



WILLY SHIH
GIOVANNI CARRARO

Low-*k* Dielectrics at IBM

Shortly after taking over semiconductor development for IBM, I had a visit from a guy who wrote for one of the big semiconductor journals ... he had been dying to interview us because we have been struggling with SiLK. Everyone knew the reason he wanted to interview us was that he realized we were struggling, and he figured he would catch us with our pants down for having failed. So he came in and he said, "Is it true that SiLK is not meeting its specs?" and I said, "Yes." He was quite surprised. He asked, "Well then what is ... what are you going to do?" I said, "Look, years ago we invented SiCOH, an alternative material, that's one of ours. We are just going to run with that. We already have people moved over and they are running 24 hour shifts! And so he stopped for a moment, and he said, "Well how do you explain the fact that IBM had decided to use SiLK, and failed to deliver it?" I said, "We had early, imperfect, and incomplete data, so we went forward with what looked at that time to be the better bet, and we were wrong." And he said, "Well, but how do you explain that you were wrong?" I said, "It's called taking a risk. You show me somebody who has never failed and I'll show you someone who will never lead!"

— Bernard Meyerson, IBM Fellow, CTO, and VP of Strategic Alliances

As Bernard ("Bernie") Meyerson leafed through the budget proposal for semiconductor research at IBM's T.J. Watson Research Center, probing questions from an HBS student who was soon to become an IBM employee caused him to think back about the company's experiences with one of its past development programs. The company's experience with low-*k* dielectrics had certainly caused controversy in the press.

SiLK and SiCOH are two chemicals that are used as dielectrics in the manufacture of semiconductors. SiLK (which stands for "Silicon Low-K") is Dow Chemical's trade name for its advanced porous semiconductor dielectric resin that is chemically a polyarylene ether (referred to as PAE). SiCOH (pronounced "psycho") is an organosilicate glass; so nicknamed by IBM researchers as the acronym for its constituent elements. Dielectrics are insulators used in the manufacture of semiconductors. Among other things, they insulate the wiring that interconnects the hundreds of millions of transistors in a modern integrated circuit (IC), and are vital to their reliable functioning. Different dielectric materials are characterized by their relative dielectric constant *k*, which is a measure of a material's ability to store electrical charge. Silicon dioxide (SiO₂), which is the major component of beach sand, was the traditional dielectric used in ICs because it had good insulating properties and the version used for wiring had a *k* = 4.1. But this *k* made it unfit for high-frequency signaling in new faster ultra large scale ICs, as it meant a fair amount of electric charge was stored in

Professor Willy Shih and Giovanni Carraro (MBA 2009) prepared this case. HBS cases are developed solely as the basis for class discussion. Cases are not intended to serve as endorsements, sources of primary data, or illustrations of effective or ineffective management.

Copyright © 2009 President and Fellows of Harvard College. To order copies or request permission to reproduce materials, call 1-800-545-7685, write Harvard Business School Publishing, Boston, MA 02163, or go to www.hbsp.harvard.edu/educators. This publication may not be digitized, photocopied, or otherwise reproduced, posted, or transmitted, without the permission of Harvard Business School.

the dielectric, and this “capacitance” would slow down the signals flowing along the wiring.¹ Scientists had estimated that k needed to be reduced to 3.0 or below for next generation ICs, and ultimately to 2.2 or below.

The company initially bet on PAE for its 130 nm chip generation,² confident of its success and its perceived superior performance, even as other major players in the industry began to pursue other strategies. But after several years of development, IBM’s PAE process encountered repeated reliability problems, causing major delays of key product programs, forcing IBM to move back to an earlier higher- k glass material for the 130 nm generation, and migrating to lower- k SiCOH for its 90 nm generation.

Now with a deep recession unfolding, there was intense pressure for a conservative spending target. Globally almost every semiconductor fabricator had excess capacity and was losing money. While the technical staff was always seeking more investment, IBM’s finance organization (and Wall Street analysts) questioned the timeliness of continued heavy R&D spending. But as the company planned to put its 90 nm process into production, and R&D focused intensely on 65 nm and beyond, Meyerson had to think carefully about how many “options” he wanted to “buy.” He had to ask himself how many projects could he fund that were off the main product roadmap, projects that would create back-up options in case the main plan went awry:

You have to ask yourself in assessing what you fund, am I up against a technical challenge or a physical challenge? There is a big difference between the two. There was a time when really it was only a technology issue ... there was a very clear path from “A” to “B.” It was only a question of which path I chose and how much I wanted to spend. There were no physical limits. You’re going from, let’s say 1 μm to 0.8 μm , and both are well within the realm of the feasible, there were no quantum mechanical effects. It was purely a technical challenge. That drives a level of investment that is dramatically lower in terms of risk mitigation, because any one of the paths will work. You may not pick the best one, but if you throw enough people at it, it’s going to work.

So now the ones for which there are physical limits ... an example here are gate oxides that are only five atoms thick ... now you are talking about a physical limit where you are getting quantum mechanical transitions and different behaviors if you try to go thinner. Here you need dramatic levels of risk mitigation because the next spin of the crank will result in a nonphysical outcome, in other words a result where you already know it cannot work. Cannot! Not might not, cannot! That’s a big difference. Your mitigation strategy has to be enormous! You have no idea whether it’s going to be high- k that will work, whether you’re going to have to stay with the conventional oxide and find 14 other tricks like back biasing transistors to reduce leakage. It’s a very different scale of a problem, and this is what is happening now with

¹ An important consideration for the interconnect was the “propagation delay” or time for signals to traverse the wiring. The propagation delay is dependent on the overall resistance R and the overall capacitance C of the wiring, and to the characteristics of the receiving transistors. A high k dielectric would mean more C , and the “ RC ” coupling would lead to propagation delays and also worsened power dissipation.

² Chip generations are referred to by a number, often called a “node,” which is related to the smallest feature size in nanometers (one billionth of a meter, or nm) of transistors imprinted in the silicon. Nodes were a standardized way of describing the feature sizes of a particular generation of semiconductor technology. The International Technology Roadmap for Semiconductor (ITRS) refers to a technology node as an “attempt to provide a single, simple indicator of overall industry progress in integrated circuit (IC) feature scaling.” Traditionally each node has been defined with the smallest half-pitch of contacted metal lines on any product. Typical nodes referred to at the time of the case were 180 nm, 130 nm, 90 nm, 65 nm, 45 nm, and 32 nm.

each generation. More and more of the attributes of a device are reaching physical as opposed to technical limits.

As far as meeting the rising costs of developing future generations of process technology, IBM had already developed a large network of close collaborations with several industry players like Advanced Micro Devices (AMD, now GlobalFoundries) and STMicroelectronics, with whom it shared technologies, resources and R&D developments. How the company thought about buying options might be influenced by this growing collective effort as well.

Introduction to Semiconductor Manufacturing

A semiconductor IC is a complicated device. Its key components, resistors, transistors and capacitors, are built entirely on a single crystal of a semiconductor material, usually silicon, and contained within a single package. ICs are at the heart of the electronics revolution, and are used in a multitude of consumer and commercial products.

Semiconductor R&D was one of the key enablers of IBM's mainframe and server computer businesses. See **Exhibit 1** for a picture of an IBM z-Series mainframe computer and several of the key ICs that power it. These were, and remain, important revenue streams for the company, and their success was directly linked to the company's ability to implement innovative system architectures in the most advanced silicon IC process technologies, producing chips that would be reliable for tens of years to come, and that were manufacturable at a reasonable cost. The path to get to the next most advanced technology was increasingly uncertain, expensive, and filled with unknowns. Every year the company was faced with making key technology decisions with relatively little data. It had to bet on the instincts of its researchers.

ICs are typically manufactured by creating a multilevel pattern of miniaturized circuitry on a pure thin slice (known as a wafer) of material, which is usually silicon. A typical IC may range in dimensions from a few millimeters on a side, to as large as a quarter. Individual ICs, referred to as "die," are produced on circular wafers that are typically 300mm in diameter, and then are sawn apart before packaging (See **Exhibit 2** for a picture of a wafer and **Exhibit 3** for a picture of an individual die).

The wafer manufacturing process is normally divided into two major blocks. The "front end of the Line" (FEOL) is dedicated to the creation of all the transistors in the body of the wafer, and the "back end of the line" (BEOL) creates the metal traces, which connect all the transistors with each other as well as with the external world (input/output pins or "contacts").

The FEOL consists of a repeated sequence of steps that modifies the electrical properties of part of the wafer surface and grows new material above selected regions. The pattern is controlled by a mask, which acts like a photographic negative, which determines where a layer of photoresist will protect certain parts of the surface (**Exhibit 4**). After all active components have been created the second phase of manufacturing (BEOL) begins, during which metal "traces" are created to establish the connection pattern of the circuits. The traces consist of vertical segments called "vias," which connect horizontal "intermetal" layers of wiring. **Exhibit 5** shows a photomicrograph of a stack of copper traces that were manufactured in the BEOL. A typical IC the size of a child's thumbnail may have several kilometers length of traces connecting all the components.

Despite the complexity of all these individual steps, the cost to process a wafer is somewhat independent of what is actually patterned on it. Larger wafers, such as the 300 mm diameter ones in use at IBM cost roughly 40% more to process than the older 200 mm wafers, but the surface areas is

2.25X as great, which means that many more ICs per wafer. Reducing the size of the circuitry leads to a smaller “die” size, and hence more ICs per wafer. Smaller die sizes also correlate with higher production yields, since a given density of spot defects are less likely to be coincident with a smaller die. More good (and smaller) die per wafer translates into higher profit per wafer. These economies of scale have been the major driver of the semiconductor industry over the past four decades, and a major source of economic success for many companies. Manufacturers have focused continuously on reducing the die size to cut unit cost, famously paying homage to Moore’s Law.³

The Quest for a Low-*k* Dielectric

For the entire history of the semiconductor industry up until the early 1990s, IC manufacturers worked to increase the speed and performance of their products by squeezing more and more transistors into a single device. They did this by scaling down the physical dimensions of the transistors, as well as by pushing the transistors closer and closer together. As the sizes and distances were compressed, the “interconnect” network got denser and denser. Not only were there more transistors to connect, but they were packed in more tightly.

Decreased die sizes and the increased transistor density began exposing the physical limitations of the technology: high power density and multiple electromagnetic effects, which had been negligible at coarser dimensions, started to become significant. For example, aluminum had been the metal of choice for the wiring because it was easy to vacuum deposit, and because aluminum was compatible with silicon. But as the wires became smaller, the metal’s relatively low durability and susceptibility to electromigration⁴ meant it could not handle high power densities. Copper was a more conductive and far more durable electrical conductor, but new technologies and process innovations were required to overcome its incompatibility with silicon. More importantly, copper wires had the innate ability to support higher electric currents than aluminum without electromigration. This limitation for aluminum and its alloys meant that aluminum wiring could not be made much smaller.

IBM led the industry in switching from aluminum to copper wiring on chips. Company researchers had recognized the switch would be inevitable, and had worked on the technology for many years. IBM’s experience in IC packaging and electroplating techniques developed for the thin-film heads of disk drives ultimately enabled it to develop a “damascene” process for plating copper traces on ICs,⁵ and it led an industry-wide switch that started with PowerPC microprocessors that it manufactured for Apple and has been followed by every microprocessor it made since 1998.

³ Moore’s Law was an empirical observation made in 1965 by Gordon Moore, one of the founders of Intel Corporation, that the number of components on the most complex integrated circuit would double each year for the next ten years. In 1975, he observed with some amazement that he was right. He foresaw a change in rate, though, to a doubling every two years. The reason this law works is that as minimum feature size scales smaller and smaller in a linear fashion, the area that a single transistor occupies on the die scales with the 2nd power. Additionally, each transistor requires less power, and because the geometry is smaller, it switches faster. Using half the feature size means 4x the number of transistors in the same space. Intel’s first memory chip had 400 transistors. Intel’s Core 2 Duo, 37 years later, incorporates 291 million transistors with a 65 nanometer feature size.

⁴ Electromigration is a process in which electrons moving through the metal actually cause some of the metal atoms to physically shift, eventually resulting in a physical break in the conductor and an open circuit.

⁵ The damascene process takes its name from a process of producing metal inlays, used by artisans in Damascus and elsewhere. Traditional aluminum conductors could be made by vacuum depositing aluminum followed by patterning and plasma etching to produce the desired pattern, and then overfilling with insulator. But the lack of practical volatile forms of copper necessitated a process of forming trenches first, overfilling with copper by electrolytic plating, followed by polishing down with chemical mechanical polishing to the level of the first insulator, leaving wires inlaid within the insulator.

While an improvement over aluminum, copper traces alone were not sufficient for future generations. The dielectric that was used to insulate the metal traces also had an important role. The traditional insulator, SiO_2 , was ideal at coarser geometries. It was compatible with other layers of the IC, and it was easy to grow layers and etch patterns with standard semiconductor process tools. SiO_2 was used both in the FEOL as an insulator between the gate electrodes of transistors, and in the BEOL as an insulator for the wiring traces.

As feature sizes shrank, different objectives began to emerge. For the FEOL it would be better to have a high k to increase the transistor performance, while in the BEOL it was better to have a low k to minimize the “parasitic” effects of excessive capacitance or charge storage.

Low- k development at IBM

In the late 1980’s through early 1990’s, IBM’s BEOL research program targeted development of Cu wiring with a polyimide⁶ low- k organic polymer as the dielectric insulator. Polyimide was a well known insulator and was used in other parts of the company for high-performance packaging interconnects. This work culminated in a seminal demonstration of multilevel Cu wiring, published in 1993. To get there, IBM had to pioneer several breakthrough technologies, including the use of Cu electrolytic plating, tantalum (Ta) based diffusion barriers, and Cu and Ta chemical-mechanical polishing (CMP). Getting there also required out-of-box thinking on how to process pristine semiconductor wafers by dumping them in buckets of plating chemicals.

While this demonstration was a technology tour de force, it had some serious limitations. The researchers could only make simple copper wires, not some of the other things like capacitors, fuses, and interfaces to other parts of the chip. Also, the polyimide insulator was mechanically weak, and it did not look like the whole process met the reliability and cost (yield) requirements for IBM chips. On top of that, IBM’s mainframe business was in the throes of change, shifting to a transistor technology sharing far more processing methods used in more commodity-like microprocessors. At that time, the company stepped back to using the more reliable SiO_2 that was already in use.

In late 1993, in response to these challenges, Dan Edelstein, now an IBM Fellow, initiated an offshoot group to re-optimize and re-engineer the Cu process as a dual damascene process with SiO_2 dielectric, optimized for CMOS integrated circuits. As the work began, it became clear that Cu yields and reliability were superior to the polyimide case, and in fact directly comparable to yields of the far more mature leading-edge aluminum wiring process running at the time. Early Cu/ SiO_2 hardware provided the first proof that chips with Cu wiring could be as robust, under stress, as aluminum, without any of aluminum’s limitations. With these results in hand, IBM Microelectronics committed to its development and qualification for manufacturing for its 220 nm node. From then on, things progressed along a relatively normal development and qualification path, meeting the dates announced in late 1997, and with the first volume production ramp in mid-1998. One aspect that did stand out was a new and highly successful multi-site and multi-division team structure; this was continued for future generations as well.

In its first production generation, the fine-level copper wires were scaled down in thickness to offer 25% lower capacitance than comparable aluminum wiring, without the need for “risky” low- k dielectrics. The coarse upper-level copper wiring was scaled upwards in thickness and width, with taller vertical spacings, for much lower resistance than was possible with aluminum wiring. Thus this “hierarchical” Cu wiring scheme leveraged the copper’s better properties while still using the

⁶ Polyimide is an organic thermosetting polymer that has very good electrical insulating characteristic.

reliable SiO₂ insulator. This hierarchical layering is evident in the different wire dimensions shown in **Exhibit 5**.

The next Cu generation that was introduced used SiOF, a fluorosilicate glass with a slightly lower dielectric constant than SiO₂. Shifting from SiO₂ to SiOF was a very incremental approach; it required no new tooling and minimal changes in the production recipes. It was seemingly a “plug-in” change to implement, though it would only produce a relatively small reduction in *k* (from 4.1 to 3.8 at the time). Significant development work had to be done to overcome new reliability issues, but nevertheless, Cu/SiOF was qualified into production in IBM’s 180 nm CMOS node. To stay ahead of industry roadmaps, IBM would then have to target a “true low-*k*” (*k* ~ 3.0) for the subsequent 130 nm generation.

When the choice of what material to use for a next generation low-*k* dielectric came up, IBM researchers had several options, not all timed ideally as far as the overall process development program was concerned. One choice was a material known as “diamond-like carbon” (DLC). This material could naturally fit into existing process flows, but it turned out to have some problems and was ruled out.

The IBM team then made the decision to focus on the promising PAE organic polymer that was available commercially. This polymer had a low intrinsic dielectric constant (~ 2.7) as well as seemingly good electrical properties, and the thermal stability to sustain all the subsequent wafer fabrication steps without damage. In addition, the material came as a liquid, having a weight and viscosity that made it suitable for precise and convenient spin-on deposition (see **Exhibit 6**).⁷ Some initial positive results built confidence in its commercial feasibility, and in early 2000 IBM announced its intent to fabricate and sell integrated circuits using the new material.

Subsequent BEOL development followed the normal R&D project framework used within IBM. Resources and funding were allocated with a roadmap tying delivery to the 130 nm node. The program was staffed, and a full planning and review cycle were implemented. This Cu/low-*k* wiring scheme became part of the “Plan of Record” (POR).

While most of the IBM Microelectronics organization was focused on the low-*k* PAE, not everybody agreed on its potential. Low-*k* organic polymers are more permeable to process chemicals, among other shortcomings. One of the most serious problems is that they expand and contract with temperature at different rates than the Si substrate and Cu wires, which suggested potential mechanical problems. They also cannot be polished directly in the presence of copper, and so require a protective “hard mask” film of dielectric barrier material to be retained everywhere. Dan Edelstein, who had played a leading role in the initial copper project, recollected:

As the Cu/PAE data started to come out, I recognized some similar problems to what we had experienced with Cu/polyimide, and which we thought were high risk for a potential manufacturing process. Based on the apparent differences in Cu/SiO₂ vs. Cu/polymer interactions and behaviors, I thought we would be better off with a low-*k* organosilicate glass

⁷ Better known as spin coating, this technique consists of dispensing a liquid solution on the center of a spinning wafer and then allowing the fluid to spread evenly by centrifugal force. As the solvent evaporates, a thin layer of the solute is deposited on the surface. Then a high-temperature cure drives off solvents and polymerizes the precursor fragments. A certain film shrinkage and tensile stress increase comes concomitantly with this cure. Spin-coating thickness ranges are quite limited within spin-speed curves. Substantially thicker or thinner coating thicknesses require new chemical formulations with different solid contents. In contrast, CVD is quite scalable to thicker or thinner by scaling deposition time and chamber reconditioning, all for the same chemical precursor gases and chamber hardware. This ability to scale easily means potentially reaching lower values of dielectric constant than with doped glasses.

(SiCOH), which was just coming on the scene, and receiving attention by some other semiconductor companies.

By this time, Alfred Grill, a close research partner of Edelstein, had developed SiCOH, an improved low-*k* organosilicate glass. More suitable than SiOF because of the absence of fluorine,⁸ SiCOH used a chemical vapor deposition (CVD) process similar to the manufacturing process used for SiO₂ but yielded much lower dielectric values, thanks to the doping impurity carbon (see **Exhibit 7** for an explanation of CVD). As it could be polished directly without a higher-*k* hardmask, the overall effective dielectric constant could match the organic case. Edelstein recalled how IBM was willing to keep the door open to SiCOH, though as a backup for a limited time:

So I got together with Al and Bill Cote, IBM's damascene inventor, and we formed a fledgling group around mid 1999 and started working on Cu/SiCOH integration. We worked on it for a year, ran into some patterning problems, but by the end we had just a bit of data - enough to confirm in our minds that Cu/SiCOH fundamentally would be more reliable. Still, we didn't have enough data or progress to compel a decision to switch. In mid-2000, the Cu/SiCOH project was put on the shelf.

Over the following two years, IBM maintained its commitment to PAE even as competitors opted for SiCOH. As technical difficulties mounted, the IBM production site (fab) in East Fishkill, NY faced increasing pressure. But it was the failure of the organic polymer to pass a key qualification step that precipitated a major crisis. Failure at qualification meant that the process would be delayed until a resolution was found. That meant all the products waiting for the new process would similarly be delayed, a financial impact that would increase rapidly with time.

From a technology management perspective, the Fishkill team faced a dilemma. At the leading (bleeding) edge of technology, one has to make design choices based on the best information available at the time. One cannot always be right, what do you do when you are wrong? Edelstein recounted a phone call he received:

I got the call to come up to Fishkill and make SiCOH work; we had to do it. The 130 nm node and initial 90 nm products were already decommitted from *low-k* to SiOF glass, and *low-k* was pushed to a later 90 nm node. I was basically given a mandate from the highest levels, pledging complete support to develop and implement SiCOH. The first thing I did was to form a skunk works team for the most critical first phase, where all the fundamentals had to be defined for the technology, it had to be debugged, and put on the path to manufacturing qualification. We did what we needed technically to solve, even proactively, any known or potential reliability problems with SiCOH. We brought to bear all IBM's experience in copper integration and reliability, including learning from the other Cu/*low-k* schemes. We went after potential or perceived weaknesses with a vengeance, so that we could hopefully make up time and schedule in a snag-free qualification. Normally, hidden weaknesses show up at this later time and threaten the schedule. There was a critical business need to get this into a certain 90 nm chip product, and we made it. The entire span, from start at zero through final manufacturing qualification, was done in two years. It normally would have taken four!

⁸ Fluorine can react with parts of the chip and change its properties.

SiCOH Fast Track Development

Rather than use the existing organizational structure, Edelstein pulled together a “skunk works” team where he tapped into engineers and technicians that he knew. He would run cover for those individuals by being the interface to upper management. Recalled Edelstein:

I handpicked people to work with me, far fewer people than the norm. It took maybe a quarter to half of the usual headcount to get to the same progress. We forged very strong cross-boundary ties among different organizations; I personally worked very hard on that one, meeting, calling or driving from site to site to meet people in person and get them involved. As for copper and other programs, it was not new to have this type of team, but I think in this case we achieved an even higher level of interaction and quality of performance. We valued open and level communications, mutual trust, and underlying belief that this was necessary for the business. Having that mandate didn’t hurt either. Still, I believe the underlying reason for our success was physical – this dielectric material was simply more compatible with Cu for our particular application.

The SiCOH skunk works development team used a very different development process. It almost “self-assembled,” and it had a significant amount of autonomy, while supported by top management. This gave the team all the resources it needed and allowed it to act quickly and make technical decision early, without having to justify them extensively to the IBM scientific community, which normally might establish “checks-and-balances.”

The traditional boundaries between divisions and competencies were quickly overridden with all of the team’s scientists dedicated to the project, even if the issues at hand didn’t call upon specific competencies at the moment. Bringing together a group that normally would stay separate increased work efficiency, reduced communication issues, increased synergies and enabled the team to deliver better results.

The open communications also built trust and respect among members of the team, which resulted in an open approach towards problem solving and analyzing failures. The focus was on proactively solving any issues rather than strongly defining personal responsibilities. Issues were analyzed from multiple vantage points, bringing not only deep understanding of the problem but also broader consequences of any decisions. Each researcher was entrusted with a level of responsibility far beyond what they would normally receive in the hierarchical power structure of a normal R&D project.

The flat organization structure was remarkable for its absence of formal authority. No doubt Edelstein himself would be considered the leader of the project, but he never had explicit positional authority. He had a mandate from top management to make SiCOH work, but no one on the team directly reported to Edelstein. But as everyone across the company possessed a shared sense of urgency, Edelstein could easily pull in anyone he thought necessary. More than being the person in charge, Edelstein was therefore the hub of a network of people who actively collaborated in making SiCOH a commercially viable option.

After the first year, the project matured from the initial flat organization/skunk works mode to a more normal development structure, with a development manager, a larger team with more process specialists, and more strongly defined personal responsibilities or “modules.” Still, it was on a very fast track in terms of schedule, and the tight Research interlock remained in play. Edelstein remained on board as the technical lead, but passed the baton on program leadership responsibilities.

The quick positive outcome for SiCOH was also enabled by the preliminary work conducted years before, which gave the team a big jump-start. They could also draw upon Research Division scientists to rapidly develop new reliability physics formalisms that were vital to its success, as well as the wide base of scientific knowledge embedded in the company, built through years of broad scientific research. Recounted Edelstein:

We got more progress with fewer people, but in fairness, this was a relatively incremental project. We already had some degree of Cu wiring development at the 90 nm node, it had been qualified already in SiOF; in other words, we were incrementally changing the insulator, but not the fabrication dimensions. In fact the *type* of insulator was still the same – a silicate glass. We were also able to build on or jump off from new low-k learning from the organic PAE program, as well as their improvements to the BEOL development organizational structure. Even the “new” reliability physics that we brought in for SiCOH was really based on well-known material science, such as fracture mechanics, adhesion, thermo-mechanical phenomena, and so on. What was new was the comprehensive application of this field to Cu/SiCOH BEOL by our scientists. Sure we had to change a lot of the integration steps and so forth, but it wasn’t entirely a new BEOL technology. Let’s say it was half of a new technology in terms of number of steps we had to change, as a reasonable way to size the problem. This sort of suggests why we might have been able to get it done in half the time.

Finally, by having executed the SiCOH project two years earlier, and by keeping up with industry advancements since then despite that project being on the shelf, we were able to restart and hit the ground running. It shortened the path because both we and the industry had accumulated more expertise.

Looking Forward

As devices shrank, and we encountered more fundamental limits, ever more innovation was required to maintain progress. Increasingly complex innovation drives nearly exponential growth of development costs with time, and so too drives the cost of one’s missteps. As such, your investment in multiple options to provide process and material options also grows dramatically, because the probability of a single failed process increases at this same exponential rate. To support multiple options, a relatively modest effort within the IBM Watson Research Center in Yorktown has grown to our Albany site where there is an entire campus devoted to pushing the limits of silicon technology. The office building alone in Albany has more space than the entire program occupied when it was at Yorktown, with fabrication facilities up there having at least 10 to 100 times the level of investment made in Yorktown so as to enable us to mitigate risk ever more deeply than we had before.

—Bernie Meyerson

Understanding what went wrong and what enabled IBM to stage a successful comeback was important to Meyerson as he contemplated the budget proposal. (A timeline of all the development events is shown in **Exhibit 8**). As technology demands and the commensurate investments became more and more aggressive at advanced nodes, it became harder to make the right technical call. In retrospect, was PAE a bad call? Was it even possible to always make the right call? Meyerson reflected:

To progress, we had to make a decision, and we had decided to go with PAE. After the fact, whether it was an inherent property of the material, or our failure to find the right “recipe”, is almost irrelevant. What is relevant is that we went with *a* material, and it did not cause a hard shutdown (because of the structure of our organization) of the work in SiCOH. That work continued in the Research Division, only its transfer into product was shelved. The people

who had invented it continued to pursue its application as a back-burner exercise, so it continued to progress. When it became evident that despite the best efforts we had at the time PAE was not going to meet the specifications we needed, we were able to flip materials literally overnight, taking the people who have been following the “second-tier roadmap” and putting them into the position of leading the charge. We simply shut down what had been the front-up process roadmap and moved on.

In IBM’s culture, allowing second-tier projects to remain alive was as important as deciding which ones would be the front-runner. SiCOH had been in the second-tier, and had SiCOH been killed when IBM committed to PAE, the company would have not been able to recover quickly. In light of the unfolding economic downturn, how many of these contingency plans could the company afford going forward, and most importantly, which ones should be picked and which ones should be dropped? These second-tier projects were generally more pure science research than specific product development work and a path to monetization was often not clear.

Historically IBM’s approach towards R&D had been discipline-based, not just product-focused. The company continued to employ a large number of scientists, whose competence spanned many fields and whose expertise was highly regarded among the scientific community. These researchers almost had an academic environment, where they were encouraged to share and challenge ideas among each other. Collaboration was also fostered across boundaries both in terms of scientific fields as well as physical company boundaries, reaching out to company partners and the broad scientific community.

Edelstein compared other leading semiconductor companies to IBM in their approaches to research and development:

Their R&D center, if they have one, doesn’t have a strong base in fundamental science. They don’t have a *systems* view of the business, which defines the targets that our R&D has to point to. Even beyond that, the breadth of a combined semiconductor, system hardware and software business is unique. So for example, the bulk of a pure semiconductor company’s inventions might focus primarily on CMOS manufacturing. In order to supplement these basics, one company might fund a lot of research through universities, and another might spend its investment running huge number of wafers to solve technical problems in a brute force fashion. That may work in many instances, but not when up against fundamental physical limits. Our strategy is much more broadly based; we have synergies and collaborations that cross many boundaries, both vertical and horizontal, and even stretch across other companies in our alliances.

IBM had already developed substantial expertise in partnering with other companies. IBM’s “radical collaboration” model was a unique and successful business model in which the company shared resources and know-how with a select group of partners to overcome the challenges of an environment in which costs and risk were raising exponentially. Recounted Meyerson:

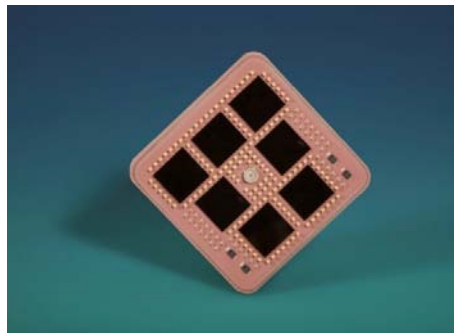
The art of the alliance game is to pool not only your financial resources, that’s just the price of entry. It is to similarly pool your intellectual capital. By pooling both your intellectual and physical capital, you effectively increase the number of options that you can pursue without breaking the bank. If you have 10 companies in an alliance, you not only can afford to maintain multiple development options until later in the development cycle, but you have the added intellectual capital required to ensure you create the very options required for such mitigation.

While effectively planning for contingencies was hard, making a decision on when to kill a non-performing first-tier project was equally tough. Any R&D project goes through up and down cycles and breakthrough events are generally unpredictable. Furthermore personal involvement of scientists and company politics can interfere with the objectiveness of the decision making process. A long history of success could lead to overconfidence, diminishing the objectiveness of internal judgment.

In the end Meyerson reflected that semiconductor processes had grown in complexity, cost and risk. Their uncertainty had grown exponentially, likewise the capital investment required. It seemed inevitable that in this environment failures and setbacks were a fact of life as by definition a leading edge decision is made in the absence of complete information.

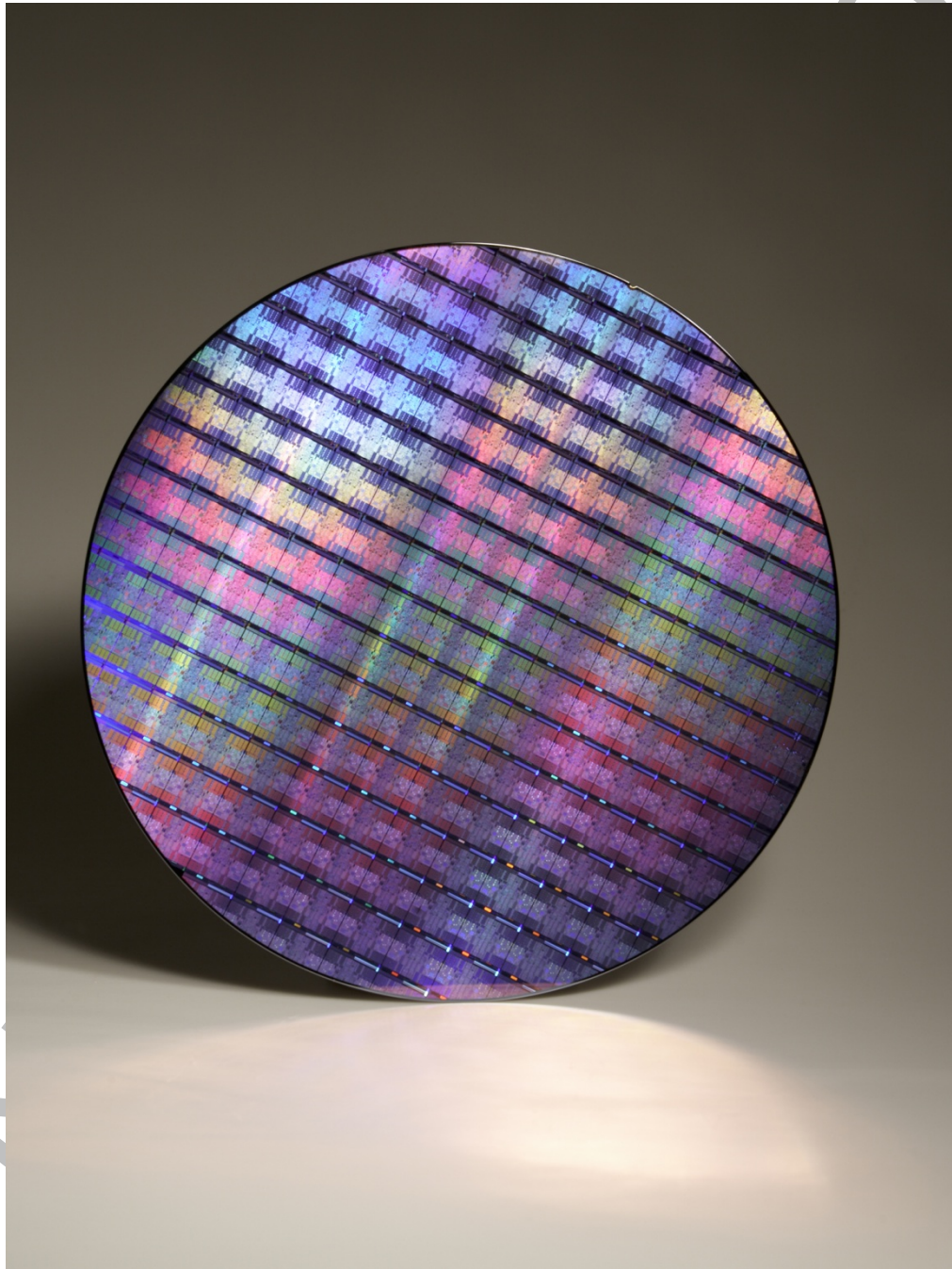
Making the wrong decision is still better than making no decision, which results in intellectual gridlock. There are people who find themselves, when faced with inadequate data, unable and/or unwilling to make a critical decision based on imperfect data. The organizations they lead are the ones that ultimately disappear into history. In business they fall by the wayside, becoming an anachronism very quickly, and they are gone. The people who survive have two fundamental traits. The first is that they are willing to accept the risk of making a leadership decision based upon imperfect data. A successful leader's other key trait is simply that they know when to punt. These are two leadership character traits that will greatly influence the success or failure of a business. Businesses must make crucial decisions on a regular basis where imperfect data means added risk, but unless you have a clairvoyant on staff, and they are in very short supply, you must accept such risk taking as inherent in maintaining leadership.

Exhibit 1 IBM z-Series Mainframe (top) and z10 ICs on a Multichip Module (bottom)



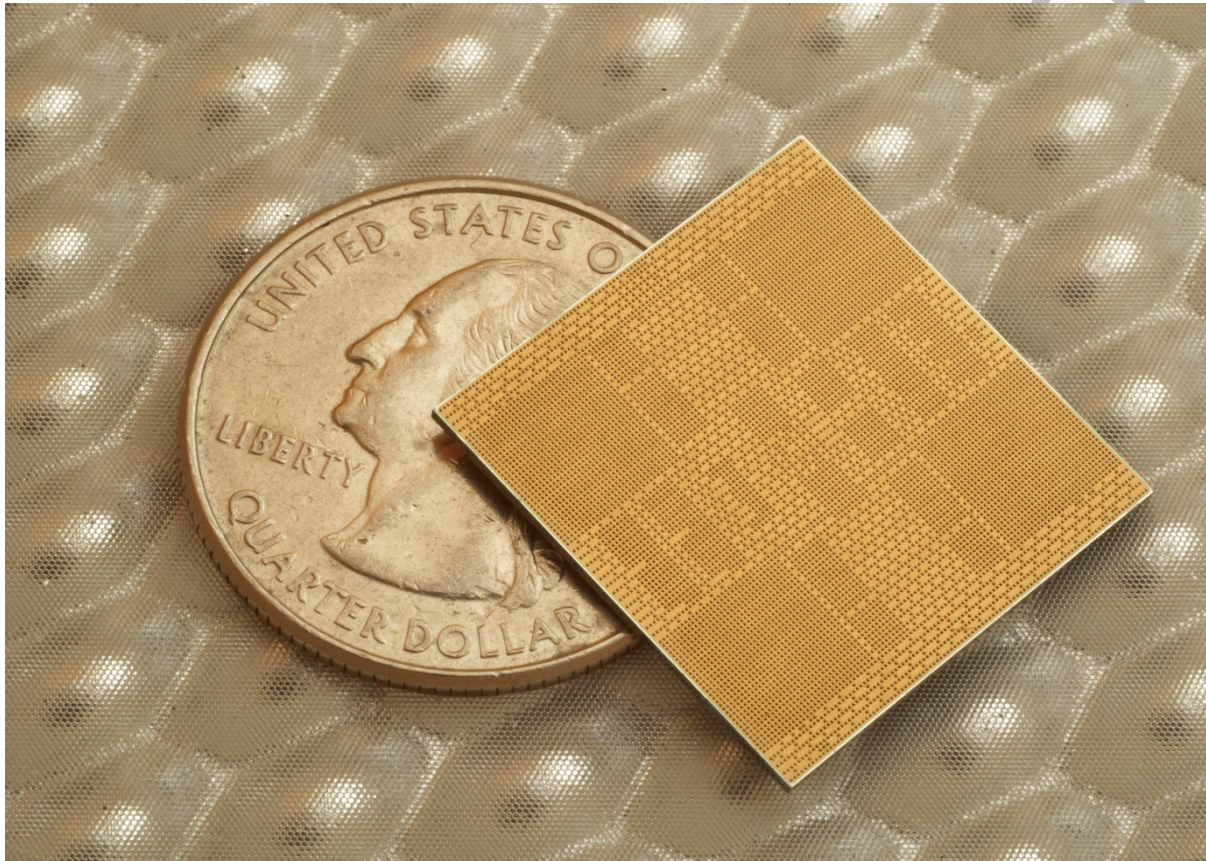
Source: IBM Corporation.

Exhibit 2 A 300 mm Wafer of IBM Power 6 Microprocessor “die”



Source: IBM Corporation.

Exhibit 3 IBM's z10 Mainframe Processor, containing 994 million transistors on a 454 mm² die



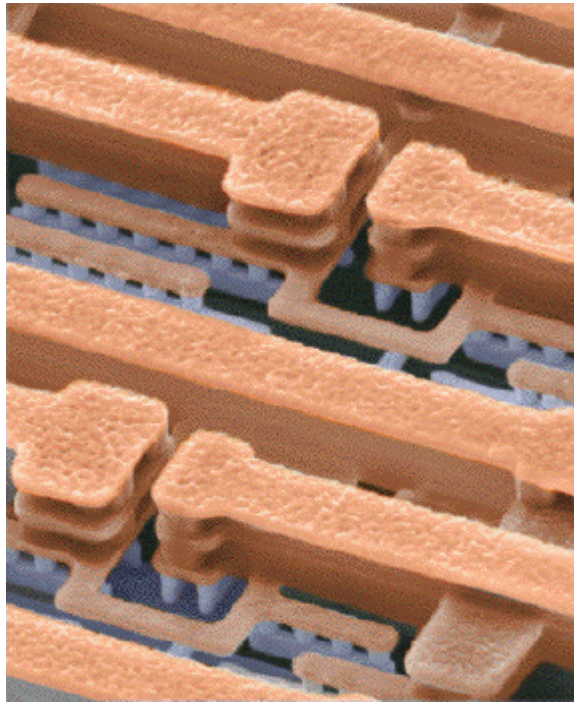
Source: IBM Corporation.

Exhibit 4 IBM Process Engineer Holding a Mask



Source: IBM Corporation.

Exhibit 5 Scanning Electron Micrograph of Copper BEOL Interconnect

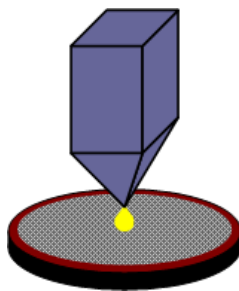


Source: IBM Corporation.

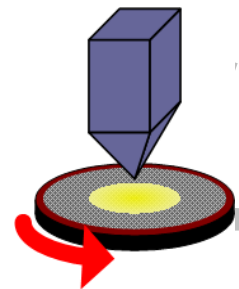
The horizontal copper conductors are known as “traces.” They connect to vertical post of copper called “vias” that together provide a three-dimensional network of conductors to interconnect transistors buried in the silicon.

Exhibit 6 Spin-Coating Process

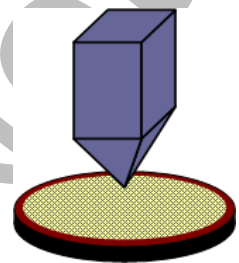
Spin-coating is a widely used process that consists of dispensing a liquid solution on the center of a spinning wafer and then allowing the fluid to spread evenly by centrifugal force. As the solvent evaporates, a thin layer of the solute is deposited on the surface. The thickness is dependent on the speed of rotation and the solution concentration. By not requiring a chemical reaction, spin coating allows the deposition of a greater variety of material, and therefore the ability to potentially reach lower values of dielectric constant, than with doped glasses. In addition controlling the solution concentration and the speed of rotation is simpler than controlling pressure, temperature and gas concentration in CVD. The spin coating process also does not require high temperatures, reducing the stress on the semiconductor wafer.

**Step 1**

A drop of solution containing the dielectric is laid on the center of the wafer.

**Step 2**

The wafer is spin at high speed; the centrifugal force spreads the solution evenly over the entire surface; the motion facilitates the solvent's evaporation, leaving a thin film of solute on the surface.

**Step 3**

Excess solution drops over the edges of the wafer and is removed.

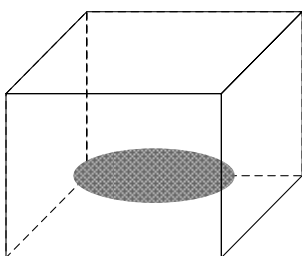
The wafer with a thin film of dielectric is ready for photolithography and etching.

Source: Prepared by case writer.

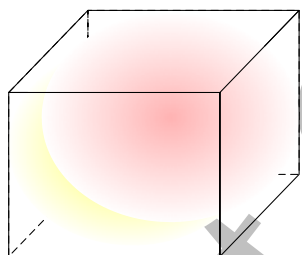
Exhibit 7 Chemical Vapor Deposition Process

Chemical Vapor Deposition (CVD), is a process for depositing a thin film of material on top of the surface of a wafer. The wafer is exposed to one or more gasses in a heated reaction chamber, and the gasses react on the substrate producing the desired thin film. This process is used to deposit SiO_2 by introducing silane (SiH_4) and oxygen (O_2) into the reaction chamber, producing an SiO_2 film and hydrogen (H_2) as byproduct. Glasses can be deposited using CVD processes by changing the composition of the reaction gasses, and lower-*k* dielectrics can be formed by doping SiO_2 with other atoms, like carbon or fluorine.

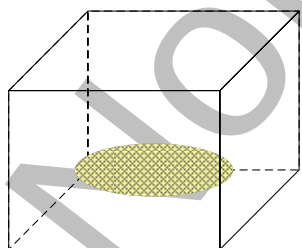
The wide spread use and familiarity with CVD process represents a major strength of this techniques. The ability to easily change gasses or the operating conditions (pressure, temperature) gives users a great deal of flexibility and capability.

**Step 1**

A wafer is inserted in a reaction chamber and the temperature is raised to $\sim 450^\circ\text{C}$

**Step 2**

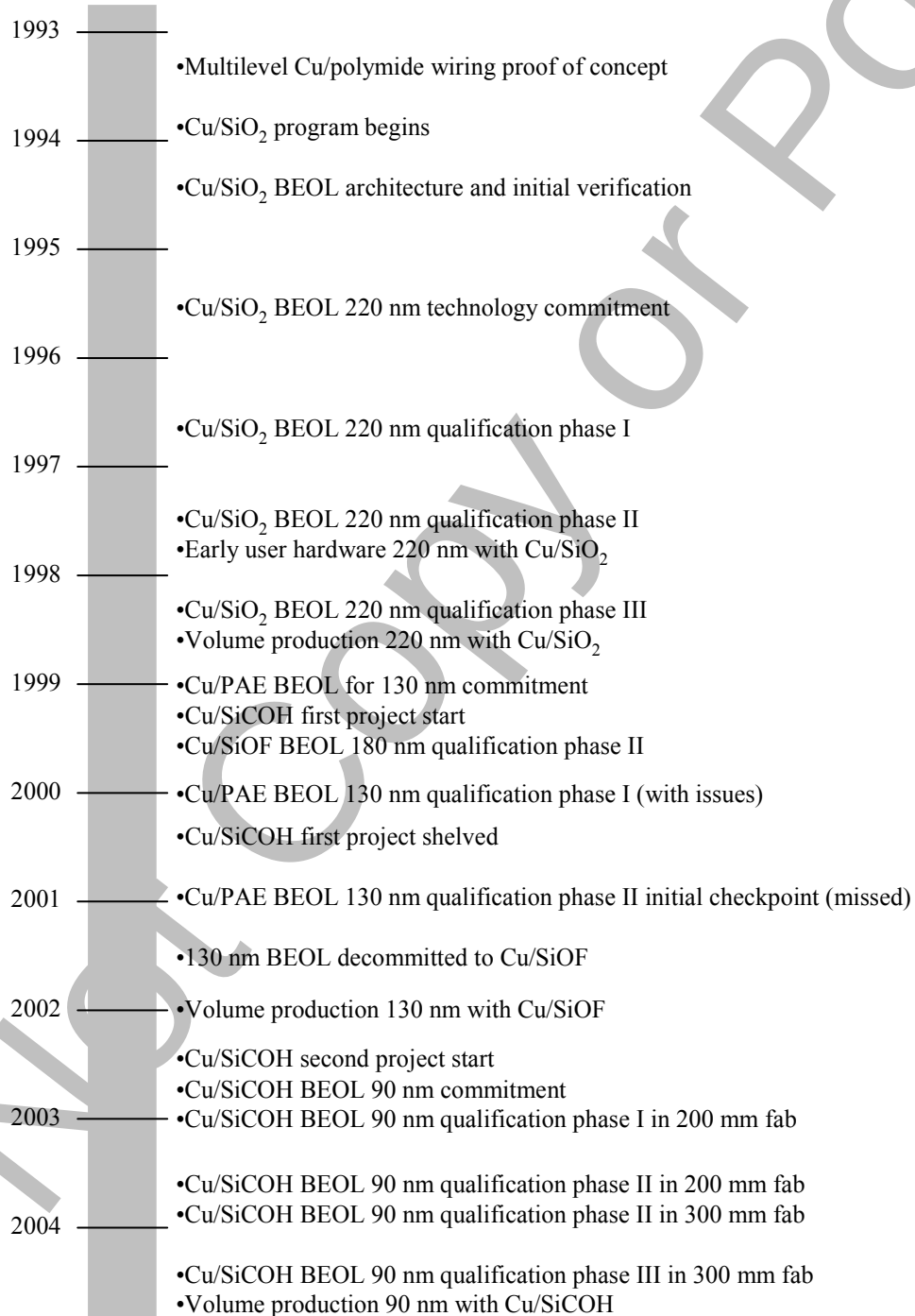
Process gasses, like silane (SiH_4) and Oxygen (O_2) are introduced into the reaction chamber; with the high temperature and pressure they combine to form an oxide on the wafer surface (SiO_2) and byproducts (H_2).

**Step 3**

Excess gasses (SiO_2 , O_2 , and H_2) are expelled and temperature is decreased.

The wafer with a thin film of dielectric is ready for photolithography and etching.

Source: Prepared by case writer.

Exhibit 8 IBM Low-*k* Dielectric Development Timeline

Source: Prepared by case writer.