



**MICROTECH  
SYSTEMS, INC.**

## Phosphoric Acid Nitride Removal

### Introduction

During the semiconductor manufacturing flow silicon nitride ( $\text{Si}_3\text{N}_4$ ) is used for the definition of the active area.  $\text{Si}_3\text{N}_4$  often denominated as pad nitride is deposited on a  $\text{SiO}_2$  layer ("pad oxide") and partially removed by anisotropic plasma processes on those areas that are used for isolation of the individual devices. Typical layer thicknesses for the pad oxide are between 50-400 Å and for the pad nitride between 200-1600 Å, depending on the application and technology.

In order to obtain fully functional devices the active area must be completely free of silicon nitride, while the isolation, consisting of thermal oxides, LOCOS (local oxidation of silicon) or STI (shallow trench isolation), remains intact. Also it is important that the underlying silicon not be attacked. In particular for IC products with reduced device topology and TEOS based shallow trench isolation the budget for oxide removal is extremely tight.. Typical selectivity requirements are

>80:1 for  $\text{Si}_3\text{N}_4$  vs.  $\text{SiO}_2$

>500:1 for  $\text{Si}_3\text{N}_4$  vs. Si

This isotropic removal process is performed with hot phosphoric acid ( $\text{H}_3\text{PO}_4$ ) etching in an immersion bath. Dry (plasma) etches cannot be used, since achievable selectivities are much lower.

Although heated hydrofluoric acid can be used for nitride removal (which is often done for wafer or monitor reclaim) this is not appropriate for product applications due to the higher etch rates of silicon oxide and therefore not meeting the selectivity requirements.

A typical process flow for selective nitride etching consists of the following steps, as shown in Figure 1. An optional short HF dip (including subsequent rinse) is often included as the first processing step to etch away a thin native oxide layer, which may have formed due to partial re-oxidation in the thermal oxidation step to form the passivation oxide or the field oxide. This layer acts like an etch barrier which will take substantial time to be removed in  $\text{H}_3\text{PO}_4$  and therefore impact process time and uniformity.

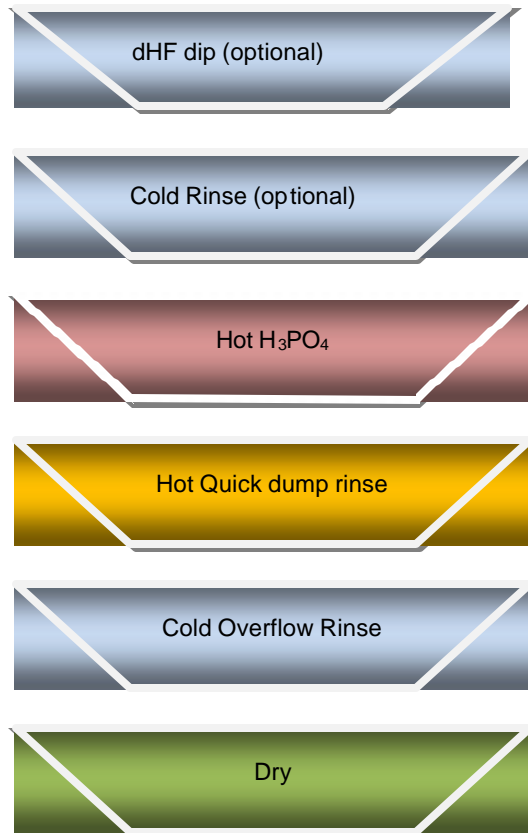
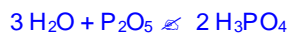


Figure 1. A typical process flow for selective nitride removal in hot phosphoric acid.

### Chemical Reactions

The most common phosphoric acid compound is ortho-phosphoric acid, which is formed when three molecules of water are added to one molecule of phosphorus pentoxide.:



Ortho-phosphoric acid is commercially available at high purity grade in concentrations of 80 or 85 %. The chemical reaction equations for silicon nitride and silicon oxide etching in these media are (1):





# MICROTECH SYSTEMS, INC.

Water hydrolyzes the silicon nitride to form hydrous silica and ammonia, which remains in solution in form of ammonium phosphate. The stoichiometry suggests that water is an integral part of the chemistry involved in the etching. During the etching of nitride hydrated silicon dioxide ( $H_2OSiO_2$ ) is formed, which impacts the silicon dioxide etching. In a fresh bath silicon oxide is removed by ortho-phosphoric acid, but as the  $SiO_2$  concentration continues to increase over time in the etching solution the etch rate of  $SiO_2$  from the wafer surface decreases. Thus the selectivity of the etching rates is dependent on bath utilization, which effects the concentration, as shown in Figure 2.

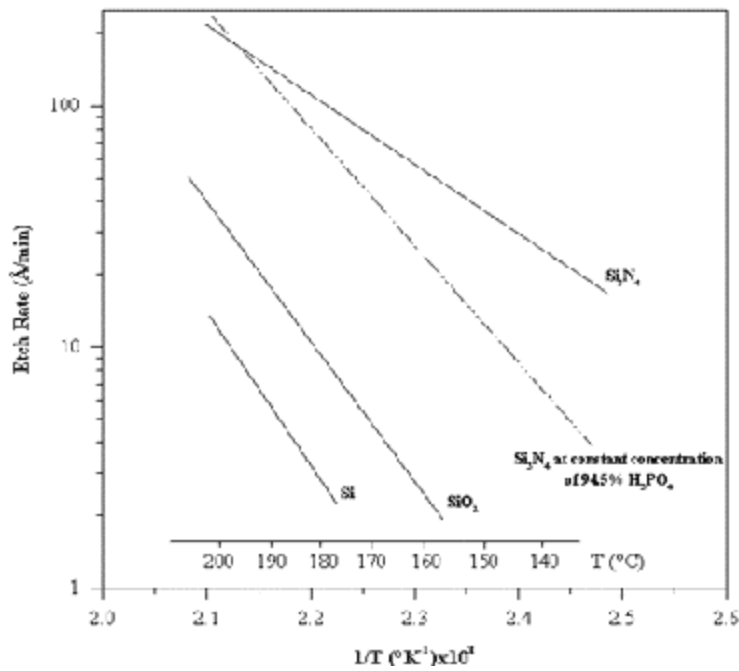


Figure 2. Etching rates of silicon nitride, silicon oxide, and silicon at different temperatures in phosphoric acid (2). The solid lines represent data of commercially available 85% ortho-phosphoric acid, while the dashed line indicated  $Si_3N_4$  etching with higher concentrations.

The etching rate of silicon nitride thus is depending on the temperature as well as the concentration typical etch rates can be seen in Figure 2 (2). The etching follows the Arrhenius equation for both the  $Si_3N_4$  as well the  $SiO_2$  etching. Also a small etch rate for  $Si$  is observed, which are caused by the non-complexed, overheated  $H_2O$  in the solution. It is desirable to perform the process at highest possible temperatures in order to reduce process times. For reasonable exposure times of 20-40 minutes, etching rates in order of 50  $\text{\AA}/\text{min}$  are required, which results in



# MICROTECH SYSTEMS, INC.

typical process temperatures of 160-165°C. Higher temperatures will result in faster removal, but also lower selectivity to SiO<sub>2</sub> and Si.

As can be seen in Figure 2, higher concentration phosphoric acid does not necessarily result in higher etch rates, confirming that H<sub>2</sub>O is intrinsically required for the process. At lower temperatures, higher water content provides better results, while at higher temperatures high concentration H<sub>3</sub>PO<sub>4</sub> appears to be more effective. Therefore, the process must be optimized and controlled for both the process temperature as well as the concentration.

This optimization cannot be done independently, since the boiling point of phosphoric acid varies with the concentration, as show in Figure 3. As delivered 85 % ortho-phosphoric acid will boil at 157°C, higher concentrations are required in order to increase process temperature above 160°C. This can be easily achieved, since H<sub>2</sub>O continuously evaporates from the heated solution. Vice versa the performance of the etch will decrease even at elevated process temperatures, if the concentration of the acid continuously increased during bath lifetime. Therefore continuous replenishment of H<sub>2</sub>O is required to maintain the etch rate (and selectivity) at a given process temperature.

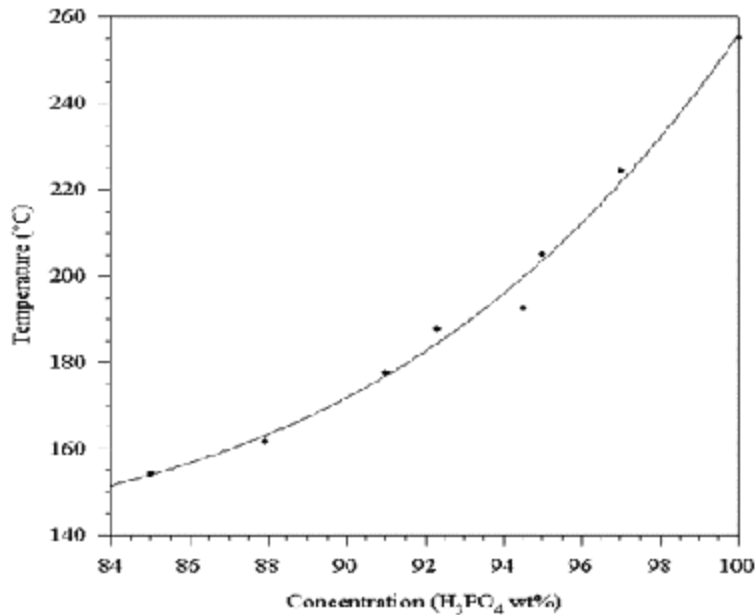


Figure 3. Boiling point temperatures of ortho-phosphoric acid vs. temperature (2).



## MICROTECH SYSTEMS, INC.

Optimally the process is run at highest possible temperatures and water concentrations resulting in the boiling point temperatures slightly above the set temperature. In this regime the bath will remain relatively quiet, allowing liquid flow control with only a small amount of bubbles and easiest exhaust handling.

### Temperature and Concentration Control

Basically there are two methods to control the nitride etch process in an immersion bath, both of which include continuous heating and water replenishment:

1. Temperature control with constant power heating and boiling point adjustment by H<sub>2</sub>O spiking.

This method, which is the most common way of control, uses powerful heaters capable of heating up the medium to the boiling point and above. The heating power is not controlled during the process (*i.e.* set at a fixed, mostly maximum set point). By varying the amount of H<sub>2</sub>O replenishment the liquid concentration is setup to boil precisely at the determined process temperature. Thus the temperature is limit controlled by the boiling point of the actual mixture.

This method provides acceptable performance results for standard applications. It is the most cost effective solution from investment point of view and can be used in automated and also manual (*i.e.* simple process control) wet benches. Disadvantages are the processing of wafers in a bubbling liquid which impact within wafer etch non-uniformity and often causes the wafers to lift out of the cassette or holder, and produces an excess amount of process vapor which needs to be prevented from re-condensation in the wet bench and efficiently ventilated. Also the entire process is controlled by a single parameter, which does not allow fine tuning of varying etch performance over the loading effect and the time.

2. Independent power and concentration control with variable power heating.

By using a heating system with variable power, controlled by a temperature sensor in the bath solution can be heated to the desired processor while the H<sub>2</sub>O replenishment is set to achieve a constant concentration of the mixture to have a boiling point slightly above the process temperature. The bath itself will remain calm and substantially free from bubble. By PID control the heating is able to compensate for process variations, *i.e.* temperature drops caused by cold wafers entering the bath and result in more reproducible etching over the entire process time. Also Si loading effects over the bath lifetime can be compensated by time or lot depending on the water spiking process. Finally vertical temperature differences from the tank bottom (where typically the heaters are located to the top (the interface to the ambient temperature air) can be minimized by flow optimization, which is only possible in a calm liquid. Disadvantages are obviously the higher efforts and component cost to support the control mechanism.

Equipment for advanced device manufacturing typically uses the latter method in order to meet the product requirements, less demanding products can be build with good results also with the first method.



### Bath Loading and Seasoning Effects

As mentioned above both the etch rate of  $\text{Si}_3\text{N}_4$  and  $\text{SiO}_2$  are depending of the amount of hydrated Si and silicon dioxide in the bath. For both species the etch rate will drop with the number of wafer processed, however at a different rate, as shown in Figure 4. Often in a fresh bath the oxide to nitride selectivity is found to be less than 1:20 while after processing several batches of wafers the required 1:80 ratio is achieved. This is caused by the formation of hydrated silicon byproducts (*i.e.*  $\text{H}_2\text{O SiO}_2$ ), which inhibits the  $\text{SiO}_2$  etching. In order to minimize the effect on the products wafers and maximize the process reproducibility main IC manufacturers run a couple of blank nitride monitors through the bath after bath change, before the process is re-qualified for production.

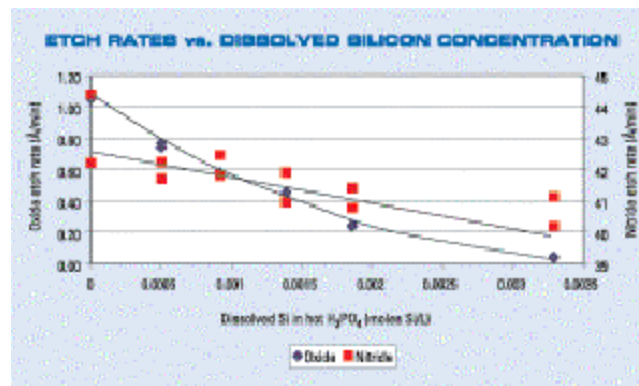


Figure 4.  $\text{Si}_3\text{N}_4$  and  $\text{SiO}_2$  etch rates (@ 160°C) as a function of Si Loading in the bath (3).

The slower rate in  $\text{Si}_3\text{N}_4$  etching is caused by the fact the any system in equilibrium will act to maintain the equilibrium (*i.e.* Le Chatelier principle). Thus the reaction strength also for the phosphoric acid with  $\text{Si}_3\text{N}_4$  will decrease if the concentration of the reaction product of hydrous silica increases. This effect can be partly compensated by incrementally increase of the water spiking up to the point when the process temperature equals the boiling point temperature, but finally the drop in etch rate requires a fresh bath. This typically determines the bath life time in production. Depending of the method of temperature and concentration control this may vary between 24 hours for boiling point controlled baths up to 72 hours with sophisticated spiking control.

Further extension of bath lifetime (and thus reduction of down time) can be achieved with partial or incremental replacement of the bath only. By adding fresh phosphoric acid by time or load the seasoning of the bath can be remained (thus the selectivity drop in a fresh bath is avoided) and the down time is minimized.

### Typical Process and Performance Parameters



## MICROTECH SYSTEMS, INC.

The hot phosphoric bath for selective nitride etching has been pretty much standardized in semiconductor manufacturing. The most important process parameters are as follows:

Initial concentration (as delivered ortho-phosphoric acid):	85 wt%
Typical concentration during production (@165°C):	87- 91 wt %
Process temperature:	160 -165 °C
Temperature tolerance (single process):	+/- 5 °C
Temperature tolerance (ideal bath)	+/- 2 °C
Typical H <sub>2</sub> O replenishment rates (200 mm)	60-90 ml/min
Bath recirculation (200 mm):	> 6 gpm
Bath lifetime	24 -72 hrs
Process time (30 nm Si <sub>3</sub> N <sub>4</sub> plus 30 % over etch)	10 – 15 min

In a seasoned bath the process in a well controlled equipment can meet the following performance:

Etch rate Si <sub>3</sub> N <sub>4</sub> at 160°C	40 – 44 Angstr/min
Etch rate Si <sub>3</sub> N <sub>4</sub> at 165°C	50 – 55 Angstr/min
Within wafer non uniformity:	5 % max /min
Wafer to wafer non-uniformity:	3 % max/min
Lot to lot non uniformity	5 % max /min
Selectivity to Oxide	>80:1
Silicon etching	< 0,1 Angstr/min
Contamination (including rinse /dry)	< 30 adder (> 0.16 µm)

### Post Etch Rinsing

As for all chemicals steps in wet immersion processing the wafers must be rinsed after treatment. Rinsing after hot phosphoric etching is one of the most critical rinsing process in the production flow for various reasons.

First, even with hot UPW (ultra pure water) in the rinse there is a temperature drop of approximately 100°C between the hot phosphoric and the rinse bath. The sudden change in temperature presents a shock to the wafer potentially causing enhanced wafer breakage (in particular if the wafers have seen significant stress before). Next, both the hydrates silica as well as the acid itself is difficult to remove from the surface in pure water. In sufficient rinsing will result in particulate contamination and phosphate blooming on the wafer after drying.



## MICROTECH SYSTEMS, INC.

Most efficient rinsing of phosphoric acid is achieved by a multi-step quick dump process, starting with hot ultra pure water (to minimized the temperature shock) and ending with one or tow cold sequences. The wafers are transferred into a full bath of heated UPW (> 65°C), which is dumped after 15-30 sec. After drain the tank is refilled from the bottom with or without top spray rinse followed by a 60-120 sec overflow. This sequence is repeated 3 to 4 times with hot water being used in the first step and cold water in the last ones. Rinsing performance can greatly be enhanced by the use of megasonics during the overflow steps of the rinse process.

### Hardware Considerations

Stable process results and high productivity are strongly dependent on a number of critical hardware related features that will be briefly summarized here:

#### Process Tank

Hot phosphoric etching is performed in a quartz tank with chemicals supplied from the bottom and returned to a recirculation through overflow. Chemical flow through the tank shall be uniform even with a batch of wafer present to ensure the same continuous exchange of process media on the entire wafer surface. High exchange flows are preferred to minimize the vertical temperature gradient as well compensate for cooling of the medium by introducing cold wafers into the bath.

The quartz tank shall not be in direct contact with heating devices and temperature at the quartz surface must be limited to a maximum of 200°C. In particular above 180°C phosphoric acid tends to etch tiny channels into the quartz material causing leakage. The heater power shall be evenly distributed over the tank to avoid local temperature difference in the medium. The overflow shall be uniform and not create any turbulence, inparticular where the medium passes the substrates.

#### Recirculation

The recirculation shall include filtration, which capable to handle high viscosity liquid. In particular when not preheated phosphoric acid is supplied during bath change the recirculation shall not be limited through the filter resistance. Water replenishment shall be done at the entrance into the recirculation loop to allow the cold medium to heat up and mix well with the acid before it re-enters the tank. The medium shall be re-introduced to the process tank through a flow manifold ensuring an even distribution of the medium across the area of the process tank (temperature and flow uniformity).

#### Bath Change

Fresh phosphoric acid is supplied at 80 or 85 % concentration from facilities, which has a boiling point temperature below the process temperature. Excess amount of water must be evaporated during the heat up cycle resulting in long heat up times. Often a preheat tank is used to prepare the medium before the actual bath exchange starts by heating up the chemistry to 80-120°C and evaporate excess water to allow a quick heating cycle in the tank



## MICROTECH SYSTEMS, INC.

to the actual process temperature. This preheat tank also can be used for partial replenishment of the bath without interrupting the production.

### Exhaust Ventilation

Significant amounts of fumes evaporate from the bath when heated. Sometimes cooling units are placed on top of the bath where these fumes are condensing. Resulting liquid which drops back into the tank and will create a significant temperature gradient impacting the top to bottom non uniformity of the etch. Therefore condensed liquid shall be fed into the recirculation line for re-heating.

The evaporation process causes a cooling of the liquid. Therefore the ventilation shall be optimized to efficiently take away the vapor to protect other equipment modules and the cleanroom environment, but shall be low enough to allow a quasi equilibrium state between liquid and gas phase. Minimized evaporation results in smallest possible temperature gradient between bottom and top of the acid in the process tank.

### Wet Cleaning Tool Configuration

The configuration for hot phosphoric nitride removal baths, with automatic wet bench configuration is shown in Figure 5. Various alternatives to this configuration are available depending on the process goals. Spiking and temperature control are also available.

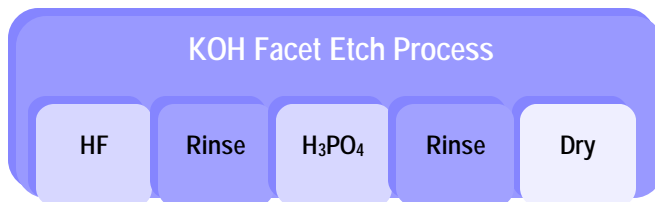


Figure 5. Example of a hot phosphoric nitride etching configuration for a wet bench.

### References

1. W. van Gelder, and V. E. Hauser, "The Etching of Silicon Nitride in Phosphoric Acid with Silicon Dioxide as a Mask"; *J. Electrochem. Soc.*, 114(8) p.869 (1967).
2. S. Clark, "Chemical etching of Silicon Nitride with hot phosphoric acid", Bold Technologies Technical Paper, [www.boldtech1.com](http://www.boldtech1.com).
3. S. J. Buffat, M. S. Lucey, M. R. Yalamanchili, and C. S. Hatcher, "Hot Phosphoric Acid APC for Silicon Nitride Etch", *Semiconductor International*, 9 p. (08/01/2002).