Notes on Moore's Law
by Marcian E. "Ted" Hoff
February 6, 2016

I first learned of what we now call Moore's Law shortly after joining Intel Corporation in 1968. Gordon invited me to his office to show a chart in which he had plotted the progress made in integrated circuit fabrication since they were invented in 1959. The chart showed that the number of components was doubly almost every year. He noted that Intel's proposed products would advance that progress, possibly even increasing the growth rate. I understand he had published his observations a few years before while still at Fairchild Semiconductor, in an effort to predict where the technology would be a decade later.

Gordon had also prepared yield curves showing the number of working chips one could expect as a function of chip size (i.e. area) for different levels of manufacturing process performance (e.g. defect density). Chip manufacturing is actually a form of printing, in which circuits are printed on a silicon substrate. For a chip to work, it must be free of defects, or at least free of any defects large enough to impact any of the chip's circuitry. Printing of books and documents for humans to read could usually be done with a resolution of a few thousanths of an inch, and humans could tolerate a speck or a bit of missing ink. Integrated circuits required much finer resolution and were quite intolerant of printing defects. At the time I joined Intel, the minimum feature size was in the order of ten microns, or about 0.0004 inch--perhaps ten times finer than book printing.

A major source of defects was dirt or dust particles, floating in the air. There are many sources of such particles, clothing lint, dust disturbed from surfaces, particles from humans and their clothing, smoke particles, etc. The method used for the IC printing process was contact printing, i.e. masks printed on glass were carefully positioned (to align with prior steps) and placed in contact with the silicon wafer. Light passing through the mask would then expose photoresist on the wafer, which would then be developed and the wafer subject to a process step. Our silicon gate process typically needed four such steps. One might consider it somewhat equivalent to color printing of a book which often involved printing with four different inks.

The masks used in such production were called working plates, and would be used only for a few exposures, because they would pick up particles of photoresist from the wafers, and thus become damaged and a source of addtional defects.

The manufacturing process actually used a fixed size wafer of silicon--at Intel in 1968 a two-inch diameter wafer was used. The form that a chip would take was typically a small rectangle, perhaps a tenth of an inch on a side. Thus there could be perhaps several hundred chips printed on a single wafer. Testing would be done after the printing steps were completed, and non-functional chips marked. The wafer would then be broken up into individual chips, those with marks discarded. The unmarked chips would be packaged and then final testing of those packaged chips would select only those undamaged by the packaging process.

It was very important to chose an optimum chip size. Gordon's yield charts showed that doubling the size of a chip would almost (but not quite) result in a yield that was the product of the smaller chip yield. Consider a wafer with 200 possible chips, yielding ten per cent. One would expect to get twenty functional chips per wafer. Double the area of the chip, and there would be less than 100 possible chips (the wafer was round and the chips rectangular, so part of the periphery of the wafer is unusable, and that part grows with chip size). With the same process conditions, the double-area chip would yield closer to one per cent (ten per cent squared), so would produce about one functional chip per wafer. Thus doubling the functionality of a chip might increase the cost more than twenty times over. Making a chip too small had its own problems. The cost of packages and packaging steps was quite significant, so could dominate the overall cost of a system that required many small chips.

Over time, great progress was made in reducing the defect densities experienced in manufacturing, thus improving yields and allowing larger chips to be made. Manufacturing rooms were closed off from the rest of the factory, human operators in those rooms were required to wear special clothing that would not shed lint particles and would cover most of the operator, particularly the hair, and special air conditioning was installed, with carefully filtered air caused to flow very smoothly from the ceiling downward, so as not to disturb any dust in the room.

Consider that the design of a complex chip might take a year, and after it was offered for sale, it might take another year for the market for that chip to develop. The use of Moore's law to predict the ultimate manufacturability of a chip was very important, and those companies that could make accurate predictions had considerable market advantage.

In my own case, it was understanding and applying Moore's Law that guided how complex a processor might be produced in an effort to perform the Busicom calculator functions. If the processor were too aggressive, it might not have been manufacturable at an acceptable price. The 4004 processor was conceived using a guideline target in the order of two-thousand transistors.

Moore's Law played a role in subsequent microprocessors as well. The target specifications for the 8008 were written about a half year after those for the 4004, and it was understood that an increase in complexity in the order of 50 percent was acceptable.

Many of the improvements in IC manufacture came from cleaning up the process and other changes such as going from contact printing to projection printing. Around the middle of the 1970's another step began to play a role. As the printing process improved, it became print finer lines and smaller transistors, i.e. to make a chip of given functionality smaller in overall area. In 1974 a group from IBM, including Bob Dennard (who was recognized by the National Inventors Hall of Fame for the first single-transistor cell DRAM) published a paper in the IEEE Journal of Solid State Circuits, which noted advantages from making circuitry smaller. The paper noted that the current minimum

feature size was about 5 microns (i.e. 5,000 nanometers) and proposed feature size to be reduced to 1 micron, such that 25 times as many circuits could be made from the same amount of silicon, each circuit would consume 1/25 of the power of the original, and would operate five times faster. Thus a five times reduction in linear dimensions gives about 125 times the performance at hopefully about the same cost.

That reduction in feature size has helped keep Moore's Law relevant to this day. Minimum feature size is now in the order of ten nanometers, and efforts are underway to see if they can be reduced further.

Another effort to keep Moore's Law alive is in the area of three-dimensional ICs. Packaging still remains a part of the cost of the final circuit, and the ability to stack layers of functionality is one way to add more function to each final package.

I believe Gordon viewed his work as observation, but the acceptance of the concept has helped set targets for the technologists of the industry, and for fifty years, their progress in meeting those targets has ensured that Gordon's observation should be considered law.