SPECIAL REPORT

Gilder Technology Report

Inside:

Message from Korea

Try to start a business from a home in the Berkshires of western Massachusetts, as my colleague Charlie Burger is trying to do, and you'll soon tumble into the gap between bandwidth and connectivity. If you are a phone company, putting terabits per second through a fiber is practical, and measured by the bit is cheap. Connecting to the fiber is the rub. Charlie's home is just six miles from the Massachusetts Turnpike, along which run several fiber optic cables streaming terabits per second of potential bandwidth, but he must dribble his bits through a dialup modem. I live even closer to bandwidth Nivana, a fiber cable running down the valley along a Tennessee natural gas pipeline a quarter mile below my house. But the bits may as well be on the moon. In fact, I get my Internet service from 23,600 miles away, through an artificial satellite moon launched by **Hughes** (GMH) Directway. Most residential users in the U.S. are not so lucky and still connect to the Internet via a dial-up service such as **AOL** (AOL) or **Earthlink** (ELNK).

In the past—and still today in benighted circles in Washington's communications bureaucracies—this narrowband plight of Americans is taken for granted. The then Chairman of the President's Council of Economic Advisers pointed out to me a few months ago that cable modems and telco digital subscriber lines (DSL) are available to ninety percent of the population. But only 20 percent take it. The problem is demand. Most people are satisfied with their TVs. They take their broadband service downstream only, as God intended, mixed with a pulsating potpurri of edifying advertisements.

In the face of news from abroad American complacency cannot last. The inventor, incubator, popularizer, and financier of Internet technology the U.S. may be. But the U.S. is no longer anywhere near the lead in applying it. In the last three years Asia has swept massively ahead of the U.S. in broadband deployment and use.

CONTINUED ON PAGE 2

As Charlie regularly points out, the Internet and its traffic are non-linear. Business and investment life are non-linear. It takes little time to turn the world of technology upside down. It happened here. Including a near hundred fold burst of new email and browser traffic in 1995 and 1996, Internet use rocketed 9,000-fold in seven years by 2002. Carrier traffic changed from docile 64 kilobit streams of voice to bursty riots of data requiring at least six to one headroom to handle. From under one percent of total traffic in 1994, IP (Internet protocol) data soared to near 60 percent in 2002. Some 5000 new Internet Service Providers (ISPs) popped up and sought ways to peer with the backbone kings such as **AT&T** (T), **Sprint** (FON), and Worldcom, who in turn were contested by upstarts **Qwest** (Q), Global Crossing, IXC, and **Level 3** (LVLT)—all once imperial companies now either gone or gimpy.

The Korean example shows that when broadband connections are deployed, the Internet will undergo a new non-linear surge comparable to the hundredfold U.S. rocket of 1995 and 1996

Amid this eruption, optical equipment suppliers had to meet demand for capacity to handle hundreds of billions of emails, hundreds of millions of web pages, then billions of them, increasingly laden with pdfs, pics, gifs and jpegs, QuickTime and Real, Macromedia Flash and **Microsoft** (MSFT) media, then a surreptitious galore of peer-to-peer MP3 music files and MPEG2 films. The few experts who knew what was going on, such as Ethernet inventor Bob Metcalfe, believed that the Net would crash under the pressure. But despite many slow crashes and fast financial debacles, the industry rose to the challenge, providing for under half a trillion dollars an infrastructure that would have cost $39 trillion to build with the technology—all those **Lucent** (LU) 5ESS switches and **Nortel** (NT) Add/Drop Multiplexers (muxes)—on sale in 1995.

Some called this amazing achievement a bubble, citing as precursors tulip mania and South Seas panic. But it was more like 1929, a policy debacle in which protectionism, taxes, and deflation aborted a decade of overwhelming accomplishment. The Roaring Twenties propelled radios, automobiles, and telephone, oil and steel into mass markets, and laid the foundations for an ascendant America that could win World War II. The accomplishments of the 1990s were comparable.

Overseas revolution

After meeting the hundredfold Internet surge of '95 and '96, American communications technology managed to handle an ongoing ramp of traffic at a pace of near doubling every year through 2003. As traffic rose by a factor of 3,000 between 1996 and 2001, telecom revenues rose 50 percent. But bureaucrats in several Federal agencies and 50 states imposed a million word regulatory stranglehold on telecom and taxed the local loop as if they wanted to stop it in its tracks like tobacco. MIT economist Jerry Hausman estimates the average tax on wireless services at 18 percent. Because of the high elasticity of demand for cellphone use, so he calculates, these taxes costs the industry $3 in revenues for every $1 collected by the government. Added to these direct taxes were the punishing indirect taxes of an incredibly destructive spectrum auction process both in the U.S. and in Europe. Meanwhile, during the very period that the global telecom infrastructure was rebuilt for the Internet, monetary authorities imposed a deflationary chokehold on the dollar.

In a five-year deflation beginning in 1996, the dollar rose between 25 and 40 percent in value against other currencies, gold, and commodities. During the same five years, the U.S. telecom industry plunged hugely into debt to transform the global network infrastructure. The chief effect of deflation is to punish debtors, who have to pay back their creditors in more valuable dollars. With a total of around 800 billion dollars of debt, increasing in value to over a trillion 1996 dollars by 2001, the communications industry sank under the load.

Hearkening to the regulators and litigators and ascribing the mostly non-existent bubble to *in*flation and *easy* money the purblind media and politicians got almost everything wrong. A thousand bankruptcies in telecom? That was a product of accounting fraud and entrepreneurial crime. A paralysis in the local loop? That stemmed from the obstinate refusal of the Bell Operating companies to share their wires with rivals. The disappearance of thousands of dot.coms launched in the expectation of a broadband world? That reflected a lack of demand for broadband. A rapidly surging stock market? That resulted from inflationary monetary policy by Alan Greenspan, trying to assure the election of Republicans. Now, after the deflation is long over, Greenspan has begun warning about a deflationary spiral that already happened and the press fumes that the FCC is too deregulatory.

The real case was summed up by [Walter] Wriston's law, "Capital goes where it is welcomed and stays where it is well treated." While Washington raged at CEOs, concocted far-fetched indictments, pummeled telecom with new rules and taxes, and supplied cover for trial lawyers, the industry's advance did not stop. Internet deployment and use accelerated sharply. The politicians, regulators and trial lawyers simply drove the capital and technology of the Internet revolution overseas, from Silicon Valley to Korea and China.

The tweedledee dums at the FCC are still proud of their broadband policies, which are said to have sated Americans with broadband. "Fastest deployment of any consumer product," they crow. Yet Korea, a country of 48 million, with half of America's per capita wealth, commands at least twenty times more per capita bandwidth, both wired and wireless, than the U.S. does. American service providers charge around $40 a month for well under one megabit per second. The Koreans charge $25 for between five and eight megabits per second. For around $30 per month, they have also already linked more than a million households with VDSL (very-high-speed DSL) connections at 13 to 20 megabits per second and plan deployment of some two million links of 50 megabits per second in the next twelve months. Rapidly deploying **Qualcomm's** (QCOM) CDMA2000 (code division multiple access) and launching the 2.4-megabit-capacity EvDO (offering an average speed of 500 kilobits per second), Koreans have even supplied *wireless* bandwidth per capita comparable to U.S. wired connections. My answer to the skeptics at the FCC: If U.S. customers similarly enjoyed bandwidth at a price per bit some 20 times lower, there would be a broadband boom in the U.S. as well.

Starting in the local loop, the difference in bandwidth ramifies back through the network. U.S. telcos supply on average one broadband DSLAM (DSL access multiplexer) slot for every 35 customers and call it broadband. Korea has provisioned its local loops with one channel for every four customers. Most U.S. telcos supply backhaul from the local loop on T-3 lines of 45 megabits per second. Koreans provide optical carriers (OC-3) at 155 megabits per second, with many links at OC-12 (622 megabits per second).

Korean broadband explosion

While the U.S. has supplied a meager form of broadband to 20 million households (20 percent of the total), Korea has connected some 11 million households (73 percent of the Korean total) with real multi-megabit pipes. While the U.S. pretends that the Internet boom was a scam and a delusion, the Koreans now run a third of their economic transactions through the Net. They execute 70 percent of their stock trades online, half of all banking transactions, and constant retail orders around the clock for everything from groceries to furniture. While the U.S. depicts Internet commerce as mostly a mirage, Korea is living the reality.

The Koreans accomplished all this in just three years. With the adjustments needed in a poorer society, the Chinese have made similar gains and now lead the world in total cellphone use and are third in use of the Internet. While the U.S. communications industry remains mired in depression, the Korean and Chinese industries are thriving. Barron's warns against the overvaluation of **Samsung** (05930.KS), the Korean colossus that is selling at 13 times earnings and 7.3 times cash flow. The *Journal* dwells portentously on an Internet bubble among Chinese dot.coms that have quadrupled in value over the last year. But while the U.S. economy eeks forward, then slips back, the Korean and Chinese economies are growing some twice as fast. While the U.S. pretends to have a stock market resurgence—the figment of a commendably reflated dollar—Korea and China are undergoing real equity expansions. U.S. economists still fool themselves that they live in a national rather than a global economy. But when the U.S. stock market goes up 12 percent and the U.S. dollar goes down 20 percent, the real effect is sharply cheaper stocks, not more expensive ones.

Originating in the U.S. is nearly all the technology—the digital subscriber lines, the DSLAMs the cable modems, the optical carriers, the CDMA wireless systems, the chip designs that made Asian broadband possible. But the Koreans and Japanese are now rapidly taking over the industry and the Chinese are rushing up from behind.

While Qualcomm has broken through in the wireless markets in both Korea and China, all of the ten companies competing for VDSL contracts in Korea are Korean

The Korean companies in the forefront of this drive are Samsung, the leader, the rapidly privatizing **Korea Telecom**, **Hanaro Telecom** (HANA), and **SKT**, CDMA pioneer and the largest Korean wireless carrier. Combining leadership both in DSL, flat panel displays, microchip memories, and CDMA handset/cameras, Samsung represents a total play in Korean bandwidth. Hanaro is the hero of the Korean saga, entering the industry to push DSL prices well below cost three years ago and forcing KT to follow. As usual, throughout the history of business, lower prices brought higher revenues and ultimate profits. "The elasticity was far greater than we thought," comments a Korea Telecom strategist. KT is now making money on broadband. Close to break even, Hanaro is rushing ahead to VDSL. The Korean government is expected to permit **Lucky Goldstar** (LG) to combine with Hanaro to create a more robust competitor for KT. Most of these Korean companies offer more solid value than the China.coms that have recently experienced fourfold gains.

The second boom?

With traffic up close to a hundredfold in three years, the Korean example shows that when the new broadband connections are deployed, the Internet will undergo a new non-linear surge comparable to the hundredfold U.S. rocket of 1995 and 1996. Igniting the boom of the late nineties in communications gear, the U.S. upsurge came from a lower bandwidth base than the later Korean one. As countries around the globe begin imitating the Korean and Chinese models, American communications suppliers will gain a second chance for major growth. But it will not be easy. While Qualcomm has broken through in the wireless markets in both Korea and China, all of the ten companies competing for VDSL contracts in Korea are Korean. Led by Samsung, some are even competing for microchip slots with **Infineon** (IFX), **Analog Devices** (ADI), **Texas Instruments** (TXN), **Metalink** (MTLK) and **Ikonos**.

American carriers managed to handle the first Internet boom with wavelength division multiplexing (WDM), putting every stream on a different color of light and merging them in an infrared band down the fiber for a hundred miles or so and then converting the dwindling signals back to electronics to do it again. R&R—recovery and regeneration and sometimes 3Rs—with retiming added—meant that the network was constantly translating light pulses into electronic streams and then back again through arrays of lasers and filters and erbium doped amplifiers and down boards of mixers and muxers, serdes (serializers and deserializers), transceivers and analog to digital converters. It all worked well enough to handle the first Internet boom. It provided explosively growing markets for the companies making the transmission lasers and pump lasers, chiefly **JDS Uniphase** (JDSU), and the semiconductor houses selling mixers, analog to digital converters, and digital signal processors, namely Texas Instruments and Analog Devices. But the second Internet boom of broadband video, wireless imaging, and ubiquitous wireless data now happening in Korea and Japan remains stillborn in the U.S. The local loop remains fractured, in a copper cast and a legal straitjacket. Backbone carriers compete on price, while the lords of the last mile maneuver in Washington.

None the less, the three-year ascent of Korea from also-ran to bandwidth colossus shows the way to a new Internet boom in the U.S. With Peter Huber's critical mass of 20 million broadband subscribers having been surpassed this spring, the transition to 100 million subscribers will occur before 2010, according to Huber, by which time the Telecosm will have undergone an all-optical transformation. But well before then it will jump to its new energy state or broadband paradigm with a rush that will be completely missed by technologists, Wall Street analysts, and companies nursing older optical technologies. It happened before.

PARADIGM ONE:
1870-1990—Bandwidth Abundance

During the pre-Internet age, telephony thrived on bandwidth abundance, at least when measured against the modest demands of voice. Bandwidth was wasted as a matter of course. Most of the capacity of a telephone network lay fallow more than 95 percent of the time as people used their phones an average of 20 minutes a day. In a world of bandwidth abundance, circuit switching—connecting the two parties over a line devoted entirely to their call—made sense.

With circuit connections, switches could even be slow. An operator could route the calls manually.

PARADIGM TWO:
1990-2003—Bandwidth Scarcity

As the Internet rose and data became dominant, users put their computers online for many hours at a time. Even as absolute bandwidth soared, it grew scarce relative to demand. Confronting a regime of bandwidth scarcity, the titans of telecom in the 1990s had to learn how to economize on bandwidth. With guidance from Bell Labs, they had mastered the secrets of statistical multiplexing—digitizing calls, distributing them in time slots, and combining many calls onto a single long-distance backbone connection. Then from the Internet they laboriously learned the rules of packet switching, cutting up every message into many packets, each bearing a separate address. While a circuit-switched phone network sets up the call in hundreds of milliseconds, a packet switched network functions like a multi-megahertz post office. The envelopes are switched not in minutes or even milliseconds but in microseconds. Load-balancing data across the network, packet switching is optimal in a regime of scarce bandwidth. It was an era of superfast switches, "grooming" the data and distributing it through the pipes.

Overlaying the redundant and voice optimized SONET facilities of the phone companies, which operated on the physical and transport layers, was a parallel system of **Cisco** (CSCO) and **Juniper** (JNPR) routers. Sixty-four kilobit SONET voice carriers bore 1550 byte Ethernet frames enveloping IP packets. With separate quality of service functions, transport protocols, and service recovery provisions, the routers managed the Internet Protocol packets on "layer three," the network layer, handling all the final IP addresses on the Internet. In this era, the hardware and software piled up in triplicate in optolectronic nodes, ISP hubs and telco central offices across the country, and Moore's law processing speed compensated for bandwidth scarcity and network complexity.

PARADIGM THREE:
2003 TO 2010?—Abundance Redux

The next paradigm shift—from today's relatively narrowband net to Peter Huber's high-speed broadband world of streaming video phone calls and billions of cell phone digital cameras—will spark yet another non-linear traffic surge and another transformation of the technology regime. That is the message from Korea and China. If another 100-fold paradigm shift were to occur during 2004 – 2005, long-haul backbone network capacity needs balloon to 188 exabytes per month (188 x 10^{18} bytes) to handle the traffic during December 2005 without disruption. But Korea took three years, so extend it another year, to December 2006. Measured in terms of today's all-optical technology, that's 363 separate **Corvis** (CORV) systems sporting 160 OC-192 lambdas apiece. By comparison, if today's U.S. long-haul Internet backbone were combined into one seamless network, just three Corvis systems would suffice.

Long-haul links are only one part of the end-to-end network. All backbone traffic must first traverse the smaller metropolitan area networks. The Great Unknown, metro traffic seems to have eluded estimates. Among RBOCs, consensus has been that only 25 percent of metro traffic passes into the long-haul networks. Based on that guesstimate, aggregate U.S. metro traffic exceeds backbone traffic by four times. Returning to our sample paradigm shift, by December 2006 total metro network traffic would equal 100 exabytes per month (2 exabytes per month in each of 50 U.S. metro areas). Thus, each metro network would need to transport an order of magnitude more traffic than today's entire U.S. Internet backbone network and each would require the bandwidth equivalent of four Corvis long-haul systems. Then all these numbers should be more than doubled again, to cover the explosive growth of traffic around the globe.

The Corvis Era

But as elegant and efficient as the Corvis technology is, national carriers will not purchase 363 Corvis systems, nor will regional networks install 200 Corvis bandwidth equivalents in metro networks nationwide any more than they will multiply giant SONET add-drop multiplexers to handle the broadband paradigm. Today's hybrid opto-electronic network will give way to a rainbow of light, and traffic will flood toward the low-cost, low-delay, coherent systems that use Corvis gear. First and foremost that means the Broadwing network that Corvis essentially stole from Broadwing for $17 million. "Listening to the technology," in the way of Carver Mead, we discover that the primary rule of a broadband network becomes: multiply lambdas (wavelengths) for connectivity, not bandwidth. The SONET ring architecture once imperative for network protection and restoration will die under the impossible burden of adding an entire ring of SONET boxes for every new lambda or wavelength (color) of light, followed by more rings of Cisco and Juniper routers, DSLAMs and cable modem terminators, and other boxes galore. Attempting to bear the Net traffic of entire cities on a few score light beams, the networks of the future will choke in a multitrillion-dollar, multi-million laser, multi-hundred-thousand box router-switch-SONET-IP electronic, optical, and protocol conversions morass. The new networks will instead require millions of addressable colors of infrared light. The bits will ride on wavelength lightpaths bearing their own routes and their own addresses.

WDM, which sends many colors of light down a single fiber thread, is ushering in a tide of fabulous bandwidth abundance. In a world of bandwidth abundance, bandwidth-wasting circuits become ideal once again. Rather than economizing on bandwidth by chopping everything into packets and multiplexing them into time slots, the mandate is to waste bandwidth. As in the old telephone system, the approach is circuits that last the duration of the "call."

In this case, the system software sets up wavelength circuits between terminals at the edge of the fiber network where the wavelengths are finally converted back into packets or launched into the fiber "cloud." But the reach of wavelength circuits will steadily expand into metro area networks, across corporate campuses, and finally into enterprises and even neighborhoods. Many of the giant routers will go away, replaced by millions of smaller routers, hubs and service nodes in homes and businesses. Cisco is preparing for that world with its low-end nodes and with Linksys and Aironet for Wi-Fi and beyond. **Intel** (INTC) is preparing with Centrino for wireless access and Gigablades for direct optical access from servers. Meanwhile, on the ever ramifying backbones, passive optical switches will shift and shuffle wavelengths scarcely faster than the operators of yore. The slow switch Corvis era will begin.

The trillion-dollar challenge that can truly unleash the Telecosm is access, last mile connections to homes and offices

Among those with a low titillation threshold in optics, the continuing onrush of the Telecosm makes for titillating reading. Just don't get taken in by it. None of this will happen. It's merely the fancy of cloud-nine cranks who refuse to accept defeat even after being proven wrong. The survey takers and market forecasters and Spitzered analysts assure us that that Corvis CEO David Huber et al are cranks. Unfortunately for the wise guys, however, it is a logical fallacy to assume their arguments are correct merely because they are getting the right answers. Anyone can win at Russian roulette—for a while.

Paradigm III leaders

David Gelernter, another Telecosmic crank, tells us that no matter how certain its eventual coming, we normally fail to envision an event whose exact time and form of arrival are unknown. We tend not to believe in the next big war or economic swing. We certainly don't believe in a repeat of the two-year, 100-fold network traffic jump of 1995-1996, and so we plan our businesses according to the current trends.

From JDSU to **Avanex** (AVNX), from **Bookham** (BKHM) to **Oplink** (OPLK), everyone talks of emerging as "a survivor of the downturn." This hackneyed phrase focuses attention backward instead of forward, turning problems into business plans and companies into pinballs bouncing among the obstacles of the day.

While demand soared in the spring of 2000, JDSU scrambled to increase production by a factor of four every 18 months. A year later, JDSU was scrambling to decrease production just as quickly. But growth is not a reflex action. It demands creativity, vulnerability, risk-taking—a vision for the future. What will be the next market or paradigm? How large? How can we create a significant advantage over the potential competition and increase revenues? Long-term investors look for return on capital, not perceived growth through cutting costs.

JDSU has introduced approximately 75 new products over the past year. A sampling includes a temperature tunable source laser, a WDM source laser for CATV, a credit-card sized EDFA (erbium-doped fiber amplifier), test and measurement instrumentation, and standard amplifiers that reduce costs and have short lead times because they are built on a simplified platform that is scalable and flexible for a wide variety of applications. All of these modules are up-to-date but none will lead the way in the broadband network. Module platform manufacturing is not uncommon in the optical components industry, and **Corning's** (GLW) components operation (now part of Avanex) has become adept at it for many product lines, now also including amplifiers, which until a year ago were virtually all more-expensive custom models. Mini EDFAs were pioneered by Nasser Peyghambarian at **NP Photonics** and by several other startups working on EDWAs (erbium doped waveguide amplifiers) and can be had now from Corning as well. Temperature-tuning of DFB (distributed feedback) lasers is a first-generation technology with limited wavelength selectivity. **Agility** and **Santur** have much more advanced tunable modules already on the market and Intel may be ready with its tunable transponders in early 2004.

Nearly half of JDSU's new products are for transmission, compared to less than 20 percent in previous years. And while the company's portfolio of transceivers and transponders is one of broadest in the industry, including products for enterprise, SAN, metro, and edge applications, the world's volume leader in fiber-optic transmission sales is still **Agilent** (A), not JDSU. JDSU has never shown a strong interest in tunable source lasers since they are too far into the future for immediate revenues and represent too much risk. The "Components Superstore" shows no signs of nearing breakthrough research in this area or in other Paradigm III technologies such as broadly tunable transponders, high-channel count multiplexers (Avanex and **Essex**), or Raman amplification (Corvis).

JDSU's pattern of growth by M&A is really "growth" by buying up someone else's customers. It is an expensive and time-consuming strategy which diverts attention and resources toward integration and slashing expenses and away from innovation. Did JDSU really grow over the years it acquired the likes of E-TEK and SDLI? We can probably never know, since growth would be hopelessly hidden in the complex accounting of acquisitions.

With the disadvantage of a $1.2 billion cash cushion and a clean balance sheet, JDSU can rest on its laurels from the boom and continue along the path of least resistance. Over past year the company has acquired LA Label to extend capabilities in product authentication and security where JDSU sees itself as a global leader. JDSU has also acquired the transceiver/transponder unit of **OptronX** to extend transmission product line in metro and short-reach applications and the data communications unit from **IBM** (IBM). Most recently, it acquired TriQuint Semiconductors' undersea pump-laser packaging technology, enabling the development of entire pump modules.

The eternal life of excess network capacity has become the zeitgeist of the Telecosm, and many companies have been seduced by it. However, when bits and bytes surge once again and functionality reemerges as the watchword of networks, carriers and OEMs will not judge their suppliers by the success of their cost containment programs or even by their profitability. In that day, the "survivors of the downturn" will be the innovators who were ready for the upturn of the broadband network. Today that means Corvis more than any other technology.

The trillion-dollar challenge that can truly unleash the Telecosm is access, last mile connections to homes and offices. The value of networks in a time of bandwidth abundance comes not from capacity but from connectivity. As Paul Green puts it, "There are terahertz of potential bandwidth at the core of the network and many gigahertz of potential bandwidth in the internal links of computers. But between them is a bottleneck, where even cable and DSL operate at speeds thousands of times slower. If this bottleneck can be broken, the entire industry will be awash in demand." The key, therefore, to the prospects of optical technologies and fiber-optic networks is the connectivity of light.

—George Gilder and Charlie Burger

The China Phenomenon

China is poised to take over the world's manufacturing. And that's not all; it will soon move into software and hardware development. Chinese enterprise will begin with simple assembly manufacturing and by serving huge internal markets. China will be a world leader in technology by combining its emerging semiconductor industry with the engineering output of its growing university system. China's rapidly growing internal markets are good for domestic and for international companies. The huge supply of cheap labor and of engineering graduates is also good for companies, but not good for individuals outside China, who may be displaced on a large scale. Westerners displace manual labor with cheaper manual labor. Manual labor is located at a specific place. When we move a plant to a cheaper place, laborers lose their jobs. With Chinese capacity expanding and the Internet providing a global communications network, we are seeing for the first time that *mental* laborers—of which engineers are the largest group—are subject to these same dynamics on a large scale.

I spent three weeks in China in November 2000. I received an invitation to visit the Three Gorges River project and the Yangtze River valley with a U.S. civil engineering delegation. (The Yangtze River runs generally west to east across much of China.) Although, I'm not a civil engineer, I like construction projects and heavy equipment, so I went. Our delegation met with Chinese planners and engineers along a route from near Wuhan, just below the dam site, to Chongqing, upriver. We visited construction sites, including the main site for the dam. There we walked among the workers (no OSHA, no hard hats).

The Three Gorges Dam is the largest construction project in the world. Its scale is breathtaking. The dam construction project employs 100,000 workers and a good portion of the world's heavy cranes. There's not much machinery between the workers and the heavy cranes. Digging a trench, which involves a backhoe and a single operator in the United States, employs a dozen workers with picks, shovels, and wheelbarrows. I saw six or eight workers using ropes and their leg and back muscles to move boulders. I saw loads of rebar (twenty-foot-long steel rods for reinforcing concrete) transported by *bicycle*. I also saw an ultra-modern cable-stayed bridge near the dam site.

As we cruised upriver from the dam site, I got the impression that China is an agrarian economy that hasn't changed in hundreds of years. The Yangtze River valley is a patchwork of tiny farms. There is no farming machinery. There are few paved roads. Smoke shrouds the countryside like an inland fog—from burning soft coal, China's universal heating and cooking fuel.

I returned in June 2002 to speak at an integrated circuit design conference. This time, I visited Zhuhai, a city on the Pearl River delta and a short ferry ride from Hong Kong. Instead of sleepy countryside, I saw a modern and booming metropolis. The contrast between inland farms and coastal cities is shocking—it's still there and it's an important part of the story.

China:
Big Markets, Big Opportunities, Big Problems

GENERAL Slightly smaller area than the U.S.
Population 1.3 billion, frugal, enterpreneurial
Literacy >80%
Engineers 700,000 engineering graduates per year

INDUSTRY 18 million people per year enter the workforce
Unskilled Labor Rate $0.60 per hour
Engineering Salary $4800 to $8800 per year plus housing, medical, and pension

GOVERNMENT Looming financial problems
Banks The four major state-owned banks are insolvent (but operating)
Pensions & Social Security Unfunded obligations may exceed GDP
Stock Market Markets in Shanghai and Shenzen are subject to manipulation

FOREIGN INVESTMENT $40 billion a year
Chip Manufacturing World's second highest foundry capacity by 2003
Other Manufacturing Also moving to China from the rest of the world

MARKETS Economy has grown at 10.9% per year since 1979
PCs 5% penetration. 18% annual growth. World's largest market by 2006
Cell Phones 167 M users. World's largest market. Will double by 2004
Internet Use 56 M users. World's largest by 2006

China is slightly smaller in area than the United States, but China has *1.3 billion people*. Its people are *industrious and entrepreneurial*. More than 80% are *literate*. They *value education* and they *value engineering*. Like any emerging country, China is both primitive and modern. It uses animals, not tractors. It skips wired telephones and goes directly to wireless. It skips generations of bridge-building evolution and goes directly to modern structures, but paved roads are still the exception. There are fifteen miles of paved road in the United States for every mile of pavement in China (2.5 million miles vs. 170 thousand miles). It adopts state-of-the-art electronics and semiconductor manufacturing and skips decades of evolvement.

Foreign investment in China averaged $40 billion a year in the late 1990s and was $45 billion by 2001. Ninety percent of foreign investment goes to coastal provinces with access to international shipping. (The Three Gorges Dam will open much of China's interior to ocean-going ships.) Technology investments get a two-year tax exemption plus three years at half of the standard 15%. Land and labor are cheap. Import and export tax policies are favorable. Exports grew 43% per year between 1985 and 1998. Internal markets are emerging.

What's not good? There are significant language and cultural barriers. Local companies have the advantage dealing with bureaucracies. Foreign companies can enter the economy by partnering with local companies, but companies complain of pressure to share intellectual property with joint-venture partners. In emerging technical markets, standards are set by the *government*. Ordinarily, deviating from international standards leads to dead-end products, but China's internal markets are large enough to make this work. Its membership in the World Trade Organization will encourage China to adopt international standards.

China's economy has grown at 10.9% per year since 1979 and should continue to grow by at least 7% per year for several decades. This rate can be sustained by the growth of internal markets and does not depend on world economic conditions. Internal PC shipments are expected to grow at 18% through 2006, when China will pass the United States to become the world's largest PC market. China had 167 million cell phone subscribers in 2002—the world's largest cell phone market. New cell phone users number *4 to 6 million a month* and will reach 320 million in 2004. With 56 million Internet users; China overtook Japan in 2002 to become the second largest, in number of users, behind the United States. By 2006, China will be number one in Internet-connected users. China's software market was $1.6 billion in 2001 and will grow at 37% per year, reaching $7.8 billion by 2006.

The booming Chinese economy has skeptics. They point to bad loans, corruption, weakening exports, and lack of visibility into financial transactions as reasons to doubt the reported growth in China's gross domestic product. But secondary measures—estimates of electricity, coal, steel, long-distance phone calls, wages, and employment—verify China's growth to within the accuracy reported by Western countries.

Semiconductors

China isn't just about circuit board assembly and internal markets. Its semiconductor manufacturing is growing rapidly, with forty new semiconductor plants announced in 2002 alone. Half of China's chip production is in Shanghai. By 2005, Shanghai will have a dozen semiconductor plants with an aggregate capacity of 500,000 wafers per month.

In March 2002, the government of Taiwan approved investment in semiconductor manufacturing in China, limited to 200-mm diameter silicon wafers and to process tolerances not better than 0.25 microns. As a result, both **Taiwan Semiconductor** (TSMC) and **United Microelectronics** (UMC) have begun to invest in foundries in China. As Taiwan's foundries move to 300-mm wafers, they will move their legacy production to China.

China's domestic markets consumed $15 billion in semiconductors in 2001. It's growing at almost 25% per year, so domestic consumption should be $35 billion by 2005. Domestic production of semiconductors was $1.2 billion in 2001, implying that China imports more than 90% of its semiconductors. Even with domestic production growing faster than 80% per year (all those new semiconductor plants coming online), it will rise to only $13 billion by 2005. China will still be importing 70% of the semiconductors for its internal markets in 2005.

Manufacturing

As China's economy shifts from centrally planned to market-driven and from agrarian to industrial, the huge population provides an infinite supply of cheap labor. Eighteen million people enter the workforce each year. Unskilled labor is *sixty cents an hour*—a quarter of what unskilled labor costs in Malaysia and an eighth of what it costs in Singapore. Unskilled labor in the United States and in Japan costs twenty times what it costs in China. As a consequence, manufacturing is shifting from Japan, Korea, Malaysia, Singapore, Thailand, Indonesia, Taiwan, the United States, and even Mexico to the coastal provinces of China. **Motorola** (MOT), **Texas Instruments** (TXN), Sony Ericsson, **Samsung** (05930.KS), and **Nokia** (NOK) are investing in cell phone production in China. **AMD** (AMD), **IBM** (IBM), and **Intel** (INTC) package chips in China.

BOOTSTRAPPING. Flextronics (FLEX) began as a "board stuffing" business in Newark, California, in 1969. Board stuffers placed and soldered electronic components on circuit boards. Silicon Valley system makers generally contracted their overflow work to independent board stuffers. The contracting company supplied the circuit boards and the electronic components to the board stuffer. Board stuffing was a labor-intensive, low margin business. Flextronics first introduced automated manufacturing to reduce labor costs, but in 1981, it moved to Singapore to further reduce operating costs.

From board stuffing, Flextronics moved up to contract manufacturing. It now ordered components and tested boards too. Next, it moved into printed-circuit-board design, computer-aided design, and front-end component testing. With these new services, a customer could approach Flextronics with just an idea; Flextronics responded with a complete manufacturing plan for the customer's approval.

By the late 1980s, Flextronics expanded its services to include *system* assembly. The company could build ready-to-ship products for its customers. Flextronics built disk and tape subsystems for **Sun Microsystems** (SUNW) and it built modems for Hayes. It built the **Microsoft** (MSFT) mouse and the original Palm Pilot. Today, Flextronics uses an "industrial park" model—locating suppliers close to the manufacturing plant to lower costs and to improve flexibility and responsiveness.

What I have described above is a "bootstrap operation." Flextronics pulled itself up by its bootstraps to move from board stuffing to manufacturing complete systems. It grew, in thirty-three years, from 2 employees to 70,000 employees in a capital-intensive business. It expanded from a local operation to an international business. It did this by starting in a business most would find uninteresting.

Flextronics builds Motorola cell phones, **Dell** (DELL) circuit boards, and **3Com** (COMS) fax modems on Chinese production lines. While Flextronics operates in twentyeight countries, its lowest-cost operations are in China. By this year, 40% of its worldwide production could be in China. Among its "manufacturing partners" are **Cisco** (CSCO), **Ericsson** (ERICY), **HP** (HPQ), Microsoft, Motorola, Nokia, **Nortel** (NT), **Philips**, **Siemens** (SI), and **Xerox** (XRX). Flextronics is in the right place at the right time with expertise in electronics manufacturing services.

In most countries, as experience and skill increase, wages rise, leading to higher costs for the manufacturer. In China, unskilled labor gets especially low wages. As their experience and skill increase, workers move to skilled jobs in other plants. China's huge population means that there are always workers entering the labor force to backfill vacant positions at low wages, keeping manufacturing costs *constant*. What Flextronics did for one industry, China will do for the world. Laborers bootstrap their skills and companies bootstrap the sophistication of their manufacturing operations. With the infinite supply of cheap labor, China can draw in manufacturing from the rest of the world, bootstrap both it and the labor force, and *never outsource anything*. In addition to exports and to existing internal demand, the bootstrapping labor force grows the internal markets for its own products. This demand for products grows with the increasing size and affluence of China's labor force.

LEGEND. You have probably never heard of **Legend Group Limited** (LGHLY.PK). It's the largest PC manufacturer in China. It has 30% of China's domestic PC market. Legend outsells its domestic rivals by at least three to one and it outsells its international rivals by at least seven to one. Legend is not just a box maker. Legend understands the support needs of a population new to personal computers. In addition to low-end PCs for homes and small businesses, Legend makes servers and notebooks. In the last quarter of 2001, it broke into the world's top ten in server sales, with servers based on Intel's Xeon microprocessor. In 2002, Legend introduced notebooks using the Pentium 4 microprocessor.

Legend partners with Intel, IBM, Microsoft, **AOL** (AOL), Siemens, and Texas Instruments. Legend distributes HP printers, Cisco routers, and IBM midrange computers. The Siemens and Texas Instruments partnerships help Legend in wireless and in handheld markets. Just two years after development began, Legend announced six multi-function cellular handsets, five were GSM handsets, one was CDMA—the smallest and lightest on the market.

Legend's partnership with AOL lets it offer simple Internet connections to the home. China's few television stations are about as entertaining as C-SPAN, so the Internet wins the contest for eyeballs.

Legend builds PCs primarily for the low-end of the domestic market. By copying techniques from successful high-volume manufacturers, Legend achieves efficiencies that equal Dell in average time from order to delivery. Legend is growing with China's domestic PC market, but with its efficient manufacturing and its low cost, it will sell its PCs internationally. The Chinese government, through Legend Group Holdings, owns 60% of Legend Group Limited.

The stories of Flextronics and of Legend are being duplicated in countless factories throughout coastal China. Flextronics and Legend are China in microcosm —bootstrapping operations from simple manufacturing to sophisticated end-products. But now the new time scale will compress decades to a few years. And there'll be no move overseas, since China will remain the cheapest place to manufacture. The endless supply of cheap labor keeps costs low even as many prosper.

Education

China will grow beyond manufacturing. Chinese society puts education on a pedestal. Families, restricted to a single child, see education as the way to betterment. As China bootstraps its manufacturing and its internal markets, it is developing its educational system to support emerging requirements for engineering talent. University programs are expanding. In 1992, only 5% of the candidates participating in the national three-day entrance exams advanced to universities. This year, 14% of the candidates will advance to universities. Though only the top 14% get to go to college, *Chinese universities produce 700,000 engineers a year—37% of college graduates.* U.S. universities, in contrast, produce only 65,000 engineers a year—6% of graduates—from a pool that's not nearly as selective.

B.S.-level engineering graduates can expect to earn $4,800 to $8,800 per year plus housing, medical, and pension. For the hiring company, these benefits add 50% to the cost of an engineer. Many of China's new engineers take jobs in Taiwan, Singapore, or the United States. For a few years, they gain experience in semiconductor manufacturing, assembly, testing, and a host of other tasks. Then they return to build and to manage manufacturing, testing, and assembly plants in China.

Software piracy

In most countries, hardware revenues are twice software revenues. In China, the ratio is nine to one. Illegal copying and counterfeiting of software have been rampant. As much as 90% of the installed base of software was illegally copied. Copies of Microsoft Windows operating systems could be had for a dollar.

Here's what got the *Chinese government* to clamp down on software piracy. China may have an infinite supply of cheap labor and it may have lots of cheap engineers and programmers, *but even cheap resources can't compete with theft and counterfeiting that avoid all expenses except copying and packaging.* Leaders in the Chinese government's State Council, realized that a *domestic* software industry couldn't develop in this environment. They issued Document 18 in 2000, "Notice of Certain Policies to Promote the Software and Integrated Circuit Industry Development." Document 18 institutes fines of ten times the product's list price for purchasers of pirated software. For *sellers*, penalties include confiscation of equipment, jail, and even execution. Capital punishment for software piracy!

With Document 18 in place and with software piracy dwindling, international software suppliers are entering the market and a domestic software industry is emerging. And the ultraselective education system motivates families to buy PCs, which signals a growth market in educational software.

Globalization

Perhaps I should have been a civil engineer or a mechanical engineer because I like big machines and heavy equipment, but I'm an electrical engineer. There's a controversy today over the H-1B visa program that brings electrical and computer engineers into the country in the midst of hard times for local engineers seeking jobs. The Institute of Electrical and Electronics Engineers, in spite of being an international organization, has taken a protectionist stance against H-1B visas. I'm against H-1B, but not for protectionist reasons. I'm against H-1B because it's temporary; we should *encourage* immigration and citizenship for these engineers. August 2002 5

H-1B opponents say companies hire foreign engineers because they're cheaper. "We should employ our citizens first." This is the idea that limiting the number of doctors and dentists keeps professionals' incomes high. The idea is to limit competition for engineering jobs by restricting immigration, visas, and even the output of engineering schools. That might work for civil engineers or for environmental engineers—as it does for doctors and lawyers and dentists—because their workplace, their work, and their customers have geographic ties that make outsourcing impractical. But the Internet makes money, status (including live camera coverage), design files, and contracts available anywhere in the world. Most electrical and computer engineering design projects are eminently portable. Circuit boards, integrated circuits, and systems can be designed and manufactured anywhere. Integrated circuit manufacturing is in Taiwan and, as we've just seen, circuit board and system manufacturing are moving to China. It's a global market, so we have three choices: our "citizen" engineers can take the jobs at competitive rates; we can give the jobs to immigrants who will work for less; or we can let the jobs go overseas. Protectionist policy won't work; raising barriers to immigration will cause the *jobs* to emigrate.

I see a great future for companies investing in manufacturing, in engineering design, and in research in China

Government

Banks. The biggest Chinese banks are state-owned (the Bank of China, the Industrial and Commercial Bank of China, the China Construction Bank, and the Agricultural Bank of China) and they are *insolvent*. The problem has two legs. The first is the Chinese workers and the second is the autocratic political system. Each year, Chinese workers' savings deposits equal 40% of the gross domestic product— the highest individual savings rate in the world. At the direction of political leaders, these state-owned banks make huge loans to state-owned enterprises. Executives at these enterprises, knowing that both the banks and the enterprises are state-owned, view these loans as free money. The cash value of these loans on the banks' balance sheets might be as little as 10% of their face value. The banks have been borrowing short (workers' demand-deposits) and lending long (enterprise loans)—a recipe for financial collapse. If depositors had investment alternatives, such as a reasonable stock market, or if they *knew* that the banks were insolvent, the withdrawals would lead to a banking collapse. China's banks are running the world's largest Ponzi scheme. The banks, which have received bad debt bailouts twice before, expect government bailouts. But this problem is too big for the Chinese government.

Stock markets. China's stock markets operate in Shenzen and in Shanghai. They are loosely regulated and they are subject to manipulation, to corruption, to deceptive accounting, and to insider trading. This situation will not improve soon. Small shareholders aren't well represented; there's a permanent ban on shareholder lawsuits. Major shareholders manipulate and defraud the small shareholders. Right now, it's just as well that the stock markets are corrupt; if the stock markets worked, the movement of even a small percentage of workers' savings into equities would cause the banks to collapse.

The central government created the stock markets to sell equities *in state-owned enterprises*. The central government is also the stock markets' regulator—a clear conflict of interest. The government needs to sell equity in state-owned enterprises to raise money for other obligations (insolvent banks and pension and social security shortfalls, for example), but it cannot do this without triggering a precipitous fall in stock prices.

Pensions and other obligations. Darn! It's those unintended consequences of central management again. China is successfully controlling its birthrate by decree. Births are down. The limit is one child per family. Birthrates have fallen below replacement levels. Parents want a male heir, so more males than females survive. The imbalance reduces the birthrate further, accelerating population aging. When everyone is old, who pays pensions and social security? Current workers' pension and social security obligations are unfunded. The system is running behind by amounts that may *exceed the gross domestic product*. The payments that are being made—a small fraction of obligations—for pensions, for unemployment insurance, for guaranteed living allowance, and for other social welfare programs are paid out of *current* revenues. This can't work for long. As time goes by, obligations dwarf revenue as the population ages.

Lessons

I've said good things about China. The boundless supply of cheap labor is great for manufacturing. The people work hard and they are entrepreneurial. Favorable tax and trade policies encourage international investment, making it attractive for foreign companies to import components, to manufacture in China, and to export to international markets. That's great for companies such as Flextronics, Legend, IBM, Intel, and Motorola that have expanding operations in the rapidly growing coastal provinces. They have access to shipping to reach international markets and they have access to fast-growing internal markets.

I expect the entire world's manufacturing to move to China. As the world's manufacturing moves to China, the growing local workforce becomes affluent—creating great internal markets for products. Burgeoning internal markets improve corporate profitability by avoiding international transportation and distribution costs.

Chinese are literate and their culture values education. China's university system is competitive, and it's turning out hundreds of thousands of engineers every year. The growth of China's university system is decreasing its dependence on foreign education. But many students still go to United States and to European universities for both undergraduate and graduate degrees and return to China. Engineering graduates frequently go to places such as Taiwan or Singapore to gain experience in semiconductor design and manufacturing and to return to China to manage local manufacturing and design operations. That's good for China and it's good for companies that want to bootstrap from low-level manufacturing into circuit board design, or into electronic systems design, or into electronics research. IBM, Intel, and Motorola have research centers in China. The pool of qualified candidates for research positions yields thousands of applications for dozens of positions, making these organizations more exclusive than their counterparts in other countries.

It sounds ideal: cheap labor, literacy, university educations, engineering talent, favorable tax and trade policy, and cultural and political alignment. I see a great future for companies investing in manufacturing, in engineering design, and in research in China. It's a great environment for Flextronics for electronics manufacturing; it's a great environment for Chartered, for TSMC, and for UMC for semiconductor manufacturing; it's a great environment for manufacturing and for internal markets for Legend Group Limited; and it's a great environment for internal markets for **Via Technologies** (2388.TW).

My positive comments do not apply to stateowned enterprises or to the Shanghai/Shenzen stock markets. China has problems with its banks, with its stock markets, with its political system, with its underfunded pension plans, and in paying its social security bill. Potential government interference in Legend's operation is a concern, since the government owns the holding company with a majority interest in Legend Group Limited.

Nevertheless, China is about to explode on the scene.

—Nick Tredennick and Brion Shimamoto

Mead's Analog Revolution

When Bill Gates launched his new XP network-based operating system at a recent trade show, this king of the digital age did not begin by entering a password or clicking icons in a pop-up window. Instead he put his finger firmly on a glowing biometric touch pad that recognized Gates, loaded his personal settings, and gave him access to his personal files and digital kingdom. The company that supplied this open sesame is called **DigitalPersona**. But its innovation in pattern matching is *analog* and its technical leader, Vance Bjorn, represents a movement originating at Caltech in the mid-1980s to transform the world of analog interfaces to the digital world.

Dominating the PC touch pad market and breaking through the financial doldrums of 2001 and early 2002 with the first successful Silicon Valley IPO in twelve months was **Synaptics** (SYNA), another primarily analog company, also with roots in this Caltech movement. **National Semiconductor** (NSM), meanwhile, is an analyst darling among single chip system companies. But the bulk of the long run value in both National and Synaptics may be their shares in another company with a Caltech pedigree, **Foveon**, whose just-announced next generation silicon imager will finally overthrow photographic film. In coming years, Foveon will be a dominant force in imagers in cameras of all kinds, still and full motion, a fast growing market of over $20 billion. Followed in the *Gilder Technology Report* since its founding in 1997, Foveon has fulfilled its promise with the single most powerful commercial technology I have seen since launching my newsletter in 1996.

> **Of all our analog companies we have long been excited by the prospects of National Semiconductor and its spinout Foveon**

But the story of this analog insurrection actually begins at least a decade earlier, in 1986, when the semiconductor industry was roaring back at last from a cataclysmic slump in which revenues dropped some 45 percent in a year. In a Caltech classroom, an eminent professor of engineering and applied science, like many in his trade, seemed to be flaunting his august connections in industry. Projecting the design of a massively parallel processor on the screen as a model for a revolution in computing, he said, "Now I've been up in Silicon Valley, talking to the guy who made this thing and…." Why was this class laughing? Don't they believe in Carver Mead, the industry's first and most profound prophet of the very large scale integration (VLSI) of digital chips, who had performed the crucial researches on which Moore's law itself was founded? But the massively parallel design he was exhibiting on the screen to such friendly hilarity was not digital. It was analog. It was a schematic of the human brain.

Mead's citation of the brain was not unusual in computer science. What was radical at the time was that rather than treating the digital computer as a possible model for an ultimately superior brain, he was offering the analog brain as a model of an incomparably more powerful computer. After twenty years as the industry's most authoritative proponent of the power of digital electronics, he was reversing direction and declaring the onset of a new era of analog. Prone to dour observations about the perceptual powers of digital computers compared to those of fruit flies, Mead maintained that what he termed "neuromorphic analog VLSI" offered the possibility of a radically more effective image processor. Later in the class, he presented the first example of such a machine, a silicon retina chip, modeled on the human eye, that could follow a rotating fan without aliasing (reversing direction like movie wheels) and could adapt to changing intensities of light. It was a significant first step toward creating a real-time imager on monolithic silicon. And among Mead's laughing students and auditors were several who would go on to form the vanguard of a broadband analog revival in the twenty-first century.

Yet, at the time, prevailing wisdom in the industry militated massively against doing a vision system in analog. Analog systems, everyone knew, could not scale to the huge densities of imager pixels. Requiring thousands of discrete devices, they loomed as too cumbersome to build and too wasteful of power.

Mead launches Synaptics

But there was Mead offering his device in plain bulk CMOS (complementary metal oxide semiconductor) silicon, slashing power by running the transistors at subthreshold voltages like the micro-power systems in digital watches. Rather than exotic discretes, the photoreceptors were created as part of the CMOS manufacturing process by turning a traditional CMOS bug into a feature. At the junction of every cell's two transistors— the complementary negative and positive devices—is a nasty potential bipolar transistor or "latch-up," a parasitic device that must be neutralized. Rather than neutralizing it, however, Mead listened to the technology and enhanced what the silicon wanted to provide.

Collecting light on its P/N junction, the latch-up transistor became an effective photoreceptor. With gain of over 1,000, the latch-up outperformed ordinary photodiodes that would have to be implemented off-chip. Integrating the photoreceptors onto the CMOS device, Mead showed the way to create analog systems that scaled like digital systems in accord with Moore's law. Within a few months, Mead launched a company called Synaptics to extend this technology to all the human senses, from hearing and imaging to touch. Joining him was Federico Faggin, the builder of **Intel's** (INTC) first microprocessors and inventor of the self-aligned silicon gate that made them possible.

But there would be no giddy ascent for Synaptics or for Mead's new analog vision.

Analog? Digital?

For forty years, thanks in large part to Mead's own pioneering work in mega-scale digital systems and their design, the reduction of all information to digital numbers has seemed the essence of progress itself. As the chip burst beyond the backplanes of computers, thoroughly analog devices like cameras, radios, television sets, telephones, ovens, airplanes, and automobiles all began moving into the digital realm. Although most radios remain dominantly analog, the music that they broadcast originates on digital disks, where increasingly video also resides. Digital satellite radio systems have emerged from **Sirius** (SIRI) and **XM Satellite Radio** (XMSR). And the digital camera continues its advance, dominating newspaper and magazine photography and making its way steadily into amateur cameras such as computer imagers. CCDs (charge coupled devices) delivering as many as 16 million pixels will theoretically saturate the eye's ability to resolve detail just as the CD saturates the ear. With digital systems triumphant even in the realm of the senses, Mead's ideas for analog VLSI incurred solid resistance even from leading analog semiconductor companies, such as **Analog Devices** (ADI) and **Texas Instruments** (TXN), **Linear Technology** (LLTC), and **Maxim Integrated Products** (MXIM).

Appealing to the senses

At the outset, Synaptics targeted the three key human senses, touch, vision, and hearing. But for the first seven years the company made little progress in Mead's agenda of neuromorphic devices or large-scale analog neural networks. Only slowly did the silicon retina chip eke forward. In May 1991, after some 20 iterations of the device in Carver's lab at Caltech, it made the cover of *Scientific American* in the form of the face of a cat registered by Mead's retinal camera. Inside, the story by Mead and his student, the late Mischa Mahowald, was confident: "The behavior of the artificial retina demonstrates the remarkable power of the analog computing paradigm embodied in neural circuits…. A neuron is an analog device: its computations are based on smoothly varying ion currents rather than on bits representing discrete ones and zeros. Yet neural systems work with basic physics rather than trying constantly to work against it…."

In a sense, however, the cat on the cover—a blurred, almost unintelligible image in one color—was a downer, belying both the confident assertions inside and the grandiose claims of the Synaptics' business plan. Captured in only 2,500 pixels, the image seemed to pose no significant threat to the Moore's law juggernaut of digital electronics that was already propelling a thriving industry of machine vision for manufacturing applications. The key technology, which would also become the basis of digital photography, was charge coupled devices—a kind of silicon bucket brigade resembling a single stretched transistor with thousands of gates between source and drain that convert incident photons into electrons and pass them on in a serial array. With a single CCD chip holding millions of pixels, even then many company laboratories were experimenting with digital cameras that offered resolutions far higher than Mead's. Few were awed by his claim that he could scale his device to digital densities a hundredfold greater than the early rendition. Some 250 thousand monochrome pixels scarcely endangered **Eastman Kodak** (EK) or **Sony** (SNE).

The right feel

Synaptics faced financial failure until as a result of ingenious mixed signal inventions by Mead student Tim Allen, the company broke through in the mid-1990s first in the realm of touch, where Mead had done little work. So superior were the company's touch pads that they quickly took over the industry. Unlike rivals **Logitech** and **Alps**, Allen used a capacitive sensing pad rather than a resistive pad to identify the placement of the finger. A patented analog converter can locate the capacitance aroused by the finger on the pad to an accuracy of around 25 microns, or a quarter the width of a human hair. A totally solid-state solution in large-scale analog, the patented device palpably outperforms all other touch pads. Assembled in Thailand and then shipped to dominant PC manufacturers on Taiwan, Synaptics' superior pads came down in price to the point that rival Logitech exited the business. With a profitable run rate, Synaptics is now fully engrossed in the touch pad business. Bursting through the tech market doldrums on January 28, 2002, Synaptics is the first vessel of Mead's analog vision.

Shortly after Synaptics' breakthrough in touch pads it became clear the company's very success would limit its focus to touch, shedding Mead's more ambitious sensory agenda. Mead eventually broke with Faggin and relinquished his role as chairman, though retaining his shares.

National Semiconductor partnership

Spun out of Synaptics was Mead's new company Foveon, 49 percent owned by National Semiconductor, a crucial partner contributing not only cash but analog chip fabs and all its intellectual property in imagers, as well as the man who created much of it, analog engineer Dick Merrill. Merrill is described by Mead as the most creative engineer he has met in all the combined disciplines of wafer fabrication, circuit design, device physics, and photography. Synaptics retained 15 percent of Foveon, likely to become that company's most valuable asset.

Inspired by Mead's retina chip created at Caltech and work at **Apple** (AAPL) by Mead associate and Foveon cofounder Richard Lyon, Foveon is the most revolutionary vehicle of massively parallel analog VLSI. From the outset, Mead set the company up to master the intricacies of a camera system that could render authentic color. The key, Mead believed at the time, was to keep all the information in analog form as long as possible.

Dramatizing the challenge is the complete absence of color silicon imagers before Foveon. Digital cameras capture images on silicon and from those images produce color photographs. But the silicon photoreceptors operate in black and white. More precisely, they measure the intensity of the light striking them, not its wavelength. They are, if you will, indiscriminate photon counters.

In conventional digital photography progress from black and white to color is achieved not by gathering more information but by throwing information away. Over the photoreceptor for each pixel (roughly speaking the smallest component point in the image: think dots per inch in your ink jet printer) is a filter letting in only red, blue, or green light (the three colors captured by the rods and cones of the human eye). Thus, at each pixel the receptor measures only the intensity of one color, the immediately neighboring pixels capturing the others.

The camera thus starts by throwing away two-thirds of the information at every point in the image. It is never recovered. The final color image is produced by an elaborate digital guessing game, an algorithmic approximation whereby speedy but expensive digital signal processors project values for nearby red and green receptors onto the blue pixels, and so on. Because the algorithms function best by incorporating information from a range of nearby pixels, the guessing game for each can require a hundred arithmetic operations, one reason that at mega pixel levels the cameras waste time and power.

Ingenious as the guessing games are, the original decision to toss away so much information permanently impairs picture quality. Notoriously, certain color patterns trigger aliasing in the form of arbitrary rainbows, checks, and whorls where nature intended a blue shirt or a plaid jumper. As always in the digital realm, the preferred way over the rainbow is to do more with Moore: as chip densities go up, add more pixels, with smaller receptors, and handle the burgeoning computational load with even faster digital signal processors (DSPs).

> **Even as we savor Foveon, we look ahead to more systemic disruption that will be unleashed by Carver Mead's analog apex, Impinj**

When digital systems exert themselves in the analog realm, however, the DSPs often hurtle at gigahertz pace blithely past crucial signposts from Mother Nature. Ultimately pixel size is limited not by Moore's law but by less tractable limits like the wavelengths of visible light—at roughly half a micron already dwarfing the smallest digital circuits—and the resolution of the human eye.

Mead was determined to avoid throwing away information. As Lyon explained, they wanted "no guessing at all." They wanted to "measure every color at every pixel." In the Foveon camera, every pixel would register real features of the image rather than digital simulations of it.

In the first generation Foveon camera that would mean tossing the filters and substituting a prism, splitting the red, green, and blue light and directing each to its own single chip imager. Then, instead of guessing, the signal processor would combine the actual red, green, and blue values for each pixel and produce the final image.

Foveon flies under the radar

The result was pictures of extraordinary quality, mocking even the best digital competitors and rivaling the Hasselblad studio cameras that establish the state of the art for film. It also meant a craft guild manufacturing process, producing handcrafted modules of glue and prisms, mirrors and multiple microchips all aligned with exquisite accuracy. At $50,000, the original sales price also marked Hasselblad as nearly their only competitor and professional studio photographers as their only market.

Fine with Mead. Let the competition scoff at the handcrafted prisms that would never enable a viable consumer product. Let them conclude that Foveon was no threat. Flying in under the radar, the Foveon team would be free to pursue the real goal, a single-chip silicon color imager that would yield the best, cheapest, and easiest to use mass market cameras ever made, shedding not only film but virtually all the precision mechanics, including ultimately the shutter itself. Left would be only lenses and silicon, a true solid-state camera. But first the imager: Mead, Lyon, and Merrill had some ideas.

Through 1997, most of the ideas resided on Dick Lyon's desk, from which a few percolated to his brain. As part of the agreement with National and Synaptics, Foveon had inherited all the intellectual property on imaging held by both companies. Most of it, Lyon recounts, could be discarded. But Merrill had been a compulsive patenter at National ("Patents are a way to do something with an idea, without too much work. You dump it on the patent attorneys.") While at National, Merrill was trying to create out of CCDs a truly differential analog technology where only the deltas are measured and the noise drops out. Existing CCDs captured electrons (the negative energy) but threw away the holes (the positive energy). He patented a CCD that could keep both the electrons and the holes, and balance them off, registering only the changes. As Carver and Mischa had pointed out in the *Scientific American* article this is one of the fundamental advantages of analog systems, which "respond to differences in signal amplitude rather than to absolute signal levels, thus largely eliminating the need for precise calibration….Because only changes and differences convey information, constant change is a necessity for neural systems—rather than a source of difficulty, as it is for digital systems."

(The DigitalPersona fingerprint touch pad exploits this analog "change" phenomenon to differentiate between latent prints left on the touch pad and a live fingertip.)

Merrill's idea was based on the fact that different colors of light penetrate silicon to different depths. Bipolar P/N junctions buried at two depths on the chip would collect either the red light or the green, creating the differential analog levels. It didn't work: because one was red and one green the twain never met in a way that could enable differential analog. Instead, Merrill thought, this idea combining National's biCMOS process and transducer skills, might be useful someday as a color imager. Three buried P/N junctions could collect all three colors at a single pixel without filters, in effect tripling the "bandwidth" of the silicon plane. As the highest frequency and energy color, blue would be captured near the surface, only a half-micron down. The less energetic green would sink one and one-half microns before it agitated the silicon enough to be absorbed. The lowest energy photons—red—would penetrate down to three microns. One chip, every color at every pixel.

Alas, Merrill was convinced that a slight overlap of the blue, green and red levels in the silicon would make the system noisy, and unusable in a high precision application. True to the habits of a lifetime, he tossed the idea to the patent attorneys and then essentially forgot it. It was Lyon, charged with mining the National IP for gold, who rediscovered it. He saw Merrill's objections, but working like Mead from biological premises, he observed that the eye is noisy in almost exactly the same way. Repeated six million times across the retina, the eye's three different cones identify the three colors with a small overlap. Lyon recommended that the engineers tweak the technology so that the overlap in the silicon correlated closely with the pattern in the eye, yielding an accurate rendition of colors as humans see them.

The magic of analog

Recalling that former Caltech scholar Tobi Delbruck, son of the Nobel Laureate, had come up with a similar idea, Mead was immediately impressed by Lyon's logic. The bipolar photo detectors repeated the original retina design, with not one but three buried junctions. If it proved manufacturable and effective, the single chip color image plane would repeat in analog the magic of the digital microcosm. It would be both better, cooler, cheaper, and lower power than its rivals, and it would scale, according to Lyon's calculations, to no fewer than 300 million pixels, far more than the eye could absorb. But whereas those excess pixels might be useless in a CCD device, the Foveon team saw that they would be the key to the solid-state camera. As chip processes improve, instead of making pixels exiguously smaller, more circuitry can be added to each, shifting currently mechanical functions like F-stop adjustment and auto focus to the silicon itself. The every-color-at-every-pixel-principle would facilitate the creation of pixel clusters, allowing Foveon to adjust dynamically the effective number and size of the pixels and thus their receptivity to light, the equivalent of allowing a photographer to change the ISO speed of his film between one shot and the next. At the optimum pixel count, the chip, requiring far fewer arithmetical operations than a standard digital device, could process images with virtually no delay, enabling a film quality video camera for motion pictures or a consumer to use a single camera for both home video and stills. With such bells and whistles increasingly providing differentiation in the camera market and expanding its margins, the Foveon imager would encompass on a single sliver of silicon essentially all the value of a modern camera except for the lens. In conventional digital cameras today, pixel proliferation is driven by marketing hype as companies try to persuade customers 16 million guesstimated dots is better than four. Transforming the pointless pixels into points of value, Foveon would use Moore's law of digital progress to rule the imaging world with an analog device.

Foveon's competitors become customers

But not as a camera company. Foveon is pursuing Clayton Christensen's innovation cycle, following an initial phase of integration (wherein quality results are so hard to come by that every aspect of a system must be integrated and optimized by the manufacturer) with a mature phase of modularity (wherein quality is so abundant that assemblers can use plug-and-play parts). By dropping prismatic handicraft for a single chip module, Foveon transforms its multi-billion dollar, market dominating, entrenched competitors into its multi-billion dollar, market dominating, entrenched customers. Sony is the world's leading digital image-maker. But Sony is the world's leading customer for digital imagemakers and reportedly in negotiations with Foveon. So is virtually every other major camera maker, though no deals have been announced except for an alliance with **Sigma**, which will deliver the first Foveon enabled cameras to market later this year.

Foveon's new silicon imagers are the single most powerful new commercial technology I have encountered since launching the GTR

On the wall outside Mead's corner office at Foveon is a dramatic symbol of the amazing advances achieved by the company: a three-foot high image of the face of a cat. Rather than the blurry monochromatic sketch offered on that old *Scientific American* cover of ten years before, the new image offers a full-color vividness and verisimilitude never before achieved in photography. Some six square feet, the image resolves every hair, whisker, glint, and gleam of the feline fur and renders the eye of the cat with a lifelike glow that gives the viewer the distinct and disturbing feeling that a formidable animal is watching him. Yet the picture is of a kitten. Enlarged to orders of magnitude its actual size, it shows not the slightest deterioration, distortion, or alias.

For the consumer today, the prime motive of digital photography is easy upload to the Web for storing and sharing. Ironically anemic digital images justify anemic compression technologies (JPEG) and anemic dial-up transmission pipes. The Foveon world demands better. It would be unsurprising, if Foveon's most important ally turned out to be an unusually large software company determined to extend its dominance of PC platforms to the Web.

Impinj

While technology investment does not get any hotter than Foveon, a private company 64 percent owned by two public companies—National Semiconductor 49 percent and Synaptics 15 percent—we look ahead to the more systemic disruption that will be unleashed Impinj and Applied Neurosciences.

Impinj is a radical innovator in self-adaptive semiconductors. Applied Neurosciences will offer a probable breakthrough venture in speech recognition. A kindred company, also associated with Mead, called **Sonic Innovations** (SNCI), emerged in 1999 and is rapidly gaining ground in the global hearing aid business. This jumble of apparently unrelated companies all embody the singular new vision unleashed by Mead some 20 years ago in his classes at Caltech and brought to diverse fruition by an amazingly ingenious cohort of his students and associates.

Of all the companies launched by Mead, it is Impinj, co-founded by Mead and his Caltech student Chris Diorio, whose technical innovations could have the broadest and deepest impact on the semiconductor landscape.

The Holy Grail of communications semiconductors is systems-on-a-chip (SoC). If you can put digital signal processing, microprocessing, network, memory—and maybe even some analog—functionality on one chip, you can dramatically lower power, cost, and real estate. And increase performance. Telecosm companies like **Broadcom** (BRCM), **Texas Instruments** (TXN), Analog Devices, and National Semiconductor are leaders at integrating components for the cable, DSL, LAN, and mobile phone markets. **Altera** (ALTR) has field programmable gate arrays (FPGAs) with up to 114,000 logic elements, 28 DSP blocks, and 10 megabits of RAM, all on a single chip.

But clearly we have a long way to go. A 3G phone from **Ericsson** (ERICY), for instance, has more than 600 discrete components, all devouring power, taking up space, and requiring interconnection on a printed circuit board. The two biggest SoC integration obstacles are dynamic random access memory (DRAM) and analog transistors. Until now, neither analog circuits nor DRAM cells have scaled with logic CMOS. In April 2002, **IBM** (IBM) broke through the DRAM barrier, building 5 megabytes onto **EZchip**'s (LNOP) world-beating 10-gigabit network processor.

Impinj too will tackle the $40 billion analog and mixed-signal SoC markets. Using a previously unknown, but ever present, transistor phenomenon discovered and patented by Diorio, Impinj can make mixed-signal systems-on-a-chip that are 30 times smaller, use 10 times less power, cost much less to build, and by some metrics perform 2 orders of magnitude better than today's leadingedge products. With its first product already released and a multitude in the pipeline, the small Seattle start-up will be challenging some of the world's largest semiconductor companies within the year.

Analog devices detect, transmit, and create real-world waves or continuous fluxes of voltages, current and phase (timing) that travel through the air or down wires. Two of the most important analog components are the aptly-named digital-to-analog converter (DAC) and analog-to-digital converter (ADC). The A to D process of converting messy highfrequency waves coming in through the air or over a wire into digital signals is by far the more difficult task. After a wave is detected by an antenna, it is passed through several layers of RF (radio frequency) and IF (intermediate frequency) downconversion and then to the ADC, all before a Texas Instruments, ADI, or **Qualcomm** (QCOM) DSP goes to work. To accurately represent an analog signal as a digital string of ones and zeros, an ADC must sample the signal at a rate at least twice its bandwidth. A 5 MHz signal, for example, must be sampled more than 10 million times each second.

In other words, required increases in accuracy impose exponentially increasing burdens of digital processing. An ideal 14-bit converter handles 16,384 voltage levels per sample, although most 14-bit or even 16-bit products deliver 12-bit effective performance. Working in the analog domain, Impinj has just taped out a “true” 16-bit, 75 Msps (millions of samples per second) ADC that stores 65,536 voltage levels. With a 90+ dB dynamic range and 75 MHz of bandwidth, it can usurp much of the IF circuitry in a cell phone or base station.

Impinj stands for *Imp*act-Ionized Hot-Electron *Inj*ection. Of course. But before you get too excited, let’s review why analog and digital components mix like oil and water.

Digital electronics scales with Moore’s law. As logic transistors shrink, they get faster, cooler, and cheaper. Their job is to register ones and zeros, and as long as they can do that, they are good enough. Digital chip designers can operate in their virtual, abstract worlds of Boolean logic—and let Moore take care of the device physics. Because of their less rigorous task (registering ones or zeros), two similarly designed logic transistors that contain real-world defects are still functionally equivalent.

In analog, however, even slight variations in theoretically identical transistors can prove deadly. “The Achilles Heel of analog is that every transistor is different,” says Diorio. As analog transistors shrink, moreover, small imperfections are magnified by the square of the reduction, creating an overwhelming incentive for analog designers to use larger transistors, entirely incompatible with the scaling magic of Moore’s law.

While a bachelor’s EE can design digital chips after a few months of training, analog chip design is an exacting “black art” that often takes 20 years to learn. *IEEE Spectrum* reports there are only 1,800 experienced analog designers working in the U.S. today. There are hundreds of thousands on the digital side. Handcrafted analog thus suffers from a technological bottleneck and a personnel bottleneck.

Analog automation

Impinj breaks these bottlenecks by “tuning” its transistors. When Diorio received his first Impinj digital-toanalog converter (DAC) chip back from **Taiwan Semiconductor Manufacturing Company** (TSMC), it essentially didn’t work. None of his chips do—initially. Impinj builds its circuits in advanced sub-micron CMOS geometries (e.g., 0.18 microns) ill-suited to analog. Mismatched, noisy, and otherwise defective transistors mean competitive leading edge products might outperform Diorio’s chips by a factor of two. But Impinj’s transistors change, they morph, they learn. Through external and self-adaptive calibration—post-fabrication—transistor performance improves enough to overtake DAC competitors by a factor of three or four—while saving 30x in space and 10x in power and enabling integration with DSP and µP logic, memory, and other components. Plus, young engineers can now design “sloppy” analog circuits and clean them up afterwards. Redesigning a product for a new geometry—say, going from 0.25 to 0.18 microns—would have taken an eight-person team a year. Now it takes three people one weekend.

Impinj can also improve upon, and replace, digital components like DSPs and just as easily integrate them onto a single-chip-system. Like the best exemplars of Moore’s law, Impinj’s “self-adaptive” silicon co-opts much of the functionality of almost every nearby component. High-precision analog in digital CMOS: this is the road toward the cell-phone-on-a-chip.

Tall tails

Carver Mead had predicted this route in *Analog VLSI* (1989): “The best way to ensure that a circuit will tolerate… variations is to have it self-compensate….Self-compensation has another advantage: as circuits age and change and shift with time, the system tunes itself up.” Trouble was, Mead didn’t know exactly how to do it. So he pointed some of his best students down one possible path.

Early on in Diorio’s doctoral program, Mead asked him and two colleagues to find out what they could do with a floating gate transistor. A floating gate is a layer of polysilicon that “floats” between layers of oxide insulator. Sealed in oxide, floating gates are free from any direct electrical contact and are therefore good stores of charge for nonvolatile memory and other uses. Mead himself had researched these devices, which were invented in the late 1960s and are the foundation of EEPROMs (electronically erasable, programmable read only memories) and flash memory. He always thought their potential went unrealized, but what Diorio found surprised him.

“Look at these little tails,” Diorio told Mead, showing him some supposed noise in his data. “The tails shouldn’t be there….What are they?” Mead didn’t know. “It’s the only time I ever caught Carver,” says Diorio.

Electrons were in fact being injected into the floating gate of the transistor. It occurred in both p-FETs and n-FETs, the two basic types of CMOS field effect transistors. Diorio soon found that by using Fowler-Nordheim tunneling at the opposite end of the floating gate, he could also expel electrons from the transistor. He could even "read and write" at the same time. If he could add and subtract electrons from a transistor at will, a transistor that is a nonvolatile memory like an EEPROM, and do so at subthreshold voltages, Diorio realized he could create a synapselike structure that could both store and process high-resolution analog information. Unlike many previous attempts in floating-gate research, Diorio's transistor could morph through the use of local feedback, not an external or awkward on-chip mechanism. He had invented a nonvolatile analog memory with locally-computed updates on a single transistor. Impactionized hot-electron injection was a major boost to Mead's vision of creating neural systems in silicon.

Although Diorio's work, patented in 1996, earned a prestigious IEEE award, and served as his 1997 Caltech doctoral thesis, it was not until 2000 that he recognized the immediate commercial potential. Racing street cars at the Laguna Seca Racetrack in Monterey, it hit him. He called Mead, and that very night they had dinner at Fresh Cream, "Monterey's best restaurant seven years running." Diorio, by that time a University of Washington professor, told Mead his silicon synapses could change the communications IC market. A week later, Mead had Diorio in the offices of Silicon Valley's Venture Law Group signing incorporation papers.

Meanwhile, a friend and former colleague of Diorio's at **TRW** (TRW), Bill Colleran, had just sold his company, Innovent, to Broadcom. Innovent developed the chips that now form Broadcom's Bluetooth and 802.11 Wi-Fi product lines. Colleran, an articulate EE Ph.D. and Harvard J.D., quickly joined his friend Diorio in Seattle as CEO of Impinj.

The Mead method

Diorio and Colleran oscillate between hinting at Impinj's enormous potential and offering practical short-term product road-maps, but they know what they've got in those 14 patents. One big idea, is a massively parallel analog front-end, previously unthinkable because massive parallelism implies massive transistor mismatch. It would divide an incoming signal into perhaps hundreds of low-and-slow pathways to greatly reduce uncorrelated errors, jitter, and thermal and substrate noise. Diorio has shown how simple digital inverters aren't really digital at all: the entire transition phase—the continuous voltage swing between the "one" and the "zero"—can be used for analog signal processing. Diorio's doctoral thesis was also a major breakthrough in the field of neural networks: it envisions his transistors as the key building blocks of *field programmable learning arrays*.

With Foveon, Synaptics, and Impinj achieving major breakthroughs, Mead's method is igniting the Telecosm's second phase. Of the key analog players—Texas Instruments, ADI, Linear, Maxim, **Fairchild** (FCS) and the rest—only National, with its key role in Foveon, is sure to play more than a bit part on the world's analog stage.

—George Gilder

The Router on a Chip

Breaking out into every type of communications gear across the Net is a new general purpose chip that could be as central to networking as the Intel microprocessor was to computing. Way back in September of 2000, when thousands of networking, chip, and dot-com companies were still plowing through their stashes of venture and IPO cash, scores of companies in this arena known as "network processors" approached the *Gilder Technology Report* with big claims. Some 50 or more net processor chip companies, from the very large—**IBM** (IBM), **Agere** (AGRb), **Motorola** (MOT), and **Intel** (INTC)—to the very small—**Silicon Access, Internet Machines,** and **Bay**—had acquired and funded some $25 billion worth of new leading edge chip designs.

But only one company caught our eye with a truly original and paradigmatic solution. We endorsed the company, though then it was little more than a group of 50 engineers with dazzling PowerPoint slides. Today that company has survived all its competitors, big and small, most of whom have left the business or ceased to exist, and now finds itself in a head-to-head competition with the lone remaining player, Intel.

Emerging as the leader among scores of companies spending a total of some $25 billion in this network processing space, is our pick from almost four years ago: **EZchip** (LNOP). (See *GTR*, September 2000, January 2002, April 2002, June 2003.) It is the most promising solution because it puts memory and processing on the same chip. Embedding memory radically increases its speed—the rate at which data can be retrieved from memory and delivered for processing. Thus, it provides the only workable answer to the dilemma of Intel scientist John Shen, who recited the *GTR*'s memory mantra to *EE Times*: "My personal view is that memory is the predominant performance bottleneck. CPU speed increases 40 to 50 percent per year. However, memory speed increases at a paltry 5 percent per year. That gap will continue to widen. Today it takes 100 to 150 clock cycles to access main memory for one to two gigahertz CPUs. That could expand to several hundred clock cycles in the foreseeable future."

EZchip seeks to usurp first generation network processors that integrate several off-the-shelf RISC (reduced instruction set computing) machines on a chip and couple them to separate co-processors. General purpose devices, RISCs cannot delve deep into packets. Other net processors use a mix of RISCs and ASICs, charging the RISC with the core processor tasks and the ASICs with specific high-speed jobs. ASICs add to the number of chip interfaces and by definition are not programmable. All first generation net processors use off-chip memory, imposing a 32- or 64-bit limit on the links—the buses—between processor and memory.

With a superior design and order-of-magnitude cost and performance advantages, EZchip is poised to prevail

Five megabytes of on-board DRAM (2 MB for the buffer and 3 MB for searching) and buses as wide as 512 bits mean EZchip can go beyond reading simple headers in network layers 2-4 and deep into the strings of text-based data in layers 5-7. Processing layers 5-7 is essential for such functions as server load-balancing and per-use accounting of web-based video or software applications.

Scalable and cascadeable, EZchip's first product, the NP-1, could process eight ports of 1 gigabit Ethernet, one port of 10 gigabit Ethernet, or one port of OC-192 SONET.

EZchip's team comes from the networking industry rather than the chip industry. Their previous project implemented Token Ring networks, which were elegant in concept but succumbed to the superior robustness and momentum of Ethernet. With a focus on 10 gigabit Ethernet, this time they have got it right, conceiving a perfect device to reduce the router to a chip and put it everywhere.

After delaying for several months to perfect the complex design and tape it out, in the winter of 2002 EZchip finally announced eight customers and a sampling date (March 2002) of its flagship NP-1 chip. Manufactured by IBM, the chip gained nearly all of its advantages—less power, less board space, more look-ups, more bandwidth—from IBM's mastery of embedded DRAM technology. From **Texas Instruments** (TI) to **Micron** (MU) and **LSI Logic** (LSI), many macho fab experts have come a cropper on the treacherous intricacies of integrating DRAM cells, optimized for large capacitance, with CMOS transistors, optimized for as little capacitance as possible.

Though numerous devices claim network processor status, real network processors have 4 tasks: classifying packets (identifying and parsing headers and fields); searching IP look-up tables (finding addresses); resolving packets (assigning destinations); and modifying packets (prioritizing, scheduling, tagging, and policing them). Processing 10-gigabit Ethernet or SONET streams at wirespeed, moreover, requires at least 320 Gbps of memory bandwidth—160 gigabits to buffer memory and 160 gigabits to look-up table memory. To perform all of these tasks at 10-gigabit wirespeed is extremely hard. Indeed, while OC-192 optics has been on the market for more than four years, only recently have companies been able to craft true 10-gigabit transceivers, and no one has produced a 10-gigabit network processor. EZchip was the first to execute all these functions in one large chip—the NP-1.

Power hungry competitors

To get a feel for the size and complexity of the NP-1, we compared it to a 2.2 GHz Intel Pentium 4. At the time, the newest Pentium had 478 pins, or external wire connectors. EZchip's NP-1 had 1,247 pins. Pentium 4 had 520 kB of SRAM (an 8-kB Level 1 cache and a 512-kB Level 2 cache). With 4.2 MB of DRAM plus 1 MB of SRAM for microcode, NP-1 had 10 times as much on-chip memory. Pentium 4 had 32 Gbps of memory bandwidth. NP-1 had 500 Gbps. Pentium 4 dissipated 55.1 Watts. NP-1 dissipated just 15 Watts (largely because it runs at 200 MHz, rather than 2.2 GHz). While they are becoming powerful enough to take on some network functionality, like administering simple encrypted VPNs, PC microprocessors are not suited to most high-speed networking tasks. A functional comparison could thus only be made among EZchip's then-numerous peers, who were claiming breakthroughs and imminent sampling of real silicon chips.

While EZchip used 1 to 10 devices, including external memory, and dissippated a total of some 20 Watts of power, its competition lagged far behind. One net processing start-up with products due in 2002, Silicon Access, used at least 21 chips consuming some 60 Watts. Another vendor's 10-gigabit "network processor" was said to be contrived using 51 chips that dissipated 154 Watts. Competitive offerings from sector-leaders IBM, **Applied Micro Circuits** (AMCC), and Agere, due later that year, required 56, 60, and 76 devices, respectively. Power dissipation might have reached 200 Watts. Imagine, then, multiplying these chip-counts and power figures by many, many Ethernet ports per switch, and you quickly reach hundreds or even thousands of chips consuming thousands of Watts. These offerings were less network processors than network mainframes.

Bart Stuck and Michael Weingarten of Signal Lake Ventures showed that one line-card produced with such an architecture would manufacture for $10,000, implying an ultimate market price of some $90,000. (Confirmation: One port on **Ciena**'s (CIEN) CoreDirector optical switch priced at $100,000.) EZchip CEO, Eli Fruchter, modestly believes the NP-1 will replace at least 10 components in a typical line-card, enabling a Cisco or Juniper or Ciena to reduce its non-optical component costs by 75 percent. Using Weingarten and Stuck's assumptions, however, building with EZchip saves you closer to 90 percent. Fruchter believed that his company was one generation ahead. Would you believe two generations?

In conjunction with 64 parallel and pipelined "Task Optimized Processors" and patent-pending search algorithms, EZchip's DRAM enables extraordinary access to look-up tables and buffer memory. Multiple busses, each from 256 bits to 512 bits wide, connect at 200 MHz to 4 DRAM cores totaling 4.2 MB to attain its 500 gigabits per second of on-chip memory bandwidth. Multi-chip solutions connect to external DRAM at 64 bits and must endure a longer path as well.

Fruchter emphasized reduced cost and power dissipation, greater port-density, and manufacturing and programming simplicity. But because the product existed only on paper, we thought he may be underestimating his device's edge in pure performance. It was difficult to believe, for instance, that alternatives using 50-plus chips could actually deliver on the goal—a faster, more robust Internet. By keeping its packets on-chip, EZchip limits their electronic lives and maximizes their photonic lives. Packets do not get lost. Traffic jams are avoided. Latency is reduced. Fewer components mean fewer points of potential failure. In the microcosm, smaller is better.

The difficulty in designing network processors derives from their broad flexibility and applicability, but that also points to their large potential market. Just as Intel Pentiums are used in low-end $1,000-desktops and high-end servers alike, net processors may someday find themselves in everything from small firewall boxes to large core routers. Outside the domains of the all-optical network but everywhere on its edges, no piece of communications equipment that stands in the data path of the Net will be immune—from storewidth appliances to 3G wireless base stations. Cahners In-Stat projects network processor unit sales will increase from approximately 2 million in 2002 to over 20 million by 2005. In the same period, revenues from these sales are expected to vault from $1 billion to over $10 billion. The size of the market will depend on how cheaply the chips can be made. But with a superior design and order-of-magnitude cost and performance advantages, we believed EZchip was poised to prevail.

The key to EZchip's apparent accomplishment is to break what might be called the Intel bottleneck. Although the world is preoccupied with the last mile bottleneck, the bandwidth bottleneck that concerned me first-and the one that gave rise to our entire system-on-a-chip (SoC) paradigm-was the bottleneck inside the computer itself. The logic-memory bottleneck.

It was clear even in the mid-1990s that although the density of both memory and logic scaled with Moore's law-doubling every eighteen months-the performance of memory, where bits are stored, lagged that of logic, where bits are processed. As *GTR* editor Nick Tredennick and Brion Shimamoto explained in a March 2002 report, "In 1981, microprocessors and DRAMs were about the same speed…" At 4.77 MHz, the clock of an IBM PC processor ticked once every 210 nanoseconds, and it could access its 64 Kb DRAM chip once every 225 nanoseconds. Virtually identical. The chips ran in synch. Over the last 20 years, however, even as DRAM capacity has kept pace with Moore's law-witness the gigabit DRAM available from **Samsung** (05930.KS) and others next year-the time it takes to search that memory has not. A vast and crippling performance chasm has opened. In 2002, leading edge microprocessors are 60 times faster than DRAMs.

As Amdahl's law tells us, a system is only as fast as its slowest component. Thus, clock speed advances, now beyond 3 GHz in the Pentium 4, are wasted. Depending on how often the Pentium 4 must go "off-chip" to access the DRAM instead of its small on-chip cache, a 100 percent increase in clock speed can result in a systemic performance increase of just a few percent. The upshot is that the clock frequency is no longer a limiting factor in PC performance. Most of us could not tell the difference between a 1 GHz computer and a 2 GHz computer, but we all know it when our RAM is doubled.

This bandwidth bottleneck takes on new meaning as we move out of the world of web surfing and spreadsheets and into the world of packet processing on the high-speed optical Internet.

There the challenge is to parse, sort, alter, and route ten billion bits of disparate, far-flung data in a single second, a task far too speedy and complicated for a Pentium. Today a router or switch is a box, a big, expensive box full of custom ASICs, memory, classifiers, and the wires and interfaces connecting them all together. Tomorrow the router could be a chip.

The company closest to creating a router on a chip is the same one that is most aggressively dissolving the bandwidth bottleneck between logic and memory, the one transcending the off-chip light-speed delay, the one potentially cutting the cost of a 10 gigabit router by 90 percent, enabling a faster, smarter, more profitable Internet. The company melding logic and memory is EZchip.

To route a 10-Gbps Internet stream, you need at least 320 Gbps of memory bandwidth, and you need to store more than one million IP look-up tables. But a Pentium 4 doesn't come close. While the link between the Pentium 4's two on-chip SRAM caches achieves 384 Gbps, it has 10 times less on-chip memory resulting in some 10 times more off-chip searches. Going off-chip to a gigabit DRAM yields just 25 Gbps. EZchip's ingenious design puts 5 MB of DRAM on the same chip as its 64 custom processors, eliminating most of the time-consuming off-chip searches, and squeezing out 500 Gbps.

Fab-u-lous IBM

Without a partner, however, EZchip's ideas would have remained in slideware. The router on a chip will not be realized without the help of embattled Big Blue. Although distracted by SEC investigations and earnings disappointments, IBM is increasingly the across-the-board leader in process technologies.

From silicon germanium to copper interconnects, and from 90-nanometer geometries to embedded DRAM, IBM has mastered silicon manufacturing and become the fab of choice for designers of advanced microchips. In a recent Electronic News article, rivals claim IBM is struggling with "yield and material issues" in its 130-nanometer process. If that is so, asks the same article, why did IBM's ASIC business grow in 2001, even as its closest competitors suffered 20 to 40 percent revenue declines?

One fabless chip company CEO summed it up: "IBM is very demanding, and they may not be the fastest, but they are meticulous-and they deliver chips that work."

EZ does it?

EZchip followed this latest trend and went to IBM to solve the embedded DRAM challenge. Logic gates must be fast, so they tolerate "leaky" transistors that never turn off completely. The threshold between a "1" and a "0" is somewhat fuzzy. But DRAM transistors need to hold their "1s" and "0s" on a capacitor for precious microseconds to read and write the stored data. Logic chips and dynamic memory chips are therefore built differently. Most on-chip memory is static random access memory (SRAM), which is highly compatible with logic circuits but is at least six times less dense than DRAM. EZchip, running at 200 MHz, can outpace a speedy 2 GHz Pentium because it chooses the path that's "low and slow." Its parallel processors more closely couple the clock frequency with the memory speed, yielding increased performance at just one-quarter the power dissipation.

Some DRAM suppliers are developing double-data-rate DRAM (DDR DRAM) ahead of the industry-standard schedule-attempting to deliver 400 MHz products this year, while DDR-333 MHZ isn't due until 2003 and DDR-II-533 MHz until 2004. Even with DDR-400 MHz, however, system level bandwidth is just 25 Gbps, an order of magnitude less than the 500 Gbps that EZchip achieves by placing the DRAM and logic on the same chip.

A year ago, EZchip still faced at least ten serious rivals. Now all its competitors have fled the field except for Intel Corporation. Although currently producing no chip remotely comparable in capability, Intel is fully committed to the network processor arena, already rules the low end, and will be a serious player as time passes. But with some 30 design wins, including more than ten large systems vendors, one in the U.S., EZchip has become the overwhelming technological leader, with impressive commercial prospects.

Laying down the law

In the course of writing my book on Foveon, I have had to consider the strategic enigmas of breakthrough innovations. In moments of weakness and enthusiasm, Mead, Merrill, Lyon and others at Foveon speak of transforming the camera business. That is not a promising goal. The reason why it is not a promising goal was identified by my partner Nick Tredennick. As a sometime amateur legislator, I would like to make a law about it.

As a law giver Carver Mead is preeminent, since he named Moore's law ("chip capabilities double roughly every 18 months"), but I rank high for naming Metcalfe's law ("the value of a network rises by the square of the number and power of the machines attached to it"). If I may resume a legislative toga for a minute and appropriate my partner's insight, I would like to announce Tredennick's law: *"Seek performance and you do not get volume. Seek volume and you get performance."*

Catchy isn't it? The essence of it is the learning curve. Making the argument with authoritative data and detail is Harvard's Clayton Christensen in his forthcoming book, *The Innovator's Solution.* Creating a high performance product is only the first step. If you make one brilliant prototype of a magical Silicon Wonderchip XXX, and then embark on an agenda of costly performance improvements, you will restrict yourself to a sparse population of elite users. In the end, this small market of demanding buyers—whether of high-end cameras or high-end routers or specialized business communications—will not be able to pay for the early rate of improvement. Meanwhile your rival—Intel, perhaps—incorporates an inferior rip-off on some underused corner of a Pentium and makes billions of units. Moving down the learning curve of the semiconductor industry with Moore's law, the Pentium will soon be doing the job more cheaply and better than your Silicon WonderchipXXX.

Foveon can achieve volume by putting the chip in every cellphone, ATM machine, airport security booth, PC monitor, convenience store corner, and digital watch. Similarly EZchip must move rapidly down the curve from its current niche of routers and switches. It must dismantle the router and put it everywhere. Promising are its some 20 design wins for a variety of metro switches, IP routers, server switches, firewalls and virtual private networks, wireless hubs, load balancers, and the trusted computing association platform. It will not finally prevail, though, until it is in every settop box, wireless router, and home entertainment hub, seamlessly shuffling, converting, and handing off the variety of Ethernet, NTSC, HDTV, USB, Firewire, DOCSIS, IPv4-v6, MPEG 2-4, Foveon, Flash, 802.11X, DIVX, FM, SCSI, Fibre Channel and DS0 connections, and distributing them among a variety of terabyte storage facilities and on sundry displays, monitors, phones, speakers, and personal digital assistants. As the only single chip network processor that operates at all seven layers of the network, from phy to presentation, EZchip can become dominant by seeking volume in the domicile where all the layers converge.

The coming challenge for all our innovators is to find strategies to build up volume. Foveon took a key first step by licensing its technology to National for low-end imager applications. EZchip is preparing to enter new businesses. Lower prices bring higher revenues and expanded markets. That is the ultimate harvest of Tredennick's law.

By late 2002, EZ was still looking for major customers but had already announced its second generation chip, the NP-1c. To be manufactured in IBM's .11-micron process, the chip would be 30 percent less expensive than the original, at $795, while delivering a 50 percent performance increase. NP-1c would enable any solution, from a small firewall box to a metro optical router, using just four chips, the EZ net processor, plus four cheap DRAM memories. Hurtling beyond Moore's law, young EZchip was delivering 300 percent gains in cost effectiveness per year.

The implications of such an advance permeate the entire networking industry.

As EZchip integrates more and more memory and processing devices onto single CMOS chips, **Cisco** (CSCO) will probably continue to purchase them in ever greater numbers. But in time they will comprise most of the value of router hardware. Cisco will become a box assembler like **Dell** (DELL). Soon enough the router will go away. It will become an Intel or a **Broadcom** (BRCM) or an EZ chip.

A rough rule of the Telecosm ordains that hardware softens on the edge of the network and software hardens at its center. The network processor represents a software intensive router. As Cisco CEO John Chambers sometimes seems to recognize, the likely outcome is that Cisco will retreat from its hardware revenue addiction to a role as a networking mutual fund and a software bastion of intellectual property. Already most of the value of Cisco boxes resides in software: its Internet Operating System, Border Gateway Protocol, its Open Shortest Path First algorithms and all the other code structures that underlie most of current Internet architecture. A street map of Cisco City, this is a rich vein indeed. But what happens when the vein turns into glass?

EZ's new abundance

An additional EZchip advantage came into focus in late August 2003 at our annual Lake Tahoe technology conference, Telecosm VII. With planned dinner speaker Ivan Seidenberg of **Verizon** (VZ) held hostage in his copper cage by his union, we asked Bob Metcalfe to run a panel on a development that gives the edge to EZchip in several new markets. That development is a new Internet Protocol, IPv6, that opens up a vast new abundance of potential connectivity.

Technologies advance through an interplay of abundances and scarcities. Entrepreneurs exploit the abundant resources to relieve the scarcities. They use oil and gas to save human muscle; they tap the compacted fossil fuels and uranium of the earth's core to preserve the arable and aesthetic spaces on the surface from visual pollution and agricultural exhaustion. On this point, Howard Hayden of the *Energy Advocate* offered the best trope of the conference. Commending the increasing practice of putting unsightly power lines under ground, he proposed a similar solution for windmills.

Both entrepreneurs and economists live in a world of scarcity. Only the entrepreneurs see the abundance beyond.

Opening the world to a new abundance of potential connectivity is IPv6. After surmounting the Y2K problem, the digital infrastructure once again faces a challenge of too few bits. A few years after the intrepid mainframe geeks inaugurated their famous two-digit shortcut for dates, they gave birth to ARPANET, the early Internet, using packets that were routed via the network control protocol (NCP). By 1981, the Vint Cerfs and Bob Kahns of the world had worked out the lower-layer protocols of what that came to be called TCP/IPv4 (transmission control protocol/Internet protocol version four). With each IP address 32 bits long, the total "address space" of the Internet totaled an apparently inexhaustible 4.3 billion potential nodes.

Within the next decade, however, an onrush of net-connected machines—including cars, phones, cameras, wireless mesh relay stations, and all manner of remote sensors—will bolt past the 4.3 billion IPv4 address limit (in reality, the effective limit is about 3.3 billion). The numbers point to an eventual address shortage as Asia comes online and more and more devices are digitized and connected to the net.

The successor to IPv4 is of course named IPv6. Version six is vastly larger than version four: the difference between 128-bit addresses and 32-bit addresses. This is the point in the speech where v6 futurists start babbling about IP addresses for all the leaves on the trees or the insulin molecules in your pancreas. Suffice it to say such talk distracts from the real and imminent importance of the expanded address space and functionality.

Although the American builders of the Internet have been slow to adopt IPv6, Japan, China, and India have moved to mandate its use. Then came a U.S. breakthrough when in August 2003 the U.S. Department of Defense announced that starting October 1 all equipment deployed into its Global Information Grid (GIG) must accommodate the 128-bit space. Because most companies want to bid for the government's vast GIG-BE (Bandwidth Expansion) projects, this announcement likely will push wide adoption among Western telecom equipment companies.

Unlike Y2K, however, this Internet transformation is not about a one-time labor-intensive patch for a problem but about real technological capability. It is the same challenge we have been writing about for at least seven years. Logic MIPS (million instructions per second) and storage bytes are abundant when it comes to *computation* but not necessarily when it comes to "wire speed" digital *communication* among billions of potential nodes at 10 gigabits per second.

When pushing packets across the network, digital telecom devices perform two key tasks. The first is a routing table look-up, which analyzes variable-length data known as longest prefix match (LPM) and then provides the destination address of the packet. The second is flow-classification, which governs the treatment the packet receives, be it type of service (ToS), time to live (TTL), or other provisioning instructions for special applications, users, security, and accounting.

Most current routing architectures from Cisco and its cohorts employ a central computer, usually a network processor or ASIC, and large numbers of content addressable memories (CAMs) and static random access memories (SRAMs). CAMs include comparison logic with each bit of storage and SRAMs provide more than ten times quicker access times than their cousin, dynamic random access memory (DRAM).

Although costing 35 times more per chip, CAMs and SRAM are just one-thirtieth as dense as DRAM and one-tenth as power-efficient. EZchip's fundamental insight was to shun high performance CAMs and SRAM in favor of slow, cheap DRAM. The new abundance was on-chip bandwidth, made possible by new semiconductor process techniques pioneered by **IBM** (IBM). Embedding very dense DRAM into a logic device, EZchip could eliminate most off-chip communications, using on-chip bus widths of 1,024 bits rather than off-chip busses of 64 bits, and drastically decreasing the round trip path-length between logic processors and memory cells.

EZchip's solution of one highly integrated network processor, the NP-1, plus four commodity DRAMs handles every proposed IPv6 implementation, from the low-end device with 125K look-up entries to the 2 million-entry core router. Where memory costs using CAMs and SRAM range from $1,500 to $23,000, memory costs for EZchip never exceed $28, the total of four DRAMs at $7 apiece. The four DRAMs, meanwhile, dissipate just 2 Watts versus a minimum of 23 Watts and a maximum of 532 Watts for conventional CAM and SRAM solutions.

Where the CAM and SRAM-based solutions struggle to perform tasks, moreover, EZchip's NP-1c-DRAM combination handles IPv6 with 88 percent of the DRAM memory to spare for future updates of software and look-up tables. Thus time in market is extended years beyond normal solutions.

EZchip's unique IPv6 capabilities have already been decisive in several of its 30 design wins, especially in Asia. In August 2003, EZchip pushed its advantage further by providing IPv6 software for its chips.

Many vendors and service providers have gotten around IP address shortages by using network address translation (NAT) boxes that create artificial IP space behind and around the official realm. An IP-addressable NAT router might sit in front of an office LAN, for instance, eliminating the need for each Ethernet switch, PC, and server behind the NAT router to have its own IP address. But the NAT patch will only take us so far. Beyond a certain point, the Internet becomes an interNAT, with cascades of new complexity, rigidity, and cumbersome hierarchical structures. With whole nations and governments mandating the switch to IPv6, the new protocol is now on its way and so is EZchip.

—George Gilder and Bret Swanson

Telecosm Lounge

The Telecosm Lounge is a diverse community of learned financial advisors, engineers, technology professionals, everyday investors, and the *Gilder* staff collaborating daily on new investment ideas.

Here's how *Gilder Technology Report* subscribers can register for FREE access to **the Telecosm Lounge,** the premier technology message board on the Net.

- Visit www.gildertech.com today
- Login using your *GTR* subscriber account number
- Click 'Telecosm Lounge' (a square registration box will appear)
- Choose a user name and use your account number as your password.
- You are now a registered member of the Telecosm Lounge!

George Gilder is a frequent participant in the Lounge, answering subscriber questions and composing essays on technology, investing, economics, and politics.

The Telecosm Lounge features a special forum called "**George Says**" featuring all of George's posts and the related subscriber discussion threads, making it easy for subscribers to keep current with George's latest commentaries and to ask questions about the most recent newsletter.

The following is a sample of the quality content you can expect to find in the Lounge. It's what keeps *GTR* subscribers coming back and why many have come to consider the Telecosm Lounge the most valuable tool in their investment arsenal.

Author: George Gilder
Subject: Network processors
Date: January 5, 2004

"I sense in EZChip (LNOP) a bias toward upmarket slots similar to Foveon's orientation to Single Lens Reflex multi-thousand dollar cameras. Thus while Foveon first attacked Canon and Sony where their defenses were strongest—among their professional camera customers—EZChip seems to be confronting Cisco and Juniper in their most cherished markets by supplying high level capabilities to various fringe router vendors. But the ultimate benefits of single chip systems come on the edge—cheap devices for high volume consumer-type applications where Cisco will not readily follow. I am not saying that EZchip has not pursued the edge slots, only that it has positioned itself chiefly as a top-of-the-line full duplex, 10 gigabit per second chip adaptable to multi-protocols and linkable to thousands of ports. That is where Cisco lives and where Cisco must respond most aggressively to a threat.

In both cases, the challengers may have had no choice: When you can produce only dozens of chips, you had better target them where they do the most good. But that possibly necessary strategy raises the risks of EZchip somewhat. Faced with a tiger like Cisco, you don't want to wound it. Use a disruptive strategy oriented toward customers who do not use routers today at all, such as home networks, DSL and cable modem aggregators, wireless LANs and WANs, cheap firewalls and concentrators, load balancers and nodes for small businesses. Bypass the tiger first, then feed it, and only try to kill when it attacks you.

I believe that LNOP is a supreme opportunity. But it is not devoid of serious risk if its customers fail to gain ground in the market."

Trouble accessing the Telecosm Lounge? Email customer support at info@gilder.com