3)

Evaluation of Continuous Batch Processing for Vertical Diffusion Furnaces

by

Douglas S. Fong

B.S., Industrial Engineering and Operations Research
University of California, Berkeley, 1992

Submitted to the Department of Materials Science and Engineering
and the MIT Sloan School of Management
in Partial Fulfillment of the requirements for the degrees of

Master of Science in Materials Science and Engineering
and
Master of Science in Management

at the
Massachusetts Institute of Technology
June 1998

Signature of Author _____
Department of Materials Science and Engineering
MIT Sloan School of Management
May, 8 1998

Certified by ____ _____
Professor Klavs Jensen, Thesis Advisor, Lammot du Pont Professor of Chemical Engineering and Professor of Materials Science and Engineering
Department of Materials Science and Engineering

Certified by _____ _____
Professor Roy Welsch, Thesis Advisor, Professor of Statistics and Management Science,
Director, Center for Computational Research in Economics and Management Science
MIT Sloan School of Management

Accepted by_____ _____
Larry Abeln
Director of the Master's Program
MIT Sloan School of Management

Accepted by_____
Linn W. Hobbs, John F. Elliott Professor of Materials
Chair, DMSE Committee on Graduate Students
Department of Materials Science and Engineering

Evaluation of Continuous Batch Processing for Vertical Diffusion Furnaces

Douglas S. Fong

Submitted to the Department of Materials Science and Engineering
and the MIT Sloan School of Management on May
in Partial Fulfillment of the requirements for the degrees of
Master of Science in Materials Science and Engineering and
Master of Science in Management

Abstract

While Intel's Copy Exactly! strategy has largely been described in the context of how technology is transferred, little is written to describe how the process is maintained. This thesis chronicles the path of a change proposal through the Copy Exactly! change management process.

The proposal was to change a work rule in the diffusion toolsets to improve operational efficiency. As the latest manufacturing process was being developed, lots were loaded into and out of the vertical diffusion furnaces in a serial manner. However, in order to achieve the run rate targets for the tool sets, a functionality of the furnace called continuous batch processing (CBP) needed to be used. This allows a batch to be loaded into the buffer rack of the furnace while a batch is being processed. Once processing on the first batch has been completed, then the tool automatically loads the second batch into the quartz boat. When using CBP, a key decision is whether or not to immediately process the next batch or wait until the measurements of the previous batch have been completed.

While analyzing the trade-offs between the two options, the efficiency gains were quantified by gathering data from the manufacturing information system rather than conducting traditional time/motion studies in the fab. This allowed for an analysis of data from hundreds of batches from three different facilities. The methodology used is described in detail.

Continuously processing batches also adds more risk if the next batch starts before the measurements from the previous batch are not analyzed first. Once the furnace starts processing, the batch is committed, so instead of losing one lot, two will be lost. This risk is analyzed from both a statistical and an engineering perspective. Finally, the benefits are weighed against the costs.

This document starts off with some general background material on semiconductor processing, and then describes the functions that are performed by the vertical diffusion furnaces. The analysis of the rewards and risks of continuous batch processing follow. Finally, the White Paper Process at Intel is described, and then the history of the white paper titled "Continuous Batch Processing for the Vertical Diffusion Furnaces in P856" is documented.

Thesis Advisors:
Klavs Jensen, Professor of Chemical Engineering and Professor of Materials Science and Engineering
Roy Welsch, Professor of Statistics and Management Science

Acknowledgments

I would like to thank Intel Corporation for sponsoring this work along with several personnel that were instrumental in assisting me with this project. The diffusion process engineering team was one group of busy people who were readily available when I needed them. For that, I want to thank Ryan Lee, Hien Nguyen, Steve Soss, and Jun-Fei Zheng. I also want to thank a couple of people from lithography that helped with another project while I was at Intel, Barbara Wang and Mark Wantanabe. My LFM cube mates, Chris Cowger and Kevin Farrelly also made this experience memorable. Thank you for all of your collaborative efforts. I would also like to thank my supervisor, Dennis Arnow, and the project sponsor, Don Myers.

I would also like to acknowledge the Leaders for Manufacturing program for providing this opportunity. The program and the people affiliated with it have far exceeded my expectations. My thesis advisors, Klavs Jensen and Roy Welsch, need to be thanked for their guidance and flexibility during this internship.

Table of Contents

List of Figures

List of Tables

This page intentionally left blank

1. Continuous Batch Processing

The Kokusai vertical diffusion furnaces used at Intel provide a functionality called continuous batch processing (CBP). CBP is a function of the vertical diffusion furnaces that utilizes a buffer within the tool, enabling it to automatically unload and load batches of wafers into the furnace. This functionality increases the operational efficiency of the tool because batches can be continuously loaded into and out of the furnace without technician supervision. CBP can be run in a gating or non-gating manner, and while non-gating provides higher throughput, it also adds more risk. Because the next batch may start processing before the metrology has been completed on the previous batch, there is the potential to lose an additional batch if the tool starts producing out of spec.

In order for the vertical diffusion furnaces to achieve the wafer run rate objectives set forth in the virtual factory[i] planning models for the latest manufacturing process, continuous batch processing needed to be used. In fact, these models projected capacity based on the use of non-gating CBP. Because CBP was not being used in the latest manufacturing process (P856), it needed to be assessed and qualified so it could be implemented across the virtual factory.

CBP enables maximum utilization of the tool because it eliminates the time that the tool sits idle while: 1) the batch is loaded into the tool, 2) a recipe has been completed and the tool waits to be unloaded, and 3) the test measurements are being taken on the completed batch. The first period is eliminated because wafers are loaded into the buffer while the furnace is processing another batch. The second and third

---

[i] Intel runs the identical process at multiple facilities to produce the same product. This group of fabs is called a virtual factory.

periods are eliminated because once a recipe is completed, the batch in the buffer is automatically loaded into the tool. A diagram of the tasks performed during processing is shown in Figure 1.

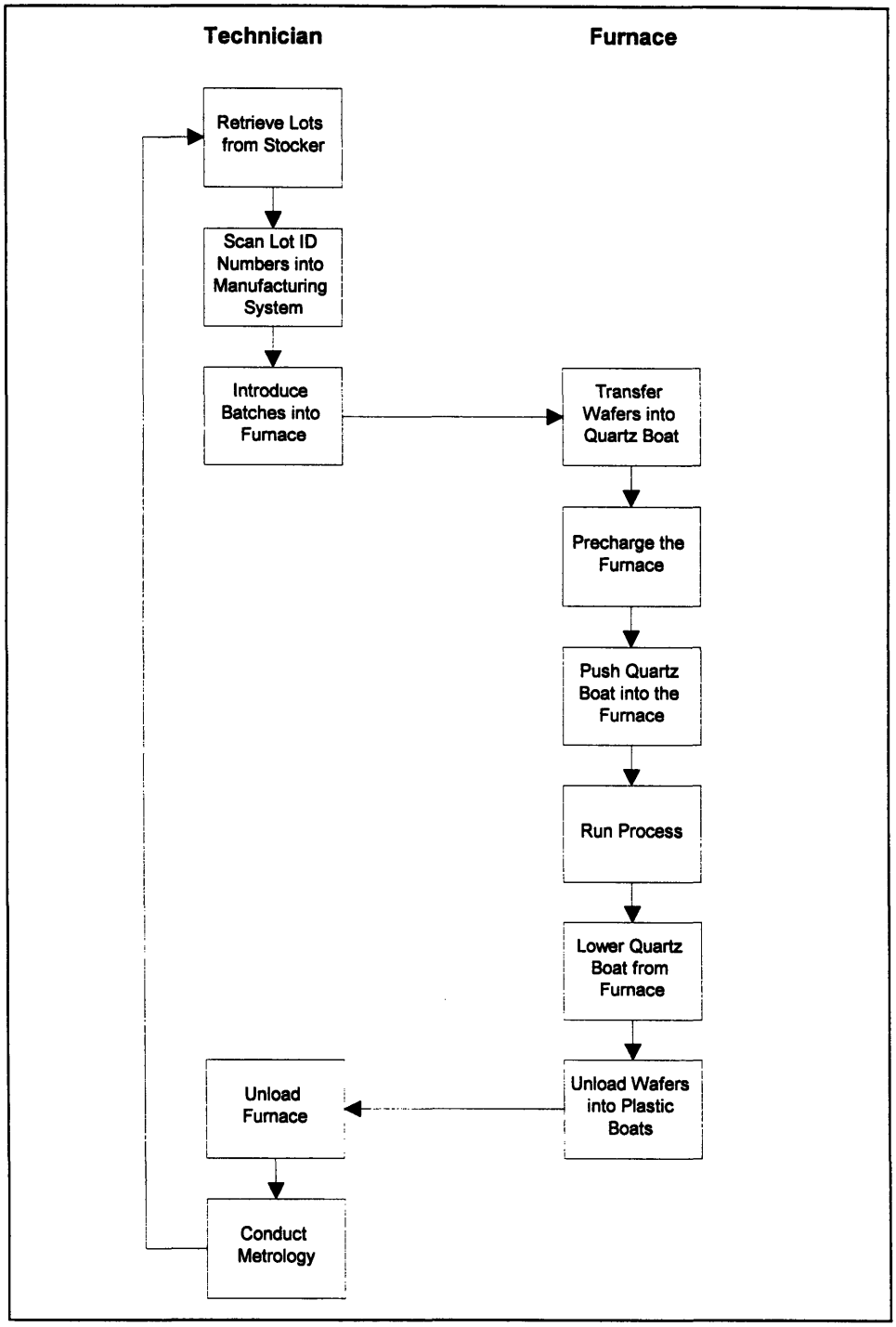

Figure 1. Tasks Performed when Processing Wafers through the VDF

1.1 Gating or Non-Gating?

There are two strategies by which the CBP functionality may be used: gating and non-gating. When run in a gating manner, the next batch will not be pushed into the furnace until the measurements on the previous batch have been completed and are determined to be in control. This method is the safest way to utilize the CBP functionality because it does not increase the risk of the process. If a batch is out-of-control, then the next batch will not be processed.

When the process is run in a non-gating manner, batches are processed regardless of whether the metrology has been completed or not. Once a recipe is complete, the tool will unload the boat, load the next batch into the boat, and start the furnace. This strategy provides the most flexibility and also ensures that the run rate targets will be met. The furnace recipes run for several hours, so two batches could be loaded at the start of the day, and then the furnaces can be run unattended. This method allows for the more flexible use of manufacturing technicians.

CBP was not being used for the latest manufacturing process being developed and needed to be qualified for this process. Before doing so, the question of whether to run it in a gating or non-gating fashion had to be decided. In the two previous process generations, CBP was used in differing ways. Two process generations ago, when the process was developed at this facility, non-gating CBP was used. However, the last generation's process used gating CBP.

1.2 Methodology

The key issue in deciding between gating and non-gating CBP is whether or not the additional operational efficiency justifies the increased risk. Operational efficiency is improved because non-gating CBP ensures that the run rate can be met while the tool is up for production. The run rate is not dependent on technician efficiency that is adversely affected if one is not available when a recipe is completed, nor is it

affected by a metrology tool not being available when needed. On the other hand, non-gating CBP will increase risk because if a batch is out of disposition (OOD), which means that it does not meet specifications, then most likely the next batch will also be OOD. Instead of one batch being scrapped, now two batches will be scrapped when the tool drifts out of specification. While the probability of the tool drifting does not increase, the consequences when this does happen doubles.

In order to determine whether or not a gating or non-gating process should be used, a risk/reward comparison was performed. The rewards were quantified by calculating the amount of time that could be saved for each batch by eliminating the three periods of idle time. This was done by gathering data from the manufacturing information system from three different facilities. By mapping the activities to existing data, information from hundreds of batches were readily available to be analyzed.

The risk was determined by analyzing the process control history of the different diffusion steps at the development fab for the first half of 1997. The total number of out-of-control and out-of-disposition occurrences was tabulated to assess the probability that a batch would be OOD. This is supplemented by an engineering assessment of the sensitivity of the process in terms of uniformity and thickness.

In the following chapter, background material on semiconductor processing is provided. Chapter 3 illustrates the functionality of the vertical diffusion furnaces and provides process models for oxidation and chemical vapor deposition. The methodology, rewards, and risks are described in Chapter 4. Finally, the journey of the white paper titled "Continuous Batch Processing for the Vertical Diffusion Furnaces in P856" through Intel's Copy Exactly change management process is described in Chapter 5.

2. Background

This chapter provides a broad overview of semiconductor processing which introduces several concepts that will be built upon in later chapters. The basic process steps of semiconductor processing are described, and then they are placed in the context of building a device. The role of monitors in the process is then explained.

Semiconductor processing is also called microfabrication because it is used to create microscopic patterns on a substrate of a semiconducting material, the most popular of which is silicon. The process creates layers of material on top of the silicon substrate. Each of the layers is produced by a series of basic steps, called loops. The basic flow in a loop is deposition of a layer, transferal of a pattern, and then etching away of the extra material. Millions of transistors are formed, and then the wiring that connects them is laid on top of these devices. The formation of the underlying devices is called the front end of processing. The creation of the wiring that connects the devices is called the back end.

In field effect transistors (FETs), each transistor consists of a source, drain, and gate. The source and drain are heavily doped regions of the semiconducting substrate that are positively charged in a p-type device and negatively charged in an n-type device. The source and drain are separated by a gate, and this gate is activated by applying a voltage to it that creates an electronic field. In an n-type device, a positive field is applied to the gate, which will then positively charge it. The gate will then attract electrons, and a strong enough field will allow movement of electrons from the source to the drain. P-type devices behave in the opposite manner.

2.1 Process Steps

There are several different process steps that are used to create the transistors in an integrated circuit. The basic steps are deposition, lithography, etching, and doping. Each is explained below.

2.1.1 Deposition

In this step, a thin film is deposited over the entire wafer surface. Methods used for this step include oxidation, chemical vapor deposition (CVD), and physical vapor deposition (PVD). Diffusion is primarily used for oxidation reactions to provide a passivation layer of silicon dioxide. $Si0_2$ is also used as a mask to protect regions of the wafer when etching features into the device. Physical vapor deposition is primarily used to apply metals to the substrate. Metals such as aluminum and tungsten are sputtered over the wafer. This process is highly directional. Chemical vapor deposition deposits a layer of material by introducing reactants in gaseous form into a chamber. When the gases reach the surface of the wafer, a chemical reaction occurs and leaves a layer on the surface. This process deposits more evenly on the surface in a non-directional manner.

2.1.2 Lithography

Lithography is used to transfer a pattern onto a wafer. During this step, a layer of light-sensitive photoresist is deposited on the wafer, and then light is shone through a mask. Current processes use either I-line ultraviolet or deep ultraviolet light (DUV) and chemically enhanced photoresists. This light changes the properties of the resist, making it more soluble in positive resists and less soluble in negative resists. A developer is applied to the wafer, and the soluble portions of the resist are removed. This uncovers specific areas of the wafer so that they may be etched or implanted.

2.1.3 Etch

During etch, portions of a layer are selectively removed. Areas may be uncovered so that they can be doped, metal lines may be etched out of a layer of aluminum, holes may be etched in a dielectric for vias, or trenches may be etched in the dielectric so they may be filled with metal. The remaining photoresist on a wafer from the lithography step protects the underlying layer from the etchants. A key performance parameter of an etchant is its selectivity. Selectivity measures the propensity of the etchant to dissolve away the target material as compared to the mask. Material can be wet-etched by immersing the wafer in an etchant or dry-etched by exposing it to an ion beam. Wet etching tends to be isotropic, which means that the etching is uniform and non-directional. This yields a semi-circular etch pattern. Dry etching tends to be more anisotropic which means that etching is directional. In today's fabrication processes, most etching is done with dry processes.

2.1.4 Doping

The substrate is doped to create areas with specific electrical properties, such as the source, drain, and gate. The process technologies used for this task are diffusion and ion implantation. In current devices, diffusion is used early in the process and in areas where a uniformly doped layer is desired. Ion implantation is the process of choice for the source and drain because it allows precise control of the amount of doping and is more directional.

When using diffusion, the entire wafer is exposed to the dopant in a high temperature environment. The areas of the substrate where the oxide has been etched away are now susceptible to the dopant. Diffusion is an isotropic process. It is also used to create silicide barrier layers to prevent aluminum from coming in direct contact with silicon.

Ion implantation uses a high-energy ion beam to drive positively or negatively charged atoms into the wafer. Once the implantation is complete, the damage to lattice is repaired with an anneal, and the ions are diffused into the material and ordered with a heat treatment.

2.2 Process Flow

The basic flow to create a transistor utilizes the various process steps to create one layer at a time. A cycle to create a layer is called a loop. Many loops are required to create a device, and since each loop utilizes much of the same tools, the process flow is called re-entrant flow. The most basic steps for an n-type MOSFET are described below.

2.2.1 Dope wells

The first step in creating a device is to create the source and drain. An insulating layer of SiO_2 is grown over the surface of a p-type wafer by oxidation. A pattern that will expose the sources and drains of the device is then transferred onto the wafer with lithography. The exposed areas are then etched away, exposing the substrate. These areas are then doped with a type III element such as boron by diffusion or ion implantation to create n-wells. The wells are then electrically activated and driven deeper into the substrate by a heat treatment. Another oxide is then deposited or grown on the surface.

2.2.2 Create gate

The next step is to create the gate. This gate is typically made of polysilicon. A pattern is transferred that exposes the gate areas with lithography. The exposed areas are then etched such that they leave a thin oxide over the substrate. This thin oxide is doped again with a type III element to tune the threshold voltage of the device. Polysilicon is then deposited over the surface with a CVD process and doped with

a type V element. Another oxide is then deposited on top of the substrate. The devices are now defined in the substrate and need to be connected.

2.2.3 Connect devices

The devices are connected by opening holes in the oxide to expose the sources, drains, and gates. This is done by transferring a pattern with lithography and etching the areas away. Barrier layers are then added to isolate the metal layers from the silicon. A metal film such as aluminum is then deposited on top of the substrate with either a CVD or PVD process. Another pattern is transferred to define the individual wires on the metal film. The remaining material is etched away, and then a dielectric is deposited to fill in the open areas.

2.3 Inspection and Monitors

Semiconductor manufacturing occurs on a microscopic level and in a highly automated fashion. In many ways, manufacturing in this industry resembles a job shop. Manufacturing floors tend to be laid out by functional area, and material moves from tool set to tool set in a non-linear manner. However, there are also strong similarities with process manufacturing. Each tool is essentially its own process and is highly automated. Technicians primarily load material and ensure that the tools are running properly. They also troubleshoot the tools when necessary.

Because material is modified at such a small scale, inspection cannot be done in a traditional sense, and testing is an extremely important tool. While many of the physical properties of semiconductor fabrication are well understood, the models used for process engineering do not provide perfect information. The cycle time to produce integrated circuits is on the order of months, and tooling is

expensive, and therefore costly to use. Therefore, it is important to ensure that all material is of good quality before passing it down the line to the next toolset.

A monitor is a test that is conducted to evaluate a tool, a set of wafers, or both. For example, a test on an etching tool would be used to determine if its removal rate is consistent. Another test could check for particles created by a lithography tool by processing some wafers through it and then counting the particles left on it. A test on a set of wafers could be to ensure the proper oxide thickness is deposited after an oxidation process.

For some monitors, the metrology results must be reviewed before any more production lots can be processed on a tool. This is known as a gating monitor since it the monitor holds the tool up until the metrology is completed. If the tool can continue to process lots before the metrology results are review, then the monitor is considered to be non-gating.

3. Vertical Diffusion Furnaces

Diffusion furnaces are batch processing tools that are used for thermal processes (Figure 2). This type of tool is called a hot wall tube because the heat is transferred by directly heating the walls of the tube, rather than directly heating the wafers on a susceptor within a tube. Wafers are transferred by a robotic arm from their cassettes into a quartz tube, which is also called a boat, that is then pushed up into the furnace. Once inside the furnace, the wafers are immersed in various gases and reactants as the boat is heated up to and held at the desired temperature. The advantage of using these tools is that they process large batches of material and consume little floor space. In an industry where factory space is extremely expensive and high throughput is essential, these advantages are very attractive. The design of the reactors also lead to extremely uniform and repeatable film production. The processes run on these tools usually last for several hours. The applications of these tools are described and then process models for oxidation and chemical vapor deposition are presented.

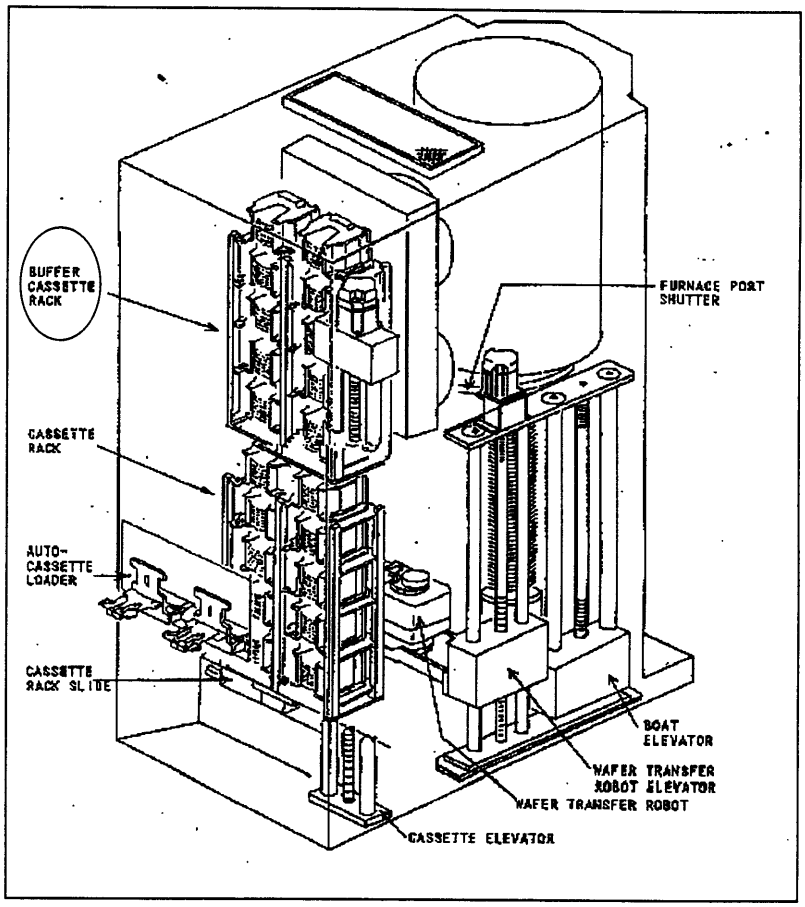

Figure 2. Vertical Diffusion Furnace

3.1 Applications

Furnaces can be used for oxidation, deposition, doping, ion activation, and annealing. However, as devices on integrated circuits have become smaller, the applications of the furnace have become more narrow, and now diffusion furnaces primarily perform oxidation and deposition along with some diffusion. Because devices and their wells have become more narrow, most doping is currently being done with ion implanters. Another result of shrinking device features is the reduction of the thermal budget due to the need to reduce the depth of the wells, so rapid thermal processing has become the process of choice for ion activation and annealing. RTP tools can provide similar functions as furnaces, but with much shorter cycle times, so they consume less of the thermal budget. While smaller device geometries have mostly reduced the workload of the diffusion furnaces, it has also provided a new need. Smaller devices require a higher degree of planarity, so now diffusion furnaces also perform deposition and reflow operations. New oxide systems have been developed that reflow at lower temperatures, and the furnaces provide the capability to deposit and reflow in one process step. The applications of diffusion furnaces are described below.

3.1.1 Oxidation

The creation of a silicon dioxide layer (SiO_2) is the main form of a mask used in microelectronic fabrication. It is a stable oxide that provides a good passivation of the silicon surface, and its growth is well understood and easy to replicate. The diffusivity of dopants such as boron and phosphorus are significantly less in SiO_2 than in Si, so it serves as an excellent mask for doping.

Silicon dioxide naturally forms on silicon when exposed to an oxidizing ambient such as oxygen or steam, known as dry and wet oxidation, respectively. This oxide is formed by the diffusion of an oxidant through the oxide to the silicon/oxide interface. Silicon is consumed, and the interface moves deeper into the silicon. At room temperature, a thin oxide of less than 20Å will form[1]. When exposed to elevated

temperatures, the oxide grows more rapidly. Therefore, this is also called a thermal oxide since it is produced by heating the substrate in a furnace.

3.1.2 Diffusion and Doping

In early semiconductor manufacturing, the wells were doped with diffusion processes. Wafers were immersed into a dopant-rich environment at elevated temperatures. The difference in dopant concentration between the surface and the substrate along with elevated temperature provide the driving force for diffusion. This process worked well for devices with junction depths of 1-3μm; however, in modern devices, the junction depths are well below 1μm, and this depth is more difficult to achieve with diffusion processes. Currently, devices are doped with ion implantation, which provides more precise control of the dopant level. It is also more directional, so thinner wells may be created.

3.1.3 Deposition

Because diffusion furnaces were designed to immerse wafers in a specifically designed chemical and thermal environment, they can also serve as chemical vapor deposition (CVD) tools. The basic mechanism for CVD processes is that reactants are delivered to the wafer surface where a chemical reaction occurs. The rate of deposition is dependent upon the reaction rate and the mass transfer rate. Ideally, processes and tools should be designed such that they are reaction rate limited rather than mass transport limited so the reaction, and thus the film thickness, may be manipulated by the temperature.

CVD is used to deposit the polysilicon (or poly) used for the gates of the devices. Polysilicon is the material of choice because its properties are easily manipulated to produce the necessary electrical properties. Resistivity is controlled by the amount of doping and the grain size of the film. Doping can reduce the resistivity of a semiconductor. On the other hand, grain boundaries increase the resistivity of

the material through several mechanisms. Because atoms move more freely at the grain boundary, the diffusivity of dopants is higher than in crystalline silicon. Therefore, the grain boundaries will attract a disproportionate amount of the dopants than will the bulk crystal. Grain boundaries also serve as traps for free carriers because they contain many incomplete bonds. Thus, smaller grains lead to higher overall resistivity.

In addition to polysilicon deposition, CVD is also used for the interlayer dielectric (ILD) that surrounds the gate. The topography of the wafer surface has sparked development of several deposition and reflow processes that reduce voiding and increase the planarity of the devices. The aspect ratio, or the ratio of height to width, of modern devices is much higher than those of earlier devices. This can lead to voiding with traditional CVD processes as shown in Figure 3. CVD is a largely conformal process where the film grows at a fairly even rate along the surface. However, as aspect ratios become higher, the film grows faster on the top of a feature than it does in a trench. Ultimately, the film on two adjacent features will come together and leave an empty space below it. This void can attract water vapor, which will adversely impact the performance of the device. Modern TEOS-ozone CVD processes eliminate voiding and produce lower surface angles.

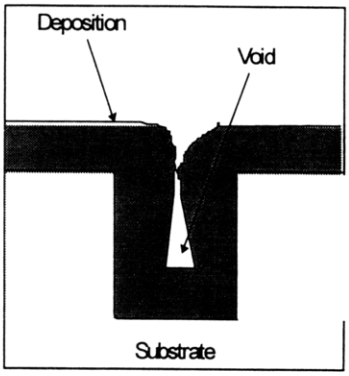

Figure 3. Voiding

Topography of the wafer surface can also lead to poor pattern transfer. As feature sizes decrease, so does the depth of focus. $TEOS/O_3$ systems can be modified by adding a dopant such that its reflow temperature is substantially decreased. The layer is then deposited and heated such that it will reflow. Materials such as phosphorous doped oxide (PSG) and borophosphorsilicate glass BPSG are examples of dopants used in this manner. PSG reflows at 1000° C. BPSG reflows at a lower temperature, deposits with less stress, and provides a more effective passivation layer. This material is the system of choice for

the dielectric used prior to the first metal deposition on 1MB and 4MB DRAMs. APCVD using TEOS offers a single-tube, deposition and reflow process.[2]

3.1.4 Ion Activation

Since ion implantation essentially thrusts ions into the substrate, the ions tend to occupy interstitial rather than substitutional sites. In order to electronically activate the ions, a thermal step is needed to provide the driving force to provide ions the mobility to move into the substitutional sites. Full activation requires temperatures in excess of 800°C.

Thermal processes allow the ions to redistribute themselves horizontally as well as vertically. Therefore, implanted species will diffuse deeper into the substrate and the gap between the wells will shrink while the ions are being activated. This will reduce the concentration of the ions near the surface and increase the effective depth of the implant.

RTP cycles of 1000°C for 10s can activate the implant as effectively as a 30 minute furnace cycle at 1000°C. The redistribution distances when using RTP also decrease from several thousand angstroms to the order of hundreds of angstroms.

3.1.5 Annealing

Ion implantation also causes great damage to the lattice of the substrate. Thermal processes are needed to provide the energy for the lattice to heal. At temperatures up to 500°C, vacancies and interstitials start to recombine. However, at higher temperatures, dislocations start to form, which can trap impurity ions. These dislocations cannot be dissolved until the anneal temperature reaches 1000°C.[3] Since the activation energy of impurity diffusion is always less than that of Si self-diffusion, the higher the anneal

temperature the better. This is another reason why RTP is usually the process of choice for annealing after implantation. RTP processes can ramp the temperature up and down faster than a furnace.

3.2 Process Models

3.2.1 Oxidation

Oxidation results from two primary mechanisms: diffusion through the existing oxide to the interface and the reaction of the oxidant with the substrate. In dry oxidation, wafers are immersed in oxygen. After the oxygen molecules diffuse through the oxide, the following reaction takes place:

$$Si\ (solid) + O_2\ (vapor) \xrightarrow{k_s} SiO_2\ (solid)$$

The grow of silicon dioxide is well understood, and for films greater than 300Å thick, the Deal-Grove linear-parabolic model accurately predicts oxide thickness over a wide range of temperatures and oxidant partial pressures. Thinner oxides grow much more rapidly than the model predicts.

The model is based on the equilibrium of three fluxes: 1) the flux of the oxidizing species through the bulk gas to the gas/oxide interface, 2) the flux of the oxidizing species diffusing through the oxide, and 3) the flux of the oxidizing species as it is consumed by the reaction at the Si/SiO_2 interface. There are numerous references that show the derivation of the model, including Wolf & Tauber and Middleman & Hochberg, so none will be given here. However, a more qualitative discussion about the fluxes and how their behavior results in the linear-parabolic form follows.

The rate of the reaction that converts silicon to silicon dioxide is proportional to the concentration of the oxidizing species at the Si/SiO_2 interface. This may be expressed as:

$$F_3 = k_s C_i$$

Therefore, if a constant concentration of the oxidant can be maintained at the interface, the film will grow at a liner rate.

However, in order for the oxidant to reach the interface, it must first diffuse through the existing oxide. As the oxide becomes thicker, the flow of the oxidants to the interface will decrease. This flux may be represented as:

$$F_2 = D\frac{(C_o - C_i)}{z}$$

where D is the diffusion coefficient; C_o and C_i are the concentrations of the oxidizing species at the gas/oxide interface and oxide/silicon interface, respectively; and z is the thickness of the oxide. The growth of the oxide starts out linearly because the film is thin and the oxidants can readily diffuse through to the interface. The oxidant is plentiful, so the reaction is not limited by its concentration but by the speed in which the reaction can take place. In this regime, the growth is reaction rate controlled. However, as the film starts to grow, and the reaction can occur as fast as the oxidizing species can reach the interface, it becomes diffusion controlled and grows in a parabolic fashion.

The final form of the Deal-Grove model is:[4]

$$t - t_0 = \frac{Z - Z_0}{k_{LIN}} + \frac{Z^2 - Z_0^2}{k_{PAR}} \qquad (3\text{-}1)$$

where: t_0 is the duration of the initial oxidation regime,

Z_0 is the initial oxide thickness at the time t_0,

k_{LIN} is the linear rate constant, and

k_{PAR} is the parabolic rate constant.

Both of the constants are dependent upon temperature, and may be expressed as follows:[5]

$$K_{LIN} = A_{LIN} \exp\left(\frac{-E_{AL}}{k_B T}\right) \tag{3-2}$$

$$K_{PAR} = A_{PAR} \exp\left(\frac{-E_{AP}}{k_B T}\right) \tag{3-3}$$

where: $A_{LIN} = 6.23 \times 10^{10}$ Å/hr, $E_{AL} = 1.23$eV/molecule, and

$A_{PAR} = 7.72 \times 10^{10}$ Å^2/hr, $E_{AP} = 2.0$ eV/molecule

for {111} silicon and 1 atm of pressure. The data for {100} silicon will be less. The surface of {111} silicon is denser, and thus provides more molecules for the reaction.

3.2.2 CVD

Chemical vapor deposition is achieved by introducing a gas into the furnace which then reacts with the surface of the substrate. There are two main mechanisms by which material is deposited: gas phase processes and surface processes. The gas phase process consists of reactants crossing the boundary layer between the gas flowing into the chamber and the substrate. The rate of this process is characterized by the diffusivity of the gas and the concentration gradient across the barrier layer and is weakly influenced

by temperature ($D \propto T^2$). Once the reactants cross the barrier layer, they may be consumed by a surface reaction. This reaction rate is highly dependent upon temperature ($R \propto e^T$).

Thus, the rate of reaction in CVD processes may be either mass transport limited or reaction rate limited. When the gas phase processes constrain the rate of deposition, the process is mass transport limited. The surface reaction can proceed faster than reactants can cross the boundary layer, so as soon as the reactants are available, they are consumed. This leads to non-uniformity because the reactants are depleted and the concentration gradient across the surface of the wafer becomes significant. The areas downstream from the gas inlet exhibit a lower growth rate. When the surface processes constrain the rate of deposition, then reactants are readily available at the surface. This causes the rate of deposition to be more uniform. Because the reaction rate is highly dependent upon temperature, the deposition rate can then be easily manipulated.

Assuming a reaction rate limited process for the system:

$$xy \xrightarrow{k} x + y$$

The rate of film growth of reactant x may be expressed as follows:

$$\Re = kP_x = k_s C_x \qquad (3\text{-}4)$$

where $\Re$ is the rate of deposition, k is the first order reaction rate constant, P_x is the partial pressure of the reactant x, k_s is the surface reaction rate, and C_x is the concentration of the reactant x.

One manner to facilitate a reaction rate dependent process is to use a low pressure CVD (LPCVD) reactor. By lowering the pressure, the diffusivity of the bulk gas through the barrier layer is greatly

increased because diffusivity is inversely proportional to pressure. Under a pressure of about 1 torr, the diffusivity of the gas species increases by a factor of 1000 over the diffusivity at atmospheric pressure. On the other hand, the lower pressure only increases the length of the boundary layer by less than the square root of the pressure. The net effect is more than an order of magnitude increase in the mass transport through the boundary layer.

Hitchman and Jensen demonstrate by using a simple LPCVD model where wafers are stacked perpendicular to the gas flow that the within wafer uniformity is characterized by single, dimensionless, parameter based on wafer radius, reaction rate constant, diffusivity, and wafer spacing.[6] The main points of the derivation are highlighted below:

The interwafer region in the mass balance is described as follows:

$$D\left[\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial c}{\partial r}\right)+\frac{\partial^2 c}{\partial z^2}\right]=0 \qquad (3\text{-}5)$$

where r is the radial coordinates of the wafer, z is the axial coordinate along the length of the reactor, and c is the concentration of the gas.

The boundary conditions are:

$$\left.\frac{\partial c}{\partial r}\right|_0 = 0, \quad c|_{r=R_w} = c_0 \qquad (3\text{-}6a)$$

$$\pm D\frac{\partial c}{\partial z} = kc \quad \text{at } z = z_l \text{ and } z = z_l + \Delta = z_{l+1} \qquad (3\text{-}6b)$$

where Δ is the wafer spacing, and k is the reaction rate constant.

In LPCVD, the axial variation of concentration around the wafer can be simplified because it will only be significant in systems with extremely fast reactions. These reactions are not appropriate for this type of tool, so the simplification is reasonable. By averaging in the axial direction, the following mass balance equation for the interwafer region may be used:

$$D\left[\frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\left(r\frac{\mathrm{d}c}{\mathrm{d}r}\right)+\frac{2}{\Delta}kc\right]=0 \tag{3-7}$$

By scaling the radial coordinates with $\xi = r/R_w$ and the concentration with $y = c/c_0$, equation 3-6 can be represented as follows:

$$\frac{1}{\xi}\frac{\mathrm{d}}{\mathrm{d}\xi}\left(\xi\frac{\mathrm{d}y}{\mathrm{d}\xi}\right)-\phi^2 y=0 \tag{3-8}$$

with:

$$\left.\frac{\mathrm{d}y}{\mathrm{d}\xi}\right|_{\xi=0}=0,\quad y|_{\xi=1}=1, \text{ and} \tag{3-9}$$

$$\phi^2=\frac{2R_w^2 k}{\Delta D} \tag{3-10}$$

where D may be replaced by the binary diffusion coefficient using Chapman-Enskog theory as:

$$D_{AB} = \frac{1.41\sqrt{T^3\left(\frac{1}{M_A}+\frac{1}{M_B}\right)}}{P\sigma_{AB}^2\Omega_D} \quad (\mathrm{cm}^2/\mathrm{s}) \tag{3-11}$$

where: M_A is the mass of compound A in grams

M_B is the mass of compound B in grams

P is the pressure in torr

σ_{AB} is the Lennard-Jones molecular diameter in angstroms $\{\sigma_{AB} = \frac{1}{2}(\sigma_A + \sigma_B)\}$

Ω_D is the collision integral which is dimensionless

Therefore, the film thickness variation between the wafer edge and a radial position is:

$$\frac{\delta(\xi)}{\delta(\xi = 0)} = y = \frac{I_0(\phi\xi)}{I_0(\phi)} \tag{3-12}$$

Where I_0 is the incomplete zeroth order Bessel function. The within wafer non-uniformity is thus characterized by ϕ^2. This parameter is known as the Thiele modulus or Danköhler number, and is a measure of the relative importance of diffusion and reaction time scales. If $\phi^2 << 1$, then the system is reaction rate limited. When $\phi^2 >> 1$, then the reaction is mass transport limited.

The film thickness uniformity may be quantified by defining a parameter, η, equal to the ratio of the average deposition rate relative to the deposition rate at the edge of the wafer. The following relationship then measures uniformity as a function of ϕ:

$$\eta = 2\int_0^1 yx\,dx \quad = \frac{2I_1(\phi)}{\phi I_0(\phi)} \tag{3-13}$$

There are several ways to reduce ϕ^2 and increase the uniformity. The temperature can be lowered, which will reduce k, but this will also reduce the deposition rate. The wafer spacing, Δ, could be increased, but then fewer wafers could be processed in each batch. The best alternative is to reduce the pressure, which will increase D.

3.3 Summary

Vertical diffusion furnaces are batch processing tools that are used for thermal processes. While the tools have historically performed many different functions, such as oxidation, deposition, doping, ion activation, and annealing, in current processes, operations performed on VDFs are mostly limited to oxidation and deposition. As feature sizes have shrunken, the processes required to produce these devices have escaped the process windows that VDFs can provide. Narrower wells have made doping by ion implantation necessary. The reduced thermal budgets of current devices require rapid thermal processing for ion activation and annealing once the wells are doped.

While the applications of vertical diffusion furnaces have declined, they still offer highly controllable process capabilities for the operations that they continue to perform. Process models are described for both oxidation and chemical vapor deposition. Oxides are grown by immersing the wafers in an oxidizing environment of pure oxygen or steam. The oxidant then diffuses through the silicon dioxide to the silicon/oxide interface where silicon is oxidized. This process is well understood for thicker oxides (>300Å) and oxide growth is described well by the Deal-Grove linear-parabolic model. This model utilizes two parameters, each dependent upon pressure and temperature, to estimate the oxide growth.

In chemical vapor deposition, reactants are introduced into the furnace in a gas phase where they may react with the surface of the substrate. These reactions may be either reaction rate limited or mass transport limited. Reaction rate limited systems are preferable because they can be controlled by adjusting the temperature and provide more uniform films. Mass transport limited systems need to ensure an equal concentration of reactants are available across the surface of the wafer if uniform films are to be produced. One technique to facilitate reaction rate limited systems is to use low pressure CVD, which greatly increases the diffusivity of the reactants through the boundary layer while modestly increasing the distance of the boundary layer. The uniformity of these processes may be estimated with a single parameter, ϕ, which is dependent upon the wafer radius, reaction rate constant, diffusivity, and wafer spacing.

4. Risk/Reward Analysis

In this chapter, the rewards of using non-gating CBP are compared to the risks associated with it to determine whether or not the process should be run in a non-gating manner. The rewards are quantified by calculating the time savings by eliminating the following three periods: 1) the batch is loaded into the tool, 2) a recipe has been completed and the tool waits to be unloaded, and 3) the test measurements are being taken on the completed batch. Data from three different facilities were used.

The risk is assessed from both a statistical and engineering perspective. The control chart data from the development fab was used to determine the probability of a lot being out-of-disposition. This probability is then used to place a cost on the risk incurred when running the process in a non-gating manner. An engineering sensitivity analysis building on the process models described in the previous chapter is also conducted. This analysis provides boundaries around the temperature and pressure parameters that must be met to ensure conformity to theoretical specifications.

4.1 Rewards

The operational efficiency was quantified by analyzing the three distinct periods of wasted time. In order to determine how much time is wasted, it is useful to understand how the data from the manufacturing information system maps to the process. This map is displayed in Figure 4. There are two main sources of data needed for this analysis: the lot history and the entity history. The map shows that the time wasted when loading the batches (Wasted Time 1, or WT1) is the difference between the CHECK EQP transaction and the VAL8VDF transaction in the entity history. The time wasted while the furnace waits to be unloaded (WT2) is the difference between the END8VDF transaction in the entity history and the MOVE OUT VDF transaction in the lot history. Finally, the time wasted while the metrology is being

conducted (WT3) is the difference between the MOVE OUT transaction from the furnace in the lot history and the MOVE OUT transaction from the metrology tool in the lot history.

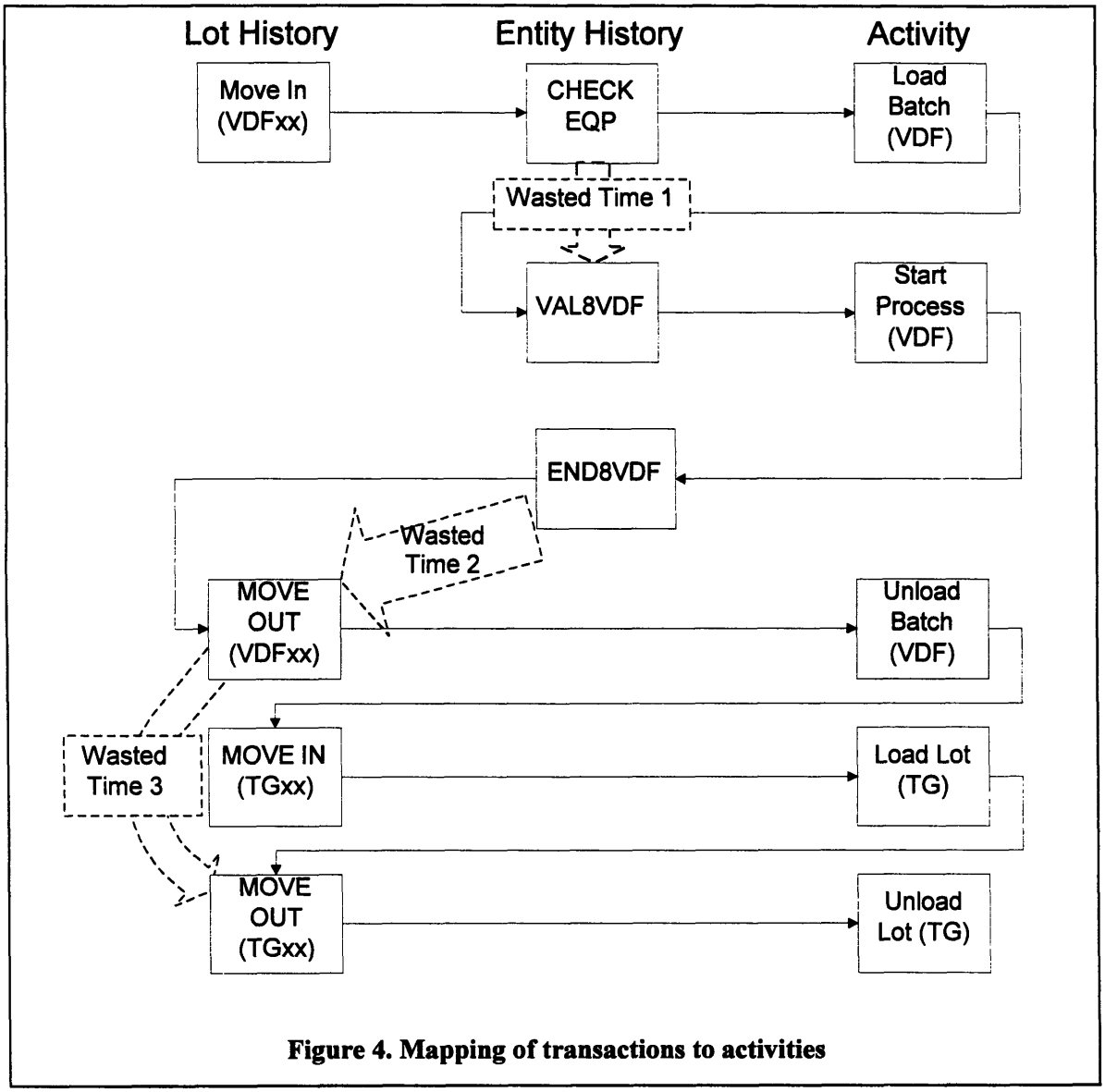

Figure 4. Mapping of transactions to activities

The challenge in calculating the wasted time periods is to sequence the two transactions properly so that the difference in the times the transactions were posted to the database correctly represent the elapsed time for the wasted time period in question. The lot history is the easiest to work with because each record contains both the lot number and the process step. Using the entity history is more difficult

because some of its records only contain the lot number, and the corresponding process step also needs to be tied to a record before the data can be used. Because of the re-entrant flow nature of semiconductor processing, a lot may be processed through the same furnace more than once. Likewise, each process step may be performed on any of several tools that have been qualified for that step.

Data was extracted out of the manufacturing database into Microsoft Access. Within Access, queries were developed that would automatically sequence the data and calculate each of the three wasted time intervals for a representative lot for each batch. The WT1 and WT3 queries ran relatively quickly. The WT2 query was much more complex and needed to be completed overnight.

The WT1 time period was relatively simple to calculate because the furnace records for the CHECK EQP and the VAL8VDF transactions all contain both the lot number and process step. The CHECK EQP transactions were extracted into a separate table as were the VAL8VDF transactions. The differences between the CHECK EQP and VAL8VDF transactions are calculated for each process step and lot. To account for multiple transactions, the minimum, non-negative difference is considered to be WT1 for the process step-lot pair.

Because there are multiple lots in each batch, one lot needs to be chosen to represent the batch. In the database, all of the lots in a batch have their transactions posted at the same time, so the minimum lot number was taken to represent the batch. This method was used for all of the wasted time periods.

The WT2 time was the most difficult period to quantify because the END8VDF transactions do not indicate which lots were in the furnace or what process step was being run. An END8VDF transaction simply notes which furnace stopped at what time. In order to tie an END8VDF transaction to a lot number, the VAL8VDF transactions were used. Since the VAL8VDF transactions contain the lot number and are the last transaction before an END8VDF transaction, the minimum, non-negative time between

the transactions for each END8VDF transaction for each furnace was extracted. The VAL8VDF time is then deleted from the record, and a lot and process step are now associated with an END8VDF event. The difference between this time and the MOVE OUT event for the corresponding lot and process step determine the WT2 time.

The WT3 time period was also relatively simple to calculate because it only requires the lot history. For each diffusion operation, there is a corresponding metrology step. For example, consider a metrology step #101 that always follows a diffusion operation #100. Every production lot must go through both of these steps. In order to get a the elapsed time for metrology on a batch, the move out times for each lot were extracted for every paired set of diffusion and metrology steps. Since there are some redundant transactions in the database, the latest time is taken from the MOVE OUT transactions of the furnace for each lot and step. This is paired with the earliest MOVE OUT transaction from the metrology tool for the corresponding metrology step, which produces the most conservative estimate.

If a lot was placed on hold or if there was an error in the furnace while a batch was running, then the data for the lot would be exaggerated. When a lot is placed on hold, it will not complete the metrology step until a disposition is made on the lot. This will lengthen the WT3 period. If a furnace encounters an error, then the lot will continue to be assigned to a furnace until the error is resolved. This will increase the WT2 period.

In order to identify when a lot was placed on hold, once all of the queries were run, another query was used to extract all of the unique lot numbers in the dataset. This lot list was used to find any HOLD LOT transactions for these lots during the time of the study. If a lot was placed on hold, then the data for that process step was not used for this analysis.

A similar technique was used to filter out the data when there was a furnace error. This was accomplished by compiling a list of all the times there was a VDF ERROR transaction in the entity history. If there was a VDF ERROR between the start of the WT1 event and the end of the MOVE OUT event for the furnace, then the data was set aside.

The benefits of using non-gating CBP at D2 were significant. The median time savings for a batch was almost forty-five minutes per cycle. The median time was used because some batches had very long WT2 periods where they were waiting to be unloaded from the furnace. The median times over all of the operations are described in Table 1. Note that the median times do not add up to forty-four minutes. This is because forty-four minutes is the median time for the sum of all wasted time periods for a batch.

Process Steps	WT1: Loading furnace	WT2: Waiting for unloading	WT3: conducting metrology	Total Wasted Time
No Particle Monitor	7.37 minutes	14.57 minutes	11.05 minutes	39.07 minutes
Particle Monitor	7.30 minutes	18.58 minutes	23.32 minutes	57.48 minutes
All VDF Steps	7.35 minutes	15.90 minutes	14.23 minutes	43.83 minutes

Table 1. Summary of Wasted Time Periods for D2.

For D2, the data set consisted of 700 batches that were run in the first half of 1997. Initially, one year of data was analyzed; however, this time period was chosen because the cycle time of the process became more stable as one would expect in a development environment.

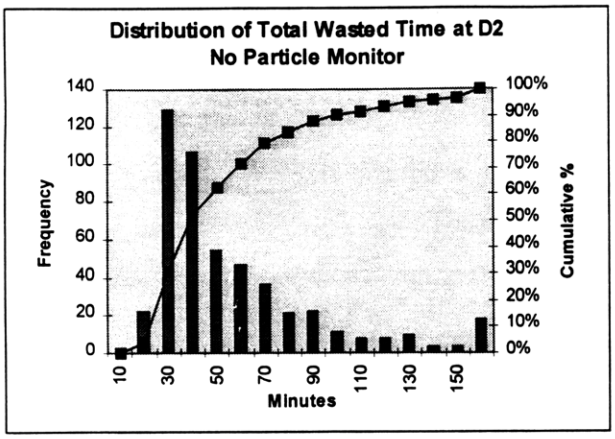

Figure 5. Distribution with no particle monitor: D2

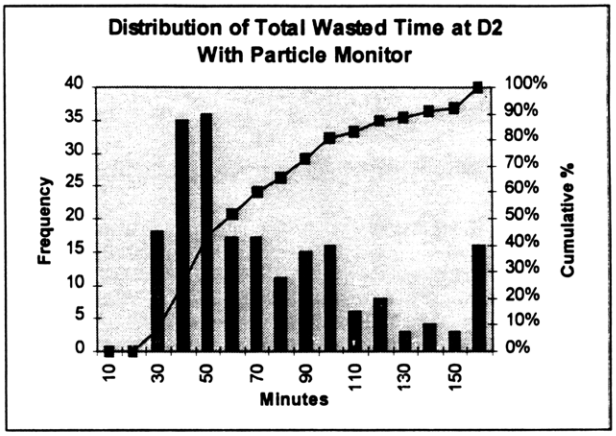

Figure 6. Distribution with particle monitor: D2

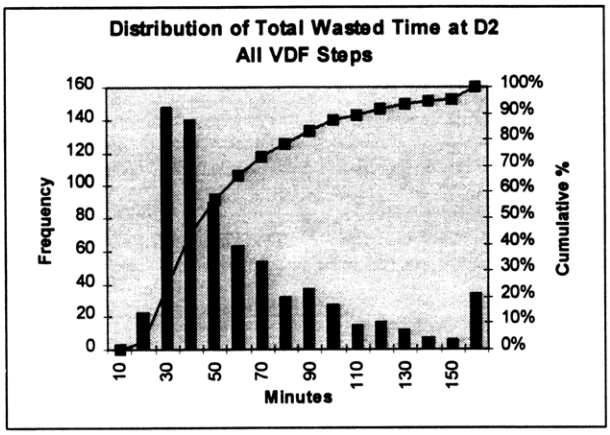

Figure 7. Distribution for all steps: D2

The above figures show the distribution of the total wasted time at the D2 facility. Figure 5 displays the distribution of the total wasted time for batches that only require a thickness monitor and do not require a particle monitor. Figure 6 shows the distribution of the total wasted time for batches also required a particle monitor. Notice how the distribution is not as smooth. This is due to the fact that two tests must be done, and there may be some additional idle time between the tests. Figure 7 shows the distribution for all of the steps.

One key difference between fabs that will affect operational efficiency is the use of automated material handling systems (AMHS) within a functional area. A functional area is made up of several bays, which are simply smaller corridors that branch off the main aisle. While the use of AMHS is commonplace to move lots of wafers between bays, the use of automation within a bay is less common. Automation is one facet of fab design that has not been fully integrated into Intel's Copy Exactly! approach, and AMHS within a bay is called Intrabay.

Fabs using Intrabay in the diffusion area will be more efficient because once a lot is completed, the automation will remove the test wafers from the furnace and present them to the technicians who are centrally located. This eliminates the time wasted while waiting for the furnace to be unloaded.

Another difference between the fabs is the type of capacity management information system used. D2 is the last fab to use an older system. Therefore, the names of the transactions used for the various states of the process differ. The transaction names are described in Table 2.

Old System	New System
CHECK EQP	VALID EQP
VAL8VDF	BEGIN PRECHG
END8VDF	WAIT METRO

Table 2. Capacity Management System Transaction Comparison.

The new system also has another transaction called END METRO that is posted when the metrology is completed. This allows this analysis to be done without the lot history, but at the expense of combining the WT2 and WT3 times. The old system did not have a transaction to indicate when the metrology was completed, so the lot history was used to determine when the metrology was completed by the time of the MOVE OUT transaction from the metrology tool. While the new system did allow the study to be one with only the entity history, the lot history was still used in this study to better compare the three different times at the different fabs.

In addition to using different automation, the other fabs all use gating CBP. Therefore, the WT1 time is eliminated. The use of gating CBP manifests itself in the data with long WT1 times because the batches are loaded, which posts a CHECK EQP transaction, then they queue until the current batch finishes. The time between the CHECK EQP and the BEGIN PRECHG transaction is much longer than normal.

A summary of the median total wasted time for Fab 11, which uses intrabay, is shown in Table 3. Because of the use of intrabay, the only wasted time is in WT3. Batches are preloaded, eliminating WT1, and they are automatically unloaded, which eliminates WT2. However, the WT3 time will be longer. This increase in WT3 is caused by the test wafers being unloaded immediately and then presented to the command center. Once the wafers arrive, they may queue here. The WT3 time at an intrabay fab is essentially the sum of the WT2 and WT3 times at a non-intrabay fab. While the WT3 time is longer at F11, it is still less than the sum of the wasted time periods at D2. For a fab using intrabay and the new capacity management system, it would be more useful to simply look at the furnace history, and the wasted time would be the time between the WAIT METRO and END METRO transactions.

Process Steps	Total Wasted Time
No Particle Monitor	26.61 minutes
Particle Monitor	30.73 minutes
All VDF Steps	26.63 minutes

Table 3. Total Wasted Time for F11

The distributions of the total wasted time at F11 are displayed in Figures 8, 9, and 10. Notice that the distributions are more tightly clustered with lower variation that those of D2. This was expected because the test wafers are delivered to the technician. This will help eliminate the time that the furnace waits for a technician to notice that processing has been completed and then unload the test wafers for measurement.

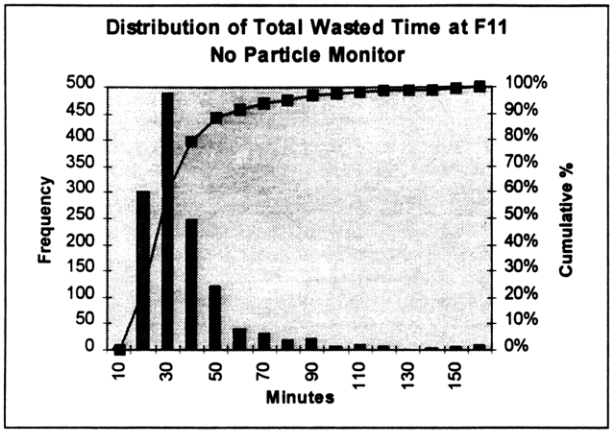

Figure 8. Distribution with no particle monitor: F11

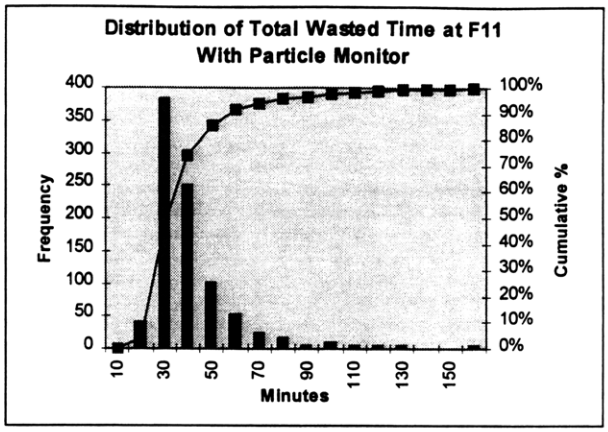

Figure 9. Distribution with particle monitor: F11

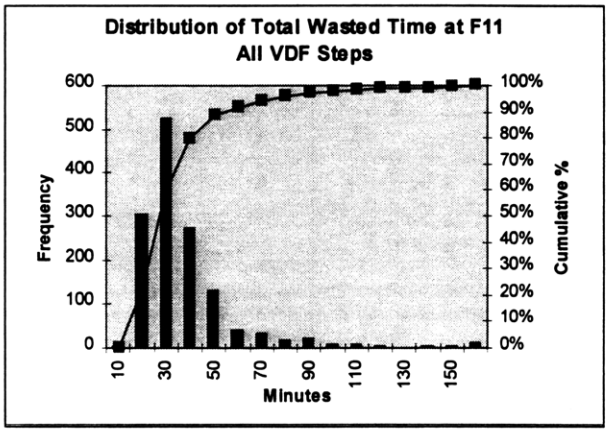

Figure 10. Distribution for all steps: F11

Additional data was gathered from F15, which is a smaller fab similar to D2 that does not use intrabay. This fab provides a better comparison to the performance of D2. Since the factory uses gating CBP, there is no WT1. However, the WT2 and WT3 times are directly comparable. A summary of the median times is provided below in Table 4. The distributions of the total wasted time are shown in Figures 11, 12, and 13.

Process Steps	WT2: Waiting for unloading	WT3: conducting metrology	Total Wasted Time
No Particle Monitor	14.57 minutes	11.05 minutes	39.07 minutes
Particle Monitor	18.58 minutes	23.32 minutes	57.48 minutes
All VDF Steps	15.90 minutes	14.23 minutes	43.83 minutes

Table 4. Summary of Wasted Time Periods for F15.

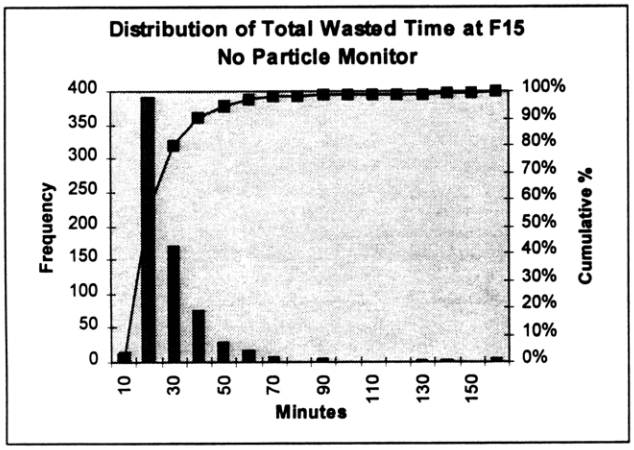

Figure 11. Distribution with no particle monitor: F15

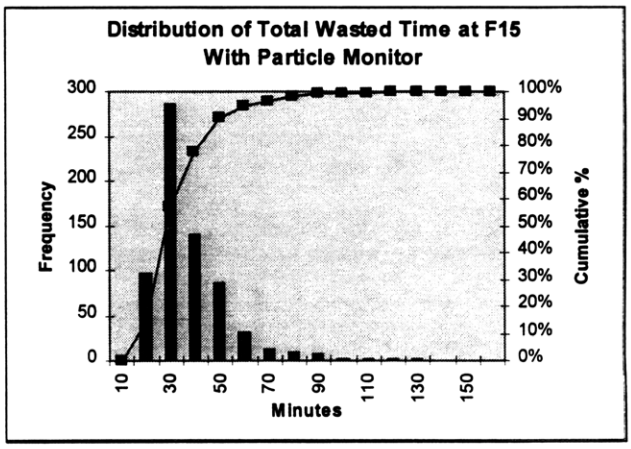

Figure 12. Distribution with particle monitor: F15

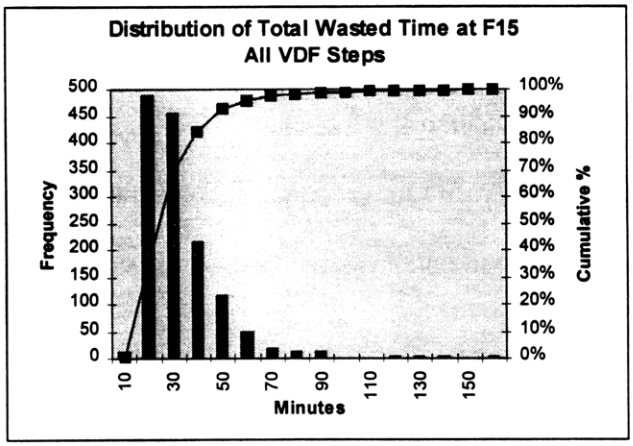

Figure 13. Distribution for all steps: F15

The figures show that the distributions for the total wasted time for F15 are also tighter than those of D2. However, unlike F11, F15 is a better comparison to D2. Neither fab uses intrabay, and they run comparable production volumes.

Another wrinkle that was discovered during this analysis is that the furnace needs to precharge before it can push the boat up into the furnace. Processing of the next batch may be stopped during this time with no negative impacts on the batch. This effectively reduces the effect of the WT2 and WT3 periods on the total wasted time when using gating CBP. If the metrology is completed before the furnace completes precharging, then there is no wasted time. Table 5 shows the amount of time that it takes to precharge the furnace for the different process types.

Furnace Type	Precharge Time
A	19 minutes
B	17 minutes
C	56 minutes
D	52 minutes

Table 5. Precharge Times

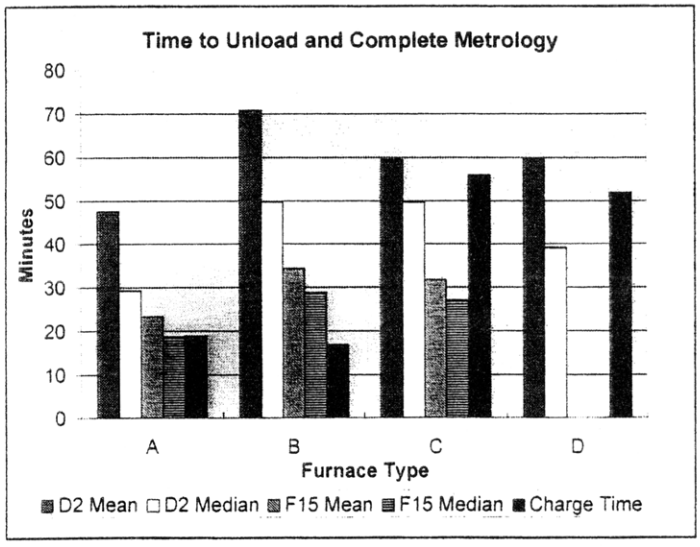

Figure 14. Comparison of WT2 + WT3 to precharge time

The data from F15 showed that, with the exception of the furnace type "B," the batches could be unloaded and measured before the furnace completed precharging. Figure 14 shows that the median times for F15 are roughly equal to the precharge time for furnace type "A." This indicates that half of the time, the furnace can be unloaded and the metrology completed before the furnace charges. Thus, gating CBP can

provide a run rate equal to non-gating CBP. This indicates that with a different staffing plan or work rules that non-gating CBP run rates can be achieved when running the process in a gating manner.

4.2 Risk

If CBP is run in a non-gating manner, then there is the risk that an additional batch will be lost if there is an OOD condition. The first batch would be scrapped under any processing work rules; however, the second batch would have been stopped after the metrology was completed if the process was run in a gating manner. Therefore, the probability of a batch being lost remains the same, but the consequences double if there is an OOD condition.

Two ways to analyze the risk will be presented: a statistical analysis and an engineering analysis. The processes in semiconductor manufacturing are managed from a largely statistical perspective. In fact, so much data is collected that Intel is rationalizing how much data it really needs. The statistical analysis presented relies on control chart data to place a probability on the failure of a batch. For an engineering perspective, process models are analyzed. From these models, boundary conditions can be established to determine how big of a process window exists to keep the process producing films that are in spec.

4.2.1 Statistical Analysis

In order to quantify the risk of losing a batch, the control chart data was analyzed from D2 for the first two quarters of the year. Of 1548 batches that were run, 59 were out-of-control, and only one was OOD. The one OOD batch was caused by some work being done on one of the gas lines. Therefore, the root cause was not equipment related. These numbers indicate that there is very little risk of losing an additional batch. However, the OOC frequency of 3.8% does call into question the control limits set for the process. Some of the control limits used in the statistical tool were not based on three sigma rules. In addition, some of the work rules allowed technicians to "tweak" the process.

The process capability (C_{pk}) measurements reinforce the conclusion that there is little risk of an OOD. Some sample data are listed in Table 6. An interesting observation is that the process step with the highest C_{pk}, Process "S", is one of the steps that require a particle check.

Process Step	Cpk
P	2.2
Q	1.7
R	1.6
S	2.5
T	1.3

Table 6. Sample C_{pk} data

With the assumption that the process is centered, the probability of an OOD can be calculated from the C_{pk}. The formula for C_{pk} for a centered process is shown below:

$$C_{pk} = \frac{USL - LSL}{6\sigma} \qquad (4\text{-}1)$$

Thus, for process step "S," where the C_{pk} is 2.5, the specification limits are $\pm 7.5\sigma$! Using the assumption that the process is normally distributed, the probability of an out-of-spec condition is:

$$2\left[\Phi(7.5) - 1\right] = 6.4x10^{-14} \qquad (4\text{-}2)$$

4.2.2 Engineering Analysis

The main tool parameters that can cause a failure are the reactant flow and the temperature. The mass flow controllers (MFC) that regulate the amount of gases that are introduced to the furnace are critical. If these fail or start to provide erroneous data to the station controller, then a proper film will not be deposited. The thermocouples that measure the temperature are also important. After all, these are

thermal processes that are highly temperature dependent. An analysis of the process models for oxidation and chemical vapor deposition highlights the effects of these variables.

4.2.2.1 Oxidation

For oxides grown in atmospheric pressure processes, the effects of an MFC failure is minimal as long as the wafers remain immersed in an oxygenated environment. However, the process is also highly dependent upon temperature. The effect of temperature differentials on the thickness of the film can be evaluated by using the Deal-Grove model described in Chapter 2. By combining equations 3-1, 3-2, and 3-3, the model can be expressed as the following (ignoring the initial conditions Z_0 and x_0):

$$t = \frac{Z}{A_{LIN}} \exp\frac{E_{AL}}{k_B T} + \frac{Z^2}{A_{PAR}} \exp\frac{E_{AP}}{k_B T} \tag{4-3}$$

Differentiating with respect to Z and T yields:

(4-4)

$$0 = \frac{1}{A_{LIN}} \exp\frac{E_{AL}}{k_B T} dZ - \frac{ZE_{AL}}{A_{LIN} k_B T^2} \exp\frac{E_{AL}}{k_B T} dT + \frac{2Z}{A_{PAR}} \exp\frac{E_{AP}}{k_B T} dZ - \frac{Z^2 E_{AP}}{A_{PAR} k_B T^2} \exp\frac{E_{AP}}{k_B T} dT$$

After collecting terms:

$$\frac{dZ}{dT} = \frac{Z\left(\frac{ZE_{AL}}{A_{LIN}} \exp\frac{E_{AL}}{k_B T} + \frac{ZE_{AP}}{A_{PAR}} \exp\frac{E_{AP}}{k_B T}\right)}{k_B T^2\left(\frac{1}{A_{LIN}} \exp\frac{E_{AL}}{k_B T} + \frac{2Z}{A_{PAR}} \exp\frac{E_{AP}}{k_B T}\right)} \tag{4-5}$$

Substituting back k_{LIN} and k_{PAR} yields:

$$\frac{dZ}{dT} = \frac{Z\left(\frac{E_{AL}}{k_{LIN}} + \frac{ZE_{AP}}{k_{PAR}}\right)}{k_B T^2\left(\frac{1}{k_{LIN}} + \frac{2Z}{k_{PAR}}\right)} \qquad (4\text{-}6)$$

For an oxide growth of 1500Å at 1000°C, $\frac{dZ}{dT}$ = 11.3 Å, or 0.75%. Therefore, to achieve a film thickness of ±5%, the temperature can vary by approximately ±7°C. This large of a process window explains why the C_{pk} of the pad ox is 2.2.

4.2.2.2 CVD

One of the most common types of CVD is the deposition of silicon that is transformed from silane gas. The net reaction that occurs is:

$$SiH_4 \xrightarrow{k_s} Si + 2H_2 \qquad (4\text{-}7)$$

Silane is broken down and transformed into silicon and hydrogen gas. The actual chemistry is not well understood, but it is believed that silane first breaks down into intermediate gases before the silicon is fully separated from the hydrogen. In this example, silane will be converted directly into silicon.

The thickness of a deposited layer will be dependent upon the reaction rate. The rate of the reaction, $\Re$, in terms of both the first order reaction rate, k, and the surface reaction rate, k_s, may be expressed as follows:[7]

$$\Re = kP_{SiH_4} = k_s C_{SiH_4} \tag{4-8}$$

where:[8]

$$k = 9.6 \text{ x } 10^4 \exp(-24000/T) \text{ cm/(s} \cdot \text{torr)} \tag{4-9}$$

Thus, the change in thickness as a function of pressure is linear. However, the change in thickness with respect to temperature is more complex and may be expressed as:

$$\frac{\partial \Re}{\partial T} = \frac{24000}{T^2} \cdot 9.6 \text{ x } 10^4 \exp\left(\frac{-24000}{T}\right) \text{ cm/(s} \cdot \text{torr)}$$

$$= \frac{2.3 \text{ x } 10^9}{T^2} \cdot \exp\left(\frac{-24000}{T}\right) \text{ cm/(s} \cdot \text{torr)} \tag{4-10}$$

For a deposition temperature of 625°C, the variation in the growth rate is 3%/°C. Therefore, in order to maintain a wafer-to-wafer thickness variation of ±5%, the temperature can only vary ±1.7°C.

In order to quantify the impact of temperature and pressure on within wafer uniformity, the parameter ϕ from Equation 3-10 is needed. Once a process has been defined, the only variables that may change during processing are k_s and D.

From Equation 4-8, k_s can be solved for as follows:

$$k_s = \frac{kP}{C} \tag{4-11}$$

$$C = \frac{P}{760 R_G T} \ (\text{mol/cm}^3) \qquad \text{where } P \text{ is in torr} \tag{4-12}$$

$$k_s = 760 k R_G T \ (\text{cm}^4/\text{s} \cdot \text{mol}) \tag{4-13}$$

The surface reaction rate for silane,

$$k_s = 760 \cdot 9.6 \text{ x } 10^4 \exp\left(\frac{-24000}{T}\right) \cdot 82 \cdot T \frac{\text{cm}^4}{\text{s} \cdot \text{mol}} \cdot \frac{2.3 \text{ g}}{\text{cm}^3} \cdot \frac{\text{mol}}{28 \text{ g}} \tag{4-14}$$

$$= 4.9 \text{ x } 10^8 \cdot T \cdot \exp\left(\frac{-24000}{T}\right) \ (\text{cm/s}) \tag{4-15}$$

By substituting Equation 3-11 and Equation 4-15 into Equation 3-10, ϕ^2 may be represented as follows:

$$\phi^2 = \frac{2 R_w^2 \cdot 4.9 \text{ x } 10^8 \cdot T \cdot \exp\left(\frac{-24000}{T}\right) \cdot P \sigma_{AB}^2 \Omega_D}{\Delta \cdot 1.41 \cdot T^{3/2} \sqrt{\frac{1}{M_A} + \frac{1}{M_B}}}$$

$$= \frac{2 R_w^2 \cdot 4.9 \text{ x } 10^8 \cdot \sigma_{AB}^2 \cdot \Omega_D}{\Delta \cdot 1.41 \sqrt{\frac{1}{M_A} + \frac{1}{M_B}}} \ T^{-1/2} \cdot \exp\left(\frac{-24000}{T}\right) \cdot P \tag{4-16}$$

In order to maintain a within wafer uniformity or 99%, ϕ^2 must be less than 0.08. With the parameters set as R_w = 10cm, σ_{AB} = 3.5Å, Ω_D = 0.8, D = 0.35 cm, MA = 32g, MB = 2g, T = 625°C, P = 0.1 torr, a uniformity of 99.7% is obtained. The temperature may rise up to 671°C, and the within wafer uniformity will still be greater than 99%. The pressure may rise up to 0.36 torr and the desired uniformity will still

be achieved. These calculations indicate that the growth rate, or wafer-to-wafer uniformity, is not as robust as the within wafer uniformity.

4.3 Reward/Risk Comparison

Now that the rewards and risks have been described, the key trade-off, assuming that the process is not constrained by theoretical tool capacity but by operational efficiency, is whether it is more costly to add more technicians or run the process in a non-gating manner. If more technicians are added, this will reduce the amount of time that a batch will wait to be unloaded by reducing the likelihood that a technician will be busy with another task. Consider the extreme where there is one technician per furnace, and there should be no WT2 time. On the other hand, if the process is run in a non-gating manner, then there is the risk of losing an additional batch of 125 wafers. The costs associated with each of these is described below.

Semiconductor manufacturing technicians are a scarce resource. They represent a skilled employee, and with the recent growth in the semiconductor industry, companies like Intel cannot hire enough skilled technicians to staff their fabs. Most technicians have a minimum of two years of college and some have four-year degrees. These technicians earn around $40,000 per year, which will translate into roughly $60,000 including benefits. Since the factory runs four shifts to achieve twenty-four hour, seven day per week coverage, the cost is multiplied by four. This cost of $240,000 will be called C_T

Calculating the cost of the risk of losing a batch is a little more difficult. The expected cost of losing an additional batch is the value of the batch multiplied by the probability of losing the batch. This probability is one minus the product of the probability that each step using non-gating CBP will be within the specification limits. This is represented by the formula below:

$$C_B = VK\left\{1 - \prod_{i=1}^{n}\left[2\Phi\left(3C_{pk_i}\right) - 1\right]\right\} \qquad (4\text{-}17)$$

where: C_B = Expected cost of losing a batch

V = Value of batch

K = number of batches run per year

n = index of process steps using non-gating CBP

C_{pki} = process capability of process step i

Therefore, non-gating CBP should be used for cases where $C_B < C_T$. An example of using non-gating CBP with process steps Q & S from Table 6 is shown below. Assumptions made are that the value of each wafer is $100,000 and that six batches can be run per day.

$$C_B = \$12{,}500{,}000 * 6 * 365 * \left\{1 - \left[2\Phi(3*2.5) - 1\right]\left[2\Phi(3*1.7) - 1\right]\right\}$$

C_B = $9,300

This is the cost of the risk for running non-gating CBP on only one furnace at full capacity for four-hour recipes. The cost will scale along with the furnace:technician ratio. A coverage ratio of 26:1 would be needed for C_B to be greater than C_T. In order for it to be cheaper to use technicians to solve a capacity problem, one technician would need to cover twenty-six furnaces.

4.4 Summary

The rewards of using non-gating continuous batch processing will vary depending upon the loading of the fab. If the diffusion functional area is running near capacity, then the use of non-gating CBP can quickly add more capacity to the area. However, if the diffusion furnaces are not the bottleneck, then there is little value in taking on the additional risk, no matter how small it may be. While the use of non-gating CBP will create cost savings and the expected value of using the strategy is positive, the incremental savings may not be compelling enough to take the additional risk.

Analyses were done to evaluate the rewards of using non-gating CBP at three different facilities. This analysis utilized the transaction data from the manufacturing information system to quantify three distinct periods of wasted time that could be saved when using CBP. It was noted that gating CBP can run at the same rate as non-gating CBP if the metrology could be completed before the furnace precharged.

The use of gating CBP creates no additional risk compared to not using CBP. However, the risk of using non-gating CBP creates additional risk over the use of gating CBP. If a batch is OOD, it is likely that the next batch will also be OOD, so two batches will be lost instead of one batch if the tool drifts too far. This risk was assessed by both statistical and engineering perspectives and is considered to be low for most process steps.

If the diffusion area starts to become capacity constrained, then a recommendation is to begin using non-gating CBP on certain process steps. The most benefit will come from using it on process step S, which runs on furnace type B. This furnace type requires both a particle and thickness monitor, and the process step precharges in the shortest amount of time. Because the process step has a high C_{pk} (2.5), there is little risk to the use of CBP, and there will be a measurable increase in the throughput of the tool. If more capacity is still needed, then additional steps should be run in a non-gating fashion. The order in which they should be run are by decreasing C_{pk} on furnace type "B" until a minimum threshold of C_{pk} is

reached that corresponds to the amount of risk production is willing to take. Once process steps on furnace type "B" are exhausted, then the same approach should be used to start running process steps on furnace type "A" in a non-gating manner. While the metrology can ideally be completed before the furnace completes precharging, there is little gap between the times.

This page intentionally left blank.

5. Process Changes at Intel

With multiple, worldwide sites, process management is crucial to manufacturing at Intel. Intel considers fabs that run the same process technology to be a "virtual factory." From a customer's perspective, the output from these factories needs to be indistinguishable from one another. In order to accomplish this end, Intel has developed a strategy called Copy Exactly (CE). The objective of this strategy is to facilitate the technology transfer, production ramp, and process management of its high volume semiconductor fabrication plants.

5.1 Overview of Copy Exactly!

The complexity of the manufacturing process drives the difficulty of the technology transfer and subsequent process management. To better manage this complexity, Copy Exactly aims to replicate the entire environment and production processes of the development fabs at the high volume fabs that run the same process technology. This includes everything from equipment and process recipes to the building itself and the lengths and bends in the pipes. The underlying philosophy is that:

- Identical systems operating with identical inputs will produce identical outputs, and
- Differing inputs with differing systems MAY allow matching outputs. As the number of inputs, number of outputs, and complexity of interactions increases, success becomes less likely.[9]

5.2 Why Copy Exactly?

As technology was transferred to new fabs, yields experienced what was called the "Intel U."[10] While the yield started out high at the development fab, as the process was transferred, yields would decrease at the

the new fab. Ultimately, the process would become stable, and its yields would match or exceed those at the development fab. However, precious time was lost while the new fab went down the learning curve. With process technology lifecycles of 2-2.5 years and product lifecycles of 3-4 years, time is a significant competitive advantage.

The "Intel U" was the result of the focus on matching outputs. During a technology transfer, the fab receiving the new process was allowed to choose its own equipment and processing techniques with the goal of matching a list of electrical and physical output parameters. Technology transfers were regarded as an opportunity to "engineer out" shortcomings in the process. While this sounds reasonable, the complexity of semiconductor processing makes this practice questionable. When process interactions are difficult to understand, then problem solving becomes more complicated. With multiple, identical environments, the root causes of problems are more easily isolated.

Although Copy Exactly was created to focus on the technology transfer and ramp up of new facilities, it is also used to manage the fabrication process throughout the virtual factory. Once the process has been transferred successfully, it is still kept in sync with the rest of the virtual factory. Any changes to the process must be implemented throughout the virtual factory. This is managed by the white paper process.

5.3 White Paper Process

The white paper process is the major mechanism used to maintain the CE system. Before any changes can be make to the process, a white paper must be created and approved. A white paper is a living document that serves as a template for process documentation. The completeness of these papers is rigidly enforced because the paper helps ensure a complete and comprehensive evaluation and implementation of any change. White papers also double as ISO documents, and over seventy-five papers are reviewed each week.

All white papers must follow a common format. The process is paper based, rather than electronically based, and a Microsoft Word template is provided to the engineers. This template ensures that engineers follow all of the necessary steps and consider all of the basic issues when proposing a change. There are several check box-like sections for general issues. There are also sections where owners are identified for different issues to ensure accountability. These include tool owners and support system owners. The format of the paper allows the review boards to quickly find the relevant information needed by that body. The major sections of a white paper include:

- Prose description of change
- Summary of the reasons for the change
- Various classifications that describe the risk level and extent of change
- A list of specific impacts that need to be considered such as environmental and safety issues
- A list of specific changes that need to be made to other documentation
- Other concerns and considerations
- Qualification plan
- Data summary
- Implementation plan
- Detailed data

In order to explain how the white paper process functions, the organizations that review these papers are first described. A process flow of how a paper is approved follows, and the role of the various organizations for each step is illustrated.

5.3.1 Organization

The process utilizes three main bodies to review any changes that are proposed. They are the Joint Engineering Team (JET), the Integration Joint Engineering Team (IJET), and the Process Change Control Board (PCCB). These groups meet each week.

5.3.1.1 JET

The JET is made up of at least one engineer from each site of the virtual factory in the same functional area. This team provides cross-site, functional expertise when evaluating proposals. Most of the technical issues are assessed within the JET as well as any cross-site issues. This group provides sponsorship for a white paper.

5.3.1.2 IJET

The IJET is a cross-site team of representatives from the Reliability and Integration organizations. This team provides a technical assessment of the proposal. Its primary focus is to assist in determining what data are needed to ensure that any changes will not adversely impact end-of-line yield and to review the data collected. The IJET is new to P856.

5.3.1.3 PCCB

The PCCB reviews and approves the qualification and implementation of changes within the virtual factory. It also serves as the repository for all of the documentation for process changes. The PCCB ensures any changes to the process have been properly evaluated and implemented. It does not evaluate the technical merit of the project and assumes that this has been done by JET and the IJET. The focus of the PCCB is more on the assessment of the risk and the qualification plan of the project.

5.3.2 Process

Each white paper begins with a proposal to the appropriate JET. Usually, these proposals are made by engineers from the JET's functional area. If the JET approves of the proposal, then a white paper is created. A white paper may pass through up to three stages: preliminary, pre pilot, and final.

5.3.2.1 Proposal

The first step in this process is to present the proposed change to the relevant JET. This is the first hurdle that the proposal must pass. If the JET approves, then it assigns a risk level to the project, with each level associated with a different approval process. The JET then determines the potential impact of the project such as output capability, defects, and throughput. The project is then placed on the JET roadmap that is essentially a prioritized portfolio of projects. The projects with the highest priority usually impact safety, output capability, and process capability.

For each white paper, the JET assesses the level of reliability risk as high, medium, or low. The levels of risk are described below:

Level	Description
High	May adversely impact reliability of the device. Change is contingent upon demonstrating no degradation of reliability by end-of-line testing.
Medium	May adversely impact end-of-line yield. End of line yield or electrical testing data is required.
Low	Neither reliability nor yield are likely to be adversely affected. In-line data is required.

The JET also determines a classification to assign to the white paper. The general classifications are: 1) Changes external to the processing environment or tightening of any specification, 2) Changes that affect the processing environment or relaxing of any specification, and 3) Changes to the structure of the device or those that require customer qualification.[11]:

Once a product is developed, few white papers fall into the third classification. Therefore, when assigning a classification to the white paper, the primary question is "Does this proposal change the wafer processing environment?" Any proposal in the first classification must be verified by a member of the PCCB. The difference in the review processes between first and second classifications is significant. White papers under the first classification do not require a preliminary white paper review by the PCCB or a review by IJET. The final white paper is ratified by the JET and then sent directly to the PCCB. After the white paper is classified by the JET, the originator is then charged with writing the preliminary white paper.

5.3.2.2 Preliminary White Paper

Once the project is placed on the roadmap, the change initiator writes a preliminary white paper. This paper includes everything but the data. Qualification and implementation plans are developed which outline the data collection methodology and the success criteria for the experiment. The paper also explains the risk and potential impact of the change. If the paper is a Class 1, then it bypasses the IJET and goes directly to the PCCB.

Otherwise, the paper is presented to the IJET. With its cross-functional and statistical expertise, the IJET primarily focuses on the potential concerns of the proposal and the qualification plan. Once approved by the IJET, the preliminary white paper is presented to the PCCB.

The primary concerns of the PCCB are the implementation plan and the areas of concern. When presenting the white paper, the originator should be able to address any of the concerns that are documented in the paper. Once the PCCB is satisfied with the proposal, then experimentation can begin.

The originator of the white paper then starts the experiment and collects the data according to the plan outlined in the white paper. Once the experiment is complete, then the originator of the white paper summarizes the data and presents it to the JET. The JET then reviews the data and determines if it will continue to sponsor the paper. No paper can go to the PCCB without JET sponsorship.

5.3.2.3 Pre-pilot White Paper

If the qualification plan requires a production pilot, then a pre-pilot white paper needs to be created. In addition to the data collected from the preliminary white paper, additional data needs to be provided to demonstrate the feasibility of the project and justify why a block of production material should be committed to the experiment. Material processed during the pilot will continue to be converted to finished goods and shipped to the customer. The paper is reviewed by the IJET and PCCB in a manner similar to that of a preliminary white paper.

Upon approval from the PCCB, the pilot may be run. These pilots tend to last around thirty days so that enough data can be gathered. Once the data has been collected and analyzed, the paper is reviewed once again by the JET. Regardless of the outcome of the pilot, a final white paper must follow all pre-pilot white papers. If the results of the pilot indicate that the proposal should not be implemented, then the final white paper should document the reasons why. If the data gathered during the pilot satisfies the success criteria and the JET decides to continue its sponsorship of the paper, then the final white paper will proceed onto its last reviews by the IJET and the PCCB.

5.3.2.4 Final White Paper

The final white paper includes all of the qualification plans and data collected on the proposal. It is the cumulative history of activities performed on the proposal. Prior to the presentation of the final white paper to the IJET, it will scrutinize the data, the adherence to the success criteria, and the implementation plan. When the originator presents the paper to the IJET, the presentation should focus on how the qualification plan was carried out and any potential concerns.

Prior to the PCCB meeting, PCCB members review the white papers for completeness. The originator will then present the final white paper and should focus on the implementation plan and the timelines for it. The PCCB will also raise any concerns that it has about the proposal, and then provide guidance on any other tasks the need to be done before the change is implemented.

Once the final white paper has been approved by the PCCB, it will be assigned a number for archival purposes. The PCCB may also assign some additional tasks that need to be done before submitting the paper to the archives. As long as all of the action items given to the originator by the PCCB are completed, then no further review will be necessary. Completion of the items is documented in the final white paper, and then an electronic copy is sent to the PCCB.

5.4 Application of the White Paper Process

The following section documents the history of the white paper titled "Continuous Batch Processing for Vertical Diffusion Furnaces in P856" through the white paper process. This proposal was a work rule change and not a process change, so a production pilot was not necessary. Therefore, the paper could pass through each body once as a final white paper.

5.4.1 Proposal

Once the data was gathered, the results were presented at the Diffusion JET. Presentations at Intel are done using the direct method. The conclusions are presented first, and then the data and observations leading to the conclusions are discussed. The team at D2 had a bias for non-gating CBP because there was little risk, and this group of engineers had used non-gating in the past. However, there was resistance from the other fabs to using non-gating because the risk was perceived to be too high, and the other fabs, particularly F11, had little to gain from taking on the additional risk. Ultimately, the JET decided to sponsor the gating option.

The JET wanted to place this paper in the first classification, even though the wafer environment changed somewhat since the wafers were taken out of the boat boxes while they queued in the buffer racks. However, other fabs were already using the buffer rack, and they have not reported any problems. The buffer rack is high in the furnace, sealed from the bay, and has filtered air passing through the tool, so it is unlikely that there would be any particle problems. Since this risk was perceived to be very low, and the process would take another week if the proposal was placed in the second classification, the decision was made to call it a Class 1.

When a PCCB member was asked to confirm that this white paper was a Class 1, he raised the issue that the environment does indeed change. He would not allow this to be considered a Class 1 unless there was some data to confirm that the wafers did not accumulate more particles while queuing in the buffer rack as opposed to queuing in a boat box. He admitted that the likelihood was low that particles presented a problem and indicated that a gross reality check would be sufficient if no data could be found. Surprisingly, although the buffer racks were used for two process generations, there was no data in any of the white papers to prove that particles were not an issue.

An experiment was set up to place a wafer in a cassette in each space of the buffer rack. A particle count was taken for each wafer before the experiment and then a set of cassettes was loaded into the furnace. A control set of wafers was also premeasured for particles and then the wafers were left in a WIP rack, on the floor and outside of the stocker, for the same amount of time. Particle counts were taken again for each set of wafers, and while the wafers in the buffer had accumulated more particles than the wafers in the WIP rack, the difference was not statistically significant.

5.4.2 IJET

The IJET meeting for this paper was a formality. The issue with particles was clear from the paper, and the data set collected during the experiment was included. The team receives the papers ahead of time, and the data met the success criteria that the particle count could not be significantly higher in the statistical sense (disproving $\mu_{wip} \leq \mu_{buffer}$). Therefore, the only discussion was about how the experiment was conducted.

5.4.3 PCCB

After the IJET meeting, one of the members of the JET brought up the point that there was the possibility that there could be some oxide growth on the wafers in the buffer rack. After all, the buffer is in close proximity to the furnace, and the elevated temperature could provide the driving force for oxide growth. The JET considered this to be a legitimate concern and another experiment was conducted before the paper was submitted to the PCCB.

To address this issue, a test lot of twenty-five wafers was split, and the oxide growth on thirteen wafers placed in the buffer rack during a testfire was compared to the growth on twelve wafers that remained in the boat box. This experiment indicated that there was no growth.

After these results were added to the white paper, it was submitted to the PCCB. For PCCB meetings, an agenda is handed out before the meeting with a five minute slot assigned for each paper. The presenter only needs to be there to present his paper. The presentation was brief. The issue of oxide growth was raised by the presenter and then addressed with the data. Then the tasks of the implementation plan were discussed, and the current status of the action items on the plan was addressed. Once the presentation was over, a horizon number was assigned to the paper, and no additional action items were assigned.

5.5 Implementation

The main tasks for implementation of gating CBP were testing the functionality of the buffer racks, making changes to the information system and station controllers, and training the technicians on the new loading procedure.

The buffer racks in the furnaces had not been used for quite some time, so each needed to be tested to ensure that they loaded and unloaded the wafer cassettes properly. As stated earlier, D2 was the last facility to implement the new capacity management system, and this needed to be completed because the transactions used by the new system were integral to the new process. In order to ensure that the tool would not push the quartz boat up into the furnace before the metrology was completed, the station controller would check for the END METRO transaction once the END PRECHG transaction was posted. If it was, then the tool would automatically push the boat into the furnace. The END METRO is posted once the metrology has been completed on the lot. The END PRECHG transaction indicates that the furnace has completed precharging and is ready to receive the boat. If the END METRO was not posted, then the tool would hold the boat. Therefore, if the metrology was completed before the furnace completed precharging, the process was automatic. If the furnace completed precharging first, then a technician would have to manually direct the tool to push the boat into the furnace. These tasks were completed by the Automation group.

After the automation was completed, the technicians on all four shifts were trained on the new process over the course of four days during the evening shift change. Since the factory runs twenty-four hours a day, seven days a week, one shift needs to come in early or another shift stays late when training is done. Dummy lots were set up in the system so that the cassettes could be introduced into the furnace and a dummy recipe could be run. A two-page handout was written to summarize the changes to the process. Training was timed such that an existing production batch would finish processing during the training.

The training attempted to simulate the loading of a production batch. A batch was physically loaded into the furnace so that the technicians could see how the station controller represented a buffer batch on the computer screen, and where the buffer batch was placed in the tool. When the production batch was complete, the manner in which the tool swaps the batches was observed. The steps that needed to be taken in case of an OOC were discussed. The main difference is that with CBP, the next batch is already loaded onto the quartz boat, and these wafers may need to be unloaded. Finally, the point that the tool would automatically push the batch into the furnace if the metrology was completed by the time the furnace finished precharging was reiterated. The amount of time that the furnace needed to precharge for the various process types was written on the training documentation. Fortunately, the work force at D2 is fairly senior, so there were several technicians on each shift that had used CBP in the past, so the implementation went smoothly.

5.6 Summary

Change within Intel's Copy Exactly! manufacturing strategy is managed by the white paper process. This process ensures a complete and comprehensive examination of all proposals before they are implemented in across the virtual factory. The white paper process utilizes three main bodies. The Joint Engineering Team (JET) provides cross-site, functional expertise. The Integration Joint Engineering Team (IJET) contributes cross-functional and statistical expertise. The Process Change Control Board (PCCB) ensures

that the all white papers follow the process and also provides assistance with implementation. All three bodies are involved with the three stages of a white paper: preliminary, pre-production, and final. The criteria for a successful proposal are largely quantitative, and the success criteria are explicitly listed in the white paper before any experimentation starts. The framework and organization of the process enable the company to provide a thorough evaluation of each proposal once it enters the white paper process. The roles of the bodies are well defined, and the structure of the white paper enable them to quickly review the portions of the proposal that are of interest to them. This provides for a quick review when the paper is presented before the respective bodies.

An example of a white paper going through the process was provided. This case demonstrates how the checks and balances of the process work. When an issue is raised, if there is not quantitative proof that addresses the issue and a simple experiment will provide the answer, then the experiment is done. This also shows how the process tends to be conservative. Once the process was started, the paper passed through smoothly. Most of the discussion about the proposal was generated during the JET. The IJET and PCCB meetings were short and direct; they covered the areas of the proposal that are outlined in the white paper process and no more.

While the White Paper Process provides an excellent mechanism to capture the learnings of the organization, it does not afford the organization with the tools to easily access these learnings. During the author's research at Intel, there were no on-line tools to research what work had been done in the past. Copies of older white papers were obtained from other process engineers who happened to have a copy. A search through the archives may have uncovered a white paper where the particle test had already been conducted, and then implementation could have happened a week earlier. After this research had concluded, the author has learned that there is an initiative at Intel to provide access to the repository of white papers in a web-based tool.

References

1. Castellano, Robert N., *Semiconductor Device Processing: Technology Trends in the VLSI Era*, Gordon and Breach Science Publishers, Philadelphia, Pennsylvania, 1993.

2. Hitchman, Michael L., and Jensen, Klavs F., *Chemical Vapor Deposition — Principles and Applications*, Academic Press, 1993.

3. Intel Corp., *P854/P856 PCCB Handbook*, October 6, 1997.

4. Middleman, Stanley, and Hochberg, Arthur K., *Process Engineering Analysis in Semiconductor Device Fabrication*, McGraw-Hill, New York, 1993

5. Ruska, W. Scott, *Copy Exactly: Three Generations of Process Transfer, Intel College of Manufacturing*, October 31, 1994.

6. Wolf, Stanley, and Tauber, Richard N., *Silicon Processing for the VLSI Era — Volume 1: Process Technology*, Lattice Press, California, 1986.

Endnotes

[1] Wolf and Tauber, p.200

[2] Castellano, pp. 192-193.

[3] Wolf and Tauber, p. 305.

[4] Middleman and Hochberg, p.264.

[5] Middleman and Hochberg, p.270.

[6] Hitchman and Jensen, pp. 165-168.

[7] Middleman and Hochberg, p. 523.

[8] Model fit to data from Hitchman and Jensen, p.185

[9] Ruska.

[10] Ruska.

[11] P854/P856 PCCB Handbook, pp19-20, 34-35.