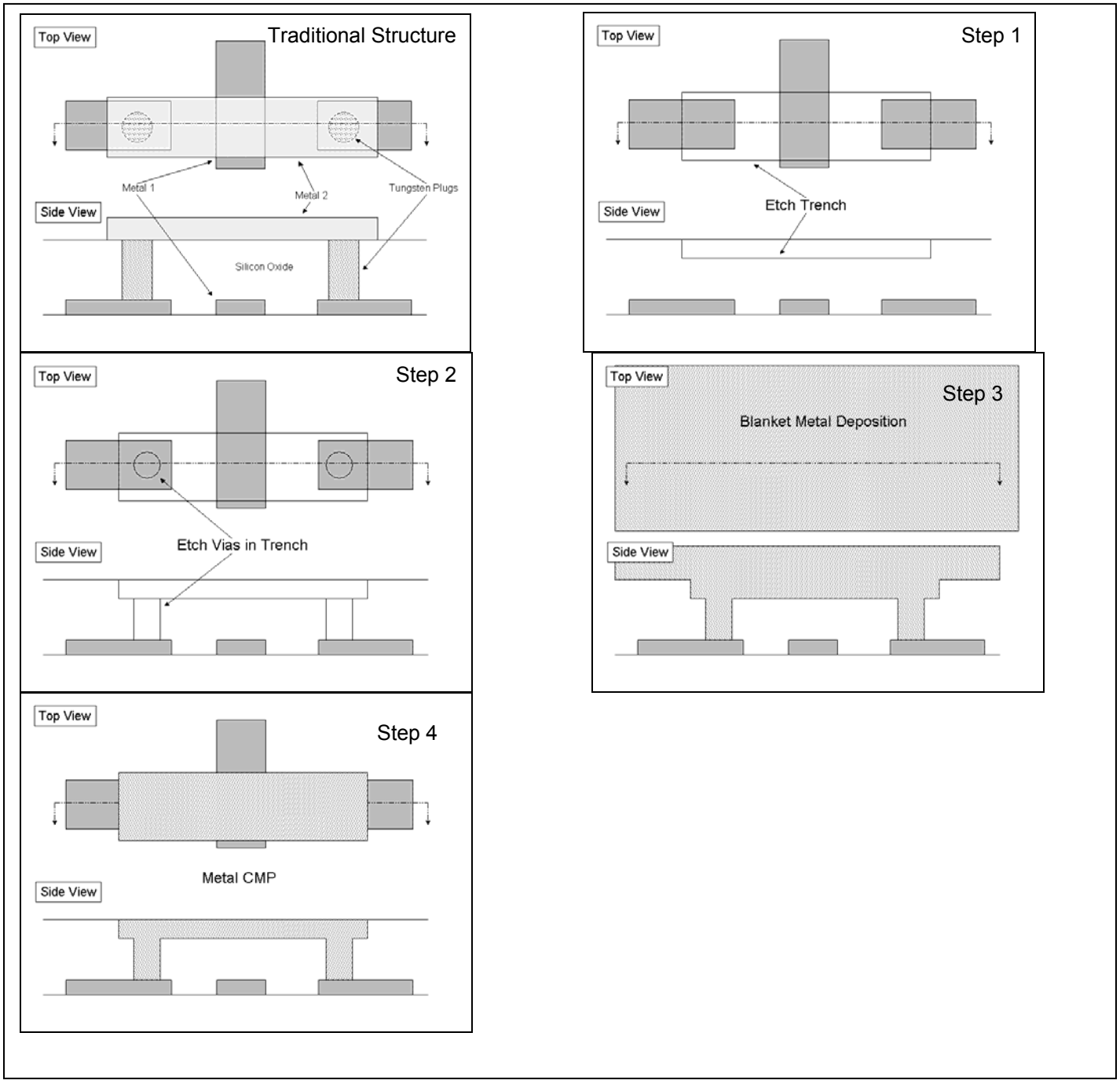






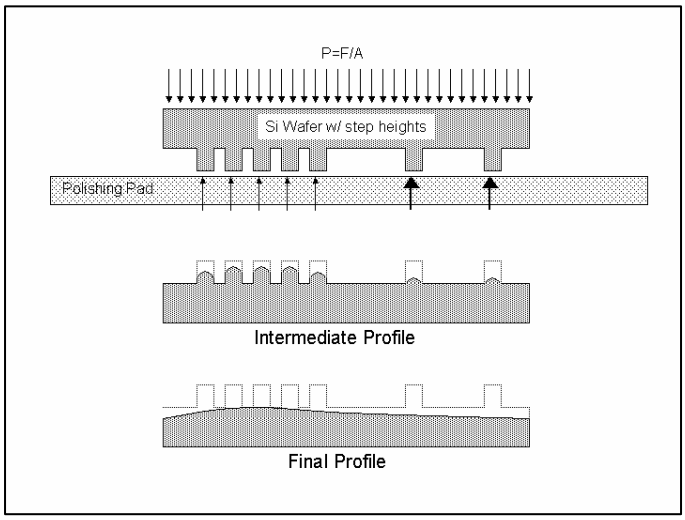
Dual Damascene CMP

In the dual damascene process, both the wiring level and the interlevel connections are created with a single polishing step. Two patterning steps are used to create features of two different depths. Blanket metal is deposited and a single CMP step is used to create the inlaid structure. This is the current process used by many IC companies to integrate copper into their circuits.



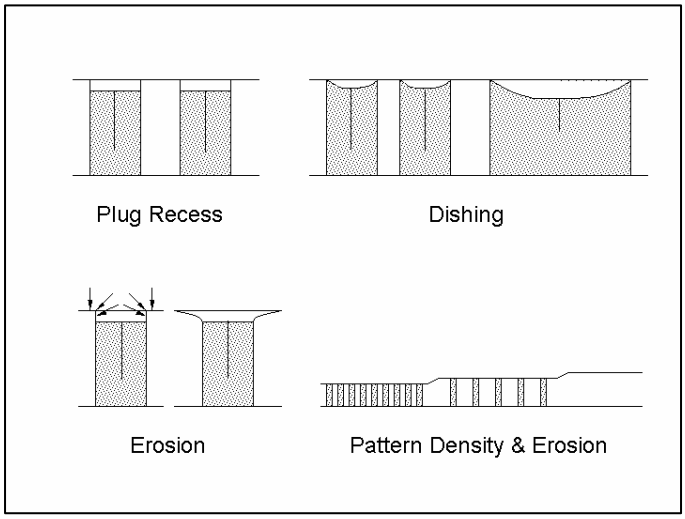
CMP Pattern Density Issues

CMP is seen by most of the semiconductor industry as critical for producing 0.35um devices and smaller, although it does suffer from some problems that need to be accounted for during the process integration. When polishing a wafer that has step features, only the top of the features touch the polishing pad, concentrating the pressure on these contact points. This increases the polishing rate above that of a blanket wafer. In addition, it causes nonuniformity in the removal rate across patterns of different densities due to variations in the pressure distribution across the pattern.



This pattern density effect on removal rate can cause problems if you have very dense and very sparse pattern in your design. Some ways to account for this is to create dummy shapes around the sparse pattern to try and match the higher pattern density.

The damascene process also suffers from several different issues. One issue is that the material being removed usually has a faster etch rate than the material the structure is inlaid in. This can cause two similar problems called plug recess and dishing. Plug recess is where the damascene structure sits slightly lower than the field area due to the faster removal rate than the field. Due to the mechanical limitation on how much the pad can deform, it is usually not excessive. In general, higher chemically active slurries will have a higher amount of recess due to the wet etching action of the slurry.



Dishing occurs when the polishing pad under the pressure of polishing, is able to deform into the damascene structure and polish it below the field area. The amount of dishing that occurs is related to the polishing pad characteristics, the size of the structure, and the polishing parameters (speed, pressure, temperature).

Erosion of the field material can also occur due to the enhanced polishing at the edges of the structures. As the damascene structure gets recessed, the corners of the field area are exposed and erode more rapidly than the unpatterned areas. This affects the local field material polishing rate and can lead to a variation in the field thickness based on the damascene pattern density. The polishing variables can be tuned to reduce this effect but usually at the expense of other parameters (removal rate, uniformity, etc.).

Glossary

Back pressure – During the polishing process, the wafer is held in place on the carrier head by the retaining ring without vacuum assistance. During polishing, air may be blown out the holes in the backing plate to affect the uniformity of the removal rate. This blowing of air is generally measured in pounds per square inch (psi). Nominal values for the back pressure are from 0 – 2 psi. If you set the back pressure too high it will blow the wafer out of the pocket.

Backing film – a cushioning polymer film attached to the backing plate with a pressure sensitive adhesive. It cushions the wafer during the polishing and compensates for slight flatness variations in the wafer or backing plate. The quality of the backing film is important to prevent uneven polishing.

Backing plate – Located in the carrier head, the backing plate is a precision flat stainless steel disk slightly larger than the wafer. It presses against the back of the wafer and transfers the polishing force to the wafer during CMP. Attached to the backing plate between the plate and the wafer is the backing film. There are many small holes in the backing plate to allow the tool to apply vacuum on the back of the wafer to transport it from the load station to the polishing pad. Vacuum is not generally applied during the polishing process.

Brushcleaner – a tool used to clean and dry wafers after CMP.

Carrier head – Tool fixture that holds the wafer against the polishing pad during the polishing operation. The carrier head is specific for the wafer you are polishing. The CNF has carrier heads for 3”, 4”, and 6” wafers.

CMP – chemical mechanical polishing, or chemical mechanical planarization. Polishing process utilizing both chemical etching and mechanical removal for nanofabrication.

Conditioning profile – varying the pad conditioning parameters over the polishing pad to affect the polishing uniformity. The Strasbaugh 6EC breaks the conditioning arm sweep into 10 zones in which you can set both the conditioning down force and dwell time for each zone. This allows you to alter the pad characteristics across the wafer track to make improvements in the removal rate uniformity.

Damascene process – CMP process in which a feature is etched in to an already planar film. The features are typically trenches or holes. A second material is then deposited on the wafer filling the feature. The CMP tool is then used to selectively remove the deposited film in the field area, leaving behind the filled feature flush with the planar film.

Diamond disk – a metal disk with embedded industrial diamond particles for conditioning the pad.

Dishing – the thinning of damascene structures below the field area due to pad deformation. Dishing is related to the structure size, pad hardness, and other polishing parameters. See also recess.

Down force – the pressure applied to the wafer during the polishing process. It is expressed in pounds per square inch (psi) with common pressures in CMP being 4 – 10 psi. Below 4 psi the wafer predominantly hydroplanes over the polishing pad and above 10 psi you risk breaking the wafer. The polishing pressure used has a large effect on uniformity and planarization.

Dual damascene – combining etched structures of different heights to fabricate both wiring and interconnects in one damascene polishing process

Erosion – thinning of the field area around damascene structures due to enhanced polishing at the feature edges. The more feature edges you have in a given area, the higher the erosion rate. Erosion issues almost always makes the actual selectivity of a given process lower than that measured on nonpatterned wafers.

Extension – the amount that the top surface of the wafer sits above the retaining ring in the carrier head. If the extension is too low, the etch rate will be reduced on the wafer edge and additionally, the retaining ring will wear down. If the extension is too high the wafer will slip out of the pocket and break during the polishing. Usually measured in mils (thousandths of an inch ≈ 25 microns). The opposite of wafer capture.

Fixed abrasive pads – pads with the abrasive component embedded into it instead of in the slurry. It is used with a chemical only liquid to perform CMP. Recent development in CMP that supposedly gives excellent selectivity.

Microscratching – micro scale scratching caused by debris on the polishing pad, agglomerated slurry particles, and pad defects. These are very hard to prevent from occurring but steps can be taken post polishing to minimize impact.

Pad conditioning – surface treatment of the polishing pad to improve removal rate stability. Hard polishing pads will glaze over from use during polishing and the removal rate will decrease over time. Rubbing the polishing pad with a diamond abrasive disk removes this top glazed surface and uncovers fresh pad material for polishing. Proper conditioning parameters can lead to very stable removal rates.

Planarization – the removal of surface topology in a nanofabricated strucuture. This was original purpose of the CMP process. Planarization can be either just

local removal of step heights, called local planarization, or it can also be uniform removal of material across a die, called global planarization. CMP is currently the only planarization process that gives global planarization.

Polishing pad – a polymer pad that the wafer is rubbed against during the CMP process. It is applied to the polishing table which rotates under the polishing arm. Slurry is dispensed on to the pad and the polishing arm pushes the carrier head against the pad to polish the wafer. Polishing pads are designed with a variety of properties and purposes.

Post-CMP cleaning – CMP slurries contain abrasive particles to perform the mechanical removal of the surface material. These abrasive particles must be removed from the wafer surface after polishing to prevent defects. After CMP, the wafers must be kept wet prior to cleaning because once the slurry is allowed to dry on the wafer, it is very hard to get off mechanically. Due to electrostatic attraction forces though, simply rinsing the wafers with water after polishing will remove little if any of those particles. Modern production equipment use wafer brushcleaners to clean and dry the wafers after CMP. These tools use PVA brushes to mechanical wipe the surface of the wafer and remove the abrasive particles. Additionally, they use dilute ammonium hydroxide to reduce the electrostatic attraction of the slurry particles to the wafer surface. It is important for any CMP process to determine how you are going to clean the wafers when you are done.

PVA – polyvinylalcohol, a soft spongy hydrophilic polymer material used in post-CMP cleaning to mechanically remove slurry particles.

Quill – or Spindle. The motor on the polishing arm that rotates the carrier against the pad. It turns in the same direction as the polishing table.

Removal Rate – the average rate at which material is removed from the surface of the wafer. Most often reported with a uniformity value as measured on a unpatterned wafer.

Retaining ring – a hard polymer ring on the carrier head that surround the wafer when it is mounted on the carrier head. The top surface of the ring sits above the wafer backing plate by a set amount to form a recessed circular area called the pocket. The depth of the pocket is important for successful polishing.

Selectivity – a ratio of removal rates between two different materials for a given CMP process. This is often determined by measuring the average removal rates on blanket film wafers. As an example, for the polysilicon damascene process, the vendor states that the slurry, when used in a specific CMP process, gives a 300:1 selectivity between polysilicon and thermal oxide. The value reported by this method is often much higher than what is achieved on a patterned wafer due to erosion effects.

Shallow Trench Isolation (STI) – Device isolation process for CMOS that utilizes CMP to allow for tighter device spacing. First, a silicon oxide and nitride layer is put down on the wafer to protect the surface from CMP damage. The wafer is etched to leave the device areas raised as 'mesas' with a lower trench area surrounding it. An oxide is deposited over the wafer thick enough to fill the trenches above the height of the mesas. An oxide CMP process is then utilized to remove the oxide over the mesas to expose the active device areas. The silicon nitride layer acts as a stopping layer for the CMP process and protects the silicon from mechanical damage during CMP.

Slurry – a mixture of abrasive and chemicals used to perform CMP. It is continuously pumped on the polishing pad during the CMP process. Most slurries remove material through a combination of chemical and mechanical methods so they are generally material specific. A given slurry may be used to polish materials other than the material it was designed for but unwanted results may occur.

Spindle – or Quill. The motor on the polishing arm that rotates the carrier against the pad. It turns in the same direction as the polishing table.

Surface roughness – a measurement of the surface irregularities. Most often expressed as a root mean square (RMS) value of the height variation in angstroms. A given polishing process will give you a set surface roughness on the small scale, but if measured over large areas it will measure higher due to the microscratching that occurs.

Touch up polish – a quick, low removal polishing step optimized for scratch removal and low surface roughness. Utilized after a primary polishing step to improve the surface finish and reduce defect levels. This type of polish was developed due to the high defect levels from the tungsten damascene process. After the tungsten CMP step, an oxide touch up or buff step would be performed to improve the silicon oxide surface qualities.

Uniformity – a measurement of how uniform the removal rate is across the wafer. It can be measured many ways but the most common is the standard deviation (SD or σ) of the measured removal rate expressed as a percentage of the removal rate. It is also sometimes called a measurement of non-uniformity.

Wafer capture – the amount of the wafer that is in the pocket usually expressed as a percentage of the wafer thickness. The opposite of wafer extension.