

# APPLICATION NOTE

## In Situ Process Monitoring of RTP Cluster Tool with FabGuard® and Transpector® CPM

### INTRODUCTION

RTP (Rapid Thermal Processing) is a key step in the salicidation process. The metal silicide layer formed by salicidation creates a high-conductivity interface between the gate polysilicon and metal contact that is essential to modern high performance CMOS devices. Unfortunately, in many cases, a detrimental oxide layer forms during RTP that remains undetected until final device test. At best, this testing reveals slower speed bin, lower value parts, and at worst yield loss. For the very common salicidation of Cobalt with a RTP Cluster Tool, the problem can be solved by the combination of an INFICON Transpector CPM RGA and INFICON FabGuard Sensor Integration and Analysis System with a return on investment time on the order of a month. Once the monitoring system is in place, additional throughput and yield-enhancing, real-time analyses can be implemented. The following benefits result:

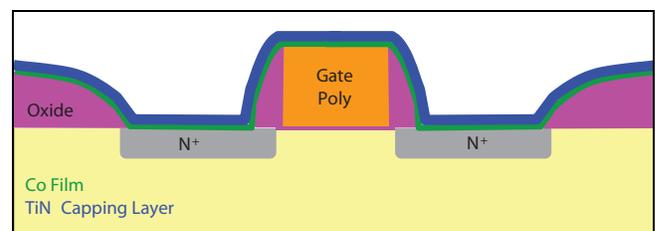
- ◆ Ability to perform rapid queue time optimization study
- ◆ Real time detection of oxide formation during RTP step
- ◆ Significant Yield Improvement (ROI time on the order of 1 month has been demonstrated)
- ◆ Demonstrated 14% increased tool throughput due to load lock optimization

### PROCESS OVERVIEW

The relevant layers of a modern gate stack are illustrated in Figure 1. A metal film, Cobalt (Co) in this application, covers the gate polysilicon. The film is capped with either a Ti or TiN capping layer. Rapid Thermal Processing (RTP) is used to form a metal silicide layer. The capping layer is subsequently removed, and the device moves on to further

metalization. Metal Silicides play a critical transition role in the gate stack between polysilicon and metal contact. Cobalt Silicide is the material of choice in devices at the 130 nm to 65 nm technology nodes. Although the sub 65 nm technology node has motivated a transition to Nickel Silicide and lower temperature RTP tools, the challenges addressed in this application are anticipated to be present in new generation processes. The need for easy removal of the capping layer, combined with the throughput mismatch of the RTP and capping layer CVD tools results in a sub-optimal process flow.

**Figure 1 Film Stack Prior to RTP Salicidation**



A capping layer is used to protect the Co film from oxygen contamination. In this regard, pure Ti has better properties than TiN. However, the Ti film is much harder to strip, and can often result in damage to the underlying Co film. This problem does not appear to be solved, and most fabs opt for TiN. The ease of removal of the TiN film comes at the cost of higher oxygen permeability, which becomes problematic when combined with the throughput mismatch of the CVD and RTP chambers. With enough time at atmospheric pressure, oxygen will diffuse to the TiN-Co interface layer, and oxidize during RTP. The resulting highly resistive oxide layer reduces the overall conductivity of the gate stack, defeating the purpose of the salicidation step. If TiN CVD to RTP salicidation queue times is long enough, significant end of line device performance impact results, to the point of limiting yield. Performing the two steps in the same cluster can eliminate the

problem, but this approach is highly inefficient. Throughput of the RTP chamber is on the order of 10x higher than that of TiN CVD, but the chamber ratio in a cluster tool is at most 3:1. To achieve an optimal ratio of chambers, separate tools must be used, and queue times must be managed.

The nature of the RTP process makes identification of excess queue times impossible prior to end of line testing. Since oxygen loading is a property of the wafer and not the process tool, the qualification run gives no information about this specific problem. In addition, no information from this process tool, or from any of the data normally gathered in prior or subsequent steps indicates that oxidation has occurred.

## APPLICATION DESCRIPTION

The solution to this problem is the combination of an INFICON Transpector CPM RGA and INFICON FabGuard Sensor Integration and Analysis System used for both process development and real-time process monitoring. Although the RGA does not directly measure oxidation, it does directly measure the concentration of oxygen absorbed by TiN cap layer of the wafer being processed. The amount of oxygen is strongly correlated to queue time and directly correlated to oxide formation during RTP.

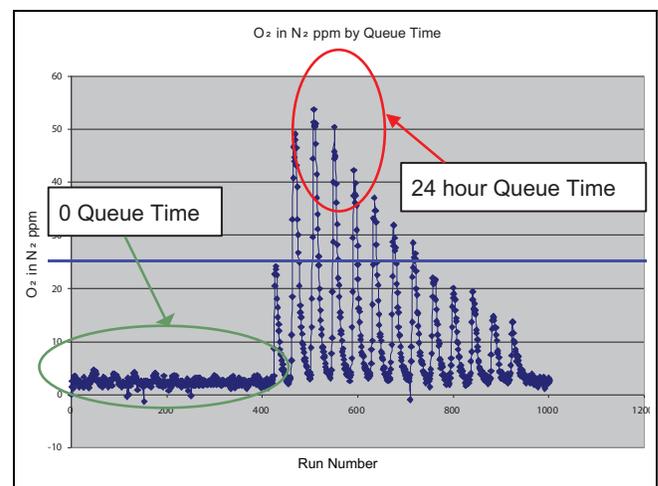
A simple process optimization study can be performed to identify maximum acceptable queue time as well as in-process O<sub>2</sub> levels. As shown in Figure 2, the longer the queue time, the higher the O<sub>2</sub> released during processing, and the worse the end of line performance. The more rigorous approach to the study is to track end of line performance as a function of total O<sub>2</sub> during RTP. In this way, precise sensitivity to TiN O<sub>2</sub> absorption as a function of device can be determined. A simpler and far less time and resource consuming approach is to use a total O<sub>2</sub> of roughly 3 or 4 times the short-queue level as a maximum acceptable level, and select the queue time based on this O<sub>2</sub> level.

Controlling queue time is not sufficient to eliminate the problem. TiN O<sub>2</sub> absorption is not completely correlated to queue time. Some percentage of wafers with acceptable queue times will still have high O<sub>2</sub> loading, impacting device performance and yield. A key advantage of controlling a sensor with FabGuard is the ease with which real-time monitoring can be

implemented. It is very straightforward to trigger the RGA to measure total O<sub>2</sub> released during the process step and associate that amount with the specific wafer being processed. Furthermore, FabGuard allows a number of fault and alarm configurations, ranging from notification to direct tool interdiction. Real-time monitoring at a minimum allows the wafers with excess oxygen to be removed from the process flow so that subsequent steps are not wasted on them. A more likely scenario is that the entire lot has excess O<sub>2</sub> absorption, and that the problem can be caught on first-wafer. The entire lot can then be diverted and possibly de-gassed. Existing implementations of this solution have resulted in return on investment times on the order of a month through yield loss prevention.

The atmospheric pressure RTP process requires a differentially pumped RGA such as the Transpector CPM. The RTP Cluster Tool application offers the further advantage that the optimal RGA placement for process chamber monitoring also gives visibility into the loadlock chamber. This can be exploited for additional throughput improving applications.

**Figure 2 Queue Time Impact on O<sub>2</sub> Concentration in RTP**



## ALTERNATIVES

Without in-situ RGA monitoring, the RTP salicidation step will be sub-optimal in a number of ways. Other sensors can be tried, but the sensitivity requirements are daunting. Both FTIR and Oxygen sensors can detect Oxygen, but the extremely small O<sub>2</sub> concentrations, on the order of 20 ppm, that result in end of line failures are below the capability of both. With no Oxygen measurement, no alternative to a

time-consuming end of line measurement study is available to determine acceptable queue times. Without in-situ monitoring, only queue time can be used as an indicator of absorbed  $O_2$ . This leads to excessively short queue times, impacting overall throughput, and some oxidation-induced end of line failures, since queue time is not a perfect predictor of oxide growth.

## IMPACTS

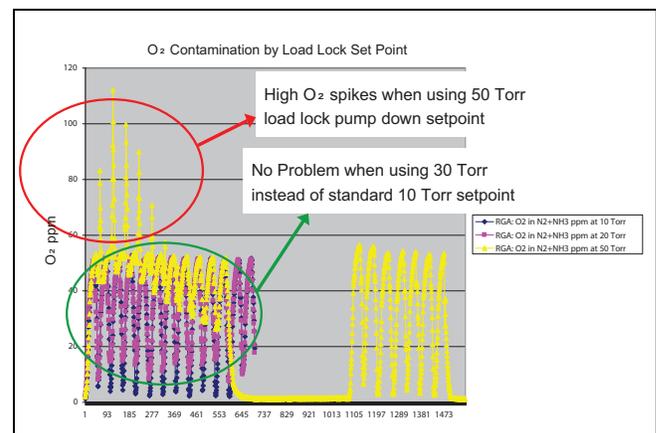
There are some capital and recurring costs involved in the solution, but ROI has been demonstrated for specific applications. In addition, INFICON offers worldwide application support to address the setup and maintenance of the new systems.

A positive consequence of the sensor installation is that several additional throughput and yield enhancing applications can easily be implemented. These applications improve ROI by adding further FDC capabilities without additional costs. Like many other process tools, the RTP cluster tool goes through a pump down and purge step prior to processing. This contamination elimination step is currently performed open loop, making it sub-optimal for both throughput and yield. From a throughput standpoint, our data has shown that default recipe settings waste significant time targeting too low a base pressure and holding for too long a time. As seen in Figure 3, the 10 Torr base pressure of the standard recipe can be raised to 30 Torr with no ill effect. All but the most contaminated wafers reach steady state levels long before the end of

the conventional step. Optimizing the load lock recipe resulted in a 14% improvement in tool throughput. Furthermore, when incoming wafers are significantly contaminated, the default recipe is insufficient to remove the contamination, and a full PM may be required to recover from the damage. The combination of wide mass sensitivity of the RGA and control power of FabGuard makes it straightforward to automatically scan for all likely contaminants at the time of pump down in order to both optimize the recipe for throughput and generate an alarm if a contaminated wafer enters the tool.

Custom applications derived from this technology are available from INFICON and adjustment from the original design may be required depending on specifications and requirements for a particular customer application.

**Figure 3 Contamination vs Load Lock Base Pressure**



[www.inficon.com](http://www.inficon.com) [reachus@inficon.com](mailto:reachus@inficon.com)

Due to our continuing program of product improvements, specifications are subject to change without notice. All trademarks are the property of their respective owners.