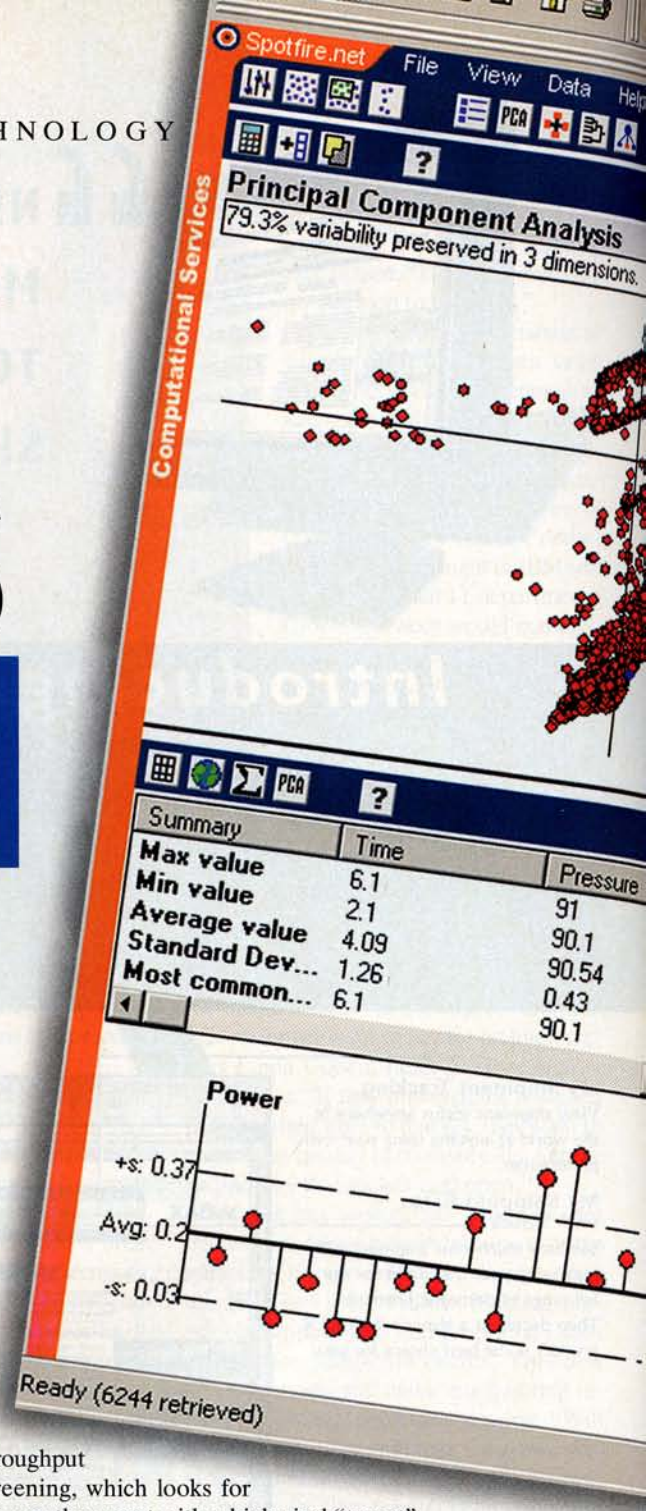
Making decisions in a flood of data



There's sunken treasure in industry's deep databases. Here's a breakthrough way drugmakers and others find it.

by Stuart F. Brown

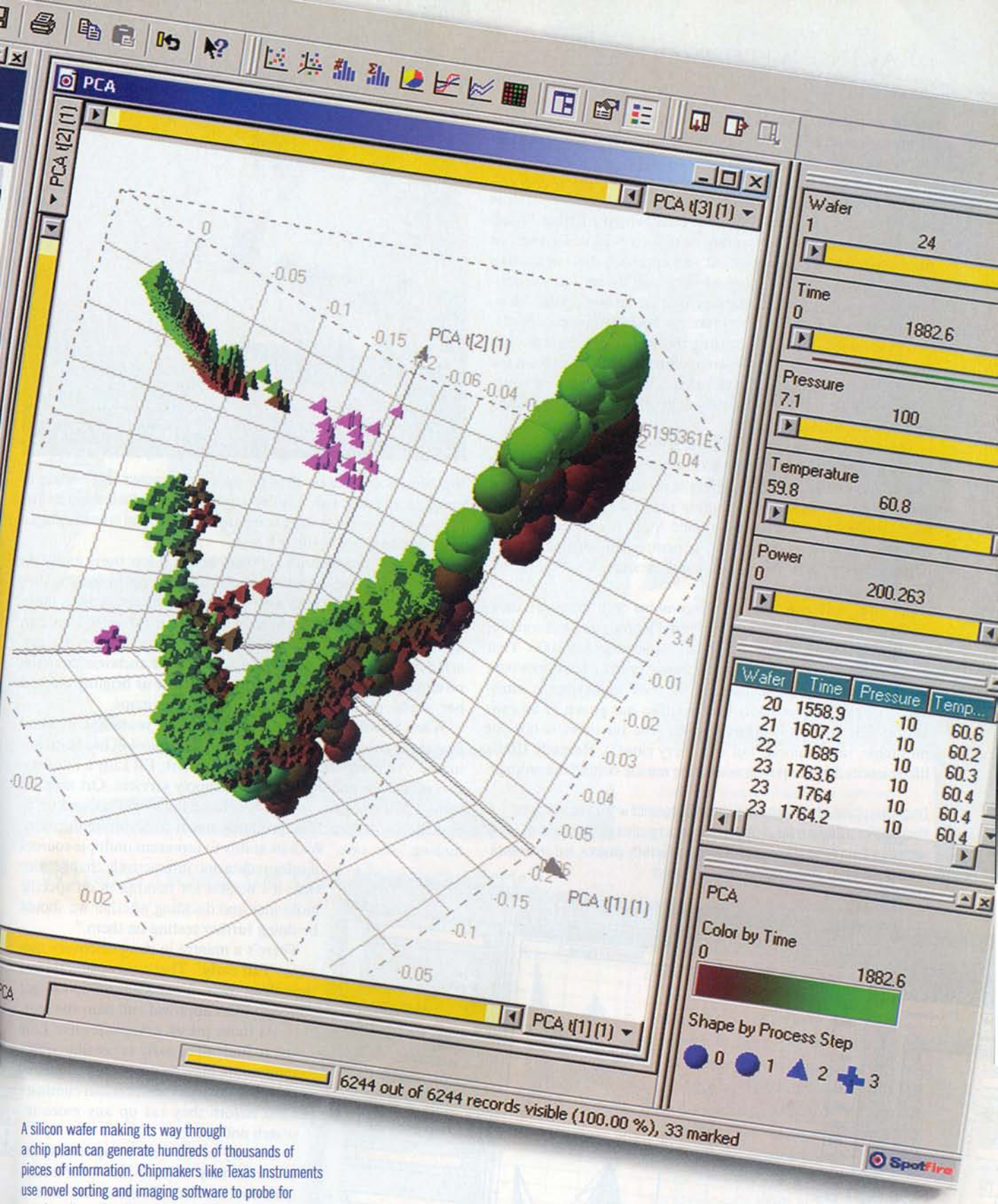
Businesses are drowning, just drowning in data. And they asked for it. Computers today make it relatively easy for industries ranging from pharmaceuticals to computer chips to oil exploration to amass vast archives of information. The data can be of the utmost importance in figuring out where to drill an oil well, or how to correct a costly flaw in a chipmaking process, or which of a zillion chemical compounds to bet on as the next blockbuster drug. But the payoffs go only to those who can somehow drag a net through tidal waves of raw information and capture the little fishes of opportunity swimming within them.

Consider the data-sifting challenges three companies face:

• Eli Lilly & Co. in Indianapolis uses an automated process called combinatorial chemistry to synthesize small but useful quantities of chemical compounds by the thousands in a few days' time. Then the chemicals are pumped into a second process called high-throughput screening, which looks for the way they react with a biological "target" substance. Hidden in the torrent of outcomes can be promising candidates for new medicines.

• Geochemists and engineers at Anadarko Petroleum, an oil-exploration and -production company in Houston, have rafts of seismic and other data available to them as they try to divine the underground locations of worthwhile quantities of oil and natural gas. Where precisely should they tell the expensive drilling crews to aim the bit? The answers are somewhere in all those data.

• Texas Instruments at its Kilby Fab in Dallas develops chipmaking methods and produces microprocessors in a sequence that routes silicon wafers through more than 400 delicate manufacturing steps over many weeks. Swarms of sensors monitor the machinery, gathering more than 140,000 pieces of information about each wafer as it progresses. Somewhere in that heap of data



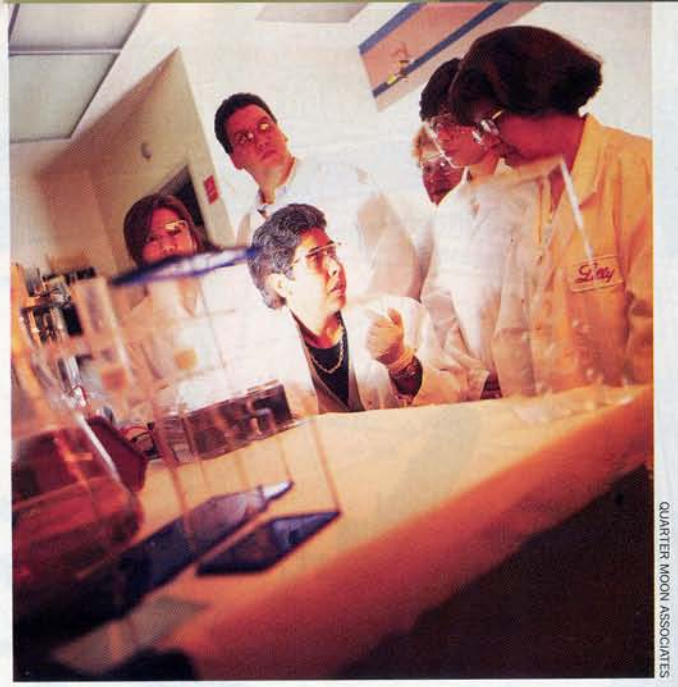
A silicon wafer making its way through a chip plant can generate hundreds of thousands of pieces of information. Chipmakers like Texas Instruments use novel sorting and imaging software to probe for warning signs of costly problems. In the main image above, the blobs are data about 24 wafers in a "3-D scatter plot." By using a mouse to adjust the yellow "sliders" at top right, troubleshooting engineers can explore the influence of variables like temperature, pressure, and power to which the wafers have been exposed, until trouble patterns emerge.

can be warnings about things going awry in the process. Detect a bug early, and a fix can often be devised before any bad chips are made.

A new idea in software is beginning to help these companies reduce the time and money they spend searching for patterns and meaning in their data oceans. Internet guru Esther Dyson compares this method to "putting on different-colored glasses, or filters, to see how things look." It's an approach that started as a doctoral thesis by Christopher Ahlberg, the 32-year-old Swedish-born founder of software company Spotfire, in Somerville, Mass. Part of his inspiration was to make probing the contents of disparate databases similar to surfing the Web with a browser. His other goal was to use visually compelling displays to present the results in ways that mere mortals can quickly and intuitively grasp. "Databases were the last area where graphic design was waiting to happen," Ahlberg says.

The databases most companies rely on were not built for ease of use. They typically use an access method called Structured Query Language, or SQL, which is so difficult to learn that many organizations employ highly paid specialists to talk to their databases. With literal-minded SQL, it's not hard to inadvertently ask a question that generates millions of hits—a huge, useless result. Worse, the databases aren't set up for users who mostly want to explore.

Spotfire's software is the first to combine both "data visualization" and powerful querying flexibility. Known as DecisionSite, the software isn't cheap—installations start at $100,000. That hasn't stopped customers in a wide range of industries from buying more than 16,000 licenses. In the four years since it introduced its product, privately held Spotfire has grown to an estimated $30 million in annual sales (the business is not yet profitable, says Ahlberg, but it is "very close"). Recently IBM's life-sciences division put its marketing muscle behind the product:



QUARTER MOON ASSOCIATES

Researchers at Eli Lilly want to decrease the time and cut the cost needed to develop drugs.

Big Blue is combining its data-management software with Spotfire's tools in a package aimed at drug companies that hope to speed up their R&D.

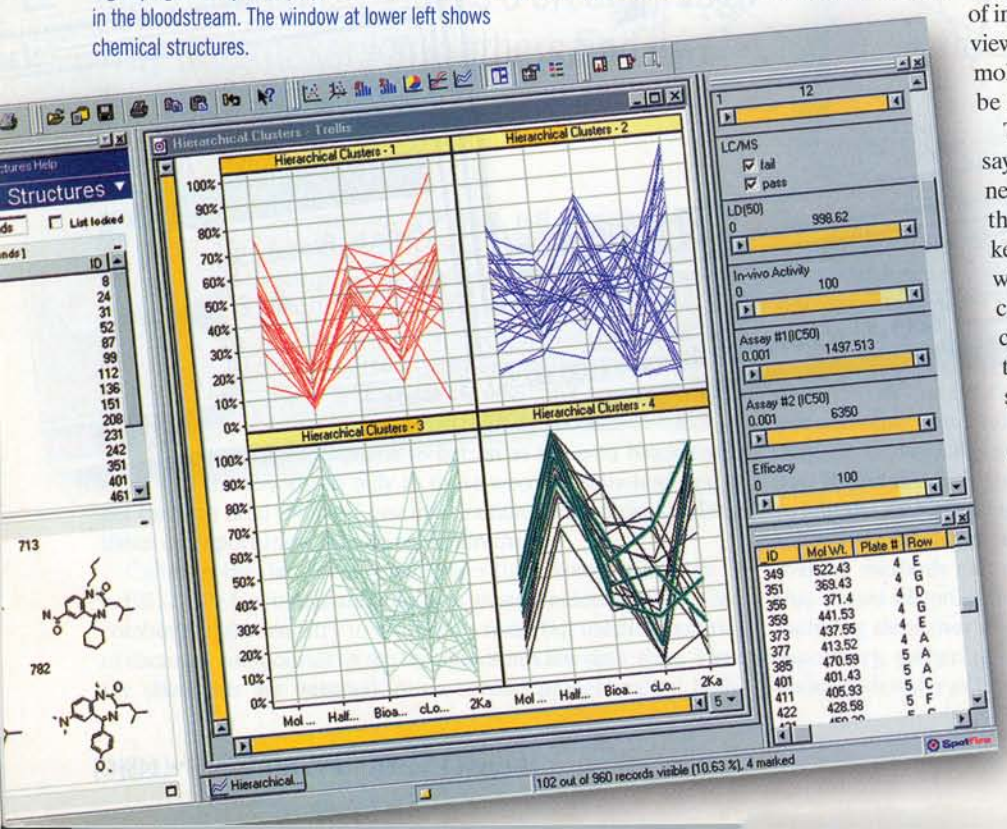
The magic in Spotfire's software is that it lets users easily do what-if comparisons of data from different sources by moving sliders on a computer screen with a mouse. In effect, it gives these data fishermen infinitely variable nets to trawl with; they can search beneath the waves for fish no longer than three inches, say, and then, with a quick adjustment, separate the anchovies from the sardines for comparison. The results appear as brightly colored bar graphs, pie charts, scatter plots, and even maps.

When Spotfire rolled out its software four years ago, it aimed first at the drug industry, where the data explosion has been immense. An early adopter was Sheldon Ort, Eli Lilly's information officer for manufacturing and supply services. Ort now has some 1,500 company scientists around the world hooked up to Spotfire's software. "We primarily use it to facilitate decision-making," Ort says. "With its ability to represent multiple sources of information and interactively change your view, it's helpful for homing in on specific molecules and deciding whether we should be doing further testing on them."

There's a mantra in drug discovery that says, "Fail early." That's because a typical new drug costs about $500 million to get through FDA approval and onto the market. At those prices, companies like Lilly want to find out as early as possible which compounds may be toxic to patients or cause unwanted side effects and eliminate them before they eat up any more research dollars.

As they study new compounds, the chemists may want to know, for example, whether the substances are able to traverse the body's blood-brain barrier. What toxicity knowledge exists about them? What's their chemical structure? How quickly does the body absorb them and how long do they persist in the bloodstream? How may they react with the other ingredients

Drugmakers must sift through thousands of compounds in the quest for cures. Here Spotfire software helps chemists analyze compounds by grouping and regrouping them by such properties as biological activity, potency, and persistence in the bloodstream. The window at lower left shows chemical structures.



of a pill? The list of key factors can be 20 or more items long. When these pieces of information are stored in different databases, as they often are, comparing them can be a hugely time-consuming process.

Using Spotfire, researchers avoid having to construct multiple queries in perfect syntax. Dragging the sliders to and fro, the user is actually launching a sequence of queries in rapid succession and seeing the outcomes expressed graphically onscreen. Lilly uses the software to conduct meetings among researchers at multiple sites who are linked on a computer network. As the person making a presentation moves the sliders on his or her screen, everyone can see the families, clusters, outliers, gaps, anomalies, and other statistical nuggets that database users fish for. Ideas can be tried out collaboratively in real time. Ort is now experimenting with using Spotfire to streamline Lilly's supply chain.

Ahlberg says he had never thought about applying his software to the energy business until Anadarko called. "I didn't see any analogy between drug discovery and the oil and gas industry. Then I realized that the oil and gas equivalent of the chemical-structure libraries that drugmakers want was maps—geographical maps." Spotfire formed an alliance with ESRI in Redlands, Calif., a prominent supplier of geographical information systems. The final touches are now being put on a version of DecisionSite incorporating a map visualizer feature for Anadarko's people to use in planning oil-field operations.

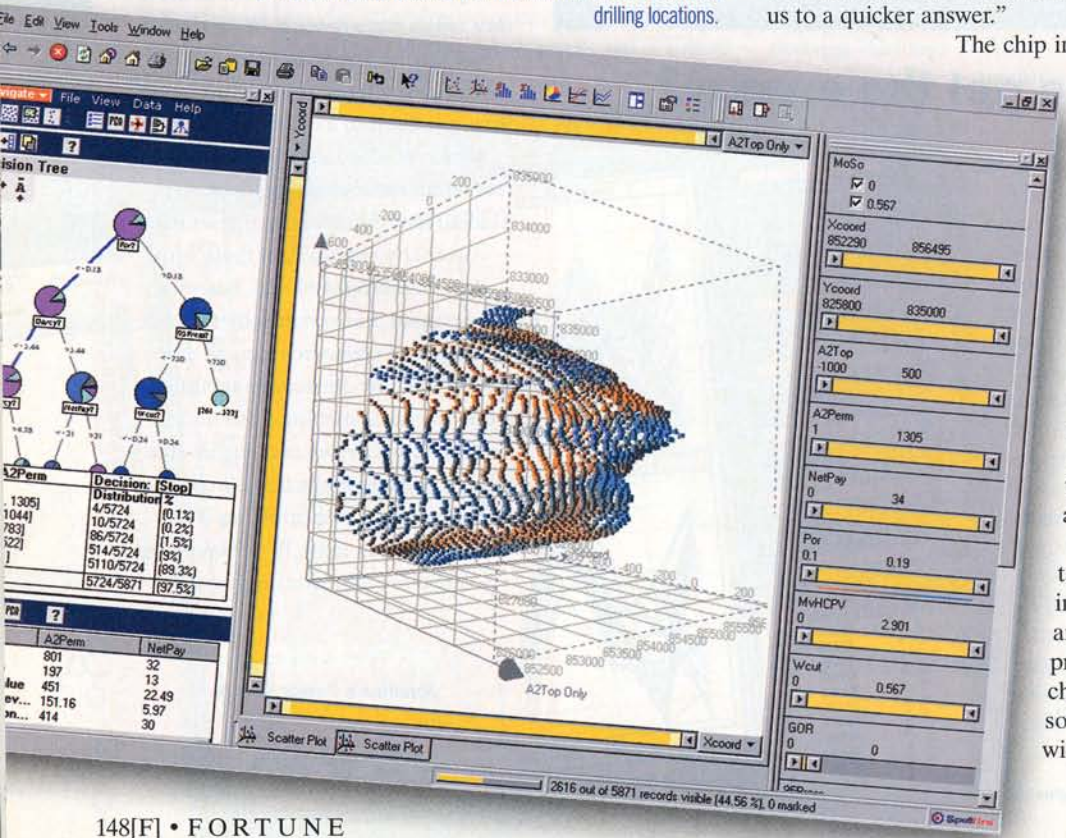
Says Ron Bain, manager of international exploration at Anadarko: "A lot of the databases our geoscientists and engineers query read out as Excel spreadsheets, and comparing 1,000 data points from a few of them is mentally difficult. Spotfire is like Excel on steroids. It does all the cross-plotting instantaneously, looking for trends."

Anadarko Petroleum's profit formula is to "drill where oil should be, not where it could be."

With oil prices up, Anadarko is starting a new well somewhere in the world every five hours. To decide where the wells should be, the company's geophysicists consult databases of magnetic data, gravity data, information from sensors in existing wells, and the results of seismic surveys of subsurface rock structures. Disk drives fill quickly in this business. A survey of a nine-square-mile block of Gulf of Mexico sea floor, for example, comprises five or six gigabytes of data; Anadarko has thousands of such survey blocks in its gigantic 20-terabyte archive.

The promise Spotfire holds is "streamlining the decision process," Bain says. "I can have a lot of information without it making me very smart or very successful. The idea here is to find oil economically and produce it. We like to say that we drill where oil *should* be, not where it *could* be. These DecisionSites guide us to a quicker answer."

The strange shape at center is a 3-D "scatter plot" of an oil field showing the underground chemical makeup at various sampling sites. Among the sliders at right, the one labeled "NetPay" links to a database about the economic performance of wells in similar fields. The decision tree helps the user select promising drilling locations.



The chip industry faces huge capital costs—new fabs start at more than $1 billion apiece—and adds immense value to its products as they flow through the complex and highly sensitive manufacturing process. Catching little mistakes before they become big ones is essential to pushing the fab's yield of good chips to a profitable level. Chipmakers rely on process-monitoring sensors, powerful inspection devices, statistical analysis, and specialized software to figure out where bugs get into the works and how to detect and avoid them.

Typical problems include contamination that ruins the chips being patterned and etched on wafers, and errors in the photolithographic process itself. Some features on TI's chips are as small as 0.13 microns, so even the slightest variations in width can spell disaster. The align-

ment of successive circuitry patterns, which can be stacked more than 25 layers deep, is also critical; imprecision can mean botched connections.

Texas Instruments gathers tons of data on the wafers while they're being made to avoid learning about mistakes the hard way—weeks after they have occurred, when it's too late to rescue the chips. The trick is to make sense of the information quickly. "We get a data tidal wave from each wafer as it passes through each step of the process," says Joe Lebowitz, director of yield and product engineering. "And it's not intuitively obvious which of the 140,000 or more parameters might be important."

Software providers offer many statistical-data-analysis programs to help meet the information-sifting needs of the semiconductor business. TI uses a number of them, yet hungers for quicker ways to find the signals hidden in all the noise. For the better part of a year the company has been evaluating and participating in the development of Spotfire's specialized DecisionSite geared to chipmakers' requirements. Lebowitz is impressed so far. "I like to make an analogy to a Rubik's Cube, which you need to look at from different sides to solve," he says. "Spotfire allows my engineers to essentially pick up their data in their hands, tumble it around, and look at different aspects of it quickly and easily. That's a big advantage."

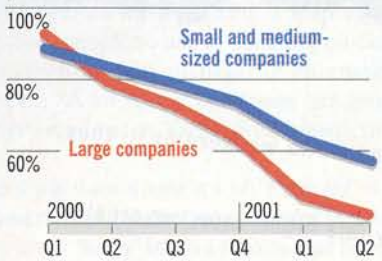
Lebowitz sees a close parallel between his group's work and that of pharmaceuticals researchers. "I want to do the exact same thing with my wafers that the drug people want to do with potential compounds," he observes. "I want to identify a bad actor early and take that material out of the line before I commit thousands more dollars." What company doesn't feel the same way? F

FEEDBACK: *sbrown@fortunemail.com. Stories from the Industrial Management & Technology section can be found at fortune.com/imt. Executives in manufacturing and research and others eligible to receive* FORTUNE*'s Industrial Edition can subscribe by calling 888-394-5472.*

THE NAM / FORTUNE MANUFACTURING INDEX

Doubters Galore

OUTLOOK Percent of manufacturers with a positive business outlook

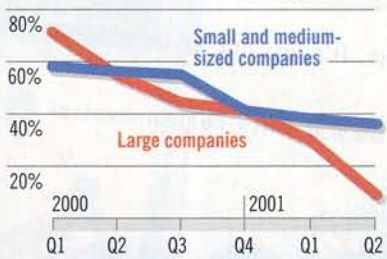


The gloom among manufacturers deepened in the second quarter, accompanied by a few signs that the worst may be over. In a poll of more than 700 companies by the National Association of Manufacturers and FORTUNE, only 42% of large companies—those employing more than 1,000—were optimistic about the next 12 months. That's a new low since the surveys began three years ago, and so is the figure for small and medium-sized companies.

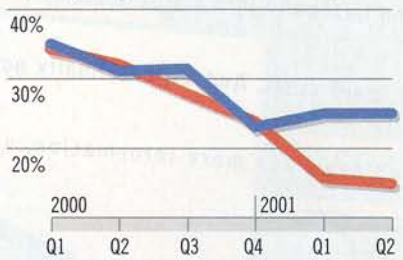
A mere 14% of large companies expect sales to rise more than 5%, less than half the percentage three months ago. This signals that the manufacturing recovery will most likely be slow, owing in part to the strong dollar's dampening effect on exports. But there are chinks of light, including a possible bottoming out in capital-spending plans by large manufacturers and signs that their inventory selloff is ending. If so, inventory shedding will be less of a drag on production in coming quarters. Smaller firms, meanwhile, are still curbing inventories. And not a single large manufacturer expects to lift employment by more than 5%.

How manufacturers view their prospects in ...

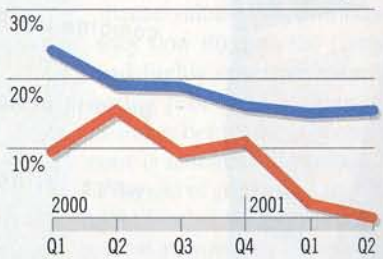
SALES Percent expecting their sales to increase more than 5% over the next year



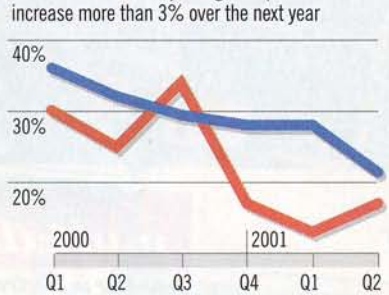
INVESTMENT Percent expecting to increase capital spending more than 5% over the next year



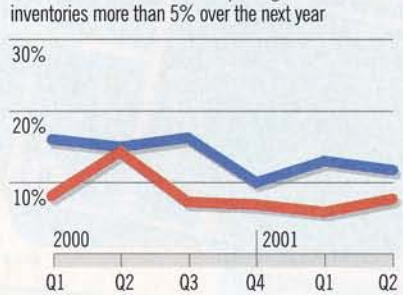
EMPLOYMENT Percent expecting their head count to rise more than 5% over the next year



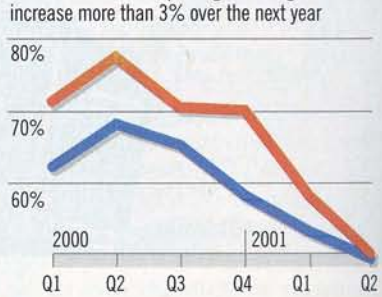
PRICES Percent expecting their prices to increase more than 3% over the next year



INVENTORIES Percent expecting to increase inventories more than 5% over the next year



WAGES Percent expecting their wages to increase more than 3% over the next year



FORTUNE CHARTS