Published in Micro Magazine, June 2003.

TAKING CONTROL 2
Process & Equipment

Using maintenance input data to increase the prediction accuracy of APC strategies

An Cao, Jill Card, and Wai Chan, IBEX Process Technology

As IC manufacturers strive to reach the next technological level and fabs become more and more expensive to build and equip, the need for advanced process control (APC) is becoming a critical component of cutting-edge semiconductor fabrication. The 2001 version of *The International Roadmap for Semiconductors* states that meeting production requirements will eventually require that process corrections be made on a lot-to-lot, wafer-to-wafer, field-to-field, and site-to-site basis. This level of granularity will require APC-enabled process tools.[1] The 2002 update includes the requirement that once a fab decides to install an APC system, it should be deployed in a few weeks. Not stated, but implied, in these requirements is that APC should be truly automated—that is, not requiring human intervention. The fact is, however, current APC solutions require significant human intervention and often cannot provide adequate process improvement.

One reason that human intervention is necessary is that processing is disrupted when tool maintenance is performed. Since a typical controller does not model maintenance events, such disruptions cause the degradation of model performance. Consequently, whenever tool maintenance is performed, the controller must be reset manually. Obviously, if a controller must be reset manually, it is impossible to reduce human intervention or to guarantee that model performance will improve.

Even the most sophisticated controllers cannot perform accurate process control around maintenance events or take tool subcomponent calibration drifts into account unless maintenance is specifically made part of the model.[2,3] This article discusses a neural-network-based APC system that incorporates maintenance input data. Based on production fab trial test and beta test data, the article illustrates how fabs can achieve both accurate and automatic control by accounting for maintenance events in their overall control strategy.

Contrasting APC Controllers

Two major types of run-to-run (R2R) control methodologies are available: proportional and model-based types. Proportional control measures one quality metric and adjusts a single process parameter based on the difference between the measured and the desired quality. Proportional control can be expanded to handle more-complex phenomena by incorporating integral and derivative calculations. By definition, proportional control is univariate. Model-based control, on the other hand, tunes the process based on a mathematical model that behaves similarly to the physical process. Model-based control takes into account the interactions among different aspects of the process, producing a richer, more robust control system.

Aside from the differences among wafers coming into the manufacturing process, three conditions affect process results: changing process parameters, changing physical conditions inside the process tool, and maintenance events. Changing process parameters, also known as set-pointed variables, are measured as trace variables by in situ sensors and include process duration, temperature, pressure, and gas and flow rates. They can cause rapid changes in process results (as reflected in quality metrics), and can improve or degrade the process. Most R2R controllers modify these parameters to optimize process results. Almost all model-based systems include trace variables and quality metrics, but virtually none keeps track of information from maintenance events.

Process results are also affected by long-term drifts caused by aging parts, calibration changes, and residue buildup inside the tool. Most of these conditions cause process drifts that proportional control can estimate directly and adjust for, but they are seldom accommodated by model-based control. When drifting variables reach their limits, the fab takes action, changing, recalibrating, and cleaning parts that affect the process. Such actions cause abrupt shifts in process performance, creating first-wafer effects and other known (but typically not modeled) changes. By not including maintenance events into their models, most APC systems are limited in their ability to improve process control.

Neural-Network-Based R2R Control

The tests discussed in this article were performed using the Dynamic Neural Controller (DNC) from IBEX Process Technology (Lowell, MA). A neural-network-based R2R system, the DNC uses data inputs from the four areas that affect process results: differences among wafers before processing, tool-state (trace) data, maintenance actions, and hardware age (odometry). The system assesses postprocess metrology data to build a model of the process using neural networks and other advanced mathematical techniques, as shown in the schematic drawing in Figure 1. It then uses that model to predict how a wafer is affected during processing and sounds an alarm when the results are out of specification. Because differences among incoming wafers, process parameters, maintenance information, and hardware age are all included in the model, the system can recommend process or maintenance correctives.

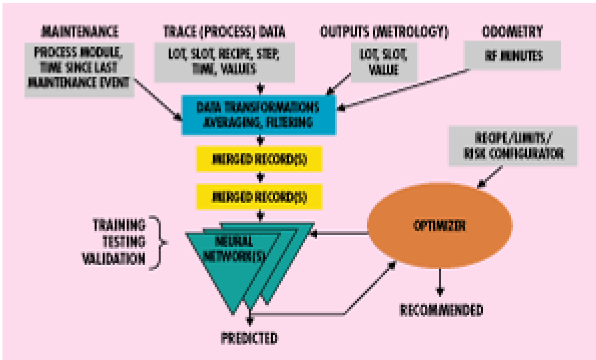

Figure 1: Schematic drawing of the DNC. The system uses neural networks and other advanced mathematical techniques to assess postprocess metrology data for building a model of the process.

Unlike most model-based systems, the DNC can model all chamber recipes together, eliminating the need to produce and maintain multiple models for individual recipes. As new data are generated, the neural networks are retrained automatically, maintaining the model's long-term stability by helping it to adapt to the dynamic changes occurring in the fab environment.

The Basics of Neural Networks. The continuing expansion of the IC market is forcing semiconductor companies to produce more chips at lower costs. As the trend toward larger wafer sizes and smaller device geometries continues, manufacturers must find new methods to increase productivity, while improving wafer quality and delivery times. However, the complex reentrant process flow during wafer fabrication complicates this effort. Multiple variables such as gas flow rates or component wear can affect output quality.

Real-time monitoring and control of manufacturing processes are essential to improving performance. Neural network technology has proven to be an effective means to control and maintain fabrication processes. It can adaptively learn the highly complex (high-dimensional), nonlinear relationships between the physical and electrical systems in the process environment.[4] Moreover, it can help predict output values (quality metrics) by recognizing learned patterns from input data (tool-state and time since last maintenance action data).

Neural network technology mimics the understood processes of the brain—especially pattern recognition and associative memory. Instead of relying on a programmed sequence of steps (e.g., if-then statements), neural networking uses relevant data to program a computer to memorize and recognize patterns. Creating an associative memory of the patterns it has learned, the computer then recognizes similar patterns, predicting future values or events.

A neural network uses mathematical algorithms to mimic the signal transmission by individual human neurons in the central nervous system and the computational capabilities of a network of these neurons.[4] The network consists of an interconnected system of "neuron" units interacting with one another through their "weight" connections.[5] For example, a basic two-layer feed-forward neural network includes input "neurons," or nodes; hidden nodes; and output nodes. The network learns to predict the output node values as a function of input training examples. One or more hidden nodes are used to distinguish complex nonlinear prediction problems.

Neural networks have been especially successful in estimating critical wafer parameters. Because they can learn over time, the networks can stabilize or even improve such estimations over time. Integrating neural networks into APC design, at least for critical processes, reduces wafer quality variability, which reduces costs, inventory, and damaged or scrapped wafers while increasing uniformity, productivity, and reliability.

Prediction Accuracy Calculations. The DNC uses two measurements of model performance. The first is a root-mean-square (rms) calculation, which is useful for comparing two models of the same parameter. The rms is calculated using the following equation:

$$\sqrt{\sum (O_i - P_i)^2 / N}$$

where O is the actual value, P is the predicted value, and N is the total number of samples.

The second measurement, prediction accuracy, is based on the limits supplied by the user. The range of the variable is partitioned into eight bins designating seven categories: the lower safety limit, the lower soft limit, the lower target limit, the target, the upper target limit, the upper soft limit, and the upper safety limit. When the predicted value falls into the same bin as the actual value, it is considered an accurate prediction. Accuracy is defined as the percent of predictions that are accurate (in the correct bin).

Testing the Neural-Network-Based Controller

To understand the impact of maintenance on model-based APC, the DNC was used to compare the prediction accuracy of models with and without maintenance data. Data used for the model came from 19 months of production fab trial test and beta test data. Most of the data were derived from a Sematech-sponsored Equipment Productivity Improvement Team (EPIT) program at STMicroelectronics in Phoenix, AZ, while other data were collected at another fab site. The tests were performed using 4520 XL etchers from Lam Research (Fremont, CA). The test data were gathered retrospectively—that is, the controller was not affecting the process.

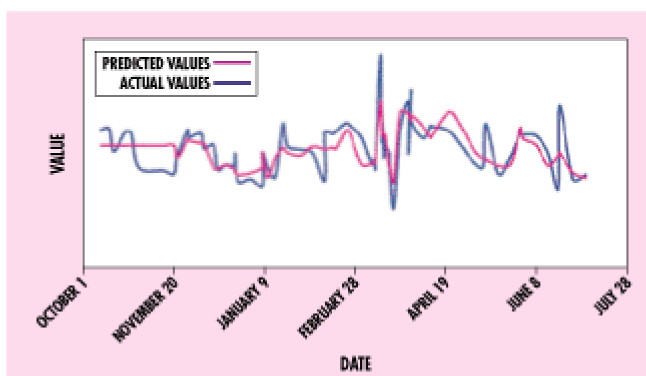

Figure 2: Comparison between predicted and actual etch-rate

values based on maintenance data only. Prediction accuracy was 53.33%.

To predict fluctuations and optimize the process, the controller used five major postetch quality metrics, two of which indicated chamber health: wafer area pressure, which indicates the spacing between the electrodes (the location of a wafer with respect to the plasma); and helium flow, which indicates, among other things, particles on the chuck. The other three metrics were derived from product or monitor wafers: film-thickness difference before and after etch, etch rate, and particle counts. In addition, the study used tool-state or trace data associated with the processing of each wafer through the etch tool.

The data from STMicroelectronics included 13 trace variables, while the data from the other facility included 9. At STMicroelectronics, 31 maintenance actions were tracked and used for calculations, while at the other facility, 14 maintenance actions were tracked and used. At STMicroelectronics, four recipes were run, while at the other facility, 12 were run. The controller routinely recorded the times when the tools underwent maintenance actions.

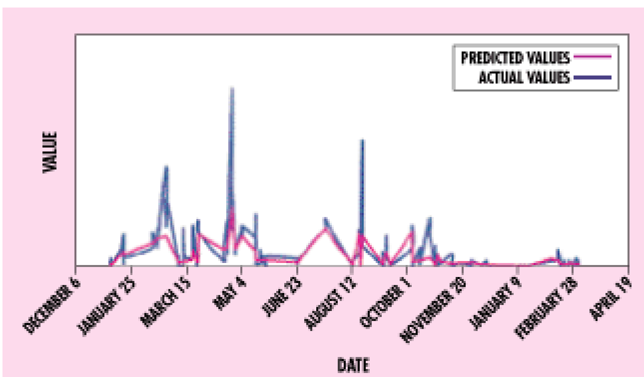

Figure 3: Comparison between predicted and actual particle-count values based on maintenance data only. Prediction accuracy was 70.48%.

During the initial controller setup, the investigators discovered that several fabs, when running monitor wafers, were not collecting tool-state (trace) data to determine the etch rate and particle counts. While it is impossible to create a process control model that can predict quality metrics using the traditional modeling approach (i.e., tool-state data) only, it was found that using maintenance data only can produce models that predict with reasonable accuracy. For example, when the only data used involved the time since the last maintenance action, the controller was able to predict the etch rate with an accuracy of 53.33%, as shown in Figure 2. And using maintenance data only, the controller was able to predict particle counts with an accuracy of 70.48%, as demonstrated in Figure 3. In contrast, as demonstrated in Table I, the exclusion of maintenance data during one beta test made it impossible for the controller to model 6 out of 14 metrics adequately in the absence of trace data.

Quality Metric	Without Maintenance	With Maintenance	
		Accuracy	RMS
Monitor-wafer particle counts	Not available	73.20	39.43
Product-wafer particle counts	Not available	70.48	7.36
Standard deviation	Not available	69.83	33.24
Etch rate 1	Not available	29.33	74.58
Etch rate 2	Not available	53.33	153.09
Particle counts	Not available	52.94	38.41

Table I: At one site, a model could not be made without maintenance data for many quality parameters because there was no tool-state data available.

In some cases, the investigators were able to develop a control model using tool-state data only that compared favorably with other models in the industry. However, the inclusion of maintenance data improved the model's accuracy. On average, predictions were 20% more accurate when maintenance data were added to the model. In fact, the inclusion of maintenance data reduced the rms error in most networks by 35% or more. Table II indicates how maintenance inputs can improve the model's prediction accuracy, and Table III shows that the use of maintenance data reduces rms error.

Quality Metric	Accuracy without Maintenance	Accuracy with Maintenance	Percentage Increase
Film thickness	15.07	24.16	60.32
WAP 1	43.30	64.58	49.15
Thickness difference	63.07	80.49	27.62
Helium flow 2	57.31	71.06	23.99
WAP 2	52.58	64.60	22.86
Product etch step height	33.00	39.00	18.18
Helium flow 1	95.04	97.80	2.90
Helium flow pressure	99.97	99.97	0.00

Table II: When maintenance history is accounted for, the model is able to predict much closer to actual results.

Quality Metric	RMS without Maintenance	RMS with Maintenance	Percentage Decrease
Helium flow 2	9.16	4.38	52.18
Helium flow 1	0.71	0.41	42.25
WAP 2	8.48	4.98	41.27
Thickness difference	285.32	169.75	40.51
WAP 1	5.89	3.86	34.47
Helium flow pressure	0.11	0.10	9.10
Film thickness	14.97	13.68	8.62
Product etch step height	0.01	0.02	–100.00

Table III: The rms improves substantially when maintenance is included in the model.

Figure 4a demonstrates that it is difficult to predict process results without understanding that maintenance has occurred. Predicted and actual film-thickness values on product wafers measured before and after etch are plotted. The green lines represent when maintenance events occurred. The pink line, indicating predicted values, tracks a fairly stable course over time, while the actual results (shown in blue) display distinct dips after maintenance events and corresponding rises between them (indicated by the red arrows).

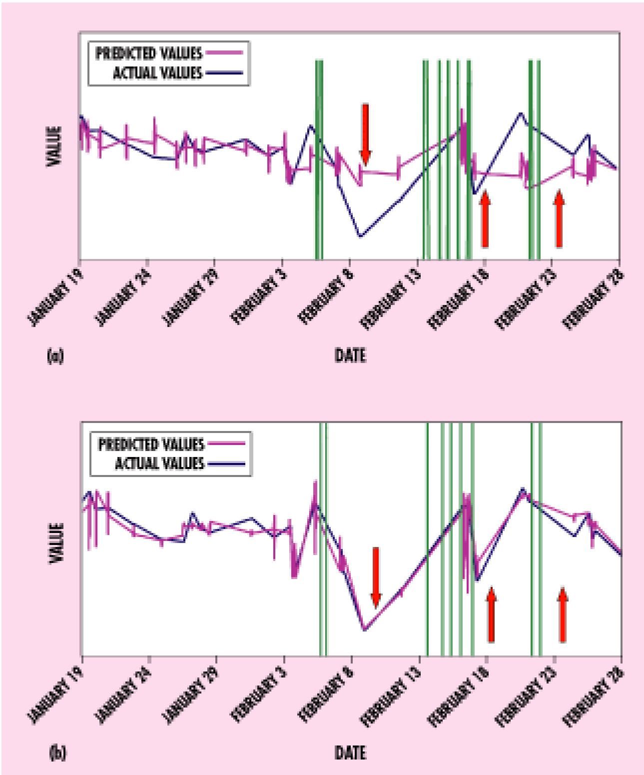

Figure 4: Comparison between predicted and actual thickness difference values (a) without and (b) with maintenance data. Without maintenance data, the model cannot predict abrupt process changes caused by maintenance actions.

In Figure 4b, the model's performance is shown with the the inclusion of maintenance actions. After the first maintenance event, predictions closely corresponded to actual results, but even beforehand, when the system appeared stable, predicted and actual values correlated well. It appears that this overall improved accuracy reflects the model's understanding of how parts aging affects the film-thickness-difference metric. Without the inclusion of maintenance data, the accuracy of the model in predicting film-thickness differences was 63.07%, while the inclusion of such data increased prediction accuracy to 80.49%, an increase of 27.62%. The rms error was reduced from 285.32 to 169.75—a reduction of 40.51%.

Helium flow, the flow of helium to the backside of the wafer, often indicates the presence of particles on the wafer chuck. It is measured inside the tool and for every wafer processed. As a result, control models of helium flow are fairly accurate with or without maintenance data, as illustrated in Figures 5a and 5b. The prediction accuracy of helium flow without the inclusion of maintenance data was 95.04%. With the inclusion of maintenance data, the prediction accuracy increased to 97.8%, an increase of 2.9%. The rms error was reduced from 0.71 to 0.41—a reduction of 42.25%.

How Maintenance Data Help the Controller to Model Tool Changes

By using maintenance data to determine when and why process excursions occur, the controller can predict both the results of short-term process shifts caused by maintenance actions and long-term drifts caused by the degradation of tool parts. Figures 4 and 5 show how maintenance information enables manufacturers to predict the results of short-term shifts, while Figure 6 shows how the inclusion of maintenance data can improve the model's ability to predict long-term drift.

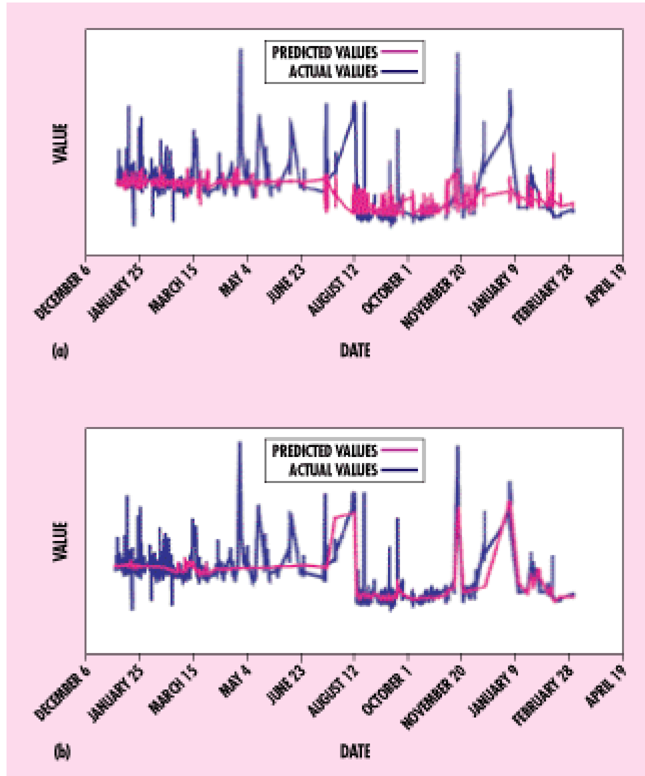

Figure 5: Comparison between predicted and actual helium flow values (a) without and (b) with maintenance data values. While helium flow can be predicted accurately without maintenance data, it can be predicted more accurately with the inclusion of maintenance data.

Figure 6a presents predicted and actual process values derived without the inclusion of maintenance data for wafer area pressure (WAP). Predicted values (represented by the pink line) remained steady throughout the entire test period, while actual values (represented by the blue line) drifted low and then high (as tracked by the yellow line). When maintenance data were included, as shown in Figure 6b, the model was able to predict the process drift (as represented by the correlation between the pink and yellow lines).

Besides improving the accuracy of the control model, the addition of maintenance data enables manufacturers to simulate a process before and after maintenance events and determine if an action has improved performance. By performing such simulations on a wafer-to-wafer basis, the controller can optimize processes by recommending beneficial maintenance actions. Initial studies have found that such recommendations can prevent wafer scrap by identifying solutions to problems long before the problems can be solved using traditional methods. For example, in one test, a minor alarm predicted the heightened presence of particles affecting 100 wafers. Although particle measurements confirmed the prediction at the end of the 100-wafer period, 200 more wafers were processed before the excursion was acknowledged, diagnosed, and corrected. The early warning could have prevented the fab from misprocessing 200 wafers.

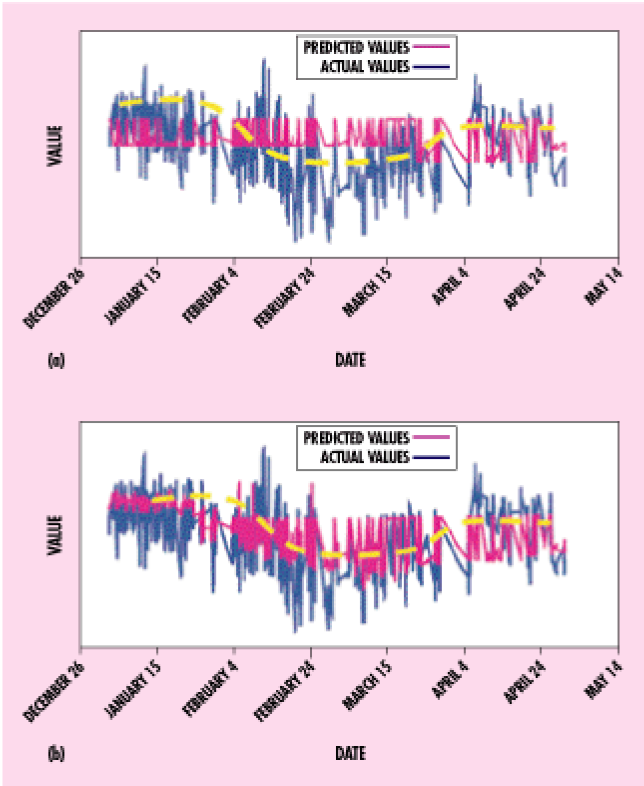

Figure 6: Comparison between predicted and actual values for wafer area pressure (a) without and (b) with maintenance data. Predicted values (pink line) remained steady throughout the entire test period, while actual values (blue line) drifted (tracked by yellow line). When maintenance data were included, the model was able to predict the process drift (correlation between the pink and yellow lines).

Another benefit of the control system is that it can help determine when maintenance should be performed, reducing the need for scheduled maintenance or emergency problem-solving measures. The investigators have found that 40–50% of all maintenance activities performed on a typical etch tool are premature or altogether unnecessary. The neural-network-based APC system helps ensure that such activities are performed at the proper time, reducing both scheduled and unscheduled downtime.

Conclusion

Most APC systems depend solely on tool-state data, resulting in the need for manual resets after each maintenance action and reducing the accuracy of control models. For APC to become fully automated, the need for manual resets must be eliminated. And to become fully functional over all wafers, APC must be highly accurate.

Including data from maintenance actions into a controller eliminates the need for manual resets. It also improves model accuracy by predicting the results of shifts and drifts caused by the state of the tool. For well-controlled tools in particular, maintenance may have more impact on process results than drifts in affecting set-pointed tool-state variables.[6] Because of the benefits of including maintenance data in model-based APC systems, it is expected that the use of such data will be a feature of all future controllers.

Acknowledgments

The authors would like to thank Anna Foard from Clear Tech (Newton, NH) for coauthoring the section in this article on neural networks. In addition, they would like to acknowledge Frank Hoppensteadt from the Center for Systems Science and Engineering at Arizona State University (ASU) in Tempe for providing information on the fundamentals of neural networking, and Jennie Si from the department of electrical engineering, also at ASU. They also wish to thank Deana Delp for providing information on neural network applications in the semiconductor manufacturing industry.

References

1. *The International Technology Roadmap for Semiconductors* (San Jose: Semiconductor Industry Association, 2001); available from Internet: http://public. itrs.net.

2. CH Cheng and J Moyne, "Intelligent Metrology and Control Systems Applied to Lithography Processes: Case Study: Lithography Overlay Run-to-Run Control" (paper presented at the 14th AEC/APC Symposium, September 7–12, 2002, Snowbird, UT).

3. R Chong et al., "Analysis of a Run-to-Run Controller on a Drifting STI Etch Process by Augmentation of the Integrated Interferometric Endpoint Detection System" (paper presented at the 14th AEC/APC Symposium, September 7–12, 2002, Snowbird, UT).

4. JP Card, M Naimo, and W Ziminsky, "Run-to-Run Process Control of a Plasma Etch Process with Neural Network Modeling," *Quality and Reliability Engineering International* 14, no. 4 (1998): 247–260.

5. S Limanond, J Si, and K Tsakalis, "Monitoring and Control of Semiconductor Manufacturing Processes," *IEEE Control Systems Magazine* 18, no. 6 (1998): 46–58.

6. J Card and L Laurin, "Using Neural Networks for Intelligent Plasma Etch Process Control," *Solid State Technology* 45, no. 11 (2002): 33–36.

An Cao, PhD, is as a consulting engineer at IBEX Process Technology (Lowell, MA), where she is responsible for the development of new algorithms. She has coauthored eight scientific papers in technical journals and is the coinventor of three patents pending in the area of neural modeling. Cao is a member of the American Association for the Advancement of Science and the Sigma Xi Society. She received a BS in biomedical engineering and a BS in electrical engineering from China's Shanghai Jiaotung University and an MS in biophysics from the University of Science and Technology of China in Hefei. She received a PhD in neuroscience and artificial intelligence from the Massachusetts Institute of Technology in Cambridge. (Cao can be reached at 978/452-0287 or acao@ibexprocess.com.)

Jill Card is founder, chairman, and chief scientist of IBEX Process Technology, where she applies advanced mathematical techniques based on evolutionary computational methods—specifically pattern recognition and associative memory—to create industrial solutions that learn and adapt over time. She has worked for more than 20 years as a statistician and applied mathematician in a variety of fields, including reliability analysis, quality control analysis, and neural network analysis for the semiconductor and other industries. Previously, Card was a consulting engineer at Digital Equipment Corp., a member of the technical staff at Bell Labs, and principal engineer at Wang Labs. She has published 11 papers and 6 proceedings presentations, and is the coinventor of seven patents pending in the area of neural control. She received a BS in biology/natural resources from Cornell University in Ithaca, NY, and an MS in theoretical and applied statistics from Florida State University in Tallahassee. (Card can be reached at 978/452-3902 orjcard@ibexprocess.com.)

Wai Chan, PhD, is director of mathematical analysis at IBEX Process Technology. He has more than 15 years of experience as a mathematician/statistician in the computer industry in manufacturing quality and reliability organizations. Before joining Ibex, he was employed at Digital Equipment Corp. as a principal engineer and as an assistant professor of mathematics at Ohio State University. Chan introduced key quality metrics and systems into the high-tech manufacturing environments at Compaq Computer and Digital Equipment. He has coauthored 14 scientific papers and is the coinventor of five patents pending in the area of neural control. He received a BS in mathematics from the University of Wisconsin in Madison, an MS in mathematics from the University of Texas in Austin, and a PhD in theoretical statistics from Florida State University in Tallahassee. (Chan can be reached at 978/452-8845 or wchan@ibexprocess.com.)